



HAL
open science

Dialogues entre les arts et les sciences : vers une phénoménologie de la création

Ivan Magrin-Chagnolleau

► **To cite this version:**

Ivan Magrin-Chagnolleau. Dialogues entre les arts et les sciences : vers une phénoménologie de la création. Sciences de l'Homme et Société. Université de Lorraine, 2023. tel-04471922

HAL Id: tel-04471922

<https://hal.univ-lorraine.fr/tel-04471922v1>

Submitted on 21 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DIALOGUES ENTRE LES ARTS ET LES SCIENCES : VERS UNE PHÉNOMÉNOLOGIE DE LA CRÉATION

**HABILITATION À DIRIGER DES RECHERCHE
SOUTENUE LE 16 JUIN 2023
À L'UNIVERSITÉ DE LORRAINE PAR**

IVAN MAGRIN-CHAGNOLLEAU

DEVANT UN JURY COMPOSÉ DE :

PROFESSEUR ROLAND HUESCA, UNIVERSITÉ DE LORRAINE, PRÉSIDENT, EXAMINATEUR

PROFESSEURE SYLVIE DALLET, UPEM, UVSQ, RAPPORTEURE

PROFESSEUR AMOS FERGOMBÉ, UPHE, RAPPORTEUR

PROFESSEUR BERNARD ANDRIEU, UNIVERSITÉ PARIS DESCARTES, RAPPORTEUR

PROFESSEURE CLAIRE LAHUERTA, UNIVERSITÉ DE LORRAINE, EXAMINATRICE

PROFESSEURE MARIE-CHRISTINE MAUREL, SORBONNE UNIVERSITÉ, EXAMINATRICE



**HUMANITÉS
NOUVELLES**



Habilitation à Diriger des Recherches

Ivan MAGRIN-CHAGNOLLEAU

Tome 1 : CV, contributions majeures et écrits choisis

Table des matières

| | |
|--|-----------|
| Table des matières | 2 |
| 1 Curriculum Vitae | 5 |
| 1.1 Expériences professionnelles en recherche | 5 |
| 1.2 Expériences professionnelles en art | 6 |
| 1.3 Groupes de recherche, sociétés savantes et organisations pro- fessionnelles | 7 |
| 1.4 Formation | 8 |
| 1.5 Responsabilités administratives et scientifiques | 9 |
| 1.6 Enseignements et séminaires | 10 |
| 1.7 Animation de la recherche | 12 |
| 1.8 Diffusion de la recherche | 17 |
| 1.9 Encadrements et supervisions d'étudiants | 18 |
| 1.10 Publications et communications scientifiques | 22 |
| 1.11 Critiques | 35 |
| 1.12 Brevets | 39 |
| 1.13 Logiciels | 40 |
| 1.14 Publications et productions artistiques | 41 |
| 2 Quelques contributions majeures | 55 |
| 2.1 Contributions majeures dans le domaine de la reconnaissance automatique du locuteur | 55 |
| 2.2 Contributions majeures dans le domaine de l'indexation so- nore et multimédia | 59 |
| 2.3 Contributions majeures dans le domaine de l'analyse des si- gnaux sismiques | 62 |
| 2.4 Contribution majeure dans le domaine de la reconnaissance automatique des langues | 63 |
| 2.5 Contribution majeure dans le domaine de la perception du langage | 64 |
| 2.6 Contribution majeure dans le domaine de la langue des signes | 65 |

| | | |
|----------|---|-----------|
| 2.7 | Contributions majeures sur la méthode de l'Actors Studio . . . | 66 |
| 2.8 | Contributions majeures en esthétique du théâtre et du cinéma | 67 |
| 2.9 | Contributions majeures en Création Recherche | 68 |
| 2.10 | Contributions majeures sur le processus créatif et sa phéno- ménologie | 70 |
| 2.11 | Contributions majeures sur l'énaction | 72 |
| 2.12 | Contributions majeures sur l'entretien d'explicitation | 73 |
| 2.13 | Contributions majeures en psychologie et psychanalyse . . . | 74 |
| 2.14 | Autres contributions en philosophie de l'art et en esthétique | 75 |
| 3 | Écrits choisis | 77 |
| 3.1 | Reconnaissance automatique du locuteur | 77 |
| 3.2 | Indexation sonore et multimédia | 184 |
| 3.3 | Analyse des signaux sismiques | 212 |
| 3.4 | Reconnaissance automatique des langues | 222 |
| 3.5 | Perception du langage | 227 |
| 3.6 | Langue des signes | 232 |
| 3.7 | La méthode de l'Actors Studio / Le jeu de l'acteur | 235 |
| 3.8 | Esthétique du théâtre et du cinéma | 295 |
| 3.9 | Création Recherche | 315 |
| 3.10 | Le processus créatif / Phénoménologie du processus créatif . | 334 |
| 3.11 | Énaction | 367 |
| 3.12 | Entretien d'explicitation | 400 |
| 3.13 | Psychologie / Psychanalyse | 434 |
| 3.14 | Autres écrits en philosophie de l'art et en esthétique | 451 |
| 3.15 | Quelques critiques de cinéma | 473 |

Chapitre 1

Curriculum Vitae

Ivan MAGRIN-CHAGNOLLEAU, artiste chercheur

Chargé de recherche au CNRS en section 35

Laboratoire PRISM UMR 7061 (Perception Représentation Image Son Musique) depuis le 1er janvier 2019 - CNRS & Aix-Marseille Université

1.1 Expériences professionnelles en recherche

- Depuis 2019 : **Affiliation principale : chargé de recherche au CNRS** au laboratoire PRISM UMR 7061 (Perception Représentation Image Son Musique), Marseille, France.
- Depuis 2020 : **Affiliation secondaire : affiliated scholar, puis presidential fellow** à Chapman University, Orange, California, USA.

Affiliations passées :

- 2010 - 2018 : **Chargé de recherche au CNRS** à l'institut ACTE UMR 8218 (Arts Créations Théories Esthétiques).
- 2001 - 2004 : **Chargé de recherche au CNRS** au Laboratoire Dynamique du langage à Lyon.

- 2000 - 2001 : **ATER** au **Laboratoire Informatique d'Avignon**.
- 1999 - 2000 : **Post-doc.** à **l'IRISA** (CNRS & INRIA), Rennes.
- 1998 - 1999 : **Chercheur associé** à **Rice University**, Houston, TX.
- 1997 - 1998 : **Chercheur** à **AT&T Labs-Research**, Florham Park, NJ.

1.2 Expériences professionnelles en art

- Depuis 2014 : **Critique** (cinéma, théâtre, performance, musique, expositions, ...)
- Depuis 2012 : **Fondateur, président et directeur artistique d'Aloha Films** : structure associative qui produit des courts métrages de fiction.
- 2011 - 2014 : **Membre du CA et responsable du cercle des scénaristes de Collectif Prod** : structure associative qui aide les jeunes à se professionnaliser dans le cinéma.
- Depuis 2009 : **Fondateur et directeur artistique de la Compagnie Aloha** : compagnie qui crée des spectacles vivants.
- Depuis 2004 : **Auteur, réalisateur, chef opérateur, monteur, étalonneur** pour le cinéma
- Depuis 2004 : **Auteur, metteur en scène, créateur lumière** pour le théâtre.
- Depuis 2003 : **Acteur** au théâtre et au cinéma.
- 2003 - 2004 : **Co-fondateur et co-directeur artistique de la Compagnie Padrivantia** : compagnie qui a créé des spectacles vivants en 2003 et 2004.

1.3 Groupes de recherche, sociétés savantes et organisations professionnelles

Groupes de recherche

- Membre du GDR 1POINT5 (Réduction de l’empreinte carbone de la recherche)
- Membre du GDR 2091 (IA, Internet et Société)
- Membre du bureau du GREX2 (Groupe de recherche sur l’explicitation 2)

Sociétés savantes

- Membre de l’APA (American Philosophical Association)
- Membre du CFC (Centre Français de la Couleur) et de l’AIC (Association Internationale de la Couleur)
- Membre de l’EASTAP (European Association for the Study of Theatre And Performance)
- Membre du WAN (Walking Artists Network)
- Membre de l’aCD (Association des chercheurs en danse)

Organisations professionnelles

- Membre de la SRF (Société des Réalisateurs de Films) et du collectif 50/50
- Membre de la SACEM
- Membre de La Réplique et de l’AAFA (Actrices Acteurs de France Associés)
- Membre du SFCC (Syndicat Français de la Critique de Cinéma) et de la FIPRESCI (Fédération internationale des critiques de films)

1.4 Formation

- 2018 - 2020 : **Formation en psychanalyse** (Association Lacanienne Internationale, Paris).
- 2017 - 2019 : **Licence de psychologie** (Paris VIII).
- 2004 - 2007 : **Actors Studio** en tant qu'acteur et metteur en scène et **Master of Fine Arts en théâtre** (New York).
- 2002 - 2004 : **DEUG de psychologie** (CNED et Paris VIII).
- 1993 - 1997 : **Thèse de doctorat en signal et image** (Télécom Paris).
- 1995 - 1996 : **D.E.A. de Phonétique** (Paris III).
- 1992 - 1993 : **D.E.A. Automatique et Traitement du Signal** (Paris XI).
- 1989 - 1992 : **Ingénieur E.N.S.E.A.** – Option Analyse et traitement du signal.

1.5 Responsabilités administratives et scientifiques

Projets scientifiques

- **Co-porteur du projet « Race, Immigration, Technology and the Creative Process in Performance »**
Projet d'1 an financé par *The France-Stanford Center for Interdisciplinary Studies* en 2013-2014. Co-porteur du projet avec Jennifer Brody du département TAPS de Stanford University. Budget de 15 k\$.
- **Co-porteur du projet RAIVES**
Projet d'indexation sonore de documents radiophoniques sur le web, d'une durée de 4 ans, financé par le CNRS dans le cadre de l'action « Société de l'information » à partir de novembre 2001. Co-porteur du projet avec Nathalie Parlangeau-Vallès. Partenaires du projet : le LORIA (Nancy) et l'IRIT (Toulouse). Budget de 800 k€.
- **Porteur du projet ALIZÉ**
Projet de réalisation d'une plateforme logiciel libre en vérification automatique du locuteur, d'une durée de deux ans, financé par le ministère de la recherche dans le cadre de l'action « Technolanguages » à partir de novembre 2002. Porteur du projet. Partenaires du projet : le LIA (Avignon), l'IRISA (Rennes), le CLIPS (Grenoble), l'IRIT (Toulouse), l'ENST (Paris), THALES (Gennevilliers), AT-LOG (Niort), CALISTEL (Montbonnot), DGA (Arcueil). Budget de 300 k€ pour les deux ans.
- **Partenaire du projet AGIR**
Projet d'indexation multimédia, d'une durée de deux ans, financé par le ministère de la technologie à partir de novembre 2003. Partenaire du projet. Budget de 400 k€ pour les deux ans.

Direction d'équipe

- **Fondateur et directeur de l'équipe de recherche EsPAS** (Esthétique de la Performance et des Arts du Spectacle) 2010-2015

Autres responsabilités

- Membre du conseil de laboratoire de l'institut ACTE 2012-2015

1.6 Enseignements et séminaires

Séminaires

- **Création et organisation du séminaire du laboratoire PRISM « Les JEUDIS de PRISM »**
2021 -
Laboratoire PRISM, 1 séance par mois.
- **Création et organisation du séminaire « Phénoménologie de la création »**
2016 - 2017
Laboratoire ACTE, 6 séances par année.
- **Création et organisation du séminaire « Esthétique expérimentale en performance et arts du spectacles »**
2015
Laboratoire ACTE, 6 séances dans l'année.
- **Organisation du séminaire du laboratoire DDL (Dynamique Du Langage)**
2002 - 2003
Laboratoire DDL, 1 séance par mois.

Enseignements

- **Cours, ateliers, masterclasses**
Interventions régulières en France et à l'étranger sous forme de cours, ateliers, et masterclasses autour de thèmes en lien avec l'art et ma recherche.
- **Cours à Chapman University en anglais**
2023
Art et éthique. (36 h)

- **Cours à l'EFAP
2022**
Créativité et innovation. (12 h)
Storytelling. (8 h)
- **Cours à Chapman University en anglais
2020**
Critique et jugement esthétique. (36 h)
- **Cours à la SWU à Bangkok en anglais
2019**
Direction d'acteur et jeu de l'acteur. (20 h)
- **Cours à l'ESRA en anglais
2019**
Écriture scénaristique, découpage, réalisation. (30 h)
- **Cours à l'université d'Évry
2009 - 2010**
Histoire de la mise en scène . (30 h)
- **Atelier de création scénique à l'université d'Évry
2009 - 2010**
Cet atelier a également représenté un volume horaire de 30 heures sur l'année. Il est décrit en détails dans une publication (voir chapitre 3).
- **École d'été ELSNET au Danemark
2002**
Ecole d'été ELSNET 2002 au Danemark sur l'évaluation des techniques et systèmes en parole et langue : cours sur l'évaluation des systèmes de reconnaissance du locuteur (7h30). Cet enseignement a donné lieu à la rédaction d'un chapitre de livre (voir publications).
- **Cours sur le traitement du signal à l'université Paris 13
2002 - 2004**
Cours sur le traitement du signal à l'université Paris 13 en troisième année du cycle d'ingénieurs en informatique (56h par an pendant deux ans).

- **Cours sur le traitement automatique de la parole à l'université Paris 13**
2001 - 2004
 Cours sur le traitement automatique de la parole à l'université Paris Nord 13 en troisième année du cycle d'ingénieurs en informatique (62h par an pendant trois ans).
- **Cours sur le traitement automatique de la parole à l'université Lyon 2**
2001 - 2004
 Cours sur le traitement automatique de la parole à l'Université Lyon 2 en DEA de sciences cognitives (18h par an pendant trois ans).
- **Cours sur le traitement automatique de la parole à l'I.N.T.**
2000 - 2004
 Cours sur le traitement automatique de la parole à l'I.N.T. en troisième année d'école d'ingénieurs (24h par an pendant quatre ans).

1.7 Animation de la recherche

Responsabilités éditoriales

- 2014- : **Fondateur et directeur de publication du journal *p-e-r-f-o-r-m-a-n-c-e*** : journal de création recherche autour des arts performatifs, des arts du spectacles, des arts visuels et des arts sonores. Membre du comité éditorial, du comité scientifique et du comité de lecture.
- 1995 - 2004 : **Relecteur régulier** pour les revues internationales suivantes : *IEEE Transactions on Signal Processing*, *IEEE Transactions on Speech and Audio Processing*, *Digital Signal Processing*, *Journal of the Acoustical Society of America*, *Journal on Applied Signal Processing*.

Groupes de recherche

- 2019- : **Membre du GDR 1POINT5** (Réduction de l'empreinte carbone de la recherche) - fondateur et animateur de l'équipe Arts&Sciences.

- 2019- : **Membre du GDR 2091** (IA, Internet et Société) - membre du groupe de travail « Intelligence artificielle, art et créativité ».
- 2017- : **Membre du bureau du GREX2** (Groupe de recherche sur l'explicitation 2)
- 2000 - 2004 : **Membre du bureau de SpLC (Speaker and Language Characterization)**, un groupe d'intérêt spécial (SIG) de l'ISCA (International Speech Communication Association) consacré à la caractérisation du locuteur et des langues. Développement et mise à jour du site web. Responsable du groupe sur la segmentation par locuteurs.
- 1999 - 2004 : **Membre fondateur du consortium ELISA**, consortium regroupant plusieurs laboratoires francophones autour du thème de la vérification du locuteur.
- 1996 - 2004 : **Membre du GDR ISIS** (Informatique, traitement du signal et des images)

Organisation de congrès, conférences, colloques

- **Membre du comité scientifique du colloque international « Enseigner les Transitions Écologiques et Sociales dans le supérieur »**
Le colloque aura lieu à Paris les 4 et 5 juillet 2023.
- **Organisateur et membre du comité scientifique du colloque international « L'apport de Pierre Vermersch à l'étude de la subjectivité »**
Le colloque a eu lieu à Marseille les 13 et 14 octobre 2022.
- **Initiateur, organisateur et membre du comité scientifique du colloque international « Francisco Varela : une pensée actuelle »**
Le colloque a eu lieu à Cerisy du 13 au 19 août 2022.

- **Organisateur et membre du comité scientifique de la journée d'étude « Méthodes collaboratives en recherche et création »**
Le colloque a eu lieu à Marseille le 17 décembre 2020.
- **Paper Chair avec Mistuko Aramaki et membre du comité de lecture du colloque international CMMR 2019 (Computer Music Multidisciplinary Research)**
Le colloque a eu lieu à Marseille du 14 au 18 octobre 2019.
- **Organisateur et membre du comité scientifique du colloque international « The science of joint improvisation 2015 »**
Le colloque a eu lieu sur trois jours à Paris en novembre 2015.
- **Organisateur et membre du comité scientifique du colloque international « En chair et en son : rencontres internationales Butō et musique acousmatique »**
Le colloque a eu lieu sur deux jours à Paris les 23 et 24 octobre 2015.
- **Initiateur, organisateur et membre du comité scientifique du colloque international « Le processus créatif dans les arts performatifs, visuels, sonores et littéraires »**
Ce colloque a eu lieu sur deux jours à Paris à la fin du mois de juin 2014.
- **Initiateur et membre du comité scientifique du colloque « Performances et animalités »**
Colloque organisé au musée de la Chasse et de la Nature à Paris du 13 au 14 novembre 2013.
- **Organisateur et animateur d'un SIG sur les neurosciences**
Dans le cadre du congrès IDEA PARIS 2013 organisé en juillet 2013, organisation et animation d'un groupe de travail sur l'impact des neurosciences en théâtre-éducation.
- **Organisateur et animateur d'une table ronde sur la réception des œuvres**
Dans le cadre du congrès IDEA PARIS 2013 organisé en juillet 2013, animation d'une table ronde sur la réception des œuvres artistiques, les prérequis à cette réception, et la façon de transmettre ces prérequis. Les invités à cette table ronde étaient : Emmanuel

Demarcy-Mota, directeur du théâtre de la Ville et du festival d'automne; Didier Deschamps, directeur du théâtre de Chaillot; Nicolas Bourriaud, directeur de l'école nationale des beaux-arts de Paris; Carole Desbarats, responsable des enfants de cinéma et directrice de la communication à l'ENS; Monique Barbaroux, directrice du centre national de la danse.

- **Membre du comité scientifique et du comité d'organisation du congrès international « IDEA PARIS 2013 »**

Le congrès IDEA PARIS 2013 a eu lieu du 8 au 13 juillet 2013. IDEA est l'organisation internationale de théâtre-éducation, qui organise un congrès international très important tous les trois ans. En 2013, c'était l'ANRAT (l'organisation française de théâtre-éducation) qui organisait le congrès à Paris. Le congrès a regroupé plus de 1000 personnes, universitaires, enseignants, artistes et étudiants. Il y avait des présentations, des ateliers, des tables rondes, des groupes de travail, etc.

- **Organisateur et coordinateur du colloque « Le corps vieillissant dans les arts de la scène »**

Sur la demande du directeur administratif du centre chorégraphique national de Caen, Bruno Joly, organisation et animation d'une journée consacrée au corps vieillissant dans les arts de la scène. Ce colloque a regroupé une dizaine d'intervenants et s'est déroulé devant un amphithéâtre d'une centaine de personnes en mars 2013.

- **Membre du comité d'organisation du colloque international « Archive vivante »**

Ce colloque international a eu lieu à l'université Paris 7 sur deux journées les 25 et 26 octobre 2012 et était organisé conjointement par Paris 1 et Paris 7. Il était sur le thème de l'archive vivante.

- **Membre du comité de programme du workshop Odyssey 2004**

Workshop sur la reconnaissance du locuteur et de la langue ayant eu lieu à Toledo, Espagne, en juin 2004. Ce workshop a regroupé environ 80 chercheurs pendant 4 jours.

Collaborations scientifiques

- **Chapman University à Orange, Californie**

En 2020, je suis parti à Chapman University pour un séjour de recherche et d'enseignement qui devait initialement durer 4 mois, mais a été écourté par la pandémie. Le séjour a néanmoins été très fructueux, et les contacts de très grande qualité. En décembre 2020, j'ai été nommé par le président de l'université « Affiliated Scholar » jusqu'à fin décembre 2023 dans un premier temps. J'ai été en contact avec des collègues de l'école de cinéma et de l'école des arts performatifs (musique, théâtre et danse), notamment Nicholas Terry avec lequel je suis en train de développer un projet de création recherche. J'y suis retourné un mois en avril 2022 et j'y retourne un mois en septembre 2022 pour préparer un séjour de recherche plus long en 2023.

- **Performance Studies à New York University**

Le champ de la performance est né aux États-Unis, et a été développé au départ par Richard Schechner à New York University. C'est donc tout naturellement que je me suis tourné vers le département des Performance Studies de New York University. J'ai effectué plusieurs séjours à New York, afin de développer des liens et des collaborations avec plusieurs membres de ce département. Richard Schechner (qui a maintenant pris sa retraite) et Allen Weiss, qui est également un spécialiste des Performance Studies à New York University, sont tous les deux devenus des amis, et nous échangeons régulièrement nos points de vue. Ils viennent d'ailleurs tous les deux assez régulièrement en France, ce qui fait que nous nous voyons souvent. J'ai aussi des échanges réguliers avec André Lepecki, Barbara Browning, Deborah Kapchan et Diana Taylor (qui a également pris sa retraite), qui font également partie du département des Performance Studies à New York University.

- **Hemispheric Institute à New York**

Diana Taylor est par ailleurs la fondatrice et l'ancienne directrice de l'Hemispheric Institute (Elle vient tout juste de prendre sa retraite). J'ai également des échanges réguliers avec son équipe.

- **Stanford University**

J'ai sollicité Jennifer Brody, directrice du département théâtre et performance à Stanford University, pour une bourse de colla-

boration auprès de l'institut France-Stanford. Nous avons obtenu cette bourse pour 2013-2014 pour travailler ensemble autour de trois thématiques : 1) Race et immigration en performance ; 2) Les nouvelles technologies en performance ; 3) Le processus créatif en performance. Ce financement nous a donc permis de nous rencontrer et de réfléchir ensemble à ces problématiques.

- **Berkeley University**

Mes séjours à San Francisco pour rencontrer Jennifer Brody et les membres de son département m'ont permis de rencontrer les membres du département Théâtre et Performance Studies à Berkeley University, notamment Catherine Cole, directrice du département, et Shannon Jackson, directrice du Arts Research Center.

- **IRIT à Toulouse**

J'ai collaboré avec Nathalie Vallès-Parlangeau sur plusieurs projets d'indexation sonore et multimédia, et je l'ai également associée, plus récemment, à une réflexion que je mène sur l'utilisation des nouvelles technologies en création artistique.

- **Ex Machina à Québec**

Enfin, j'ai noué des liens avec l'équipe d'Ex Machina, la structure de Robert Lepage, metteur en scène québécois, lors de mon séjour chez eux en août 2013.

1.8 Diffusion de la recherche

- **Article dans le journal des grandes écoles** consacré à l'éducation dans les universités (numéro de septembre 2015).
- **Interview pour le journal 19-20 de France 3 région Rhône-Alpes** sur le thème de la vérification de l'identité par la voix (16 février 2003).
- **Article dans le quotidien régional Le Progrès, Lyon**, à propos du séminaire donné dans le cadre des déjeuners de l'institut des sciences de l'homme à Lyon (12 décembre 2002).

- **Séminaire dans le cadre des déjeuners de l'institut des sciences de l'homme à Lyon** : "Comment reconnaître une personne à partir de sa voix" (22 novembre 2002).

1.9 Encadrements et supervisions d'étudiants

- **M1 Lyess Zaïdat 2023**
Mémoire en M1 (Aix-Marseille Université) intitulé : « Composition musicale assistée par ordinateur ».
Co-encadrement à **50%** avec Javier Elipe Gimenez (maître de conférences AMU à PRISM) et Marcelo Caetano (post-doc à PRISM). Suite du stage de Felipe Ariani.
- **M1 Felipe Ariani 2022**
Mémoire en M1 (Aix-Marseille Université) intitulé : « Composition musicale assistée par ordinateur ».
Co-encadrement à **50%** avec Javier Elipe Gimenez (maître de conférences AMU à PRISM) et Marcelo Caetano (post-doc à PRISM).
- **M1 Clément Delarbre 2022**
Mémoire en M1 (Aix-Marseille Université) intitulé : « Recherche sur l'atmosphère au cinéma ».
Co-encadrement à **50%** avec Natacha Cyrulnik et Rémi Adjiman (maîtres de conférences AMU à PRISM).
- **Thèse Laura Potrovic 2014 - 2017**
« Ce qu'un corps peut devenir : cartographie entre danse et philosophie »
Laura Potrovic a fait une thèse sur de la phénoménologie du corps en mouvement.
Co-encadrement de l'ensemble de sa thèse à **70%**, avec Josette Féral (université Paris 3) et Sibila Petlevski (université de Zagreb).
Membre du jury.

- **Thèse Carla Pessolano 2014 - 2015**
« **Subjectivité poétique, credo poétique et résistance dans les pratiques de la scène** »
Co-encadrement de sa première année à **80%** avec Paola Ranzini (Laboratoire ACTE).
- **Thèse Julie Talland 2014 - 2015**
« **La numérimorphose dans le spectacle vivant : des prothèses numériques aux corps synesthètes** »
Co-encadrement de sa première année à **80%** avec Paola Ranzini (Laboratoire ACTE).
- **Thèse Virginie Vincent 2013 - 2015**
« **Le motif relationnel de la déspatialisation du théâtre** »
Co-encadrement de ses deux premières années à **80%** avec Paola Ranzini (Laboratoire ACTE).
- **Thèse Simona Polvani 2013 - 2015**
« **La performativité dans l'œuvre et la pratique artistique de Gao Xingjian** »
Co-encadrement de ses deux premières années à **60%**, avec Pascale Weber (université Paris 1) et Gerardo Guccini (université de Bologne).
- **Thèse Belkacem Fergani 2003 - 2004**
« **Utilisation des approches temps-fréquence en indexation sonore** »
Encadrement de sa première année à **100%**.
- **Post-doc David Janiszek 2002 - 2004**
Sur le thème de l'identification et de la vérification automatique des langues.
Encadrement à **60%** avec François Pellegrino (Laboratoire DDL).
- **Thèse Johnny Mariethoz 2002 - 2004**
« **Algorithmes d'apprentissage discriminants en vérification du locuteur** »
Encadrement de ses deux premières années à **80%** avec Hélène Paugam-Moisy (Institut de Sciences Cognitives de Lyon).

- **Thèse Michel Dutat 1997 - 2000**
« Utilisation des approches statistiques pour l'identification de la langue »
Encadrement de l'ensemble de sa thèse à **80%** avec Frédéric Bimbot (IRISA).
Membre du jury.
- **DEA (M2) Abdenour Bounsiar 2002 - 2003**
Mémoire de DEA STIR (Université Rennes 1) intitulé : « Etude d'un modèle intonatif pour la vérification automatique de la langue ».
Encadrement à **100%**.
- **DEA (M2) Abderrezak Rachedi 2002 - 2003**
Mémoire de DEA Informatique (Université de Savoie) intitulé : « Suivi de la parole dans les documents sonores ».
Encadrement à **100%**.
- **DEA (M2) Yannick Uebel 2002 - 2003**
Mémoire de DEA de sciences du langage (université Lyon 2) intitulé : « Comment incorporer des informations linguistiques et cognitives dans des systèmes de traitement automatique de la langue des signes française : une étude préliminaire ».
Co-encadrement à **80%** avec Colette Grinevald (Laboratoire DDL).
- **Master of Sciences (M2) Regunathan Radhakrishnan 2001 - 2002**
Travail de master of sciences (équivalent DEA) en traitement du signal et des images (Polytechnic Institute of Brooklyn, New York, Etats-Unis) intitulé : « Multimedia Indexing of Video Documents ». Encadrement à **100%**.
- **Maîtrise (M1) Karine Grande 2002 - 2003**
Mémoire de maîtrise de sciences cognitives (Université Lyon 2) intitulé : « Dénomination d'objets manipulables en langue des signes française ».
Co-encadrement à **70%** avec Colette Grinevald (Laboratoire DDL).

- **Maîtrise (M1) Tristan Cénier 2001 - 2002**

Mémoire de maîtrise de Biologie Cellulaire et Physiologie (Université Lyon 1) intitulé : « Recherche d'une capacité cérébrale à traiter spécifiquement les langues naturelles par les potentiels évoqués ».

Co-encadrement à **20%** avec Fanny Meunier (laboratoire DDL) et Mélissa Barkat (Laboratoire DDL).

- **Fin d'études ingénieur ENSERG (M1) Lional Aguilhon 2001 - 2002**

Mémoire de fin d'études de cycle ingénieur à l'ENSERG (Université Joseph Fournier à Grenoble) intitulé : « Comment exploiter une grande quantité de données en vérification du locuteur ».

Encadrement à **100%**.

- **Licence (L3) Leonid Synuykov 2001 - 2002**

Mémoire de licence de Mathématiques Appliquées aux Sciences Humaines et Sociales (Université Lyon 2) intitulé : « Restauration cognitive du langage perçu à l'envers ».

Co-encadrement à **30%** avec Fanny Meunier (Laboratoire DDL) et Mélissa Barkat (Laboratoire DDL), particulièrement sur la partie informatique du stage.

1.10 Publications et communications scientifiques

Inédit présenté en vue de l'HDR

1. Ivan Magrin-Chagnolleau. *La méthode de l'Actors Studio*. À paraître en 2024.

Directions de numéros de revues avec comité de lecture

Le journal *p-e-r-f-o-r-m-a-n-c-e* est le journal que j'ai créé et dont je suis le directeur de publication.

1. *p-e-r-f-o-r-m-a-n-c-e*, 7 : *Art and Human*, 2023. Co-édité avec Fabrice Métais.
2. *p-e-r-f-o-r-m-a-n-c-e*, 6 : *Creation Research*, 2022.
3. *p-e-r-f-o-r-m-a-n-c-e*, 5 : *Healing Art*, 2019. Co-édité avec Michel Sicard.
4. *p-e-r-f-o-r-m-a-n-c-e*, 4 : *The Creative Process*, 2017.
5. *p-e-r-f-o-r-m-a-n-c-e*, 3 : *Cuisine & Performance*, 2016. Co-édité avec Allen Weiss.
6. *p-e-r-f-o-r-m-a-n-c-e*, 2 : *Theatre Education*, 2015. Co-édité avec Joëlle Aden.
7. *p-e-r-f-o-r-m-a-n-c-e*, 1 : *The Artist Researcher*, 2014. Co-édité avec Ludivine Allegue.

Chapitres de livres

1. Ivan Magrin-Chagnolleau. **Expérience phénoménologique et énaïtive d'une œuvre de poésie spatialisée visuelle et sonore.** Dans *Corporéité et construction du sens dans la poésie visuelle : énaïtion, perception, expérimentation*. Sous presse.
2. Ivan Magrin-Chagnolleau. **La méthode de l'Actors Studio : Genèse, Pratique et Phénoménologie.** Dans *Être en jeu, être en scène*, édité par Laurent Berger et Arianna de Sanctis. Presses universitaires de la Méditerranée. Sous presse.
3. Ivan Magrin-Chagnolleau. **Chance and the Creative Process.** Dans *Chance, Calculation and Life*, édité par Thierry Gaudin, Marie-Christine Maurel, Jean-Charles Pomerol. ISTE Éditions, pp. 169-183, 2021.
4. Ivan Magrin-Chagnolleau. **Actors Studio.** *Encyclopædia Universalis*, 2021.
5. Ivan Magrin-Chagnolleau. **Hasard et Création.** Dans *Le hasard, le calcul et la vie*, édité par Thierry Gaudin, Marie-Christine Maurel, Jean-Charles Pomerol. ISTE Éditions, Collection « Systèmes d'information, web et société », pp. 161-175, 2021. ISBN : 978-1-78405-724-4.
6. Ivan Magrin-Chagnolleau, Ruth Nahoum, Marcus Weisen. **Le projet sipapu : vers une micro-phénoménologie de la création.** Dans *Les enjeux cognitifs de l'artefact esthétique : Tome 2*, édité par Xavier Lambert. L'Harmattan, Collection « Ouverture Philosophique », pp. 187-194, 2019. ISBN : 978-2-343-18776-1.

7. Ivan Magrin-Chagnolleau. **L'énaction dans la création artistique : théâtre, cinéma et performance.** Dans *Action, Énaction*, édité par Xavier Lambert. L'Harmattan, Collection « Ouverture Philosophique », pp. 213-224, 2017. ISBN : 978-2-343-11403-3.
8. Ioana Zamfir et Ivan Magrin-Chagnolleau. **The Madness of Making Sense : Madness and Integrative Knowledge.** Dans *Schizo : The Liberatory Potential of Madness*, édité par Irina Lyubchenko et Fiona Ann Papps, IDN Press, pp. 155-163, 2016. ISBN : 978-1-848-88460-1.
9. Anne Sachs, Ivan Magrin-Chagnolleau. **Du chaos à la création : entretien avec Ivan Magrin-Chagnolleau.** Dans *Danse Éducation : Échos issus du congrès IDEA Paris 2013*, pp. 46-50, 2016. ISBN : 978-2-951-99010-4.
10. Ivan Magrin-Chagnolleau. **Réflexions autour de l'expertise d'une œuvre artistique.** Dans *Expertise artistique*, édité par Maryvonne Saison. Ministère de la Culture, 2015.
11. Ivan Magrin-Chagnolleau. **Évaluation des systèmes de vérification du locuteur.** Dans *Évaluation des systèmes de traitement de l'information*. Édité par Stéphane Chaudiron. Hermès, 2004. ISBN : 978-2-746-20862-9.

Articles de revues avec comité de lecture

Hors journal *p-e-r-f-o-r-m-a-n-c-e*

1. Magali Boutrais, Ivan Magrin-Chagnolleau, Maryse Maurel, Patricia Rottement. **Le rôle du corps dans la mise en évocation, le remplissage intuitif et la présentification dans le cadre d'un entretien d'explicitation.** *Explicititer*, 137, pp. 37-41, 2023.
2. Felipe Ariani, Marcelo Caetano, Javier Elipe Gimeno, Ivan Magrin-Chagnolleau. **Créativité assistée par ordinateur : composer la musique d'un film en utilisant uniquement sa courbe de luminosité extraite automatiquement.** *Arts et Sciences*, Vol. 7, No. 1, pp. 12-21, 2023.

3. Ivan Magrin-Chagnolleau. **Mise en évocation : Entretien d'explicitation et méthode de l'Actors Studio.** *Explicititer*, Hors-série (1), pp. 53-61, 2023.
4. Ivan Magrin-Chagnolleau. **Le rôle de l'auto-explicitation en recherche.** *Explicititer*, Hors-série (1), pp. 105-111, 2023.
5. Ivan Magrin-Chagnolleau. **Reflections on the impact of creation research, practice-based research, and practice as research, and how to define and maximize that impact.** *Media Practice and Education*, 2022, 23(2), pp. 105-111.
6. Ivan Magrin-Chagnolleau. **The Relationship between Photo Retouching and Color Grading.** *Color Culture and Science Journal*, 2022, 14(1), pp. 79-86.
7. Magali Boutrais, Marie-Hélène Lachaud, Jocelyne LeBlanc, Ivan Magrin-Chagnolleau, Anne Philippon. **Pratiquer l'entretien d'explicitation à distance : petit ABC.** *Explicititer*, 130, pp. 51-60, 2021.
8. Ivan Magrin-Chagnolleau. **Futurs utopiques et dystopiques : comment la fiction nous aide à penser l'évolution.** *Arts et Sciences*, Vol. 5, No. 1, pp. 1-11, 2021.
9. Frédéric Bimbot et al. **A Tutorial on Text-Independent Speaker Verification.** *EURASIP Journal on Applied Signal Processing*. 2004 :4, 430–451.
10. Ivan Magrin-Chagnolleau, Geoffrey Durou, and Frédéric Bimbot. **Application of Time-Frequency Principal Component Analysis to Text-Independent Speaker Identification.** *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 5, July 2002.
11. Ivan Magrin-Chagnolleau and Geoffrey Durou. **Application of Time-Frequency Principal Component Analysis to Speaker Verification.** *Digital Signal Processing*, Vol. 10, No. 1-3, pp. 226-236,

January/April/July 2000.

12. The Elisa Consortium. **The ELISA Systems for the NIST'99 Evaluation in Speaker Detection and Tracking.** *Digital Signal Processing*, Vol. 10, No. 1-3, pp. 143-153, January/April/July 2000.
13. Frédéric Bimbot, Ivan Magrin-Chagnolleau, and Luc Mathan. **Second-Order Statistical Measures for Text-Independent Speaker Identification.** *Speech Communication*, 17(1-2), pp. 177-192, August 1995.

Dans le journal *p-e-r-f-o-r-m-a-n-c-e*

1. Rémi Adjiman, Rosine Bénard O'Kelly, Marine Brun-Fanzetti, Natacha Cyrulnik, Clément Delarbre, Javier Elipe-Gimeno, Agnès Horellou, Ivan Magrin-Chagnolleau, Jean-Pierre Moreau, Laetitia Petit, Isabelle Singer. **Approcher la notion d'atmosphère à travers le film *All These Creatures* de Charles Williams.** *p-e-r-f-o-r-m-a-n-c-e*, 6, 2022.
2. Annie Abrahams, Ivan Magrin-Chagnolleau, Alix de Morant, Daniel Pinheiro, Muriel Piqué. **Interview about Distant Movement (English version).** *p-e-r-f-o-r-m-a-n-c-e*, 6, 2022.
3. Annie Abrahams, Ivan Magrin-Chagnolleau, Alix de Morant, Daniel Pinheiro, Muriel Piqué. **Entretien au sujet de Distant Movement (version française).** *p-e-r-f-o-r-m-a-n-c-e*, 6, 2022.
4. Ivan Magrin-Chagnolleau. **Entretien avec Miryam Bouchard, réalisatrice.** *p-e-r-f-o-r-m-a-n-c-e*, 6, 2022.
5. Ivan Magrin-Chagnolleau. **Interview with Inesa Kurklietytė, Film Director.** *p-e-r-f-o-r-m-a-n-c-e*, 6, 2022.
6. Ivan Magrin-Chagnolleau. **Interview with Bragi Thor Hinriks-son, Film Director.** *p-e-r-f-o-r-m-a-n-c-e*, 6, 2022.
7. Ivan Magrin-Chagnolleau. **Théâtre, création et psychanalyse : autour de la mise en scène de Premier Amour de Beckett.** *p-*

e-r-f-o-r-m-a-n-c-e, 5, 2019.

8. Ivan Magrin-Chagnolleau. **Entretien avec le psychologue Nader Aghakhani.** *p-e-r-f-o-r-m-a-n-c-e*, 5, 2019.
9. Ivan Magrin-Chagnolleau. **Création, jeu et contemplation : l'art comme exercice spirituel.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.
10. Ivan Magrin-Chagnolleau. **La relation réalisateur - compositeur : un exemple de processus créatif relationnel éactif.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.
11. Ivan Magrin-Chagnolleau. **Le Processus Créatif de Robert Lepage.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.
12. Ivan Magrin-Chagnolleau. **Entretien avec le metteur en scène Robert Lepage.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.
13. Ivan Magrin-Chagnolleau. **L'acteur et les émotions au théâtre et au cinéma.** *p-e-r-f-o-r-m-a-n-c-e*, 3, 2016.
14. Ivan Magrin-Chagnolleau. **Rituals and Performance in Gastronomy.** *p-e-r-f-o-r-m-a-n-c-e*, 3, 2016.
15. Ivan Magrin-Chagnolleau. **Entretien avec le chef Bernard Bordaries.** *p-e-r-f-o-r-m-a-n-c-e*, 3, 2016.
16. Ivan Magrin-Chagnolleau. **L'Héritage de Lee Strasberg.** *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.
17. Ivan Magrin-Chagnolleau. **L'incorporation animale chez l'acteur.** *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.
18. Ivan Magrin-Chagnolleau. **Du chaos à la création : une expérience de pédagogie éactive pour expérimenter le processus créatif en performance.** *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.

19. Ivan Magrin-Chagnolleau. **In Praise of the Artist Researcher.** *p-e-r-f-o-r-m-a-n-c-e*, 1, 2014.
20. Ivan Magrin-Chagnolleau. **Éloge de l'artiste chercheur.** *p-e-r-f-o-r-m-a-n-c-e*, 1, 2014.

Conférences invitées

1. Ivan Magrin-Chagnolleau and Nathalie Parlangeau-Vallès. **Audio Indexing : What has been accomplished and the Road Ahead.** *Proceedings of JCIS 2002*, pp. 911-914, Durham, North Carolina, USA, March 2002 (Invited paper).

Colloques à comité de lecture

1. Ivan Magrin-Chagnolleau. **A Phenomenological Approach to Color Theory and Aesthetics.** *Proceedings of AIC 2022*, June 2022, Toronto, Canada.
2. Ivan Magrin-Chagnolleau. **Handling Color in Photography and Film : From Photo Retouching to Color Grading.** *Proceedings of AIC 2021*, August 2021, Milan, Italy.
3. Ioana Zamfir and Ivan Magrin-Chagnolleau. **The Madness of Making Sense : Madness as Perception of the Paradoxical Dimension of Reality.** *Proceedings of Madness : Probing the Boundaries*. September 2014.
4. Ivan Magrin-Chagnolleau. **The Use of Color in Theater and Film.** *Proceedings of AIC2013*.
5. Jean-François Bonastre, Frédéric Bimbot, Louis-Jean Boë, Joseph P. Campbell, Douglas A. Reynolds, and Ivan Magrin-Chagnolleau. **Person Authentication by Voice : A Need for Caution.** *Proceedings of EUROSPEECH 2003*, Geneva, Switzerland, September 2003.
6. Karine Grande, Ivan Magrin-Chagnolleau, Annie Risler, and Colette Grinevald. **La dénomination d'objets manipulables en LSF au**

sein d'un discours narratif : peut-on parler de classificateurs verbaux ? *Proceedings of Journées d'études internationales - La linguistique de la LSF : Recherches actuelles*, Lille, France, September 2003.

7. Nathalie Parlangeau-Vallès, Jérôme Farinas, Dominique Fohr, Irina Illina, Ivan Magrin-Chagnolleau, Odile Mella, Julien Piquier, Jean-Luc Rouas, and Christine Sénac. **Audio Indexing on the Web : A Preliminary Study of Some Audio Descriptors.** *Proceedings of SCI 2003*, Orlando, Florida, July 2003.
8. Daniel Moraru, Sylvain Meignier, Laurent Besacier, Jean-Francois Bonastre, and Ivan Magrin-Chagnolleau. **The ELISA Consortium Approaches in Speaker Segmentation During the NIST 2002 Speaker Recognition Evaluation.** *Proceedings of ICASSP 2003*, Hong Kong, China, April 2003.
9. Ivan Magrin-Chagnolleau, Melissa Barkat, and Fanny Meunier. **Intelligibility of Reverse Speech in French : A Perceptual Study.** *Proceedings of ICSLP 2002*, Denver, Colorado, USA, September 2002.
10. Sylvain Meignier, Jean-François Bonastre, and Ivan Magrin-Chagnolleau. **Speaker Utterances Tying Among Speaker Segmented Audio Documents Using Hierarchical Classification : Towards Speaker Indexing of Audio Documents.** *Proceedings of ICSLP 2002*, Denver, Colorado, USA, September 2002.
11. Fanny Meunier, Tristan Cénier, Melissa Barkat, and Ivan Magrin-Chagnolleau. **Mesure d'intelligibilité de segments de parole à l'envers en français.** *Proceedings of JEP 2002*, Nancy, France, June 2002.
12. Sylvain Meignier, Jean-François Bonastre, and Ivan Magrin-Chagnolleau. **Appariement de locuteurs entre des documents sonores préalablement segmentés en utilisant la classification hiérarchique.** *Proceedings of JEP 2002*, Nancy, France, June 2002.

13. Melissa Barkat, Fanny Meunier, and Ivan Magrin-Chagnolleau. **Intelligibility of Time-Reversed Speech in French.** *Workshop on Temporal Integration in the Perception of Speech*, Aix-en-Provence, France, April 2002.
14. Jean-François Bonastre, Ivan Magrin-Chagnolleau, Stefan Euler, François Pellegrino, Régine André-Obrecht, John Mason, and Frédéric Bimbot. **SPeaker and Language Characterization (SpLC) : A Special Interest Group (SIG) of ISCA.** *Proceedings of EUROSPEECH 2001*, Aalborg, Denmark, September 2001.
15. Ivan Magrin-Chagnolleau, Guillaume Gravier, and Raphaël Blouet for the ELISA consortium. **Overview of the 2000-2001 ELISA Consortium Research Activities.** *Proceedings of the Workshop 2001 : A Speaker Odyssey*, Chania, Crete, Greece, June 2001.
16. Mouhamadou Seck, Ivan Magrin-Chagnolleau and Frédéric Bimbot. **Experiments on Speech Tracking in Audio Documents using Gaussian Mixture Modeling.** *Proceedings of ICASSP 2001*, Salt Lake City, Utah, United States, May 2001.
17. Ivan Magrin-Chagnolleau, Guillaume Gravier, Mouhamadou Seck, Olivier Boeffard, Raphaël Blouet, and Frédéric Bimbot. **A Further Investigation on Speech Features for Speaker Characterization.** *Proceedings of ICSLP 2000*, Beijing, China, October 2000.
18. Michel Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot. **Language Recognition using Time-Frequency Principal Component Analysis and Acoustic Modeling.** *Proceedings of ICSLP 2000*, Beijing, China, October 2000.
19. Ivan Magrin-Chagnolleau and Geoffrey Durou. **Application of Vector Filtering to Pattern Recognition.** *Proceedings of ICPR 2000*, Barcelona, Spain, September 2000.
20. Ivan Magrin-Chagnolleau and Frédéric Bimbot. **Indexing Telephone Conversations by Speakers Using Time-Frequency Principal Component Analysis.** *Proceedings of ICME 2000*, New York, United States, July-August 2000.

21. Michel Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot. **Analyse en composantes principales temps-fréquence : application à la reconnaissance de la langue.** *Proceedings of JEP 2000*, Aussois, France, June 2000.
22. Ivan Magrin-Chagnolleau. **Utilisation du filtrage vectoriel en reconnaissance des formes.** *Proceedings of RFIA 2000*, Paris, France, February 2000.
23. Ivan Magrin-Chagnolleau and Richard G. Baraniuk. **Empirical mode decomposition based frequency attributes.** *Proceedings of the 69th SEG Meeting*, Houston, Texas, United States, November 1999.
24. Ivan Magrin-Chagnolleau, Hyeokho Choi, Rutger van Spaendonck, Philippe Steeghs and Richard G. Baraniuk. **Multiscale texture segmentation of dip-cube slices using Wavelet-domain hidden Markov trees.** *Proceedings of the 69th SEG Meeting*, Houston, Texas, United States, November 1999.
25. The Elisa Consortium. **The ELISA'99 Speaker Recognition and Tracking Systems.** *Proceedings of the 1999 IEEE Workshop on Automatic Identification Advanced Technologies*, Summit, New Jersey, United States, October 1999.
26. Ivan Magrin-Chagnolleau and Geoffrey Durou. **Time-Frequency Principal Components of Speech : Application to Speaker Identification.** *Proceedings of EUROSPEECH 99*, Vol. 2, pp. 759-762, Budapest, Hungary, September 1999.
27. Ivan Magrin-Chagnolleau, Aaron E. Rosenberg, and S. Parthasarathy. **Detection of Target Speakers in Audio Databases.** *Proceedings of ICASSP 99*, Vol. 2, pp. 821-824, Phoenix, United States, March 1999.

28. John Choi, Don Hindle, Julia Hirschberg, Ivan Magrin-Chagnolleau, Christine Nakatani, Fernando Pereira, Amit Singhal, and Steve Whittaker. **SCAN - Speech Content Based Audio Navigator : A System Overview**. *Proceedings of ICSLP 98*, Sydney, Australia, December 1998.
29. Aaron E. Rosenberg, Ivan Magrin-Chagnolleau, S. Parthasarathy, and Qian Huang. **Speaker Detection in Broadcast Speech Databases**. *Proceedings of ICSLP 98*, Sydney, Australia, December 1998.
30. John Choi, Don Hindle, Julia Hirschberg, Ivan Magrin-Chagnolleau, Christine Nakatani, Fernando Pereira, Amit Singhal, and Steve Whittaker. **An Overview of the AT&T Spoken Document Retrieval System**. *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, United States, December 1998.
31. Ivan Magrin-Chagnolleau, Joachim Wilke, and Frédéric Bimbot. **A Further Investigation on AR-Vector Models for Text-Independent Speaker Identification**. *Proceedings of ICASSP 96*, Vol. 1, pp. 401-404, Atlanta, United States, May 1996.
32. Ivan Magrin-Chagnolleau, Jean-François Bonastre, and Frédéric Bimbot. **Effect of Utterance Duration and Phonetic Content on Speaker Identification using Second-Order Statistical Methods**. *Proceedings of EUROSPEECH 95*, Vol.1, pp. 337-340, Madrid, Spain, September 1995.

Communications non publiées

1. Ivan Magrin-Chagnolleau. « **L'Émergence Énactive du Langage Poétique** ». Conférence internationale « *Langage et Énaction* », organisée du 1er au 3 juin 2016 à Clermont-Ferrand, France.
2. Ivan Magrin-Chagnolleau. « **Vers une phénoménologie de l'émergence en improvisation et composition musicales** ». Conférence internationale « *Émergence en Musique* », organisée du 9 au 11 mai 2016 à Plaisir-Grignon, France.

3. Ivan Magrin-Chagnolleau. « **Oser l'innovation en pédagogie : une expérience de pédagogie énaactive au cœur de l'université** » dans le cadre de la conférence internationale « *Éduquer et former au monde de demain* », organisée du 6 au 7 avril 2016 à Chamalières, France.
4. Ivan Magrin-Chagnolleau. « **Towards a Phenomenology of Creation** ». International Symposium « *Practice Research Symposium* », organized on February 5th, 2016, in Plymouth, UK.
5. Ivan Magrin-Chagnolleau. « **Towards a Neurophenomenological Approach of Joint Improvisation** » dans le cadre du colloque international « *The Science of Joint Improvisation* », organisé du 04 et 06 novembre 2015.
6. Ivan Magrin-Chagnolleau. « **Performance et neurosciences : une rencontre enfin possible** » dans le cadre de la journée d'étude « *Neuroscènes* », organisée à la MSH Paris Nord le 26 octobre 2015.
7. Ivan Magrin-Chagnolleau. « **Improviser le lâcher prise** » dans le cadre du colloque international « *En Chair et en Son : rencontres internationales Butō et musique acousmatique* », organisé les 23 et 24 octobre 2015.
8. Ivan Magrin-Chagnolleau. « **Révolutionner les regards par l'empathie au cinéma** » dans le cadre de la journée d'étude « *Les figures artistiques, littéraires et cinématographiques à l'œuvre dans les révolutions* », organisé par Sylvie Dallet à la Maison des Sciences de l'Homme le 7 octobre 2015 dans le cadre de son séminaire « *Éthiques et Mythes de la Création* ».
9. Ivan Magrin-Chagnolleau. « **Être authentique et vivant, héritage du Living Theater** » dans le cadre de la journée d'étude « *L'Héritage critique du Living Theater en Europe* », organisée par Emmanuel Wallon au 104 le 13 juin 2015 dans le cadre des chantiers d'Europe 2015 et du workshop européen autour de la critique multidisciplinaire.

10. Ivan Magrin-Chagnolleau. « **Les mythes de la terre-merre et de l'arbre de vie au cinéma** » dans le cadre de la journée d'étude « *Les arts au chaudron des mythes* », organisé par Sylvie Dallet à la Maison des Sciences de l'Homme le 3 juin 2015 dans le cadre de son séminaire « *Éthiques et Mythes de la Création* ».
11. Ivan Magrin-Chagnolleau. « **Création éactive** » et présentation d'un entretien avec Michel Le Van Quyen dans le cadre de la journée « *Sticky Recipe* » qui fait partie de « *La Psyché de l'univers - Hommage à Francisco J. Varela* », une proposition d'art recherche de Raphaële Jeune en février 2015 à Rennes.
12. Ivan Magrin-Chagnolleau et Christine Leroy. « **De-frag-men-tation : New Technologies and the Moving Body** ». Conférence invitée dans le cadre de la journée « *Technology and Performance* » à Stanford University le 28 février 2014.
13. Ivan Magrin-Chagnolleau. « **Art & Spirituality** ». Conférence invitée dans le cadre du symposium « *Art & Consciousness* » organisé à Aarhus le 26 novembre 2013.
14. Ivan Magrin-Chagnolleau. « **Jouer (avec) l'animal** ». Colloque « *Performances et animalités* » organisé au musée de la Chasse et de la Nature à Paris du 13 au 14 novembre 2013.
15. Ivan Magrin-Chagnolleau. « **La question migratoire et identitaire dans 'Cœur' de Robert Lepage** ». Colloque « *Créations géoartistiques en Méditerranée : Frontières et mobilités* » organisé à Gabès, Tunisie, du 8 au 10 novembre 2013.
16. Ivan Magrin-Chagnolleau. « **The Impact of Contemplative Practices on Creativity and the Creative Process** ». European Symposium of the Mind & Life Institute on « *Personal and Societal Change from the Contemplative Perspective* » organisé à Berlin du 10 au 13 octobre 2013.

1.11 Critiques

Critique de cinéma

- Membre du SFCC (Syndicat français de la critique de cinéma) et de la FIPRESCI (Fédération internationale des critiques de films)
- Membre du jury de la critique au festival Off-Courts de Trouville en 2018
- Membre du jury de la critique au festival Schlingel des films pour la jeunesse à Chemnitz en Allemagne en 2018, 2019 et 2021
- Membre du jury de la critique pour le prix du meilleur film arabe de l'année en 2020, 2021 et 2022
- Membre du jury pour le prix SFCC du meilleur premier film français et du meilleur premier film étranger de l'année en 2019
- Présentateur de la WebTV au festival Off-Courts de Trouville en 2019, 2020 et 2021

Critiques

- **Aftersun** by Charlotte Wells. *p-e-r-f-o-r-m-a-n-c-e*, 2023.
- **Official Competition** by Mariano Cohn and Gastón Duprat. *p-e-r-f-o-r-m-a-n-c-e*, 2022.
- **The Story of my Wife** by Ildikó Enyedi. *p-e-r-f-o-r-m-a-n-c-e*, 2022.
- **Illusions perdues** by Xavier Gianoli. *p-e-r-f-o-r-m-a-n-c-e*, 2021.
- **My Very Own Circus** by Miryam Bouchard. *p-e-r-f-o-r-m-a-n-c-e*, 2021.
- **Birta** by Bragi Thor Hinriksson. *p-e-r-f-o-r-m-a-n-c-e*, 2021.
- **Phantom Owl Forest** by Anu Aun. *p-e-r-f-o-r-m-a-n-c-e*, 2019.

- **Spread Your Wings** by Nicolas Vanier. *p-e-r-f-o-r-m-a-n-c-e*, 2019.
- **Nevada** by Laure de Clermont-Tonnerre. *p-e-r-f-o-r-m-a-n-c-e*, 2019.
- **Sun** by Jonathan Desoindre. *p-e-r-f-o-r-m-a-n-c-e*, 2019.
- **Meltem** by Basile Doganis. *p-e-r-f-o-r-m-a-n-c-e*, 2019.
- **Nice Girls Don't Stay for Breakfast** by Bruce Weber. *p-e-r-f-o-r-m-a-n-c-e*, 2019.
- **La Chute de l'Empire Américain** by Denys Arcand. *p-e-r-f-o-r-m-a-n-c-e*, 2019.
- **Meerkat Moonship** by Haneke Schutte. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **Pony and Birdboy** by Mari Rantasila. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **The Big Bath** by Gilles Lellouche. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **Little Tickles** by Andréa Bescond and Éric Métayer. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **Happy as Lazzaro** by Alice Rohrwacher. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **Image Book** by Jean-Luc Godard. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **Yomeddine** by A.B. Shawky. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **Rafiki** by Wanuri Kahiu. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **Wildlife** by Paul Dano. *p-e-r-f-o-r-m-a-n-c-e*, 2018.
- **Loveless (Faute d'amour)** by Andrey Zvyagintsev. *p-e-r-f-o-r-m-a-n-c-e*, 2017.

- **It's Only the End of the World** by Xavier Dolan. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **I, Daniel Blake** by Ken Loach. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **Loving** by Jeff Nichols. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **American Honey** by Andrea Arnold. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **Hands of Stone** by Jonathan Jakubowisz. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **Timecode** by Juanjo Gimenez. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **Notre Petite Sœur** de Kore-Eda Hirokazu. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **Mia Madre** de Nanni Moretti. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **Youth** de Paolo Sorrentino. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **Au delà des montagnes** de Jia Zhang-Ke. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **La Glace et le Ciel** de Luc Jacquet. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **Magic in the Moonlight** de Woody Allen. *p-e-r-f-o-r-m-a-n-c-e*, 2014.
- **Rétrospective Sergio Leone à la Cinémathèque Française.** *p-e-r-f-o-r-m-a-n-c-e*, 2014.
- **Birdman** d'Alejandro González Iñárritu. *p-e-r-f-o-r-m-a-n-c-e*, 2014.
- **À 14 Ans** d'Hélène Zimmer. *p-e-r-f-o-r-m-a-n-c-e*, 2014.

Critique de spectacles

- **Acte sans paroles** de Dominique Dupuy. *p-e-r-f-o-r-m-a-n-c-e*, 2014.

Critique musical

- **Karine Deshayes, Delphine Haidan & Thomas Palmer.** *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **Trio SR9.** *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **Orchestre National d'Île-de-France.** *p-e-r-f-o-r-m-a-n-c-e*, 2016.

Critique d'art

- **Les Luminales** de Maryza Jonath et Jean-Pierre Quéryn. *p-e-r-f-o-r-m-a-n-c-e*, 2016.

Recensions de livres

- **Crowdfunding for Filmmakers, 2nd Edition** by John T. Trigonis. *p-e-r-f-o-r-m-a-n-c-e*, 2017.
- **Making the Magic Happen** by Peter D. Marshall. *p-e-r-f-o-r-m-a-n-c-e*, 2017.
- **Jazz Vocal Improvisation** by Mili Bermejo. *p-e-r-f-o-r-m-a-n-c-e*, 2017.
- **Being Professional** by Adam Coplan. *p-e-r-f-o-r-m-a-n-c-e*, 2017.
- **Audio Post Production for Film and Television** by Mark Cross. *p-e-r-f-o-r-m-a-n-c-e*, 2017.
- **Making It Big in Shorts** by Kim Adelman. *p-e-r-f-o-r-m-a-n-c-e*, 2017.
- **Berklee Jazz Standards For Solo Piano** by Robert Christopherson, Hey Rim Jeon, Ross Ramsay, Tim Ray. *p-e-r-f-o-r-m-a-n-c-e*, 2017.
- **Beat by Beat : A Cheat Sheet for Screenwriters** by Todd Klick. *p-e-r-f-o-r-m-a-n-c-e*, 2017.

- **The Coffee Break Screenwriter, 2nd Edition** by Pilar Alessandra. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **You Can Act on Camera** by D.W. Brown. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **The Director's Six Senses** by Simone Bartesaghi. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **Master of the Cinematic Universe** by John Bucher and Jeremy Casper. *p-e-r-f-o-r-m-a-n-c-e*, 2016.
- **Why Does The Screenwriter Cross The Road** by Joe Gilford. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **Notes to Screenwriters** by Vicki Peterson and Barbara Nicolosi. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **Pour Roland Barthes** de Chantal Thomas. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **The Authentic Actor** by Michael Laskin. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **Film+Video Budgets** by Maureen A. Ryan. *p-e-r-f-o-r-m-a-n-c-e*, 2015.
- **The Lee Strasberg Notes** de Lola Cohen. *p-e-r-f-o-r-m-a-n-c-e*, 2014.
- **Perfecting Your English Pronunciation** by Susan Cameron. *p-e-r-f-o-r-m-a-n-c-e*, 2014.

1.12 Brevets

- Qian Huang, Ivan Magrin-Chagnolleau, Sarangarajan Parthasarathy, and Aaron Edward Rosenberg. **Multimedia Search Apparatus and Method for Searching Multimedia Content Using Speaker Detection by Audio Data**. United States Patent 6 405 166, 2002.

- Qian Huang, Ivan Magrin-Chagnolleau, Sarangarajan Parthasarathy, and Aaron Edward Rosenberg. **Multimedia Search Apparatus and Method for Searching Multimedia Content Using Speaker Detection by Audio Data.** United States Patent 6 317 710, 2001.

1.13 Logiciels

- **Modélisation d'une séquence de vecteurs par un mélange de gaussiennes.**
Développé en C++.
- **Empirical mode decomposition and Hilbert spectrum.**
Développé sous Matlab en collaboration avec Richard Baraniuk de Rice University, Houston, Texas, Etats Unis.
- **Filtrage vectoriel d'une séquence de vecteurs.**
Développé en C++.

1.14 Publications et productions artistiques

Livres et catalogues artistiques

- Ivan Magrin-Chagnolleau and Roberto Pieraccini. *Dptych*. Aloha Edition, 2023. Photographie.
- Ivan Magrin-Chagnolleau. *Haiku 1*. Aloha Edition, 2023. Poésie et photographie.
- Ivan Magrin-Chagnolleau. *Handscapes*. Aloha Edition, 2020. Photographie.
- Ivan Magrin-Chagnolleau. *Choices*. Aloha Edition, 2019. Édition d'un recueil de pièces de théâtre produites et mises en scène à New York en 2017. Contient la pièce *Missed Opportunities*.
- Ivan Magrin-Chagnolleau. *Muir Woods Spirits*. Aloha Edition, 2017. Photographie.
- Sylvie Dallet, Marie-José Geffray, Ivan Magrin-Chagnolleau, Gilbert Schoon. *Les Arts ForeZtiers : Catalogue de l'édition 2015*. Aloha Edition et Institut Charles Cros, 2015.

Théâtre, opéras et comédies musicales

- **2023 Rencontres intimes avec l'anthropocène (plusieurs auteurs)**
Producteur, Metteur en scène, Lumières, Costumes
Théâtre forum
Dans le cadre de mon travail au sein de l'équipe Arts&Sciences de labos1point5.

- **2020 Deliciae (Jonathan Bell)**
Ténor soliste
Théâtre musical
Production Festival Viva Villa et Casa Velasquez - Avignon

- **2019 Mit Allen Augen (Jonathan Bell)**
Ténor soliste
Théâtre musical
Production CMMR 2019 - Marseille

- **2017 Choices (Plusieurs auteurs)**
Producteur, Metteur en Scène, Lumières, Costumes
Théâtre
Production Compagnie Aloha - New York
Au bénéfice de l'association Life for Paris qui s'occupe des victimes des attentats terroristes du 13 novembre 2015

- **2013 Premier Amour (Samuel Beckett)**
Producteur, Metteur en Scène, Lumières, Costumes
Théâtre
Production Compagnie Aloha - Paris
Au bénéfice de la recherche sur la moelle épinière et de l'institut des jeunes aveugles

- **2013 Cœur (Robert Lepage & Co.)**
Observateur
Théâtre - Mise en Scène de Robert Lepage
Production Ex Machina - Québec

- **2011 Contes magiques et écologiques (Eliane Arnoldy)**
Producteur, Metteur en Scène, Costumes
Théâtre
Production Compagnie Aloha - Ville de Courbevoie

- **2011 The Blue Room (David Hare)**
Coproducteur, Acteur
Théâtre - Tous les rôles masculins - Mise en Scène Colin Reese
Production Compagnie Aloha - Words Up! - Paris

- **2010 La leçon (Eugène Ionesco)**
Producteur, Metteur en Scène, Acteur
Théâtre - Rôle principal du professeur
Production Compagnie Aloha - Petit théâtre odyssee
Ville de Levallois-Perret

- **2010 Du chaos à la création**
Metteur en Scène
Théâtre expérimental
Production Université d'Evry - Théâtre de l'Agora - Evry

- **2010 Zoroastre (Jean-Philippe Rameau)**
Lumières
Opéra - Mise en scène de Renaud-Marie Leblanc
Production Baroques Graffiti - Théâtre de Lenche - Marseille

- **2009 Le choix d'Hercule (Hippolyte Wouters)**
Coproducteur, Metteur en Scène
Opéra-Théâtre - Création française
Production Compagnie Aloha - Baroques Graffiti - Opéra 84
Festival côté cour - Aix-en-Provence

- **2009 Il Trovatore (Giuseppe Verdi)**
Observateur
Opéra - Mise en Scène d'Elijah Moshinsky, Direction musicale de Carlo Rizzi - Roberto Alagna & Sondra Radvanovsky
Production Royal Opera House - London

- **2008 King Arthur (Henry Purcell)**
Observateur
Opéra
Mise en Scène de Corinne and Gilles Benizio, Direction musicale d'Hervé Niquet
Production Opéra of Montpellier

- **2007 The Tetralogy (Richard Wagner)**
Observateur
Opéra
Mise en Scène de Robert Wilson, Direction musicale de Christoph Eischenbach
Production Châtelet

- **2007 Uncle (Jonny & Troy Schremmer)**
Acteur
Théâtre - Rôle principal de Joe, le mari peu communicant
Mise en Scène Ron Leibman
Production Actors Studio Drama School - New York

- **2007 151 Cabaret**
Ténor soliste
Cabaret Comédies Musicales
Interprète plusieurs airs de Miss Saigon et The Secret Garden
Mise en scène Keith Buhl
Production Actors Studio Drama School - New York

- **2007 Cuba, Si! (Terrence McNally)**
Metteur en Scène
Théâtre
Production Random Acts – Actors Studio Drama School - New York

- **2007 Death and the Maiden (Susan Kim)**
Acteur
Théâtre - Rôle principal de Lawrence, messenger de mort
Mise en Scène Brian Cichocki
Production Random Acts – Actors Studio Drama School - New York

- **2007 The Odyssey (Mary Zimmerman)**
Acteur
Théâtre - Rôles de Hermès et Alcinous - Mise en Scène Allegra Pickett
Production Actors Studio Drama School - New York

- **2006 Christmas at Sweet Rhythm**
Chanteur
Cabaret Jazz - Interprète plusieurs standards de Jazz - Mise en scène Kamal Scott
Production Actors Studio Drama School & New School for Jazz and Contemporary Music - New York

- **2006 Black Snow (Keith Reddin)**
Acteur
Théâtre - Plusieurs rôles - Mise en Scène Glynis Rigsby
Production Actors Studio Drama School - New York
- **2006 An Honest Arrangement (David Wiener)**
Metteur en Scène, Lumières, Costumes
Théâtre
Production The Mercy Plays – Center Stage - New York
Au bénéfice de l'association Mercy Ships
- **2006 Hill (Amy Schultz)**
Metteur en Scène, Lumières, Costumes
Théâtre - Création
Production Actors Studio Drama School - New York
- **2006 Appetite (Caren Skibell)**
Acteur
Théâtre - Création - Rôle du Dr. Horst, psychologue pour chiens
Mise en Scène Layne Racowsky
Production Actors Studio Drama School - New York
- **2006 An Honest Arrangement (David Wiener)**
Metteur en Scène, Lumières, Costumes
Théâtre - Création newyorkaise
Production American Globe Theater, New York
Prix de la meilleure pièce et du meilleur acteur
- **2006 Hospice (Pearl Cleage)**
Metteur en Scène, Lumières, Costumes
Théâtre
Production Actors Studio Drama School - New York
- **2005 Exile Here (Paul DeSenna)**
Metteur en Scène, Lumières, Costumes
Théâtre - Création
Production Actors Studio Drama School - New York

- **2005 The Just : Act IV (Albert Camus)**
Metteur en Scène, Lumières, Costumes
Théâtre
Production Bank Street Theater - New York

- **2005 The Blue Room (David Hare)**
Acteur
Théâtre - Rôle principal de l'aristocrate - Mise en Scène Marcia Haufrecht
Production Actors Studio Drama School - New York

- **2005 Found Objects (Carla Ching)**
1er assistant Mise en Scène
Théâtre - Création - Mise en Scène de Denyse Owens
Production Actors Studio Drama School - New York

- **2005 Athena (Laconia Koerner)**
1er assistant Mise en Scène
Théâtre - Création - Mise en Scène d'Alexandra Hoge
Production Actors Studio Drama School - New York

- **2004 The God of Hell (Sam Shepard)**
Assistant de plateau
Théâtre - Création
Mise en Scène de Lou Jacobs en présence de l'auteur, avec Tim Roth
Production Actors Studio Drama School - New York

- **2004 As You Like It (William Shakespeare)**
Assistant de plateau
Théâtre - Mise en Scène de Stuart Vaughan
Production Actors Studio Drama School - New York

- **2004 Diabolus In Musica**
Co-Auteur, Co-Metteur en Scène, Lumières, Costumes, Acteur
Comédie musicale - Création - Rôle principal du vagabond (ténor)
Production Compagnie Padrivantia – Théâtre des clochards célestes
- Lyon

- **2004 Chansons Divines et Démoniaques**
Co-Metteur en Scène, Lumières, Ténor soliste
Concert musical - Création
Production Compagnie Padrivantia – Théâtre des clochards célestes
- Lyon

- **2004 Les sœurs pâquerettes**
Metteur en Scène
Comédie musicale - Création - de et avec Myriam Masson et Marion Chomier
Tournée en France

- **1990-1996 Ut Musica Poesis**
Ténor soliste
Ensemble vocal dirigé par Frédéric Albou
Musique de la Renaissance et de la fin du Moyen Âge

- **1990-1993 Ensemble Françoise Herr**
Ténor soliste
Ensemble vocal dirigé par Françoise Herr
Musique baroque et création d'un opéra contemporain

- **Pièces de théâtre**
 - Somewhere in Minneapolis
 - Missed Opportunities (produite en 2017 et publiée en 2019)
 - Si seulement tu pouvais m'entendre ...

Cinéma

- **2023 Rencontres intimes avec l’anthropocène**
Réalisateur, Acteur, Monteur, Étalonneur
Interprétation de l’un des récits
Production Aloha Films
- **2022 Walter**
Auteur, Réalisateur, Acteur, Monteur, Étalonneur
Rôle principal de Walter, le fantôme
Co-écrit et co-réalisé avec Emcy Ly
Production Studio Off-Courts
- **2021 C’est juste une p’tite ballade**
Producteur, Auteur, Réalisateur, Directeur de la photographie, Monteur, Étalonneur
Clip réalisé pour l’artiste Michel Delaigue
Production Aloha Films
- **2020 Emmanuelle**
Producteur, Auteur, Réalisateur, Directeur de la photographie, Monteur, Étalonneur
Clip réalisé pour l’artiste Archipol
Production Aloha Films
- **2020 Corps perdus 2**
Producteur, Auteur, Réalisateur, Directeur de la photographie, Acteur, Monteur, Étalonneur, Compositeur
Rôle du chamane - Co-écrit et co-réalisé avec les patients de l’hôpital de jour de Villiers-sur-Marne
Production Aloha Films
- **2019 Corps perdus 1**
Producteur, Auteur, Réalisateur, Directeur de la photographie, Acteur, Monteur, Étalonneur, Compositeur
Rôle du chamane - Co-écrit et co-réalisé avec les patients de l’hôpital de jour de Villiers-sur-Marne
Production Aloha Films

- **2019 Émotiononou**
Producteur, Auteur, Réalisateur, Directeur de la photographie, Acteur, Monteur, Étalonneur, Compositeur
Rôle du paranoïaque et rôle de l'homme heureux - Co-écrit et co-réalisé avec les patients de l'hôpital de jour de Villiers-sur-Marne
Production Aloha Films

- **2017 Colocs**
Producteur, Auteur, Réalisateur, Directeur de la photographie, Acteur, Monteur, Compositeur
Rôle principal de Pierre - Co-écrit et co-réalisé avec Florian Mazier
Production Aloha Films

- **2015 19H47**
Étalonneur
Écrit et réalisé par Philomène Chabert
Production École de la Cité

- **2013 Nomen Mysticum**
Producteur, Acteur, Monteur
Rôle principal d'Hacha - Écrit et réalisé par Florian Mazier
Production Aloha Films

- **2013 La Galette des Rois**
Producteur, Réalisateur, Acteur, Monteur, Compositeur
Rôle principal du SDF - Écrit par Vinciane Mokry
Production Aloha Films
20 sélections en festivals

- **2013 C'est la Fête !**
Producteur, Auteur, Réalisateur
Directeur de la photographie, Acteur, Monteur
Rôle principal de l'homme
Production Aloha Films
10 sélections en festivals

- **2012 Death of a Tune**
Acteur
Rôle principal de Grégory Dupont
Écrit et réalisé par Vinayak Radhakrishnan
Production EICAR, Paris

- **2009 El Hombre Muerto**
Acteur
Rôle principal de l'homme - Écrit et réalisé par Adriana Calvo
Production EICAR, Paris
- **2007 Never Forever**
Compositeur
Écrit et réalisé par Emmanuella Aristil
Production Brooklyn College, New York
- **2007 Romance Largetto**
Acteur, 1er Assistant Réalisateur
Rôle principal du mari - Écrit et réalisé par Pablo Goldberg
Production New School University, New York
- **2007 Wonderland**
1er Assistant Réalisateur
Écrit et réalisé par Carlos Leal
Production Les danseurs sénégalais, Belgique
- **2006 Mask Project Vol. 1**
Acteur
Rôle principal de l'homme - Écrit et réalisé par Jonathan Ashley
Production New York University, New York
- **2005 Ten Steps**
Acteur
Rôle principal de l'homme - Écrit et réalisé par Pablo Goldberg
Production New School University, New York
Sélection officielle du 11ème Los Angeles Latino International Film Festival
- **2005 Amniotic Trip**
Acteur
Rôle principal de l'homme - Écrit et réalisé par Pablo Goldberg
Production New School University, New York

- **2004 Un Après-Midi Ordinaire**

Auteur, Réalisateur, Directeur de la photographie, Acteur, Monteur

Rôle principal de l'homme

Production CNRS, Lyon, France

- **Scénarios**

Longs métrages

— Change Me Into Zeus's Daughter

(avec Barbara Robinette Moss)

— Les larmes de Saint-Pierre

— Saturnin

Séries

— Max and Nina

— Coma Man

— The French

— The Consultants

Courts métrages

— C'est la fête! (produit)

— Colocs (co-écrit avec Florian Mazier - produit)

— Missed Opportunities

— First Do No Harm

— L'Orgue

— Si seulement tu pouvais m'entendre ...

— L'attente

Compositions musicales et créations sonores

- **2023 Acousmatic #1**
Album regroupant plusieurs compositions de musique acousmatique
- **2019 Thinking Machine**
Morceau composé pour acousmonium dans le cadre du festival « En chair et en son : rencontres internationales Butō et musique acousmatique »
- **2018 Aerial 01**
Morceau composé pour acousmonium dans le cadre du festival « En chair et en son : rencontres internationales Butō et musique acousmatique »
- **2015 L’oracle des arbres**
Une installation sonore dans les arbres
Dans le cadre de la biennale d’art contemporain *Les Arts ForeZtiers 2015* organisée par Sylvie Dallet.
- **2015 Un corps de lumière au milieu des ténèbres**
Morceau composé pour acousmonium dans le cadre du colloque international et festival « En chair et en son : rencontres internationales Butō et musique acousmatique »
- **2015 Harmonie**
Musique pour le court métrage *Colocs*
- **2013 On The Streets**
Musique pour le court métrage *La Galette des Rois*
- **2010-2012 The Aquarians**
Diverses compositions / improvisations au sein du groupe *The Aquarians* formé avec Florian Mazier
- **2005 Never Forever**
Musique pour le court métrage *Never Forever*

Photographie

- **2023 Diptych** avec Roberto Pieraccini
Publication d'un livre de photographies avec l'intégralité de la série.
Aloha Edition.
- **2020 Handscapes**
Publication d'un livre de photographies avec l'intégralité de la série.
Aloha Edition.
- **2020 Haikus**
Série de photographies obtenues par transposition du concept de haiku à la photographie.
- **2020 Mutations**
Série de photographies obtenues par variations à partir d'une photo carrée d'un mur blanc.
- **2019 Kaléidophysis**
Série de photographies obtenues à partir de l'art vidéo *Kaléidophysis*.
- **2018 Croisés Boisés**
Série de photographies obtenues à partir de l'art vidéo *Croisés Boisés*.
- **2017 Muir Woods Spirits**
Publication d'un livre de photographies avec l'intégralité de la série.
Aloha Edition.
- **2015 Bois Résonnables**
Exposition de neuf photographies de la série « Muir Woods Spirit », de neuf photographies de la série « L'Esprit du Bois de Vincennes », de neuf photographies de la série « Bonsaï Spirit », de dix photographies de la série « L'Esprit du Forez ».
Dans le cadre de la biennale d'art contemporain Les Arts ForeZtiers 2015 organisée par Sylvie Dallet.
- **2014 Entre-Mondes**
Exposition d'une photographie de la série « Muir Woods Spirit ».
Commissaire de l'exposition : Michel Sicard.
Patronage Laïque Jules Vallès à Paris

Performances

- **2016 Phytosophia**
Performance musicale et sonore proposée par Célio Paillard
Ivan Magrin-Chagnolleau, création sonore, Frédéric Mathevet,
création sonore et Oud, Célio Paillard, création sonore
Biennale d'art contemporain Les Arts ForeZtiers 2016.
- **2014 The Exorcist Antiphon Dub**
Performance musicale et sonore proposée par Frédéric Mathevet
Ivan Magrin-Chagnolleau, voix, Frédéric Mathevet, piano, Denis
Bernardi, basse
École Louis Lumière - Cité du Cinéma à Saint-Denis
- **2014 De-frag-men-ta-tion #3**
Performance de danse vidéo
Ivan Magrin-Chagnolleau, vidéo, Christine Leroy, danse
Stanford University, Californie

Art Vidéo

- **2016 Zodiaque poétique : Bélier**
Montage vidéo et sonore autour de la poétique du zodiaque :
exploration du bélier
Dans le cadre du colloque international Vagabonds des Étoiles
organisé par l'université de Versailles Saint-Quentin et le CNES.
- **2016 Kaléidophysis**
Montage vidéo et sonore utilisant des images de nature et des effets
kaléidoscopiques
Dans le cadre de la biennale d'art contemporain Les Arts ForeZtiers
2016 organisée par Sylvie Dallet.
- **2015 Croisés Boisés**
Montage vidéo et sonore de cheminements dans le bois de Vincennes
et les monts du Forez
Dans le cadre de la biennale d'art contemporain Les Arts ForeZtiers
2015 organisée par Sylvie Dallet.

Chapitre 2

Quelques contributions majeures

2.1 Contributions majeures dans le domaine de la reconnaissance automatique du locuteur

La reconnaissance automatique du locuteur consiste à modéliser la parole d'un locuteur à l'aide d'un modèle le plus souvent statistique, puis à reconnaître (identifier ou vérifier) ce locuteur à partir d'un échantillon de sa parole.

Plus précisément, à partir d'échantillons de parole d'une personne, on extrait des paramètres (reposant le plus souvent sur une analyse de type temps-fréquence de ces échantillons telle qu'une analyse spectrale ou une analyse cepstrale). Puis, à partir de ces paramètres, on effectue une modélisation qui repose souvent sur des modèles statistiques (de type modèles multi-gaussiens ou encore de type modèles de Markov).

J'ai travaillé continuellement sur ce thème de 1993 à 2004. Cela a été notamment le sujet de ma thèse, mais aussi des années qui ont suivi, jusqu'à mon changement d'orientation thématique vers la philosophie de l'art et l'esthétique.

Mes contributions majeures dans ce domaine sont les suivantes :

- **Proposition d'un formalisme qui permet d'unifier toutes les approches de modélisation, que j'ai appelé « filtrage vectoriel », qui est une généralisation vectorielle du filtrage classique en traitement du signal. J'ai proposé également une façon d'utiliser ce filtrage vectoriel dans le cadre plus général de la reconnaissance des formes.**

- . Ivan Magrin-Chagnolleau. **Approches statistiques et filtrage vectoriel de trajectoires spectrales pour l'identification du locuteur indépendante du texte.** Thèse ENST, 1997.
- . Ivan Magrin-Chagnolleau. **Utilisation du filtrage vectoriel en reconnaissance des formes.** Proceedings of RFIA 2000, Paris, France, February 2000.
- . Ivan Magrin-Chagnolleau and Geoffrey Durou. **Application of Vector Filtering to Pattern Recognition.** Proceedings of ICPR 2000, Barcelona, Spain, September 2000.

- **Dans le cadre de ce formalisme, proposition d'une nouvelle approche de modélisation, que j'ai appelée « Analyse en composantes principales temps-fréquence » et participation, avec Frédéric Bimbot, à l'établissement d'autres approches :**

- . Ivan Magrin-Chagnolleau. **Approches statistiques et filtrage vectoriel de trajectoires spectrales pour l'identification du locuteur indépendante du texte.** Thèse ENST, 1997.
- . Frédéric Bimbot, Ivan Magrin-Chagnolleau, and Luc Mathan. **Second-Order Statistical Measures for Text-Independent Speaker Identification.** Speech Communication, 17(1-2), pp. 177-192, August 1995.
- . Ivan Magrin-Chagnolleau, Joachim Wilke, and Frédéric Bimbot. **A Further Investigation on AR-Vector Models for Text-Independent Speaker Identification.** Proceedings of ICASSP 96, Vol. 1, pp. 401-404, Atlanta, United States, May 1996.
- . Ivan Magrin-Chagnolleau and Geoffrey Durou. **Time-Frequency Principal Components of Speech : Application to Speaker Identification.** Proceedings of EUROSPEECH 99, Vol. 2, pp. 759-762, Budapest, Hungary, September 1999.

- . Ivan Magrin-Chagnolleau, Guillaume Gravier, Mouhamadou Seck, Olivier Boeffard, Raphaël Blouet, and Frédéric Bimbot. **A Further Investigation on Speech Features for Speaker Characterization**. Proceedings of ICSLP 2000, Beijing, China, October 2000.
- . Ivan Magrin-Chagnolleau and Geoffrey Durou. **Application of Time-Frequency Principal Component Analysis to Speaker Verification**. Digital Signal Processing, Vol. 10, No. 1-3, pp. 226-236, January/April/July 2000.
- . Ivan Magrin-Chagnolleau, Geoffrey Durou, and Frédéric Bimbot. **Application of Time-Frequency Principal Component Analysis to Text-Independent Speaker Identification**. IEEE Transactions on Speech and Audio Processing, Vol. 10, No. 5, July 2002.

- **J'ai travaillé, avec Jean-François Bonastre, sur une méthode pour prendre en compte des connaissances linguistiques expertes dans les systèmes d'identification automatique du locuteur.**

- . Ivan Magrin-Chagnolleau, Jean-François Bonastre, and Frédéric Bimbot. **Effect of Utterance Duration and Phonetic Content on Speaker Identification using Second-Order Statistical Methods**. Proceedings of EUROSPEECH 95 , Vol.1, pp. 337-340, Madrid, Spain, September 1995.

- **Création en France, en collaboration avec Frédéric Bimbot et Jean-François Bonastre, d'un consortium de recherche fédérant la recherche de tous les laboratoires français travaillant sur la vérification et l'identification automatiques du locuteur, le consortium ELISA.**

- . The Elisa Consortium. **The ELISA'99 Speaker Recognition and Tracking Systems**. Proceedings of the 1999 IEEE Workshop on Automatic Identification Advanced Technologies, Summit, New Jersey, United States, October 1999.
- . The Elisa Consortium. **The ELISA Systems for the NIST'99 Evaluation in Speaker Detection and Tracking**. Digital Signal Processing, Vol. 10, No. 1-3, pp. 143-153, January/April/July 2000.
- . Ivan Magrin-Chagnolleau, Guillaume Gravier, and Raphaël Blouet for the ELISA consortium. **Overview of the 2000-2001 ELISA Consortium Research Activities**. Proceedings of the Workshop 2001 : A Speaker Odyssey , Chania, Crete, Greece, June 2001.

- . Daniel Moraru, Sylvain Meignier, Laurent Besacier, Jean-Francois Bonastre, and Ivan Magrin-Chagnolleau. **The ELISA Consortium Approaches in Speaker Segmentation During the NIST 2002 Speaker Recognition Evaluation**. Proceedings of ICASSP 2003, Hong Kong, China, April 2003.

- J'ai monté et porté le projet **ALIZÉ**, projet de réalisation d'une plateforme logiciel libre en vérification automatique du locuteur, d'une durée de deux ans, financé par le ministère de la recherche dans le cadre de l'action « Technolanguages » à partir de novembre 2002. Partenaires du projet : le laboratoire DDL (Lyon), le LIA (Avignon), l'IRISA (Rennes), le CLIPS (Grenoble), l'IRIT (Toulouse), l'ENST (Paris), THALÈS (Gennevilliers), ATLOG (Niort), CALISTEL (Montbonnot), DGA (Arcueil). Budget de 300 k€ pour les deux ans.

- J'ai initié et supervisé la rédaction d'un article état de l'art sur la vérification et l'identification automatiques du locuteur, en collaboration avec des chercheurs du Lincoln Laboratory du MIT, qui étaient à l'époque les leaders dans le domaine, avec les français. L'article a fait référence et a été à ce jour cité 1229 fois (et toujours cité aujourd'hui, cité 30 fois en 2022).

- . Frédéric Bimbot et al. **A Tutorial on Text-Independent Speaker Verification**. EURASIP Journal on Applied Signal Processing. 2004 :4, 430–451.

- J'ai proposé un chapitre de livre qui reprenait l'état de l'art de l'évaluation des systèmes de vérification du locuteur, chapitre qui faisait suite à la participation comme enseignant à une école d'été sur l'évaluation des systèmes de traitement de l'information.

- . Ivan Magrin-Chagnolleau. **Évaluation des systèmes de vérification du locuteur**. Dans Evaluation des systèmes de traitement de l'information. Edité par Stéphane Chaudiron. Hermès, 2004.

2.2 Contributions majeures dans le domaine de l'indexation sonore et multimédia

L'indexation multimédia consiste à segmenter puis indexer automatiquement un flux multimédia. Par exemple, on peut vouloir segmenter et indexer automatiquement par locuteurs des bulletins d'information, qu'ils soient diffusés à la radio ou à la télé, et cela afin de pouvoir faire un archivage automatique, ou bien on peut vouloir segmenter et indexer des conversations téléphoniques.

Mes contributions majeures dans ce domaine sont les suivantes :

- **J'ai travaillé, pendant mon post-doc à AT&T aux États-Unis, et en collaboration avec Aaron Rosenberg, à l'application de certaines méthodes de vérification automatique du locuteur à la segmentation et l'indexation d'un flux audio ou audiovisuel en locuteurs.**

- . Aaron E. Rosenberg, Ivan Magrin-Chagnolleau, S. Parthasarathy, and Qian Huang. **Speaker Detection in Broadcast Speech Databases**. Proceedings of ICSLP 98, Sydney, Australia, December 1998.
- . Ivan Magrin-Chagnolleau, Aaron E. Rosenberg, and S. Parthasarathy. **Detection of Target Speakers in Audio Databases**. Proceedings of ICASSP 99, Vol. 2, pp. 821-824, Phoenix, United States, March 1999.

- **Cette approche a donné de si bons résultats qu'elle a donné lieu au dépôt de deux brevets.**

- . Qian Huang, Ivan Magrin-Chagnolleau, Sarangarajan Parthasarathy, and Aaron Edward Rosenberg. **Multimedia Search Apparatus and Method for Searching Multimedia Content Using Speaker Detection by Audio Data**. United States Patent 6 405 166, June 2002.

- . Qian Huang, Ivan Magrin-Chagnolleau, Sarangarajan Parthasarathy, and Aaron Edward Rosenberg. **Multimedia Search Apparatus and Method for Searching Multimedia Content Using Speaker Detection by Audio Data**. United States Patent 6 317 710, 2001.

- Cela m'a aussi permis de participer au premier projet au monde d'indexation multimédia, qui était le projet SCAN d'AT&T.

- . John Choi, Don Hindle, Julia Hirschberg, Ivan Magrin-Chagnolleau, Christine Nakatani, Fernando Pereira, Amit Singhal, and Steve Whittaker. **SCAN - Speech Content Based Audio Navigator : A System Overview**. Proceedings of ICSLP 98, Sydney, Australia, December 1998.
- . John Choi, Don Hindle, Julia Hirschberg, Ivan Magrin-Chagnolleau, Christine Nakatani, Fernando Pereira, Amit Singhal, and Steve Whittaker. **An Overview of the AT&T Spoken Document Retrieval System**. Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, United States, December 1998.

- De retour en France, j'ai alors monté et porté, avec Nathalie Vallès-Parlangeau, le premier projet d'indexation sonore en France, le projet RAIVES, projet d'indexation sonore de documents radiophoniques sur le web, d'une durée de 4 ans, financé par le CNRS dans le cadre de l'action « Société de l'information » à partir de novembre 2001. Partenaires du projet : le LORIA (Nancy), l'IRIT (Toulouse), le LIA (Avignon), l'IRISA (Rennes). Budget de 800 k€.

J'ai également participé, deux ans plus tard, au premier projet d'indexation multimédia en France, le projet AGIR, projet d'indexation multimédia en partenariat avec l'INA, d'une durée de deux ans, financé par le ministère de la technologie à partir de novembre 2003. Budget de 400 k€ pour les deux ans.

- . Ivan Magrin-Chagnolleau and Frédéric Bimbot. **Indexing Telephone Conversations by Speakers Using Time-Frequency Principal Component Analysis**. Proceedings of ICME 2000, New York, United States, July-August 2000.

- . Sylvain Meignier, Jean-François Bonastre, and Ivan Magrin-Chagnolleau. **Appariement de locuteurs entre des documents sonores préalablement segmentés en utilisant la classification hiérarchique.** Proceedings of JEP 2002, Nancy, France, June 2002.
- . Sylvain Meignier, Jean-François Bonastre, and Ivan Magrin-Chagnolleau. **Speaker Utterances Tying Among Speaker Segmented Audio Documents Using Hierarchical Classification : Towards Speaker Indexing of Audio Documents.** Proceedings of ICSLP 2002, Denver, Colorado, USA, September 2002.
- . Ivan Magrin-Chagnolleau and Nathalie Parlangeau-Vallès. **Audio Indexing : What has been accomplished and the Road Ahead.** Proceedings of JCIS 2002, pp. 911-914, Durham, North Carolina, USA, March 2002 (Invited paper).
- . Nathalie Parlangeau-Vallès, Jérôme Farinas, Dominique Fohr, Irina Illina, Ivan Magrin-Chagnolleau, Odile Mella, Julien Pinquier, Jean-Luc Rouas, and Christine Sénac. **Audio Indexing on the Web : A Preliminary Study of Some Audio Descriptors.** Proceedings of SCI 2003, Orlando, Florida, July 2003.

2.3 Contributions majeures dans le domaine de l'analyse des signaux sismiques

Après mon post-doc à AT&T, j'ai effectué un autre post-doc aux États-Unis, à Rice University à Houston au Texas, dans l'équipe de Richard Baraniuk. Ce dernier est en effet l'un des spécialistes des transformations temps-fréquence, et en particulier des ondelettes, et je voulais élargir mes connaissances dans ce domaine. Le post-doc étant financé par Elf, j'ai utilisé ces approches dans le domaine de la géophysique.

Mes contributions majeures dans ce domaine sont les suivantes :

- **Application d'une technique de décomposition des signaux, l'EMD (Empirical Mode Decomposition), à des signaux sismiques.** J'ai développé un programme matlab que j'ai mis à disposition de la communauté scientifique, et qui a été depuis très largement utilisé (je reçois aujourd'hui encore de temps en temps des messages de chercheurs du monde entier qui utilisent encore mon programme).

- . Ivan Magrin-Chagnolleau and Richard G. Baraniuk. **Empirical mode decomposition based frequency attributes.** Proceedings of the 69th SEG Meeting, Houston, Texas, United States, November 1999.
- . **Empirical mode decomposition and Hilbert spectrum.** Programme développé sous Matlab en collaboration avec Richard Baraniuk de Rice University, Houston, Texas, Etats Unis.

- **Application des ondelettes et d'une modélisation Markovienne pour effectuer une segmentation multi-échelle des données sismiques.** Cette approche a donné de très bons résultats et a été ensuite reprise par d'autres chercheurs en géophysique.

- . Ivan Magrin-Chagnolleau, Hyeokho Choi, Rutger van Spaendonck, Philippe Steeghs and Richard G. Baraniuk. **Multiscale texture segmentation of dip-cube slices using Wavelet-domain hidden Markov trees.** Proceedings of the 69th SEG Meeting, Houston, Texas, United States, November 1999.

2.4 Contribution majeure dans le domaine de la reconnaissance automatique des langues

La reconnaissance automatique des langues consiste à développer des modèles statistiques qui permettent ensuite d'identifier automatiquement la langue d'un enregistrement sonore ou audio-visuel. Les modèles sont appris sur des échantillons sonores des langues à modéliser.

Ma contribution majeure dans ce domaine est la suivante :

- **Co-encadrement à 80 % (avec Frédéric Bimbot, 20 %) de la thèse de Michel Dutat. Application de l'analyse en composantes principales temps fréquence à l'identification automatique de la langue.**

- . Michel Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot. **Analyse en composantes principales temps-fréquence : application à la reconnaissance de la langue.** Proceedings of JEP 2000 , Aussois, France, June 2000.
- . Michel Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot. **Language Recognition using Time-Frequency Principal Component Analysis and Acoustic Modeling.** Proceedings of ICSLP 2000, Beijing, China, October 2000.
- . Michel Dutat. **Caractérisation de la langue parlée par modèles de séquences d'événements acoustiques.** Thèse ENST, 2002.

2.5 Contribution majeure dans le domaine de la perception du langage

Lors de mon affectation au laboratoire Dynamique du Langage à Lyon, j'ai travaillé avec Melissa Barkat et Fanny Meunier sur la perception du langage.

Notre contribution majeure dans ce domaine est la suivante :

- Nous avons travaillé sur l'intelligibilité de segments de parole à l'envers en français. Nous avons pour cela testé et validé pour la première fois sur le français une expérience qui avait été faite sur l'anglais américain, à savoir que si on retourne à l'envers des petits morceaux de paroles systématiquement le long d'un énoncé sonore, et si la longueur de ces morceaux reste inférieure à un certain seuil (quelques dizaines de ms), cela ne nuit en rien à la compréhension de l'énoncé. Cela confirme que les informations fines au niveau temporel ne contribuent pas à la fabrication du sens. Cela explique notamment pourquoi on arrive à comprendre quelqu'un dans le bruit même si on ne capte pas absolument tout.

- . Melissa Barkat, Fanny Meunier, and Ivan Magrin-Chagnolleau. **Intelligibility of Time-Reversed Speech in French**. Workshop on Temporal Integration in the Perception of Speech, Aix-en-Provence, France, April 2002 (Abstract).
- . Fanny Meunier, Tristan Cénier, Melissa Barkat, and Ivan Magrin-Chagnolleau. **Mesure d'intelligibilité de segments de parole à l'envers en français**. Proceedings of JEP 2002, Nancy, France, June 2002.
- . Ivan Magrin-Chagnolleau, Melissa Barkat, and Fanny Meunier. **Intelligibility of Reverse Speech in French : A Perceptual Study**. Proceedings of ICSLP 2002, Denver, Colorado, USA, September 2002.

2.6 Contribution majeure dans le domaine de la langue des signes

Au laboratoire Dynamique du Langage, j'ai également collaboré avec Colette Grinevald sur la langue des signes.

Notre contribution majeure dans ce domaine est la suivante :

- Nous avons montré que la langue des signes, au même titre que d'autres langues, notamment en Afrique et en Amérique centrale, utilisait des classificateurs verbaux. C'est un des arguments qui a permis de faire reconnaître la langue des signes comme une langue à part entière.

- . Karine Grande, Ivan Magrin-Chagnolleau, Annie Risler, and Colette Grinevald. **La dénomination d'objets manipulables en LSF au sein d'un discours narratif : peut-on parler de classificateurs verbaux ?** Proceedings of Journées d'études internationales - La linguistique de la LSF : Recherches actuelles, Lille, France, September 2003.

2.7 Contributions majeures sur la méthode de l'Actors Studio

En 2004, j'ai choisi de changer mon domaine et mes activités de recherche, et de m'intéresser à l'art et à tout ce qui s'y rattache (philosophie de l'art, esthétique, phénoménologie de la création, création éactive, etc.).

J'ai notamment développé une expertise internationale autour de la méthode de l'Actors Studio, méthode de jeu de l'acteur développée par Lee Strasberg à New York.

- Mes études de théâtre pendant 3 ans à l'Actors Studio, un travail mené depuis sur les archives de Lee Strasberg, l'auteur de la méthode de l'Actors Studio, et une poursuite de la pratique de cette méthode, à la fois comme acteur, metteur en scène, et coach d'acteurs, m'ont permis de devenir le spécialiste en France et à l'international de la méthode de l'Actors Studio. Dans le cadre de la préparation de mon HdR, j'ai terminé un inédit en français sur ce sujet, ce qui constituera le premier livre écrit en français sur cette méthode (sans compter les traductions françaises de livres américains).

Mes contributions majeures dans ce domaine sont les suivantes :

- . Ivan Magrin-Chagnolleau. *La méthode de l'Actors Studio*. Inédit écrit dans le cadre de mon HdR. À paraître en 2024.
- . Ivan Magrin-Chagnolleau. **La méthode de l'Actors Studio : Genèse, Pratique et Phénoménologie**. Dans *Être en jeu, être en scène*, édité par Laurent Berger et Arianna de Sanctis. Sous presse. (téléchargé 1346 fois sur HAL à ce jour)
- . Ivan Magrin-Chagnolleau. **Actors Studio**. *Encyclopædia Universalis*, 2021.
- . Ivan Magrin-Chagnolleau. **L'acteur et les émotions au théâtre et au cinéma**. *p-e-r-f-o-r-m-a-n-c-e*, 3, 2016.
- . Ivan Magrin-Chagnolleau. **L'Héritage de Lee Strasberg**. *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.
- . Ivan Magrin-Chagnolleau. **L'incorporation animale chez l'acteur**. *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.

2.8 Contributions majeures en esthétique du théâtre et du cinéma

- Création en 2010 et direction jusqu'à fin 2015 de l'équipe de recherche EsPAS (Esthétique de la Performance et des Arts du Spectacle). Il s'agissait de la première équipe de recherche en France entièrement consacrée aux études de la performance (Performance Studies).

- Exploration autour de la couleur, de son utilisation au théâtre et au cinéma, et de sa phénoméologie.

- . Ivan Magrin-Chagnolleau. **A Phenomenological Approach to Color Theory and Aesthetics.** *Proceedings of AIC 2022*, Jun 2022, Toronto, Canada.
- . Ivan Magrin-Chagnolleau. **The Relationship between Photo Retouching and Color Grading.** *Color Culture and Science Journal*, 2022, 14(1), pp. 79-86.
- . Ivan Magrin-Chagnolleau. **Handling Color in Photography and Film : From Photo Retouching to Color Grading.** *Proceedings of AIC 2021*, Aug 2021, Milan, Italy.
- . Ivan Magrin-Chagnolleau. **The Use of Color in Theater and Film.** *Proceedings of AIC 2013*, Jul 2013, Newcastle upon Tyne, United Kingdom.

2.9 Contributions majeures en Création Recherche

Mes contributions majeures dans ce domaine sont les suivantes :

- Le développement d'un travail de recherche épistémologique très approfondi autour de la thématique de la création recherche. Il s'agit de proposer un cadre rigoureux et scientifique de ce qui constitue la création recherche, ce qui entre dans son cadre et ce qui n'y entre pas. En particulier, l'accent est mis sur les connaissances qui sont développées par les artistes praticiens, sur les thèmes de recherche que ces connaissances abordent, sur les conditions de leurs restitutions au sein de la communauté scientifique, et en quoi ces connaissances sont spécifiques d'une articulation entre recherche et pratique et non d'une recherche purement théorique.

- . Ivan Magrin-Chagnolleau. **Reflections on the impact of creation research, practice-based research, and practice as research, and how to define and maximize that impact.** *Media Practice and Education*, 2022, 23(2), pp. 105-111.
- . Ivan Magrin-Chagnolleau. **Éloge de l'artiste chercheur.** *p-e-r-f-o-r-m-a-n-c-e*, 1, 2014.

- **Création en 2014 et direction de publication du journal *p-e-r-f-o-r-m-a-n-c-e***, journal de création recherche dans les arts performatifs, les arts du spectacles, les arts visuels et les arts sonores, avec comité scientifique et de lecture internationaux.

- **Direction ou co-direction des 7 numéros du journal parus à ce jour ou en cours de parution :**

- . *p-e-r-f-o-r-m-a-n-c-e*, 7 : **Art and Human**, 2023. Co-édité avec Fabrice Métais.
- . *p-e-r-f-o-r-m-a-n-c-e*, 6 : **Creation Research**, 2022.
- . *p-e-r-f-o-r-m-a-n-c-e*, 5 : **Healing Art**, 2019. Co-édité avec Michel Sicard.
- . *p-e-r-f-o-r-m-a-n-c-e*, 4 : **The Creative Process**, 2017.
- . *p-e-r-f-o-r-m-a-n-c-e*, 3 : **Cuisine & Performance**, 2016. Co-édité avec Allen Weiss.

2.9. CONTRIBUTIONS MAJEURES EN CRÉATION RECHERCHE 69

- . *p-e-r-f-o-r-m-a-n-c-e, 2* : **Theatre Education**, 2015.
Co-édité avec Joëlle Aden.
- . *p-e-r-f-o-r-m-a-n-c-e, 1* : **The Artist Researcher**, 2014.
Co-édité avec Ludivine Allegue.

2.10 Contributions majeures sur le processus créatif et sa phénoménologie

- Je travaille également sur le processus créatif. J'ai choisi d'aborder le processus créatif dans sa dimension phénoménologique, c'est-à-dire le processus créatif comme une expérience vécue. Je suis en train de développer une expertise sur ce sujet, due d'une part à ma propre pratique artistique, d'autre part à mes connaissances en philosophie, et en particulier en phénoménologie. C'est dans le cadre de ce travail sur la phénoménologie de la création que je me suis intéressé à l'entretien d'explicitation, que j'utilise dans le cadre de cette recherche. Je questionne également la dimension esthétique de ce processus créatif. Je questionne enfin la dimension spirituelle de ce processus créatif, étant donné que de nombreux artistes s'y réfèrent.

- . Ivan Magrin-Chagnolleau. **A Phenomenological Approach to Color Theory and Aesthetics.** *Proceedings of AIC 2022*, Jun 2022, Toronto, Canada.
- . Ivan Magrin-Chagnolleau. **Chance and the Creative Process.** *Chance, Calculation and Life*, ISTE Editions, pp. 169-183, 2021, 9781786306678.
- . Ivan Magrin-Chagnolleau. **Hasard et Création.** Dans *Le hasard, le calcul et la vie*, édité par Thierry Gaudin, Marie-Christine Maurel, Jean-Charles Pomerol. ISTE Éditions, Collection « Systèmes d'information, web et société », pp. 161-175, 2021.
- . Ivan Magrin-Chagnolleau. **Théâtre, création et psychanalyse : autour de la mise en scène de Premier Amour de Beckett.** *p-e-r-f-o-r-m-a-n-c-e*, 5, 2019.
- . Ivan Magrin-Chagnolleau, Ruth Nahoum, Marcus Weisen. **Le projet sipapu : vers une micro-phénoménologie de la création.** Dans *Les enjeux cognitifs de l'artefact esthétique : Tome 2*, édité par Xavier Lambert. L'Harmattan, Collection « Ouverture Philosophique », pp. 187-194, 2019. ISBN : 978-2-343-18776-1.
- . Ivan Magrin-Chagnolleau. **Le Processus Créatif de Robert Lepage.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.
- . Ivan Magrin-Chagnolleau. **La relation réalisateur - compositeur : un exemple de processus créatif relationnel éactif.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.

- . Ivan Magrin-Chagnolleau. **Entretien avec le metteur en scène Robert Lepage.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.
- . Ivan Magrin-Chagnolleau. **Création, jeu et contemplation : l'art comme exercice spirituel.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.
- . Anne Sachs, Ivan Magrin-Chagnolleau. **Du chaos à la création : entretien avec Ivan Magrin-Chagnolleau.** Dans *Danse Éducation : Échos issus du congrès IDEA Paris 2013*. 2016. ISBN : 978-2-951-99010-4.
- . Ivan Magrin-Chagnolleau. **Entretien avec le chef Bernard Bordinaries.** *p-e-r-f-o-r-m-a-n-c-e*, 3, 2016.
- . Ivan Magrin-Chagnolleau. **Du chaos à la création : une expérience de pédagogie énaactive pour expérimenter le processus créatif en performance.** *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.
- . Ivan Magrin-Chagnolleau. **The Use of Color in Theater and Film.** *Proceedings of AIC2013*.

2.11 Contributions majeures sur l'énaction

- Je mène aussi une recherche depuis plusieurs années sur l'application de concepts issus des sciences cognitives à la création artistique et à l'enseignement artistique, et en particulier l'énaction, l'émergence, et l'autopoïèse proposées par Francisco Varela. Dans le cadre de ce travail, j'ai initié et organisé un colloque international à Cerisy en août 2022 : « Francisco Varela : une pensée actuelle ».

- . Ivan Magrin-Chagnolleau. **Expérience phénoménologique et énaactive d'une œuvre de poésie spatialisée visuelle et sonore.** Dans *Corporéité et construction du sens dans la poésie visuelle : énaaction, perception, expérimentation*. Sous presse.
- . Ivan Magrin-Chagnolleau. **La relation réalisateur - compositeur : un exemple de processus créatif relationnel énaactif.** *p-e-r-f-o-r-m-a-n-c-e*, 4, 2017.
- . Ivan Magrin-Chagnolleau. **L'énaction dans la création artistique.** Dans *Action, Énaaction*, édité par Xavier Lambert. L'Harmattan, Collection « Ouverture Philosophique », pp. 213-224, 2017. ISBN : 978-2-343-11403-3.
- . Ivan Magrin-Chagnolleau. **Du chaos à la création : une expérience de pédagogie énaactive pour expérimenter le processus créatif en performance.** *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.

2.12 Contributions majeures sur l'entretien d'explicitation

- J'ai aussi démarré depuis quelques années un travail autour de l'entretien d'explicitation. Appréhendé d'abord comme un outil méthodologique pour approcher la subjectivité, et donc la phénoménologie d'un point de vue pratique, l'entretien d'explicitation est devenu lui-même un objet de recherche.

- . Magali Boutrais, Ivan Magrin-Chagnolleau, Maryse Maurel, Patricia Rottement. **Le rôle du corps dans la mise en évocation, le remplissage intuitif et la présentification dans le cadre d'un entretien d'explicitation.** *Explicititer*, 137, pp. 37-41, 2023.
- . Ivan Magrin-Chagnolleau. **Mise en évocation : Entretien d'explicitation et méthode de l'Actors Studio.** *Explicititer*, 2023, Hors-série (1), pp.53-61.
- . Ivan Magrin-Chagnolleau. **Le rôle de l'auto-explicitation en recherche.** *Explicititer*, 2023, Hors-série (1), pp.105-111.
- . Jocelyne Leblanc, Anne Philippon, Ivan Magrin-Chagnolleau, Marie-Hélène Lachaud, Magali Boutrais. **Pratiquer l'entretien d'explicitation à distance : petit ABC.** *Explicititer*, 2021, 130, pp. 51-60.

2.13 Contributions majeures en psychologie et psychanalyse

- J'ai aussi démarré depuis quelques années une réflexion sur le soin psychique, et notamment l'utilisation de l'art dans ce domaine. J'ai notamment travaillé pendant un an et demi dans un hôpital de jour, en y animant des ateliers de pratiques artistiques (vidéo et musique).

- . Ivan Magrin-Chagnolleau. **Artistic Workshops with patients of a Psychiatric Day Hospital in France.** *Creative Arts in Education and Therapy*, sous presse.
- . Ivan Magrin-Chagnolleau. **Théâtre, création et psychanalyse : autour de la mise en scène de Premier Amour de Beckett.** *p-e-r-f-o-r-m-a-n-c-e*, 5, 2019.
- . Ivan Magrin-Chagnolleau. **Entretien avec le psychologue Nader Aghakhani.** *p-e-r-f-o-r-m-a-n-c-e*, 5, 2019.
- . Ioana Zamfir et Ivan Magrin-Chagnolleau. **The Madness of Making Sense : Madness and Integrative Knowledge.** Dans *Schizo : The Liberatory Potential of Madness*, édité par Irina Lyubchenko et Fiona Ann Papps. IDN Press, 2016. ISBN : 978-1-848-88460-1.
- . Ioana Zamfir and Ivan Magrin-Chagnolleau. **The Madness of Making Sense : Madness as Perception of the Paradoxical Dimension of Reality.** *Proceedings of Madness : Probing the Boundaries*. September 2014.

2.14 Autres contributions en philosophie de l'art et en esthétique

- Je développe aussi depuis 2014 une activité consistante de critique, activité qui m'amène à approfondir les fondements théoriques de la critique, en lien avec des questions esthétiques. J'ai rejoint début 2018 le Syndicat Français de la Critique de Cinéma (SFCC) ainsi que la fédération internationale des critiques de cinéma (FIPRESCI) et je prépare et vais éditer deux livres sur la critique de cinéma en France, et sur la critique de cinéma dans le monde. J'ai également mis au point un cours intitulé « La critique comme formation du regard esthétique », cours que j'ai enseigné en 2020 à Chapman University, en Californie.

- . Ivan Magrin-Chagnolleau. **Réflexions autour de l'expertise d'une œuvre artistique.** Dans *Expertise artistique*, édité par Maryvonne Saison. Ministère de la Culture, 2015.
- . Voir aussi quelques unes de mes critiques de cinéma à la fin de cet ouvrage.

- Pour terminer ce tour de mes contributions majeures dans le domaine de la philosophie de l'art et de l'esthétique, je voudrais mentionner une réflexion autour de la dimension politique de l'art. C'est une question qui m'intéresse depuis longtemps mais sur laquelle je n'avais pas produit d'écrits. La participation à un colloque sur l'évolution m'a permis de mettre en chantier une première publication dans ce domaine.

- . Ivan Magrin-Chagnolleau. **Futurs utopiques et dystopiques : comment la fiction nous aide à penser l'évolution.** *Arts et Sciences*, Vol. 21-5, No. 1, pp. 1-11, 2021.

Chapitre 3

Écrits choisis

3.1 Reconnaissance automatique du locuteur

Effect of Utterance Duration and Phonetic Content on Speaker Identification using Second-Order Statistical Methods (1995)

Ivan Magrin-Chagnolleau, Jean-François Bonastre, and Frédéric Bimbot. **Effect of Utterance Duration and Phonetic Content on Speaker Identification using Second-Order Statistical Methods.** *Proceedings of EUROSPEECH 95*, Vol.1, pp. 337-340, Madrid, Spain, September 1995.

4 pages

EFFECT OF UTTERANCE DURATION AND PHONETIC CONTENT ON SPEAKER IDENTIFICATION USING SECOND-ORDER STATISTICAL METHODS

Ivan MAGRIN-CHAGNOLLEAU[°]

Jean-François BONASTRE*

Frédéric BIMBOT[°]

[°]Télécom Paris (E.N.S.T.) – Dépt. Signal, C.N.R.S. URA 820
46, rue Barrault – F-75634 Paris Cedex 13 – FRANCE – European Union
email: ivan@sig.enst.fr and bimbot@sig.enst.fr

*Laboratoire d'Informatique Université d'Avignon
33, rue Louis Pasteur – F-84000 Avignon – FRANCE – European Union
email: jfb@univ-avignon.fr

ABSTRACT

Second-order statistical methods show very good results for automatic speaker identification in controlled recording conditions [2]. These approaches are generally used on the entire speech material available. In this paper, we study the influence of the content of the test speech material on the performances of such methods, i.e. under a more analytical approach [3]. The goal is to investigate on the kind of information which is used by these methods, and where it is located in the speech signal. Liquids and glides together, vowels, and more particularly nasal vowels and nasal consonants, are found to be particularly speaker specific: test utterances of 1 second, composed in majority of acoustic material from one of these classes provide better speaker identification results than phonetically balanced test utterances, even though the training is done, in both cases, with 15 seconds of phonetically balanced speech. Nevertheless, results with other phoneme classes are never dramatically poor. These results tend to show that the speaker-dependent information captured by long-term second-order statistics is consistently common to all phonetic classes, and that the homogeneity of the test material may improve the quality of the estimates.

1. INTRODUCTION

In this paper, the influence of phonetic content on the performance of a speaker identification system is investigated. Second-order statistical methods are chosen because they provide very good results with a low quantity of computations and with a restricted quantity of speech material, provided recording conditions and channel distortions are controlled [2]. The goal of this work is to study how the performances of this family of approaches vary with the phonetic content of the test material.

Several experiments were previously reported on the relative speaker discriminating properties of phonemes. In particular, Eatock et al. [5] used a VQ codebook based approach and concluded that nasals and vowels provided the best performances on an English language database. Le Floch et al. [9] investigated the properties of AR-vector models and concluded that vowels, diphthongs and nasals provided the best performances, also on an English language database.

The specificity of the work reported in this paper comes from the fact that second-order statistical methods are used and that the training and test material are heterogeneous. The general experimental framework is the following: 15 seconds of training material coming from phonetically balanced sentences, are used to build a reference model for

each speaker. Then, specific speech segments are selected from other phonetically balanced sentences in order to build a test pattern which is strongly biased towards a particular phoneme or phoneme class.

The database used for our experiments contains 67 cooperative speakers recorded in a slightly noisy environment. For each speaker, 50 French phonetically balanced sentences, i.e. approximately 3 minutes of speech are recorded in a single session. This volume of speech allows to have a sufficient number of occurrences for most phonemes.

The three second-order statistical measures used in the experiments are described in section 2.. Details on the speech database and on the signal analysis are given in section 3.. Section 4. reports preliminary experiments on utterance duration, and section 5. the experiments on the phonetic content. Finally, section 6. concludes on this work and gives some perspectives.

2. SPEAKER IDENTIFICATION MEASURES

The 3 speaker identification methods used in this work are inspired from statistical tests on covariance matrices [1], computed on acoustic parameters. The first two speaker similarity measures are directly derived from maximum likelihood Gaussian classifiers [7]. The third one is based on a sphericity test between covariance matrices [8]. The measures, asymmetric in their original form, are symmetrised as a weighted sum of the asymmetric measure and its dual term. In previous work [2], this procedure was shown to improve performances.

Let X and Y denote two covariance matrices of a reference speaker and of a test speaker respectively, corresponding to the covariance of some spectral vectors computed along a sentence. Let \bar{x} and \bar{y} denote the means of the spectral vectors. Let M and N denote the number of spectral vectors used to estimate the covariance matrices and mean vectors, and p the dimension of the spectral vectors. The mathematical expression of the three measures that we used in our experiments, are given in Table 1. Note that, once the covariance matrices X and Y are inverted, and that their determinant is evaluated, the computation of the measures requires very few operations.

3. DATABASE AND SIGNAL ANALYSIS

3.1. Database

The corpus is composed of read phonetically balanced sentences [4], the phonetic transcription of which can be found in the BDFON database. The sentences are prompted on a screen. Recording begins and ends automatically using a speech activity detector. Each sentence is recorded through a SHURE SM10A microphone, and digitized at a 16 kHz

$$\begin{aligned}
\mu_G(X, Y) &= \frac{1}{p} \left[\frac{M}{M+N} \operatorname{tr}(YX^{-1}) + \frac{N}{M+N} \operatorname{tr}(XY^{-1}) - \frac{M-N}{M+N} \log \left(\frac{\det Y}{\det X} \right) \right] \\
&\quad + \frac{1}{p} \left[(\bar{y} - \bar{x})^T \left[\frac{M}{M+N} X^{-1} + \frac{N}{M+N} Y^{-1} \right] (\bar{y} - \bar{x}) \right] - 1 \\
\mu_{Gc}(X, Y) &= \frac{1}{p} \left[\frac{M}{M+N} \operatorname{tr}(YX^{-1}) + \frac{N}{M+N} \operatorname{tr}(XY^{-1}) - \frac{M-N}{M+N} \log \left(\frac{\det Y}{\det X} \right) \right] - 1 \\
\mu_{Sc}(X, Y) &= \frac{M}{M+N} \log \operatorname{tr}(YX^{-1}) + \frac{N}{M+N} \log \operatorname{tr}(XY^{-1}) - \frac{1}{p} \frac{M-N}{M+N} \log \left(\frac{\det X}{\det Y} \right) - \log p
\end{aligned}$$

Table 1. Expressions of the three symmetrized second-order statistical measures (“tr” denote the trace and “det” the determinant of a matrix).

sampling frequency on 16 bits by an OROS AU22 board. The recording equipment was set up in the corridor of a university, i.e with a non-negligible background noise. The recordings are single-session. 67 speakers took part to the experiments, mostly students. They each recorded approximately 3 minutes of speech.

3.2. Signal Analysis

The speech analysis module extracts filterbank coefficients in the following way: a Winograd Fourier Transform is computed on Hamming windowed signal frames of 31.5 ms (i.e 504 samples) at a frame rate of 10 ms (160 samples). For each frame, spectral vectors of 24 Mel-Scale Triangular-Filter Bank coefficients are then calculated from the Fourier Transform power spectrum, and expressed in logarithmic scale. Covariance matrices and mean vectors are finally computed from these spectral vectors.

4. UTTERANCE DURATION

The first set of experiments investigates on the influence of utterance duration for second-order statistical methods. This allowed us to choose meaningful training and test durations for the second part of the work. Several durations for training and test are chosen: 15, 10, 6, 3 and 2 seconds for training; 10, 6, 3, 2 and 1 second for testing. For each speaker, all sentences are randomly concatenated together. The silences at the beginning and the end of sentences are not removed, but they generally do not exceed 0.1 second. First, a certain amount of speech is selected for training (for example 15 seconds). Then, the rest is segmented in several test portions, each portion having a predetermined duration, until 20 test portions are obtained unless the speech material is exhausted. As a result of this experimental design, the experiments are text-independent.

Percentages of correct identifications are given in Table 2. The best measure is always μ_G , which confirms once again that the average of the spectral vectors is a significant source of speaker specific information, for good quality contemporaneous speech.

Even though the measures are symmetric, a clear asymmetry of the results can be observed: for example, with a training of 10 seconds and a test of 2 seconds, the percentage of correct identifications with μ_G is 95.3 %, while with a training of 2 seconds and a test of 10 seconds, it only reaches 88.4 %. In fact, if a training of 2 seconds is poorly representative of the speaker (for instance, if it contains a lot of silence), this has an effect on all the tests utterances from this very speaker, which affects considerably the overall score. If, conversely, a test of 2 seconds is of poor quality, it only causes 1 mistake, which has little impact on the global performance.

In the second part of our experiments, we chose a 15 second training duration, which is a guaranty to have a reliable training and a sufficient phonetic coverage. The test dura-

tion is chosen to 1 second for two reasons. Firstly, if a longer duration is chosen, the number of tests for a given phoneme is less significant. Secondly, if the percentage of correct identifications is too high, we felt that comparisons may lack statistical significance.

| Test Duration | | Training Duration | | | | |
|----------------|------------|-------------------|-------------|-------------|-------------|-------------|
| | | 15 s | 10 s | 6 s | 3 s | 2 s |
| 10 s (1095) | μ_G | 99.9 | 99.7 | 99.3 | 94.3 | 88.4 |
| | μ_{Gc} | 99.8 | 99.7 | 98.1 | 92.0 | 85.5 |
| | μ_{Sc} | 99.8 | 99.7 | 98.3 | 91.1 | 85.3 |
| 6 s (1294) | μ_G | 99.9 | 99.5 | 98.8 | 93.0 | 87.2 |
| | μ_{Gc} | 99.6 | 99.0 | 96.8 | 88.6 | 82.6 |
| | μ_{Sc} | 99.5 | 99.1 | 97.0 | 88.5 | 82.2 |
| 3 s (1340) | μ_G | 98.7 | 97.3 | 95.3 | 87.0 | 80.2 |
| | μ_{Gc} | 97.8 | 96.0 | 92.0 | 82.5 | 74.8 |
| | μ_{Sc} | 98.1 | 96.5 | 93.0 | 82.4 | 74.0 |
| 2 s (1340) | μ_G | 97.5 | 95.3 | 91.7 | 83.3 | 74.3 |
| | μ_{Gc} | 94.8 | 92.9 | 86.3 | 75.6 | 66.1 |
| | μ_{Sc} | 95.7 | 92.9 | 86.8 | 76.3 | 66.0 |
| 1 s (1340) | μ_G | 87.5 | 83.9 | 76.0 | 65.3 | 57.0 |
| | μ_{Gc} | 79.5 | 73.5 | 65.1 | 53.3 | 47.8 |
| | μ_{Sc} | 83.6 | 79.0 | 71.7 | 57.9 | 50.7 |

Table 2. Speaker identification: results in percentage of correct identifications for different training and test durations. The numbers of tests for each test duration is indicated in parentheses.

5. PHONETIC CONTENT

5.1. Segmentation

In order to study the influence of phonetic content on the performances of the second-order statistical methods, we used an automatic system for segmenting the speech material into specific phonemes or phonetic classes. This system of automatic localisation is based on a bottom-up acoustic-phonetic decoder, which is speaker independent [3], [10], [6]. For each sentence, this decoder proposes a set of weighted phonetic hypotheses. These hypotheses are then aligned with the phonetic transcription of the sentence, by a left-right alignment algorithm. In order to obtain a high localisation accuracy, the algorithm is tuned with a high level of rejection for uncertain alignments: in this experiment, 55 % of the sentences were rejected. As a consequence, the phonetic events selected by this procedure can be considered as highly reliable, and quite typical in their category. The localisation algorithm gives a small kernel for a recognized phoneme. So the phoneme segments were extended to 5 frames before and 5 frames after the kernel. Therefore, the segments are not only composed with frames of the given phoneme, as they may also include a small proportion of transitions.

5.2. Experimental Protocol

For this set of experiments, training material consists of all speech frames derived from 15 seconds of phonetically balanced sentences, i.e. no specific phonetic events are selected. It is in fact the exact same material as the one used in the previous experiment on the influence of utterance duration (see section 4.), with training durations of 15 seconds.

For the tests, specific phonetic classes and phonemes are selected on the rest of the speech material, by the procedure described in section 5.1.. For a particular phoneme, all the frames labeled as this given phoneme are concatenated together, and this material is divided into as many tests of 1 second as possible.

5.3. Description of the phoneme classes

Results for all phoneme classes and phonemes are not presented in this article: we chose to report only on those for which more than 40 tests were carried out.

A first class, referred to as *All*, contains all the phonemes. The experiments with test data from this class is however slightly different from the one in section 4., with 15 seconds for training and 1 second for testing: speech material in the class *All* is composed of acoustic material clearly identified as a phoneme by the segmentation process, and in particular, it does not contain silences, pauses, or other non-linguistic events.

The other classes are: *Vowels* (which contains oral and nasal vowels but not glides), *Oral Vowels*, *Nasal Vowels*, *Consonants* (which contains also glides), *Non-Nasal Consonants* (which contains all the consonants except the nasal consonants), *Nasal Consonants*, *Stop Consonants*, *Fricatives* and *Liquids+Glides* (which form a single class).

5.4. Results

Results for various phoneme classes and individual phonemes are given in Table 3 and Table 4 respectively. The percentage of correct identifications on all the tests is given, as well as the average of the correct identifications per speaker. The number of tests is also indicated.

5.5. Comments

Quite surprisingly, measure μ_G still performs better than μ_{G_c} and μ_{S_c} : even though the mean vector within a phonetic class is expected to be strongly class dependent (and therefore not to match the training mean vector), it still keeps some consistence across phonetic classes.

The results for the class *All* outperform slightly those for the 15 s \times 1 s experiment of section 4., probably because the speech material is, in the second case, more reliable.

A second observation is that the results for each phoneme class is higher than the result of the class *All*. It is also the case for most of the phonemes. This tends to show that a phonetically homogeneous test material benefits to the overall speaker identification performance, even though the training material does not share the same character.

In more detail, *Vowels* give better results than *Consonants*. *Nasal Vowels* outperform *Vowels* and *Oral Vowels*. *Non-Nasal Consonants*, and more particularly *Stop Consonants* or *Fricatives*, give lower performances than all *Consonants* together, whereas *Liquids+Glides* and *Nasal Consonants* yield higher scores. Note that the class *Liquids+Glides* gives the best results altogether, except with μ_{S_c} for which the classes *Nasal Consonants* and *Nasal Vowels* perform better. In what concerns individual phonemes, the best scores with

| Phonemes | <i>All</i> (1334) | | <i>Vowels</i> (1247) | |
|-------------|------------------------------|------|----------------------------------|------|
| | I | M | I | M |
| μ_G | 90.6 | 90.5 | 97.1 | 97.0 |
| μ_{G_c} | 80.8 | 80.6 | 92.3 | 92.1 |
| μ_{S_c} | 83.5 | 83.4 | 92.0 | 91.7 |
| Phonemes | <i>Oral Vowels</i> (1206) | | <i>Nasal Vowels</i> (262) | |
| | I | M | I | M |
| μ_G | 96.0 | 95.9 | 98.1 | 98.5 |
| μ_{G_c} | 91.3 | 91.2 | 89.3 | 90.5 |
| μ_{S_c} | 90.5 | 90.5 | 93.1 | 92.9 |
| Phonemes | <i>Consonants</i> (1247) | | <i>Non-Nasal Cons.</i> (1186) | |
| | I | M | I | M |
| μ_G | 96.2 | 96.2 | 94.7 | 94.8 |
| μ_{G_c} | 91.1 | 91.0 | 89.4 | 89.3 |
| μ_{S_c} | 91.6 | 91.3 | 89.0 | 89.0 |
| Phonemes | <i>Nasal Cons.</i> (390) | | <i>Stop Cons.</i> (693) | |
| | I | M | I | M |
| μ_G | 96.9 | 97.7 | 94.2 | 95.7 |
| μ_{G_c} | 85.9 | 87.6 | 91.2 | 92.6 |
| μ_{S_c} | 95.1 | 93.9 | 91.3 | 92.7 |
| Phonemes | <i>Fricatives</i> (486) | | <i>Liquids + Glides</i> (277) | |
| | I | M | I | M |
| μ_G | 92.2 | 92.8 | 98.9 | 98.8 |
| μ_{G_c} | 83.7 | 84.8 | 92.4 | 92.3 |
| μ_{S_c} | 86.2 | 86.9 | 92.4 | 91.4 |

Table 3. Speaker identification with speech material selected from specific phonetic classes. Training is composed of 15 s of phonetically balanced speech and test is composed of 1 s of phonetically biased speech. I = Global correct identification score, M = Average correct identification score over all speakers. The number of tests for each test configuration is indicated in parentheses.

μ_G are obtained with the phonemes /o/, /d/, /ε/, and /ã/, whereas /k/, /ʒ/, /u/, /s/ and /y/ give the poorest performance levels.

6. CONCLUSION

Our experiments on the effect of the phonetic content on speaker identification using second-order statistics, tend to show that, on our database, the phonetic homogeneity of the test material is usually a significant factor of improvement, even if the training material is heterogeneous. The phonetic classes that yield particularly good results are *Liquids+Glides*, *Vowels* (and more particularly *Nasal Vowels*), and *Nasal Consonants*, but results for other classes are never dramatically poor. There may therefore exist some kind of speaker-dependent *tie* between acoustic distributions across phonemes that is captured by the second-order statistical methods. This hypothesis has to be confirmed on other types of speech data, in particular noisy speech and non-contemporaneous recordings.

REFERENCES

- [1] T. W. Anderson. *An Introduction to Multivariate Analysis*. John Wiley and Sons, 1958.
- [2] F. Bimbot, I. Magrin-Chagnolleau, and L. Mathan. Second-order statistical measures for text-independent

speaker identification. *Speech Communication*, 17(1-2):177-192, Aug. 1995.

- [3] J.-F. Bonastre and H. Méloni. Automatic speaker recognition and analytic process. In *Proceedings of EUROSPEECH 93*, volume 1, pages 441-444, Sept. 1993. Berlin, Germany.
- [4] P. Combescure. Vingt listes de dix phrases françaises phonétiquement équilibrées. Note technique, CNET (Centre Lannion A), Oct. 1980.
- [5] J. P. Eatock and J. S. Mason. A quantitative assessment of the relative speaker discriminating properties of phonemes. In *Proceedings of ICASSP 94*, volume 1, pages 133-136, Apr. 1994. Adelaide, Australia.
- [6] P. Gilles. *Décodage phonétique de la parole et adaptation au locuteur*. PhD thesis, Université d'Avignon et des Pays de Vaucluse, Jan. 1993.
- [7] H. Gish. Robust discrimination in automatic speaker identification. In *Proceedings of ICASSP 90*, volume 1, pages 289-292, Apr. 1990. New Mexico, United-States.
- [8] Y. Grenier. *Identification du locuteur et adaptation au locuteur d'un système de reconnaissance phonémique*. PhD thesis, ENST, 1977.
- [9] J.-L. Le Floch, C. Montacié, and M.-J. Caraty. Investigations on speaker characterization from Orphée system technics. In *Proceedings of ICASSP 94*, volume 1, pages 149-152, Apr. 1994. Adelaide, Australia.
- [10] H. Méloni and P. Gilles. Décodage acoustico-phonétique ascendant. *Traitement du Signal*, 8(2):107-114, 1991.

| | | | | |
|------------|-------------|------|-------------|------|
| Phonemes | /i/ (174) | | /e/ (105) | |
| | I | M | I | M |
| μ_G | 92.0 | 92.2 | 96.2 | 97.5 |
| μ_{Gc} | 79.9 | 81.3 | 86.7 | 88.0 |
| μ_{Sc} | 81.0 | 82.5 | 87.6 | 88.9 |
| Phonemes | /ε/ (258) | | /y/ (66) | |
| | I | M | I | M |
| μ_G | 97.3 | 98.2 | 87.9 | 90.6 |
| μ_{Gc} | 89.9 | 90.8 | 84.8 | 89.8 |
| μ_{Sc} | 90.7 | 93.0 | 86.4 | 92.2 |
| Phonemes | /ə/ (329) | | /a/ (378) | |
| | I | M | I | M |
| μ_G | 96.7 | 97.1 | 95.8 | 95.7 |
| μ_{Gc} | 91.8 | 91.8 | 87.8 | 86.9 |
| μ_{Sc} | 91.2 | 90.9 | 90.7 | 89.3 |
| Phonemes | /o/ (87) | | /u/ (75) | |
| | I | M | I | M |
| μ_G | 97.7 | 97.3 | 85.3 | 87.9 |
| μ_{Gc} | 87.4 | 88.3 | 73.3 | 77.4 |
| μ_{Sc} | 89.7 | 89.6 | 76.0 | 81.2 |
| Phonemes | /ɑ/ (108) | | /p/ (113) | |
| | I | M | I | M |
| μ_G | 97.2 | 97.4 | 91.2 | 92.9 |
| μ_{Gc} | 89.8 | 89.4 | 82.3 | 84.3 |
| μ_{Sc} | 94.4 | 92.9 | 87.6 | 88.7 |
| Phonemes | /t/ (196) | | /k/ (102) | |
| | I | M | I | M |
| μ_G | 90.8 | 92.3 | 82.4 | 86.0 |
| μ_{Gc} | 86.7 | 87.1 | 78.4 | 84.4 |
| μ_{Sc} | 89.3 | 90.4 | 78.4 | 81.0 |
| Phonemes | /d/ (111) | | /s/ (194) | |
| | I | M | I | M |
| μ_G | 97.3 | 97.5 | 87.1 | 89.0 |
| μ_{Gc} | 91.0 | 91.9 | 78.9 | 80.4 |
| μ_{Sc} | 92.8 | 93.8 | 84.0 | 84.3 |
| Phonemes | /v/ (42) | | /ʒ/ (48) | |
| | I | M | I | M |
| μ_G | 90.5 | 89.2 | 83.3 | 82.9 |
| μ_{Gc} | 76.2 | 77.0 | 77.1 | 76.0 |
| μ_{Sc} | 76.2 | 75.7 | 75.0 | 73.6 |
| Phonemes | /m/ (158) | | /n/ (149) | |
| | I | M | I | M |
| μ_G | 95.6 | 96.4 | 96.6 | 94.4 |
| μ_{Gc} | 79.7 | 80.3 | 87.2 | 88.0 |
| μ_{Sc} | 95.6 | 94.5 | 94.6 | 94.8 |

Table 4. Speaker identification with speech material selected from specific phoneme realisations. Training is composed of 15 s of phonetically balanced speech and test is composed of 1 s of phonetically biased speech. I = Global correct identification score, M = Average correct identification score over all speakers. The number of tests for each test configuration is indicated in parentheses.

**Second-Order Statistical Measures for
Text-Independent Speaker Identification (1995)**

Frédéric Bimbot, Ivan Magrin-Chagnolleau, and Luc Mathan. **Second-Order Statistical Measures for Text-Independent Speaker Identification.** *Speech Communication*, 17(1-2), pp. 177-192, August 1995.

25 pages

Second-Order Statistical Measures for Text-Independent Speaker Identification

Frédéric BIMBOT, Ivan MAGRIN-CHAGNOLLEAU and Luc MATHAN

Ecole Nationale Supérieure des Télécommunications

E.N.S.T. / Télécom Paris - Département Signal

C.N.R.S. - URA 820

46, rue Barrault

75634 PARIS cedex 13

FRANCE, European Union

E-mail: bimbot@sig.enst.fr and ivan@sig.enst.fr

Speech Communication, Vol. 17, No. 1–2, pp. 177–192, August 1995.

Second-Order Statistical Measures for Text-Independent Speaker Identification

Abstract

This article presents an overview of several measures for speaker recognition. These measures relate to second-order statistical tests, and can be expressed under a common formalism. Alternate formulations of these measures are given and their mathematical properties are studied. In their basic form, these measures are asymmetric, but they can be symmetrized in various ways. All measures are tested in the framework of text-independent closed-set speaker identification, on 3 variants of the TIMIT database (630 speakers) : TIMIT (high quality speech), FTIMIT (a restricted bandwidth version of TIMIT) and NTIMIT (telephone quality). Remarkable performances are obtained on TIMIT but the results naturally deteriorate with FTIMIT and NTIMIT. Symmetrization appears to be a factor of improvement, especially when little speech material is available. The use of some of the proposed measures as a reference benchmark to evaluate the intrinsic complexity of a given database under a given protocol is finally suggested as a conclusion to this work.

Abstandsmaße basierend auf statistischen Methoden zweiter Ordnung zur textunabhängigen Sprecheridentifizierung

Zusammenfassung

Dieser Artikel beschreibt mehrere Abstandsmaße der Sprechererkennung. Diese Abstandsmaße beziehen sich auf Tests basierend auf statistischen Methoden zweiter Ordnung und können unter einem gemeinsamen Formalismus betrachtet werden. Alternative Formalismen werden vorgestellt und ihre mathematischen Eigenschaften untersucht. In ihrer ursprünglichen Form sind diese Abstandsmaße asymmetrisch. Sie können jedoch auf vielfältige Weise in eine symmetrische Form umgewandelt werden. Alle Abstandsmaße werden im Rahmen einer textunabhängigen Sprechererkennung einer geschlossenen Sprechergruppe an drei Variationen der TIMIT-Sprachdatenbank (630 Sprecher) getestet : TIMIT (Sprache mit hoher Aufnahmequalität), FTIMIT (eine Version von TIMIT mit eingeschränkter Bandbreite) und NTIMIT (Telephonqualität). Beachtenswerte Ergebnisse wurden mit TIMIT erreicht, die sich mit FTIMIT und NTIMIT verschlechtern. Es stellt sich heraus, daß die Symmetrisierung einen Verbesserungsfaktor darstellt, vor allem, wenn wenig Sprachmaterial vorhanden ist. Die Verwendung einiger der vorgeschlagenen Abstandsmaße als Referenzvergleich zur Evaluierung der Komplexität einer gegebenen Sprachdatenbank unter einem gegebenen Protokoll wird am Ende dieser Arbeit vorgeschlagen.

Mesures statistiques du second ordre pour l'identification du locuteur indépendante du texte

Résumé

Cet article présente un ensemble de mesures pour la reconnaissance du locuteur. Ces mesures reposent sur des tests statistiques du second ordre, et peuvent être exprimées sous un formalisme commun. Différentes expressions de ces mesures sont proposées et leurs propriétés mathématiques sont étudiées. Dans leur forme la plus simple, ces mesures ne sont pas symétriques, mais elles peuvent être symétrisées de différentes façons. Toutes les mesures sont testées dans le cadre de l'identification du locuteur indépendante du texte en ensemble fermé, sur 3 versions de la base de données TIMIT (630 locuteurs) : TIMIT (parole de très bonne qualité), FTIMIT (version filtrée de TIMIT) et NTIMIT (qualité téléphonique). Des performances remarquables sont obtenues sur TIMIT, mais les résultats se dégradent naturellement avec FTIMIT et NTIMIT. La symétrisation apparaît comme un facteur d'amélioration, plus particulièrement lorsque l'on dispose de peu de parole. Il est finalement suggéré, comme conclusion à ce travail, d'utiliser certaines mesures proposées comme méthodes de référence pour évaluer la complexité intrinsèque d'une base de données quelconque, sous un protocole donné.

1 Introduction

1.1 A brief overview

Recent experiments [2] [16] [5] [17] using vector Auto-Regressive models for speaker recognition confirm and further develop work carried out by Grenier [13]. The vector AR approach provides excellent results on a subset of the TIMIT database (420 speakers), in a text-independent mode : with a training of 5 sentences (approximately 15 seconds) and tests of 1 sentence (approximately 3 seconds), closed-set identification scores reported by Montacié [17] are of 98.4 %, and reach 100 % when using 5 sentences for testing. By incorporating a discriminant analysis, the 98.4 % score improves to 99.3 %. On the same database, other approaches have been recently tested, in particular Neural Network based methods. For instance, Rudasi and Zahorian propose binary discriminative networks [20], and reach a 100 % identification score, with 47 speakers, 5 sentences for training and 5 others for testing. Bennani [3] reports experiments with a modular TDNN-based architecture which provides 100 % correct identification for more than 100 TIMIT speakers, using about 15 seconds for training and less than 1 second for testing.

An other method used by Hattori [14] is based on predictive networks. Under this approach, a neural network is trained, for each speaker, to predict a speech frame given the 2 previous ones. During recognition, the identified speaker is the one corresponding to the network with the lowest prediction error. With the best variant, Hattori obtains 100 % correct identifications on 24 speakers (from TIMIT), with about 15 seconds for training and 9 seconds for testing.

Still on TIMIT database, Reynolds [19] shows that a Gaussian Mixture speaker model (with 32 Gaussian distributions with diagonal covariance matrices) leads to a very high level of identification performance : 99.7 % for 168 speakers, using 8 sentences for training and 2 for testing. As discussed by Furuï [9], the Gaussian Mixture approach shares strong similarities with the Vector Quantization based approaches [22] and with the Ergodic HMM based methods [18] [21]. It is therefore very likely that these approaches would also provide excellent results on TIMIT.

1.2 Motivation

In spite of the fact that all approaches mentioned in this brief overview were tested on the same database, it is still difficult to have a clear idea of their relative performances. Among the factors of variability between the experiments are the speech signal pre-processing, the type of acoustic analysis, the length of training and test utterances and of course the number of speakers for which the results are reported.

A systematic comparison of any new approach with all pre-existing methods, under the exact same protocol, is theoretically possible but practically unfeasible ; not only owing to the amount of work involved, but also because it may be very difficult to reproduce in detail a specific algorithm for which all needed information may not be publicly available, or which is sensitive to initialization conditions. Moreover, it can be argued that such or such database is easy and non-discriminant (which may very well be the case for TIMIT), but we lack reliable tools to evaluate the intrinsic difficulty of a database.

A possible way to address this problem of evaluation is the use of a common algorithm as a reference benchmark to evaluate the complexity of a given database under a given protocol [7]. Desirable properties for such a reference method are its relative efficiency and robustness, but also its easy implementation and its absolute reproductibility [6].

The work reported in this article is dedicated to similarity measures between speakers which are derived from statistical tests, with an underlying Gaussian speaker model. The theoretical formulation of these measures illustrates their straightforward reproductibility, while the experimental results evaluate their efficiency on several databases : namely TIMIT (high quality speech), FTIMIT (a 0-4 kHz version of TIMIT) and NTIMIT (telephone quality speech).

In parallel to this large scale evaluation, we discuss the possibility of using one or two of the proposed approaches as systematical benchmarks, in order to provide baseline performance for any database and protocol. Such reference scores would give an idea of the degree of complexity of a given task, and the improvement obtained by any other method would indicate the benefits of a more elaborate speaker model.

1.3 Outline

Three families of measures are investigated in this paper, namely :

- log-likelihood based measures
- sphericity test based measures
- relative eigenvalue deviation measures

In section 3, we present all measures under a common formalism (defined in section 2), and we study their mathematical properties. In their original forms, these measures are not symmetric, and we describe, in section 4, some possibilities to symmetrize them. Section 5 is dedicated to the description of our evaluation protocol, and to the corresponding results. In section 6, we discuss, the possibility of using some of the measures as reference methods.

2 Notation, definitions, properties

2.1 A Gaussian model per speaker

Let $\{x_t\}_{1 \leq t \leq M}$ be a sequence of M vectors resulting from the p -dimensional acoustic analysis of a speech signal uttered by a speaker \mathcal{X} . For instance : filter-bank coefficients, linear prediction coefficients, cepstrum coefficients,... Under the hypothesis of a Gaussian speaker model, the vector sequence $\{x_t\}$ can be summarized by its mean vector \bar{x} and its covariance matrix X , i.e.

$$\bar{x} = \frac{1}{M} \sum_{t=1}^M x_t \quad \text{and} \quad X = \frac{1}{M} \sum_{t=1}^M (x_t - \bar{x}) \cdot (x_t - \bar{x})^T \quad (1)$$

Similarly, for a speaker \mathcal{Y} , a parameterized speech utterance $\{y_t\}$ of N vectors can be modeled by \bar{y} and Y , with

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N y_t \quad \text{and} \quad Y = \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) \cdot (y_t - \bar{y})^T \quad (2)$$

Vectors \bar{x} and \bar{y} are p -dimensional, while X and Y are $p \times p$ symmetric matrices. Throughout this article, a speaker \mathcal{X} (respectively \mathcal{Y}) will be represented by \bar{x} , X and M , (respectively \bar{y} , Y and N). We will also denote

$$\begin{aligned}\delta &= \bar{y} - \bar{x} \\ \Gamma &= X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \\ \rho &= \frac{N}{M}\end{aligned}$$

where $X^{\frac{1}{2}}$ is the symmetric square root matrix of X . Note that, when swapping \mathcal{X} and \mathcal{Y} , vector δ becomes $-\delta$, matrix Γ becomes Γ^{-1} and real number ρ becomes $1/\rho$.

2.2 Second-order statistical measures

We focus on similarity measures μ between speakers \mathcal{X} and \mathcal{Y} which can be expressed as a function

$$\mu(\mathcal{X}, \mathcal{Y}) = \phi(\bar{x}, X, M, \bar{y}, Y, N) \quad (3)$$

The measures μ that we investigate are derived from statistical hypothesis testing. They are constructed so that they are non-negative, i.e.

$$\forall \mathcal{X}, \forall \mathcal{Y}, \quad \mu(\mathcal{X}, \mathcal{Y}) \geq 0 \quad (4)$$

and they satisfy the property

$$\forall \mathcal{X}, \quad \mu(\mathcal{X}, \mathcal{X}) = 0 \quad (5)$$

In their basic forms, the measures are non-symmetric, but we propose several ways to symmetrize them, so that

$$\forall \mathcal{X}, \forall \mathcal{Y}, \quad \mu(\mathcal{X}, \mathcal{Y}) = \mu(\mathcal{Y}, \mathcal{X}) \quad (6)$$

2.3 Relative eigenvalues

We will denote as $\{\lambda_i\}_{1 \leq i \leq p}$ the eigenvalues of matrix Γ , i.e. the roots of the equation

$$\det[\Gamma - \lambda I] = 0 \quad (7)$$

where \det denotes the determinant, and I the unit matrix. Matrix Γ can be decomposed as

$$\Gamma = \Theta \Lambda \Theta^{-1} \quad (8)$$

where Λ is the $p \times p$ diagonal matrix of the eigenvalues, and Θ the $p \times p$ matrix of the eigenvectors. Classically, the eigenvalues λ_i are sorted in decreasing order when i increases.

Solutions of equation (7) are known as the eigenvalues of Y relative to X . Because X and Y are positive matrices, all eigenvalues λ_i are positive. Note also that the eigenvalues of X relative to Y (i.e. the eigenvalues of Γ^{-1}) are $\{1/\lambda_i\}_{1 \leq i \leq p}$.

2.4 Mean functions of the eigenvalues

Three particular functions of the eigenvalues λ_i are used in this article :

$$\text{The arithmetic mean : } a(\lambda_1, \dots, \lambda_p) = \frac{1}{p} \sum_{i=1}^p \lambda_i \quad (9)$$

$$\text{The geometric mean : } g(\lambda_1, \dots, \lambda_p) = \left(\prod_{i=1}^p \lambda_i \right)^{1/p} \quad (10)$$

$$\text{The harmonic mean : } h(\lambda_1, \dots, \lambda_p) = \left(\frac{1}{p} \sum_{i=1}^p \frac{1}{\lambda_i} \right)^{-1} \quad (11)$$

Because all eigenvalues λ_i are positive, it can be shown that

$$a \geq g \geq h \quad (12)$$

with equality if and only if all λ_i are equal. Moreover, swapping \mathcal{X} and \mathcal{Y} turns a into $1/h$, g into $1/g$ and h into $1/a$. In other words,

$$a\left(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_p}\right) = \frac{1}{h(\lambda_1, \dots, \lambda_p)} \quad (13)$$

$$g\left(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_p}\right) = \frac{1}{g(\lambda_1, \dots, \lambda_p)} \quad (14)$$

$$h\left(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_p}\right) = \frac{1}{a(\lambda_1, \dots, \lambda_p)} \quad (15)$$

2.5 Computation of a , g and h

Given that the trace (denoted tr) satisfies $tr(AB) = tr(BA)$ and that the determinant (denoted det) verifies $det(AB) = det A \cdot det B$, we have the following properties :

$$a(\lambda_1, \dots, \lambda_p) = \frac{1}{p} tr \Lambda = \frac{1}{p} tr \Gamma = \frac{1}{p} tr (YX^{-1}) \quad (16)$$

$$g(\lambda_1, \dots, \lambda_p) = (det \Lambda)^{1/p} = (det \Gamma)^{1/p} = \left(\frac{det Y}{det X} \right)^{1/p} \quad (17)$$

$$h(\lambda_1, \dots, \lambda_p) = \frac{p}{tr(\Lambda^{-1})} = \frac{p}{tr(\Gamma^{-1})} = \frac{p}{tr(XY^{-1})} \quad (18)$$

These equations show that functions a , g and h can be computed directly from X , Y , X^{-1} , Y^{-1} , $det X$ and $det Y$, without extracting explicitly the eigenvalues λ_i , nor calculating the matrix square roots of X and Y . Moreover, $tr(YX^{-1})$ and $tr(XY^{-1})$ can be computed without calculating the full matrix product, but only the diagonal elements of the product.

3 Second-order statistical measures

3.1 Gaussian likelihood measure

3.1.1 Definition

By supposing that all acoustic vectors extracted from the speech signal uttered by speaker \mathcal{X} are distributed like a Gaussian function, the likelihood of a single acoustic vector y_t uttered by speaker \mathcal{Y} is classically

$$G(y_t | \mathcal{X}) = \frac{1}{(2\pi)^{\frac{p}{2}} (\det X)^{\frac{1}{2}}} e^{-\frac{1}{2}(y_t - \bar{x})^T X^{-1}(y_t - \bar{x})} \quad (19)$$

If we assume that all vectors y_t are independent observations, the average log-likelihood of $\{y_t\}_{1 \leq t \leq N}$ can be written

$$\begin{aligned} \overline{G_{\mathcal{X}}}(y_1^N) &= \frac{1}{N} \log G(y_1 \dots y_n | \mathcal{X}) = \frac{1}{N} \sum_{t=1}^N \log G(y_t | \mathcal{X}) \\ &= -\frac{1}{2} \left[p \log 2\pi + \log(\det X) + \frac{1}{N} \sum_{t=1}^N (y_t - \bar{x})^T X^{-1}(y_t - \bar{x}) \right] \end{aligned} \quad (20)$$

By replacing $y_t - \bar{x}$ by $y_t - \bar{y} + \bar{y} - \bar{x}$ and using the property

$$\frac{1}{N} \sum_{t=1}^N (y_t - \bar{y})^T X^{-1}(y_t - \bar{y}) = \text{tr}(YX^{-1}) \quad (21)$$

we get

$$\overline{G_{\mathcal{X}}}(y_1^N) + \frac{p}{2} \log 2\pi = -\frac{1}{2} \left[\log(\det X) + \text{tr}(YX^{-1}) + (\bar{y} - \bar{x})^T X^{-1}(\bar{y} - \bar{x}) \right] \quad (22)$$

and

$$\begin{aligned} \frac{2}{p} \overline{G_{\mathcal{X}}}(y_1^N) + \log 2\pi + \frac{1}{p} \log(\det Y) + 1 \\ = \frac{1}{p} \left[\log\left(\frac{\det Y}{\det X}\right) - \text{tr}(YX^{-1}) - (\bar{y} - \bar{x})^T X^{-1}(\bar{y} - \bar{x}) \right] + 1 \end{aligned} \quad (23)$$

Therefore, if we define the Gaussian likelihood measure μ_G as

$$\mu_G(\mathcal{X}, \mathcal{Y}) = \frac{1}{p} \left[\text{tr}(YX^{-1}) - \log\left(\frac{\det Y}{\det X}\right) + (\bar{y} - \bar{x})^T X^{-1}(\bar{y} - \bar{x}) \right] - 1 \quad (24)$$

$$= \frac{1}{p} \left[\text{tr} \Gamma - \log(\det \Gamma) + \delta^T X^{-1} \delta \right] - 1 \quad (25)$$

$$= a - \log g + \frac{1}{p} \delta^T X^{-1} \delta - 1 \quad (26)$$

we have

$$\underset{\mathcal{X}}{\text{Argmax}} \overline{G_{\mathcal{X}}}(y_1^N) = \underset{\mathcal{X}}{\text{Argmin}} \mu_G(\mathcal{X}, \mathcal{Y}) \quad (27)$$

3.1.2 Properties of μ_G

Matrix X^{-1} being, like X , positive definite, $\delta^T X^{-1} \delta \geq 0$. Moreover, we have $\log g \leq g - 1$ and $a \geq g$. Therefore, $a - \log g - 1 \geq 0$ and $\mu_G(\mathcal{X}, \mathcal{Y}) \geq 0$. Measure $\mu_G(\mathcal{X}, \mathcal{Y}) = 0$ if and only if all eigenvalues λ_i are equal to 1 and δ is the null vector, i.e. if and only if $X = Y$ and $\bar{x} = \bar{y}$. However, $\mu_G(\mathcal{X}, \mathcal{Y})$ is non-symmetric, its dual term being

$$\mu_G(\mathcal{Y}, \mathcal{X}) = \frac{1}{h} + \log g + \frac{1}{p} \delta^T Y^{-1} \delta - 1 \neq \mu_G(\mathcal{X}, \mathcal{Y}) \quad (28)$$

3.1.3 A variant of μ_G

When dealing with noisy or distorted speech, the mean vectors \bar{x} and \bar{y} may be strongly influenced by the channel characteristics, while covariance matrices X and Y are usually more robust to variations between recording conditions and transmission lines [11]. Thus, the difference $\delta = \bar{y} - \bar{x}$ may be a misleading term in μ_G .

A Gaussian likelihood measure on the covariance matrices only, denoted here μ_{Gc} , can therefore be derived from the previous likelihood measure as

$$\mu_{Gc}(\mathcal{X}, \mathcal{Y}) = \frac{1}{p} \left[\text{tr}(Y X^{-1}) - \log \left(\frac{\det Y}{\det X} \right) \right] - 1 \quad (29)$$

$$= \frac{1}{p} [\text{tr} \Gamma - \log(\det \Gamma)] - 1 \quad (30)$$

$$= a - \log g - 1 \quad (31)$$

This measure can be expressed as a function of the eigenvalues λ_i of matrix Γ . However, it does not require an explicit extraction of the eigenvalues. It has the same properties as measure μ_G . In particular, it is still non-symmetric, since

$$\mu_{Gc}(\mathcal{Y}, \mathcal{X}) = \frac{1}{h} + \log g - 1 \neq \mu_{Gc}(\mathcal{X}, \mathcal{Y}) \quad (32)$$

3.2 Arithmetic-geometric sphericity measure

3.2.1 Definition

As presented by Anderson [1], a likelihood function for testing the proportionality of a covariance matrix Y to a given covariance matrix X is

$$S(Y | X) = \left[\frac{\det(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}})}{\left(\frac{1}{p} \text{tr}(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}})\right)^p} \right]^{\frac{N}{2}} = \left[\frac{\det \Gamma}{\left(\frac{1}{p} \text{tr} \Gamma\right)^p} \right]^{\frac{N}{2}} \quad (33)$$

This expression results from the combination of two criteria : one on the diagonality of matrix Γ , and a second one on the equality of the diagonal elements of Γ , given that Γ is diagonal.

By denoting as $\overline{S_X}(y_1^N)$ the average likelihood function for the sphericity test,

$$\overline{S_X}(y_1^N) = \frac{1}{N} \log S(Y | X) \quad (34)$$

and by defining

$$\mu_{Sc}(\mathcal{X}, \mathcal{Y}) = \log \left[\frac{\frac{1}{p} \text{tr } \Gamma}{(\det \Gamma)^{1/p}} \right] \quad (35)$$

$$= \log \left[\frac{\frac{1}{p} \text{tr}(Y X^{-1})}{\left(\frac{\det Y}{\det X}\right)^{1/p}} \right] \quad (36)$$

$$= \log \left(\frac{a}{g} \right) \quad (37)$$

we have

$$\underset{\mathcal{X}}{\text{Argmax}} \overline{S_X}(y_1^N) = \underset{\mathcal{X}}{\text{Argmin}} \mu_{Sc}(\mathcal{X}, \mathcal{Y}) \quad (38)$$

Measure μ_{Sc} appears as the logarithm of the ratio of the arithmetic and the geometric means of the eigenvalues of Y relative to X . As for measure μ_{Gc} , μ_{Sc} derives from a test on the covariance matrices only. It can be expressed as a function of the eigenvalues λ_i , but it does not require the search for the eigenvalues. The use of the arithmetic-geometric sphericity test for speaker recognition was initially proposed by Grenier [12], in the framework of text-dependent experiments.

3.2.2 Properties of μ_{Sc}

Since $a \geq g$, it is obvious that $\mu_{Sc}(\mathcal{X}, \mathcal{Y}) \geq 0$. Measure $\mu_{Sc}(\mathcal{X}, \mathcal{Y}) = 0$ if and only if all eigenvalues λ_i are equal, i.e. if and only if X and Y are proportional. In particular, $\mu_{Sc}(\mathcal{X}, \mathcal{X}) = 0$, but $X = Y$ is not a necessary condition. Finally, μ_{Sc} is not symmetric, and

$$\mu_{Sc}(\mathcal{Y}, \mathcal{X}) = \log \left(\frac{g}{h} \right) \neq \mu_{Sc}(\mathcal{X}, \mathcal{Y}) \quad (39)$$

3.3 Absolute deviation measure

3.3.1 Definition

The expression of μ_{Gc} and μ_{Sc} as functions of the eigenvalues λ_i are :

$$\mu_{Gc}(\mathcal{X}, \mathcal{Y}) = \frac{1}{p} \sum_{i=1}^p (\lambda_i - \log \lambda_i - 1) \quad (40)$$

$$\mu_{Sc}(\mathcal{X}, \mathcal{Y}) = \log \left(\frac{1}{p} \sum_{i=1}^p \lambda_i \right) - \frac{1}{p} \sum_{i=1}^p \log \lambda_i \quad (41)$$

As a matter of fact, it is possible to construct other metrics to measure the dissimilarity between speakers, through their covariance matrices. Any function of the eigenvalues λ_i , which is non-negative, and which takes the zero value when all eigenvalues are equal to unity, is a possible choice.

This approach was proposed by Gish [10], who constructed a measure which is based on the total absolute deviation of the eigenvalues from unity. The generic expression of this measure, which we will denote as μ_{Dc} , is

$$\mu_{Dc}(\mathcal{X}, \mathcal{Y}) = \frac{1}{p} \sum_{i=1}^p |\lambda_i - 1| \quad (42)$$

In this formulation, measure μ_{Dc} is the average absolute deviation of the eigenvalues λ_i from unity. Gish showed that robustness can be gained by removing large eigenvalues from the summation, because they may correspond to “abnormalities in small dimensional subspaces”.

3.3.2 Properties of μ_{Dc}

It can be easily checked that measure μ_{Dc} is non-negative, and that it is null if and only if covariance matrices X and Y are equal. The measure is non-symmetric, since

$$\mu_{Dc}(\mathcal{Y}, \mathcal{X}) = \frac{1}{p} \sum_{i=1}^p \left| \frac{1}{\lambda_i} - 1 \right| \neq \mu_{Dc}(\mathcal{X}, \mathcal{Y}) \quad (43)$$

4 Symmetrization

4.1 Motivation

All measures reviewed in the previous section have the common property of being non-symmetric. In other words, the roles played by the training data and by the test data are not interchangeable. However, our intuition would be that a similarity measure should be symmetric.

The asymmetry of measures μ_G , μ_{Gc} and μ_{Sc} can be explained by the following fact. These measures are based on statistical tests which suppose that the reference speaker model \mathcal{X} is exact, while the test model \mathcal{Y} is an estimation. But in practice, both reference and test models are estimates. Therefore, it is natural to search for a symmetric expression of originally asymmetric tests.

Moreover, it can be foreseen that the reliability of a reference model is dependent on the number of data that was used to estimate its parameters. This is experimentally confirmed by the discrepancies that can be observed in speaker identification performances, between $\mu(\mathcal{X}, \mathcal{Y})$ and $\mu(\mathcal{Y}, \mathcal{X})$, all the more as M and N , the number of reference and test vectors, are disproportionate (i.e. when $\rho = N/M$ is very different from 1).

4.2 Symmetrization procedures

A first possibility for symmetrizing a measure $\mu(\mathcal{X}, \mathcal{Y})$, is to construct the average between the measure and its dual term :

$$\mu_{[0.5]}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \mu(\mathcal{X}, \mathcal{Y}) + \frac{1}{2} \mu(\mathcal{Y}, \mathcal{X}) = \mu_{[0.5]}(\mathcal{Y}, \mathcal{X}) \quad (44)$$

For instance, the Gaussian likelihood measure, symmetrized in this manner, becomes

$$\mu_{G[0.5]}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \left[a + \frac{1}{h} + \frac{1}{p} \delta^T (X^{-1} + Y^{-1}) \delta \right] - 1 \quad (45)$$

which, for the covariance only measure, simplifies into

$$\mu_{Gc[0.5]}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \left(a + \frac{1}{h} \right) - 1 \quad (46)$$

while the arithmetic-geometric sphericity measure becomes proportional to the arithmetic-harmonic sphericity measure [4] :

$$\mu_{Sc[0.5]}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \log \left(\frac{a}{h} \right) \quad (47)$$

This procedure of symmetrization can improve the classification performance, compared to both asymmetric terms taken individually. This is the case when training and test patterns have comparable length. However, we observed an inefficiency, or a degradation of the performance when the lengths differed significantly ($\rho \not\approx 1$). When training and test patterns are obtained from speech utterances with very different lengths, it turns out that $\mu(\mathcal{X}, \mathcal{Y})$ performs better than $\mu(\mathcal{Y}, \mathcal{X})$ when $\rho \leq 1$, and conversely. In other words, when the amount of training data is significantly lower than the amount of test data, it is preferable to model the test data and compute the average likelihood of the training data for the test model, rather than doing the opposite.

In the lack of a rigorous theoretical framework, we have limited our investigations to empirical trials. We have postulated an arbitrary form for more general symmetric measures μ , i.e. linear combinations of the asymmetric terms, weighted by coefficients that are function of the number of training and test vectors (respectively M and N) :

$$\mu_{[\psi_{MN}]}(\mathcal{X}, \mathcal{Y}) = \psi_{MN} \cdot \mu(\mathcal{X}, \mathcal{Y}) + \psi_{NM} \cdot \mu(\mathcal{Y}, \mathcal{X}) \quad (48)$$

with

$$\psi_{MN} + \psi_{NM} = 1 \quad (49)$$

We have limited our tests to 2 particular functions ψ_{MN} and ψ_{NM} , namely :

$$\psi_{MN} = \alpha_{MN} = \frac{\sqrt{M}}{\sqrt{M} + \sqrt{N}} = \frac{1}{1 + \sqrt{\rho}} \quad (50)$$

$$\psi_{NM} = 1 - \alpha_{MN} = \frac{\sqrt{N}}{\sqrt{M} + \sqrt{N}} = \frac{\sqrt{\rho}}{1 + \sqrt{\rho}} \quad (51)$$

and

$$\psi_{MN} = \beta_{MN} = \frac{M}{M+N} = \frac{1}{1+\rho} \quad (52)$$

$$\psi_{NM} = 1 - \beta_{MN} = \frac{N}{M+N} = \frac{\rho}{1+\rho} \quad (53)$$

A similar approach was used by Montacié on AR-vector model residuals [17]. Note that, when $M \geq N$, $\rho \leq 1$ and therefore $0.5 \leq \alpha_{MN} \leq \beta_{MN}$.

We will not give the detailed expression of each measure, for each set of weights in this text. As an example, measure μ_G weighted by β_{MN} becomes

$$\mu_{G[\beta_{MN}]}(\mathcal{X}, \mathcal{Y}) = \frac{M \cdot \mu_G(\mathcal{X}, \mathcal{Y}) + N \cdot \mu_G(\mathcal{Y}, \mathcal{X})}{M+N} \quad (54)$$

$$\begin{aligned} &= \frac{1}{1+\rho} a - \frac{1-\rho}{1+\rho} \log g + \frac{\rho}{1+\rho} \frac{1}{h} \\ &\quad + \frac{1}{p} \delta^T \left(\frac{X^{-1} + \rho Y^{-1}}{1+\rho} \right) \delta - 1 \end{aligned} \quad (55)$$

$$\begin{aligned} &= \frac{1}{p} \left[\frac{1}{1+\rho} \text{tr}(YX^{-1}) - \frac{1-\rho}{1+\rho} \log \left(\frac{\det Y}{\det X} \right) + \frac{\rho}{1+\rho} \text{tr}(XY^{-1}) \right] \\ &\quad + \frac{1}{p} \left[(\bar{y} - \bar{x})^T \left(\frac{X^{-1} + \rho Y^{-1}}{1+\rho} \right) (\bar{y} - \bar{x}) \right] - 1 \end{aligned} \quad (56)$$

The symmetry of this expression can easily be checked.

Even though they are empirical, the symmetrizations using α_{MN} and β_{MN} provide generally better results than the symmetrization with weights equal to $\frac{1}{2}$. The optimal expression for symmetrized measures can certainly be derived from estimation theory, but it is not a trivial problem.

An exception to the general approach was applied to measure μ_{Dc} , since we experienced that it was slightly more efficient to symmetrize $\log \mu_{Dc}$ as above, instead of μ_{Dc} itself. However, $\log \mu_{Dc}$ can not be considered as a measure in the mathematical sense, since it is not non-negative. Therefore,

$$\log [\mu_{Dc[\psi_{MN}]}(\mathcal{X}, \mathcal{Y})] = \psi_{MN} \cdot \log [\mu_{Dc}(\mathcal{X}, \mathcal{Y})] + \psi_{NM} \cdot \log [\mu_{Dc}(\mathcal{Y}, \mathcal{X})] \quad (57)$$

which is equivalent to

$$\mu_{Dc[\psi_{MN}]}(\mathcal{X}, \mathcal{Y}) = \mu_{Dc}(\mathcal{X}, \mathcal{Y})^{\psi_{MN}} \cdot \mu_{Dc}(\mathcal{Y}, \mathcal{X})^{\psi_{NM}} \quad (58)$$

5 Experiments and results

5.1 Task

We have tested the measures described in this article, in the framework of closed-set text-independent speaker identification. There is a single reference per speaker (composed of a mean vector \bar{x} , a covariance matrix X and a number of data M). All test utterances are different from all training utterances, and all training utterances are different from one another. Each measure is evaluated as regards its classification ability using a 1-nearest neighbour decision rule. The possibility of rejection is not taken into account : the test speaker is always part of the set of references.

5.2 Databases

For our experiments, we used TIMIT and NTIMIT databases. TIMIT [8] contains 630 speakers (438 male and 192 female), each of them having uttered 10 sentences. Two sentences have the prefix “sa” (sa1 and sa2). Sentences sa1 and sa2 are different, but they are the same across speakers. Three sentences have the prefix “si” and five have the prefix “sx”. These 8 sentences are different from one another, and different across speakers. Sentences “sa” and “si” have an average duration of 2.9 seconds. Sentences “sx” have an average duration of 3.2 seconds. The speech signal is recorded through a high quality microphone, in a very quiet environment, with a 0-8 kHz bandwidth. The signal is sampled at 16 kHz, on 16 bits, on a linear amplitude scale. Moreover, all recordings took place in a single session (contemporaneous speech).

The NTIMIT database [15] was obtained by playing TIMIT speech signal through an artificial mouth installed in front of the microphone of a fixed handset frame and transmitting this input signal through a different telephone line for each sentence (local or long distance network). The signal is sampled at 16 kHz, but its useful bandwidth is limited to telephone bandwidth (approximately 300-3400 Hz). Each sample is represented on 16 bits (linear).

5.3 Signal analysis

Each sentence is analysed as followed : the speech signal is decomposed in frames of 504 samples (31.5 ms) at a frame rate of 160 samples (10 ms). A Hamming window is applied to each frame. The signal is not pre-emphasized. For each frame, a Winograd Fourier Transform is computed and provides 252 square module values representing the short term power spectrum in the 0-8 kHz band.

This Fourier power spectrum is then used to compute 24 filter bank coefficients. Each filter is triangular (except the first and last ones which have a rectangle trapezoidal shape). They are placed on a non-uniform frequency scale, similar to the Bark/Mel scale. The central frequency of the 24 filters are, in Hz : 47, 147, 257, 378, 510, 655, 813, 987, 1178, 1386, 1615, 1866, 2141, 2442, 2772, 3133, 3529, 3964, 4440, 4961, 5533, 6159, 6845, and 7597. Each filter covers a spectral range from the central frequency of the previous filter to the central frequency of the

next filter, with a maximum value of 1 for its own central frequency. For each frequency, only 2 filters (maximum) are non-zero, and their magnitudes add up to 1.

We finally take the base 10 logarithm of each filter output and multiply the result by 10, to form a 24-dimensional vector of filter bank coefficients in dB. For the TIMIT database, all 24 coefficients are kept, from which we compute, for each utterance a 24-dimensional mean vector and a 24×24 (symmetric) covariance matrix.

In order to simulate, for some of the experiments, a low-pass filtering of the speech signal in the 0-4 kHz band, we have simply discarded the last 7 coefficients of the 24-dimensional vectors obtained from the full band signal. The last filter, with index 17, has a central frequency of 3529 Hz, and becomes zero above 3964 Hz. This is the approach we used for NTIMIT database, since the useful bandwidth does not exceed 4000 Hz for these data. We also used this approach on TIMIT, in order to obtain results corresponding to a 0-4 kHz bandwidth, without the telephone line variability. We will refer to these data as FTIMIT data. Under these analysis conditions, each mean vector is 17 dimensional, while covariance matrices are 17×17 (symmetric) matrices.

5.4 Training and test protocols

We use 2 training protocols, namely a “long training” and a “short training”.

- For the “**long training**”, we use all 5 “sx” sentences concatenated together as a single reference pattern for each speaker. The average total duration of a “long training” pattern is **14.4 seconds**. A single reference (mean vector \bar{x} , covariance matrix X and number of vectors M) is computed for each speaker from all speech frames, represented as filter bank coefficients. In particular, no speech activity detector is used to remove silent speech portions.
- For the “**short training**”, we only use the first 2 “sx” sentences in alphanumeric order, in the same way as for the “long training”. The average total duration of a “short training” is **5.7 seconds** (including silences).

For the tests, we also have 2 distinct protocols : a “long test” and a “short test”.

- For the “**long test**”, all “sa” and “si” sentences (5 in total) are concatenated together as a single test pattern, for each speaker. In this framework, we therefore have a single test pattern per speaker, i.e. **630 test patterns** altogether. In average, each pattern lasts **15.9 seconds**.
- For the “**short test**”, each “sa” and “si” sentences are tested separately. The whole test set thus consists of $630 \times 5 =$ **3150 test patterns**, of **3.2 seconds** each, in average.

Even though the “sa” sentences are the same for each speaker, these sentences are used in the test set. Therefore, the experiments can be considered as totally text-independent.

5.5 Experiments

In the experiments reported in this article, we have systematically tested the 4 families of measures :

- μ_G
- μ_{Gc}
- μ_{Sc}
- μ_{Dc}

in two asymmetric forms :

- $\mu(\mathcal{X}, \mathcal{Y})$
- $\mu(\mathcal{Y}, \mathcal{X})$

as well as in the three symmetric forms proposed in section 4 :

- $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$
- $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$
- $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$

These evaluations were carried out on :

- TIMIT (24 filter bank coefficients between 0 and 8000 Hz)
- FTIMIT (first 17 filter bank coefficients from TIMIT between 0 and 4000 Hz)
- NTIMIT (17 filter bank coefficients between 0 and 4000 Hz)

It is reasonable to expect that the 3 databases, in this order, correspond to an increasing degree of difficulty.

In each case, we give the results for 4 possible training \times test protocols, corresponding to various typical values of the total amount of speech material per speaker $T = M + N$, of the ratio ρ between test and training material, and therefore to different weighting factors α_{MN} and β_{MN} :

- **“long-long”** protocol : long training \times long test
 $\bar{T} \approx 3000$ cs, $\bar{\rho} \approx 1.10$, $\bar{\alpha}_{MN} \approx 0.48$, $\bar{\beta}_{MN} \approx 0.49$
- **“short-long”** protocol : short training \times long test
 $\bar{T} \approx 2150$ cs, $\bar{\rho} \approx 2.79$, $\bar{\alpha}_{MN} \approx 0.26$, $\bar{\beta}_{MN} \approx 0.37$
- **“long-short”** protocol : long training \times short test
 $\bar{T} \approx 1750$ cs, $\bar{\rho} \approx 0.22$, $\bar{\alpha}_{MN} \approx 0.82$, $\bar{\beta}_{MN} \approx 0.68$
- **“short-short”** protocol : short training \times short test
 $\bar{T} \approx 900$ cs, $\bar{\rho} \approx 0.56$, $\bar{\alpha}_{MN} \approx 0.64$, $\bar{\beta}_{MN} \approx 0.57$

5.6 Results

The results are organized in 3 sets of 4 tables. The first set of tables (numbered I.1, I.2, I.3 and I.4) corresponds to results for TIMIT, the second set (Tables II.1 to II.4) for FTIMIT and the third set (Tables III.1 to III.4) for NTIMIT. The first table of each set (i.e. Tables I.1, II.1 and III.1) reports the results obtained for the “long-long” protocol, while the second one (I.2, II.2 and III.2) reports those for the “short-long” protocol. Similarly, the third and fourth tables of each set correspond respectively to the “long-short” and “short-short” protocols. In each table, the results relative to a given family of measures are organized in columns. The first line corresponds to the scores of both asymmetric terms (each cell is subdivided into 2), while the second, third and fourth lines show the results for the various symmetric forms. All results are given in terms of percentage of correct identification. Depending on this percentage S , and on the number of test patterns n , we give in Table 0 the half-width of the 95 % confidence interval, which is calculated as :

$$\pm 2 \sqrt{\frac{S \cdot (100 - S)}{n}}$$

Note that this quantity is the same for a score S and for $100 - S$.

| score : | S 100-S | 95 5 | 85 15 | 75 25 | 65 35 | 55 45 |
|----------------------|-----------|---------|---------|---------|---------|---------|
| long test, n = 630 | | ± 1.7 % | ± 2.8 % | ± 3.5 % | ± 3.8 % | ± 4.0% |
| short test, n = 3150 | | ± 0.8 % | ± 1.3 % | ± 1.5 % | ± 1.7 % | ± 1.8% |

Table 0 : *Half-width of the 95 % confidence interval for different values of the identification score S in %, corresponding to the long and short test protocols.*

We will not comment in detail each performance figure in Tables I, II and III, but we will rather try to underline several global trends.

For all measures, $\mu(\mathcal{X}, \mathcal{Y})$ and $\mu(\mathcal{Y}, \mathcal{X})$ perform differently. The term $\mu(\mathcal{X}, \mathcal{Y})$ performs better when the training speech material has a longer duration than the test material, and conversely. The discrepancy between the performances of the asymmetric terms is especially obvious for measure μ_{Dc} .

With non-distorted speech (TIMIT and FTIMIT), measure μ_G outperforms measure μ_{Gc} and all other measures on covariance matrices only. On the opposite, when channel variability is present (NTIMIT), the use of the mean vectors is, as expected, detrimental to the results.

In their asymmetric forms, the most efficient measure among the covariance-only measures is measure μ_{Sc} . However, when symmetrisation is applied, the performances tend to level-off, with a slight advantage for μ_{Dc} .

Among the symmetrization procedures that we tested, the most efficient one seems to be the one using weights β_{MN} and β_{NM} for μ_G , μ_{Gc} and μ_{Sc} , whereas α_{MN} and α_{NM} appear to be preferable for $\log \mu_{Dc}$.

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|-------|------------|--------|------------|-------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 100 % | 100 % | 100 % | 99.8 % | 100 % | 100 % | 99.5 % | 99.8 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 100 % | | 100 % | | 100 % | | 100 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 100 % | | 100 % | | 100 % | | 100 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 100 % | | 100 % | | 100 % | | 100 % | |

Table I.1: *long training (5 sentences ≈ 14.4 s) – long test (5 sentences ≈ 15.9 s)*
 $\bar{T} \approx 3000$ cs, $\bar{\rho} \approx 1.10$, $\bar{\alpha}_{MN} \approx 0.48$, $\bar{\beta}_{MN} \approx 0.49$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|--------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 93.2 % | 99.4 % | 86.7 % | 97.1 % | 94.9 % | 96.4 % | 73.3 % | 92.1 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 98.1 % | | 94.6 % | | 95.7 % | | 95.1 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 98.7 % | | 95.7 % | | 96.2 % | | 97.0 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 99.2 % | | 96.5 % | | 96.0 % | | 97.0 % | |

Table I.2: *short training (2 sentences ≈ 5.7 s) – long test (5 sentences ≈ 15.9 s)*
 $\bar{T} \approx 2150$ cs, $\bar{\rho} \approx 2.79$, $\bar{\alpha}_{MN} \approx 0.26$, $\bar{\beta}_{MN} \approx 0.37$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|--------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 97.9 % | 89.7 % | 96.2 % | 78.8 % | 97.3 % | 93.6 % | 83.6 % | 59.4 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 97.2 % | | 93.9 % | | 97.0 % | | 97.3 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 98.4 % | | 97.1 % | | 97.3 % | | 97.6 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 98.4 % | | 97.6 % | | 97.6 % | | 94.8 % | |

Table I.3: *long training (5 sentences ≈ 14.4 s) – short test (1 sentence ≈ 3.2 s)*
 $\bar{T} \approx 1750$ cs, $\bar{\rho} \approx 0.22$, $\bar{\alpha}_{MN} \approx 0.82$, $\bar{\beta}_{MN} \approx 0.68$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|--------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 83.8 % | 78.2 % | 73.5 % | 64.9 % | 81.9 % | 77.7 % | 52.9 % | 45.4 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 89.7 % | | 82.2 % | | 82.7 % | | 84.4 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 90.1 % | | 83.4 % | | 83.0 % | | 84.2 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 89.7 % | | 83.6 % | | 83.3 % | | 80.1 % | |

Table I.4: *short training (2 sentences ≈ 5.7 s) – short test (1 sentence ≈ 3.2 s)*
 $\bar{T} \approx 900$ cs, $\bar{\rho} \approx 0.56$, $\bar{\alpha}_{MN} \approx 0.64$, $\bar{\beta}_{MN} \approx 0.57$

Tables I.1, I.2, I.3, I.4 :

Text-independent speaker identification – TIMIT database (630 speakers).
The results are given in percentage of correct identification.

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|--------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 98.4 % | 99.4 % | 96.1 % | 98.3 % | 97.6 % | 98.1 % | 90.5 % | 95.4 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 99.4 % | | 97.9 % | | 97.9 % | | 98.3 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 99.5 % | | 97.9 % | | 97.9 % | | 98.6 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 99.5 % | | 97.9 % | | 97.8 % | | 98.4 % | |

Table II.1: *long training (5 sentences ≈ 14.4 s) – long test (5 sentences ≈ 15.9 s)*
 $\bar{T} \approx 3000$ cs, $\bar{\rho} \approx 1.10$, $\bar{\alpha}_{MN} \approx 0.48$, $\bar{\beta}_{MN} \approx 0.49$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|-------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 78.7% | 88.6% | 63.2 % | 77.0 % | 72.9 % | 76.4 % | 44.1 % | 65.6 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 87.0 % | | 76.8 % | | 76.4 % | | 76.7 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 87.9 % | | 77.5 % | | 76.2 % | | 77.6 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 89.0 % | | 77.8 % | | 76.4 % | | 76.2 % | |

Table II.2: *short training (2 sentences ≈ 5.7 s) – long test (5 sentences ≈ 15.9 s)*
 $\bar{T} \approx 2150$ cs, $\bar{\rho} \approx 2.79$, $\bar{\alpha}_{MN} \approx 0.26$, $\bar{\beta}_{MN} \approx 0.37$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|--------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 81.4 % | 67.9 % | 70.0 % | 49.8 % | 70.7 % | 66.3 % | 48.1 % | 33.3 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 81.8 % | | 67.3 % | | 70.4 % | | 72.2 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 84.2 % | | 71.8 % | | 71.7 % | | 73.1 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 83.6 % | | 72.6 % | | 72.0 % | | 64.4 % | |

Table II.3: *long training (5 sentences ≈ 14.4 s) – short test (1 sentence ≈ 3.2 s)*
 $\bar{T} \approx 1750$ cs, $\bar{\rho} \approx 0.22$, $\bar{\alpha}_{MN} \approx 0.82$, $\bar{\beta}_{MN} \approx 0.68$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|-------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 54.7% | 49.7% | 39.8 % | 32.2 % | 42.6 % | 41.2 % | 23.1 % | 20.6 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 61.4 % | | 43.9 % | | 44.4 % | | 46.5 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 61.8 % | | 45.3 % | | 44.5 % | | 46.8 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 61.4 % | | 45.8 % | | 44.4 % | | 43.6 % | |

Table II.4: *short training (2 sentences ≈ 5.7 s) – short test (1 sentence ≈ 3.2 s)*
 $\bar{T} \approx 900$ cs, $\bar{\rho} \approx 0.56$, $\bar{\alpha}_{MN} \approx 0.64$, $\bar{\beta}_{MN} \approx 0.57$

Tables II.1, II.2, II.3, II.4 :

Text-independent speaker identification – FTIMIT database (630 speakers).
The results are given in percentage of correct identification.

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|--------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 45.3 % | 50.3 % | 59.5 % | 63.0 % | 66.0 % | 64.9 % | 41.0 % | 51.0 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 49.4 % | | 63.0 % | | 66.4 % | | 67.9 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 49.0 % | | 63.0 % | | 66.5 % | | 68.4 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 49.4 % | | 63.6 % | | 66.5 % | | 68.6 % | |

Table III.1: *long training (5 sentences ≈ 14.4 s) – long test (5 sentences ≈ 15.9 s)*
 $\bar{T} \approx 3000$ cs, $\bar{\rho} \approx 1.10$, $\bar{\alpha}_{MN} \approx 0.48$, $\bar{\beta}_{MN} \approx 0.49$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|--------|------------|--------|------------|--------|------------|--------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 17.6 % | 24.9 % | 22.2 % | 31.0 % | 28.4 % | 29.7 % | 12.4 % | 22.5 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 24.4 % | | 29.5 % | | 29.8 % | | 30.3 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 24.8 % | | 30.5 % | | 30.0 % | | 30.8 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 25.7 % | | 31.3 % | | 30.5 % | | 30.8 % | |

Table III.2: *short training (2 sentences ≈ 5.7 s) – long test (5 sentences ≈ 15.9 s)*
 $\bar{T} \approx 2150$ cs, $\bar{\rho} \approx 2.79$, $\bar{\alpha}_{MN} \approx 0.26$, $\bar{\beta}_{MN} \approx 0.37$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|--------|------------|--------|------------|--------|------------|-------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 20.7 % | 13.8 % | 25.4 % | 13.5 % | 26.3 % | 23.0 % | 14.1 % | 5.2 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 19.3 % | | 23.4 % | | 25.2 % | | 25.2 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 21.1 % | | 25.4 % | | 26.1 % | | 26.4 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 21.4 % | | 26.1 % | | 26.7 % | | 20.5 % | |

Table III.3: *long training (5 sentences ≈ 14.4 s) – short test (1 sentence ≈ 3.2 s)*
 $\bar{T} \approx 1750$ cs, $\bar{\rho} \approx 0.22$, $\bar{\alpha}_{MN} \approx 0.82$, $\bar{\beta}_{MN} \approx 0.68$

| Measures | | μ_G | | μ_{Gc} | | μ_{Sc} | | μ_{Dc} | |
|---|---------------------------------|---------|-------|------------|-------|------------|--------|------------|-------|
| $\mu(\mathcal{X}, \mathcal{Y})$ | $\mu(\mathcal{Y}, \mathcal{X})$ | 10.1 % | 9.0 % | 12.0 % | 8.9 % | 13.7 % | 12.8 % | 6.4 % | 3.1 % |
| $\mu_{[0.5]}(\mathcal{X}, \mathcal{Y})$ | | 11.7 % | | 13.7 % | | 14.3 % | | 14.4 % | |
| $\mu_{[\alpha_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 11.7 % | | 14.2 % | | 14.4 % | | 14.8 % | |
| $\mu_{[\beta_{MN}]}(\mathcal{X}, \mathcal{Y})$ | | 11.6 % | | 15.0 % | | 14.4 % | | 12.7 % | |

Table III.4: *short training (2 sentences ≈ 5.7 s) – short test (1 sentence ≈ 3.2 s)*
 $\bar{T} \approx 900$ cs, $\bar{\rho} \approx 0.56$, $\bar{\alpha}_{MN} \approx 0.64$, $\bar{\beta}_{MN} \approx 0.57$

Tables III.1, III.2, III.3, III.4 :

Text-independent speaker identification – NTIMIT database (630 speakers).
The results are given in percentage of correct identification.

The positive effect of symmetrization is important when little speech material is available. The most significant differences are observed for the short training \times short test protocol. Table IV gives orders of magnitude of the relative error rate reduction between the asymmetric measures and their best symmetric version. If S is the percentage of correct identification for the asymmetric measure and S' is the percentage of correct identification for the symmetric measure, the relative error rate reduction is calculated as :

$$\frac{S' - S}{100 - S}$$

This relative improvement is given for the two protocols using short duration test data only. For the two others, the statistical significance of the observed differences are too small to be conclusive, given the larger confidence interval.

| measure | μ_G | μ_{Gc} | μ_{Sc} | μ_{Dc} |
|---------|--------------|--------------|--------------|--------------|
| TIMIT | $\sim 30 \%$ | $\sim 40 \%$ | $\sim 10 \%$ | $\sim 75 \%$ |
| FTIMIT | $\sim 15 \%$ | $\sim 10 \%$ | $\sim 5 \%$ | $\sim 40 \%$ |
| NTIMIT | $\sim 1 \%$ | $\sim 1 \%$ | $< 1 \%$ | $\sim 10 \%$ |

Table IV : *Order of magnitude of the relative error rate reduction between asymmetric and symmetric measures. Results for short test protocols only.*

These results show that symmetrization improves covariance-only measures (μ_{Gc} , μ_{Sc} and μ_{Dc}) as the task becomes intrinsically less difficult (TIMIT > FTIMIT > NTIMIT), and as the original asymmetric measures perform less well ($\mu_{Dc} < \mu_{Gc} < \mu_{Sc}$). On the other hand, when the Gaussian speaker model is not powerful enough for the task (NTIMIT), or when the asymmetric measure is quite efficient (μ_{Sc}), symmetrization is less useful.

6 Discussion

Our evaluations show that remarkable performances can be obtained on the TIMIT database for text-independent closed-set speaker identification (630 speakers) by second-order statistical measures, i.e. with a very simple underlying speaker model. Therefore, TIMIT is certainly an easy database for speaker recognition, and the measures exposed in this article work very well, on this database. Naturally, their overall performances degrade with more adverse conditions : a significant amount of speaker characteristics seems to be contained in the 4–8 kHz band, since FTIMIT results are significantly worse than TIMIT results. The effect of telephone channel distortion and variability are the cause of an even more severe drop on NTIMIT recognition scores. The effect of temporal drift owed to multisession recordings can not be studied with TIMIT derived data, but it is easy to predict an additional negative role of this factor on the performances. If second-order statistical measures are clearly efficient for relatively simple tasks, they are obviously not the ultimate solution to speaker recognition for any kind of applications.

6.1 Beyond the performances

However, second-order statistical measures have several advantages. They are simple to implement and easy to reproduce. Moreover, Gaussian likelihood measures (μ_G and μ_{Gc}) in their asymmetric forms are particular cases of several general approaches frequently used in text-independent speaker recognition. A 1-Gaussian speaker model is equivalent to a Vector Quantization codebook with 1 entry associated with a Mahalanobis distance. It is also equivalent to any kind of Hidden Markov Model (Left-to-Right, Ergodic,...) with 1 state and 1 Gaussian distribution. It is a particular case of a k-Gaussian Mixture model with $k = 1$. Finally, the likelihood criterion is often used on vector prediction residuals obtained from linear or connectionist models for which the identity model (0^{th} -order prediction) is a particular case.

Therefore, μ_G and μ_{Gc} are at the intersection of several classical approaches, which are extensions of this basic model in various directions (variations of the distance measure, use of more or less strong temporal constraints, refinement of the speaker distribution model, filtering of the acoustic parameters,...). Given the extreme simplicity of the second-order statistical measures, we therefore suggest that any speaker recognition task could be systematically benchmarked by one or two of these measures, in order to obtain a reference score indicating the intrinsic complexity of the chosen database and protocol. In particular, the preprocessings, the acoustic analysis, the training and test splitting of the data, and the decision strategy for the method under test should be identically used for the benchmark method.

6.2 A possible reference approach

Even though asymmetric Gaussian likelihood based measures do not systematically perform better than other second-order statistical measures, μ_G and μ_{Gc} may be preferable as reference benchmark measures in two cases : when they are compared with other asymmetric approaches (which is the case for VQ, HMM and Gaussian Mixtures), and when the length of training material is significantly higher than the length of test material. The choice between μ_G and μ_{Gc} should be guided by the processing that is applied to the data for the system under evaluation : whether, for this particular protocol, the long term average is subtracted or not to the acoustic parameters. Measures $\mu_{G[\beta_{MN}]}$ or $\mu_{Gc[\beta_{MN}]}$ can also be implemented simply and could be systematically tested. However, the lack of theoretical justification for these measures, and the relatively small improvement they provide as soon as a reasonable amount of speech material is available, make it more debatable. Nevertheless, if the approach under test is formally symmetric, it would be fair to compare it to a symmetric reference measure.

7 Conclusion

The goal of this work has been multiple. Firstly, to investigate the properties and performances of simple speaker recognition approaches, to compare them and to identify their limits. Our large scale evaluation on TIMIT, on a low-pass filtered version of TIMIT and on NTIMIT illustrates clearly that speech quality and quantity are major factors of performance, and that on high quality contemporaneous speech, simple and fast methods can be extremely efficient. For instance, this type of approach may prove sufficient for applications such as the automatic speaker labeling of radio or television recordings, for which the signal quality is constant and the voice drift relatively marginal.

Secondly, our work illustrates the extreme caution with which any conclusion can be drawn on the merit of a given method outside of any point of comparison. Since it may not be feasible to compare any new method with all state-of-the-art approaches, it is at least desirable to benchmark the task with a simple and general reference approach.

Thirdly, we believe that second-order statistical tests and measures, based on the Gaussian likelihood scoring are a good compromise as reference measures, since they are easy to implement, simple to reproduce, inexpensive in computation and light in storage requirements. Moreover, they appear, in their asymmetric forms, as simpler versions of more elaborate approaches. Though symmetrization is not a systematic factor of improvement, symmetric versions of the measures could be tested as well, especially in comparison with other symmetric methods. More theoretical work on symmetrization is however needed to find optimal symmetric forms.

The systematical use of a reference method in order to calibrate the complexity of a speaker recognition task can only result from a consensus between researchers both on the concept of a benchmark evaluation by a common approach and on the choice of the reference algorithm itself. We hope that this article will contribute to widen the concertation that had started during the SAM-A European ESPRIT project, dedicated to speech assessment methodology.

References

- [1] T. W. Anderson. *An Introduction to Multivariate Analysis*. John Wiley and Sons, 1958.
- [2] T. Artières, Y. Bennani, P. Gallinari, and C. Montacié. Connectionist and conventional models for text-free talker identification tasks. In *Proceedings of NEURONIMES 91*, 1991. Nîmes, France.
- [3] Y. Bennani. Speaker identification through a modular connectionist architecture: evaluation on the TIMIT database. In *ICSLP 92*, volume 1, pages 607–610, Oct. 1992. Banff, Canada.
- [4] F. Bimbot and L. Mathan. Text-free speaker recognition using an arithmetic-harmonic sphericity measure. In *EUROSPEECH 93*, volume 1, pages 169–172, Sept. 1993. Berlin, Germany.
- [5] F. Bimbot, L. Mathan, A. de Lima, and G. Chollet. Standard and target-driven AR-vector models for speech analysis and speaker recognition. In *ICASSP 92*, volume 2, pages II.5–II.8, Mar. 1992. San Francisco, United-States.
- [6] F. Bimbot, A. Paoloni, and G. Chollet. *Assessment Methodology for Speaker Identification and Verification Systems*. Technical report – Task 2500 – Report I9, SAM-A ESPRIT Project 6819, 1993.
- [7] G. Chollet and C. Gagnoulet. On the evaluation of speech recognizers and data bases using a reference system. In *ICASSP 82*, volume 3, pages 2026–2029, May 1982. Paris, France.
- [8] W. M. Fisher, G. R. Doddington, and K. M. Goudie-Marshall. The DARPA speech recognition research database : specifications and status. In *Proceedings of the DARPA workshop on speech recognition*, pages 93–99, Feb. 1986.
- [9] S. Furui. An overview of speaker recognition technology. In *Workshop on automatic speaker recognition, identification and verification*, pages 1–9, Apr. 1994. Martigny, Switzerland.
- [10] H. Gish. Robust discrimination in automatic speaker identification. In *ICASSP 90*, volume 1, pages 289–292, Apr. 1990. New Mexico, United-States.
- [11] H. Gish, M. Krasner, W. Russell, and J. Wolf. Methods and experiments for text-independent speaker recognition over telephone channels. In *ICASSP 86*, volume 2, pages 865–868, Apr. 1986. Tokyo, Japan.
- [12] Y. Grenier. *Identification du locuteur et adaptation au locuteur d'un système de reconnaissance phonémique*. PhD thesis, ENST, 1977.
- [13] Y. Grenier. Utilisation de la prédiction linéaire en reconnaissance et adaptation au locuteur. In *XIèmes Journées d'Etude sur la Parole*, pages 163–171, May 1980. Strasbourg, France.

- [14] H. Hattori. Text-independent speaker recognition using neural networks. In *ICASSP 92*, volume 2, pages 153–156, Mar. 1992. San Francisco, United-States.
- [15] C. Jankowski, A. Kalyanswamy, S. Basson, and J. Spitz. NTIMIT: a phonetically balanced, continuous speech, telephone bandwidth speech database. *ICASSP 90*, Apr. 1990. New Mexico, United-States.
- [16] C. Montacié, P. Deléglise, F. Bimbot, and M.-J. Caraty. Cinematic techniques for speech processing: temporal decomposition and multivariate linear prediction. In *ICASSP 92*, volume 1, pages 153–156, Mar. 1992. San Francisco, United-States.
- [17] C. Montacié and J.-L. Le Floch. AR-vector models for free-text speaker recognition. In *ICSLP 92*, volume 1, pages 611–614, Oct. 1992. Banff, Canada.
- [18] A. B. Poritz. Linear predictive Hidden Markov Models and the speech signal. In *ICASSP 82*, pages 1291–1294, May 1982. Paris, France.
- [19] D. A. Reynolds. Speaker identification and verification using Gaussian Mixture speaker models. In *Workshop on automatic speaker recognition, identification and verification*, pages 27–30, Apr. 1994. Martigny, Switzerland.
- [20] L. Rudasi and S. A. Zahorian. Text-independent talker identification with neural networks. In *ICASSP 91*, volume 1, pages 389–392, 1991. Toronto, Canada.
- [21] M. Savic and S. K. Gupta. Variable parameter speaker verification system based on hidden markov modeling. In *ICASSP 90*, volume 1, pages 281–284, Apr. 1990. Albuquerque, New Mexico, United-States.
- [22] F. K. Soong, A. E. Rosenberg, and B.-H. Juang. A Vector Quantization approach to speaker recognition. Technical Report 3, AT&T, Mar. 1987.

A Further Investigation on AR-Vector Models for Text-Independent Speaker Identification (1996)

Ivan Magrin-Chagnolleau, Joachim Wilke, and Frédéric Bimbot. **A Further Investigation on AR-Vector Models for Text-Independent Speaker Identification.** *Proceedings of ICASSP 96*, Vol. 1, pp. 401-404, Atlanta, United States, May 1996.

4 pages

A FURTHER INVESTIGATION ON AR-VECTOR MODELS FOR TEXT-INDEPENDENT SPEAKER IDENTIFICATION

Ivan MAGRIN-CHAGNOLLEAU Joachim WILKE Frédéric BIMBOT

Télécom Paris (E.N.S.T.), Dépt. Signal – C.N.R.S., URA 820
46, rue Barrault – 75634 Paris cedex 13 – FRANCE – European Union
email: ivan@sig.enst.fr and bimbot@sig.enst.fr

ABSTRACT

In this paper, we investigate on the role of dynamic information on the performances of AR-vector models for speaker recognition. To this purpose, we design an experimental protocol that destroys the time structure of speech frame sequences, which we compare to a more conventional one, i.e. keeping the natural time order. These results are also compared with those obtained with a (single) Gaussian model. Several measures are systematically investigated in the three cases, and different ways of symmetrisation are tested. We observe that the destruction of the time order can be a factor of improvement for the AR-vector models, and that results obtained with the Gaussian model are merely always better. In most cases, symmetrisation is beneficial.

1. INTRODUCTION

Auto-Regressive (AR) Vector Models have been a significant subject of interest in the field of Speaker Recognition [1] [2] [3] [4] [5] [6] [7]. Whereas the idea of modeling a speaker by an AR-vector model estimated on sequences of speech frames is common to these works, the way to measure the similarity between two speaker models is addressed very differently. Secondly, the use of AR-vector model is often motivated by the belief that such an approach is an efficient way to extract dynamic speaker characteristics, as opposed to static characteristics such as the distribution of speech frame parameters.

In this paper we report on a systematic investigation on similarity measures between AR-vector speaker models obtained as simple combinations of canonical quantities. We also design a protocol in order to examine the role of dynamic information on the performance of the AR-vector approach : we destroy the natural time order of speech frames by shuffling them randomly, and we evaluate the AR-vector approach on these temporally disorganised data. We finally compare both previous approaches to a (single) Gaussian Model [8] [9] [10] [11].

2. DEFINITIONS AND NOTATION

Let $\{\mathbf{x}_t\}_{1 \leq t \leq M}$ be a sequence of p -dimensional vectors. Let us define the centered vectors $\mathbf{x}_t^* = \mathbf{x}_t - \bar{\mathbf{x}}$ where $\bar{\mathbf{x}}$ is the mean vector of $\{\mathbf{x}_t\}$.

Let us denote \mathcal{X}_0 the covariance matrix of $\{\mathbf{x}_t\}$:

$$\mathcal{X}_0 = \frac{1}{M} \sum_{t=1}^M (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_t - \bar{\mathbf{x}})^T = \frac{1}{M} \sum_{t=1}^M \mathbf{x}_t^* \cdot \mathbf{x}_t^{*T}$$

We also define as \mathcal{X}_k the lagged covariance matrices :

$$\mathcal{X}_k = \frac{1}{M} \sum_{t=k+1}^M \mathbf{x}_t^* \cdot \mathbf{x}_{t-k}^{*T} \quad \text{with } k = 1, \dots, q$$

and the Toeplitz matrix X :

$$X = \begin{bmatrix} \mathcal{X}_0 & \mathcal{X}_1 & \dots & \mathcal{X}_q \\ \mathcal{X}_1^T & \mathcal{X}_0 & \dots & \mathcal{X}_{q-1} \\ \vdots & \vdots & \dots & \vdots \\ \mathcal{X}_q^T & \mathcal{X}_{q-1}^T & \dots & \mathcal{X}_0 \end{bmatrix}$$

A q -th order AR-vector model of sequence $\{\mathbf{x}_t^*\}$ is classically written as :

$$\sum_{i=0}^q A_i \cdot \mathbf{x}_{t-i}^* = \mathbf{e}_t \quad \text{with } A_0 = I_p$$

where $\{A_i\}$ is a set of $q+1$ matrix prediction coefficients, and \mathbf{e}_t is the prediction error vector. $\{A_1, \dots, A_q\}$ are obtained by solving the vector Yule-Walker equation [12]. With $A = [A_0 \dots A_q]$, the covariance matrix of the residual of $\{\mathbf{x}_t^*\}$ filtered by A is :

$$E_X^{(A)} = AXA^T$$

Similarly, for a signal $\{\mathbf{y}_t\}_{1 \leq t \leq N}$ with model B , we will denote :

$$E_Y^{(B)} = BYB^T$$

If we now consider :

$$\begin{aligned} E_X^{(B)} &= BXB^T \\ E_Y^{(A)} &= AY A^T \end{aligned}$$

these matrices can be interpreted as the covariance matrix of the filtering of $\{\mathbf{x}_t^*\}$ by B , and vice-versa. As A is obtained by minimising $tr(E_X^{(A)})$ and B by minimising $tr(E_Y^{(B)})$, we have $tr(E_X^{(B)}) \geq tr(E_X^{(A)})$ and $tr(E_Y^{(A)}) \geq tr(E_Y^{(B)})$.

Let us finally define $\Gamma_X^{(B/A)}$ and $\Gamma_{Y/X}^{(A)}$ as :

$$\begin{aligned} \Gamma_X^{(B/A)} &= \left(E_X^{(A)}\right)^{-\frac{1}{2}} \cdot E_X^{(B)} \cdot \left(E_X^{(A)}\right)^{-\frac{1}{2}} \\ \Gamma_{Y/X}^{(A)} &= \left(E_X^{(A)}\right)^{-\frac{1}{2}} \cdot E_Y^{(A)} \cdot \left(E_X^{(A)}\right)^{-\frac{1}{2}} \end{aligned}$$

| function f | a | $\log a$ | g | $\log g$ | $a - \log g - 1$ | $\log(a/g)$ | $a - g$ |
|---|-------------|--------------|-------------|-------------|-------------------|------------------|--------------|
| AR-vector model - spectral frames in their natural time order | | | | | | | |
| $f_X^{(B/A)} f_Y^{(A/B)}$ | 16.8 8.6 | 16.8 8.6 | 16.2 7.6 | 16.2 7.6 | 19.1 10.8 | 23.8 19.4 | 22.2 17.5 |
| symmetrised | 3.5 • | 4.1 • | 4.1 • | 4.1 • | 3.2 • | 7.9 • | 7.3 • |
| $f_{Y/X}^{(A)} f_{X/Y}^{(B)}$ | 75.6 51.4 | 75.6 51.4 | 88.3 73.0 | 88.3 73.0 | 15.2 34.3 | 7.6 18.7 | 15.2 14.6 |
| symmetrised | 6.0 * | 4.8 * | 12.4 * | 4.8 * | 5.4 ° | 7.0 ° | 6.0 ° |
| AR-vector model - spectral frames in a random time order | | | | | | | |
| $f_{X'}^{(B'/A')} f_{Y'}^{(A'/B')}$ | 2.5 56.5 | 2.5 56.5 | 4.1 58.1 | 4.1 58.1 | 2.5 56.2 | 4.1 55.9 | 3.5 54.6 |
| symmetrised | 3.5 ° | 3.5 ° | 5.7 ° | 5.7 ° | 2.5 ° | 4.1 ° | 4.1 ° |
| $f_{Y'/X'}^{(A')} f_{X'/Y'}^{(B')}$ | 42.5 45.4 | 42.5 45.4 | 98.1 82.9 | 98.1 82.9 | 1.3 22.9 | 1.0 6.7 | 3.2 8.9 |
| symmetrised | 4.8 * | 2.2 * | 46.7 * | 12.7 * | 2.9 ° | 1.0 ° | 1.6 ° |
| Gaussian model | | | | | | | |
| $f_{Y_0/X_0}^{(I)} f_{X_0/Y_0}^{(I)}$ | 37.5 47.0 | 37.5 47.0 | 98.4 98.4 | 98.4 98.4 | 0.6 7.9 | 0.6 3.2 | 2.9 6.4 |
| symmetrised | 3.8 * | 1.3 * | 97.1 * | 99.4 * | 1.0 ° | 0.6 ° | 1.0 ° |

Table 1. TIMIT - Speaker identification error rates

where $E^{\frac{1}{2}}$ is the symmetric square root matrix of E . The first matrix can be interpreted as the covariance matrix of $\{\mathbf{x}_i^*\}$ filtered by B relative to the one of $\{\mathbf{x}_i^*\}$ filtered by A , and the second one as the covariance matrix of $\{\mathbf{y}_i^*\}$ filtered by A relative to the one of $\{\mathbf{x}_i^*\}$ filtered by A .

3. SPEAKER MODELS

The purpose of this paper is to investigate on different ways of using an AR-vector model for speaker identification. A speaker is characterised by a second-order AR-vector model ($q = 2$) estimated on some speech material training. The matrix prediction coefficients $\{A_1, A_2\}$ are obtained by solving the vector Yule-Walker equation in the case $q = 2$:

$$[A_1 \ A_2] \cdot \begin{bmatrix} \mathcal{X}_0 & \mathcal{X}_1 \\ \mathcal{X}_1^T & \mathcal{X}_0 \end{bmatrix} = -[\mathcal{X}_1^T \ \mathcal{X}_2^T]$$

- A first model is a 2nd-order AR-vector model trained on speech frames presented in their natural time order. Therefore, the model of \mathcal{X} is $\{A, X\}$.
- A second model is a 2nd-order AR-vector model trained on the same speech frames as previously, but presented in a random time order. Each speaker \mathcal{X} is characterised by $\{A', X'\}$ which are obtained in the same way as $\{A, X\}$, after speech frames have been randomly shuffled.

Gaussian speaker model is also tested as a reference model. In this second framework, a speaker \mathcal{X} is represented by the covariance matrix \mathcal{X}_0 . It is equivalent to a 0th-order AR-vector model, i.e. $A = [A_0] = I_p$ and $X = [\mathcal{X}_0]$, which we will denote as $\{I, X_0\}$.

4. SIMILARITY MEASURES

We consider now 2 speakers \mathcal{X} and \mathcal{Y} , and we present a general formalism for expressing similarity measures between their AR-vector models.

Two families of similarity measures are investigated :

$$f_X^{(B/A)}(\mathcal{X}, \mathcal{Y}) = f\left(\Gamma_X^{(B/A)}\right)$$

$$f_{Y/X}^{(A)}(\mathcal{X}, \mathcal{Y}) = f\left(\Gamma_{Y/X}^{(A)}\right)$$

The first family can be interpreted as a measure between two models (A and B), via their influence on the same vector signal (X). This family of measures (which we will refer to as VI), generalises the Itakura measure to the vector case [13]. Examples of such measures are proposed in [4] and [6]. On the opposite, the second family can be viewed as a measure between two signals (X and Y) filtered by a common model (A). Some of the IS measures proposed in [3] [5] belong to this family. Note also that setting $\{A, X\} = \{I, X_0\}$ allows to construct a similar family of measures for the Gaussian model.

The function f is chosen equal to a combination of the following canonical quantities :

$$a(\Gamma) = \frac{1}{p} \text{tr}(\Gamma)$$

$$g(\Gamma) = [\det(\Gamma)]^{\frac{1}{p}}$$

It can be shown that a and g are positive and that $a \geq g$. Moreover these quantities can be computed very efficiently [11]. The composed functions $a - \log g - 1$ and $\log(a/g)$ are respectively the Maximum-Likelihood measure [9] and the Arithmetic-Geometric Sphericity measure [8].

As these measures are not symmetric, different symmetrisations can be applied on the original measures. Given $f_X^{(B/A)}$ and $f_Y^{(A/B)}$, we define :

$$f_X^{(B/A)*} = \frac{1}{2} f_X^{(B/A)} + \frac{1}{2} f_Y^{(A/B)}$$

$$f_X^{(B/A)^\circ} = \frac{\bar{M}}{\bar{M} + \bar{N}} f_X^{(B/A)} + \frac{\bar{N}}{\bar{M} + \bar{N}} f_Y^{(A/B)}$$

$$f_X^{(B/A)^\bullet} = \frac{\bar{N}}{\bar{M} + \bar{N}} f_X^{(B/A)} + \frac{\bar{M}}{\bar{M} + \bar{N}} f_Y^{(A/B)}$$

| function f | a | $\log a$ | g | $\log g$ | $a - \log g - 1$ | $\log(a/g)$ | $a - g$ |
|---|-------------|---------------|-------------|-------------|--------------------|--------------------|---------------|
| AR-vector model - spectral frames in their natural time order | | | | | | | |
| $f_X^{(B/A)} \mid f_Y^{(A/B)}$ | 38.7 30.2 | 38.7 30.2 | 37.1 29.5 | 37.1 29.5 | 42.5 35.2 | 51.1 50.8 | 49.5 49.5 |
| symmetrised | 24.8 • | 25.1 • | 24.8 • | 24.4 • | 26.3 • | 35.6 • | 33.3 • |
| $f_{Y/X}^{(A)} \mid f_{X/Y}^{(B)}$ | 93.3 86.0 | 93.3 86.0 | 96.5 94.6 | 96.5 94.6 | 44.1 69.8 | 41.6 39.1 | 49.2 39.1 |
| symmetrised | 23.5 * | 21.3 * | 32.4 * | 25.4 * | 24.4 ° | 34.6 ° | 33.0 ° |
| AR-vector model - spectral frames in a random time order | | | | | | | |
| $f_{X'}^{(B'/A')} \mid f_{Y'}^{(A'/B')}$ | 35.9 82.2 | 35.9 82.2 | 36.8 81.3 | 36.8 81.3 | 32.4 83.5 | 34.6 82.2 | 34.3 81.6 |
| symmetrised | 39.1 ° | 39.1 ° | 40.0 ° | 40.0 ° | 34.3 ° | 33.3 ° | 33.3 ° |
| $f_{Y'/X'}^{(A')} \mid f_{X'/Y'}^{(B')}$ | 78.7 71.4 | 78.7 71.4 | 98.4 93.7 | 98.4 93.7 | 15.9 43.8 | <u>13.3</u> 21.6 | 20.3 27.3 |
| symmetrisation | 21.9 * | <u>14.6</u> * | 69.8 * | 52.4 * | <u>14.0</u> ° | <u>13.3</u> ° | <u>14.3</u> ° |
| Gaussian model | | | | | | | |
| $f_{Y_o/X_o}^{(I)} \mid f_{X_o/Y_o}^{(I)}$ | 77.1 71.8 | 77.1 71.8 | 98.4 98.4 | 98.4 98.4 | <u>14.6</u> 27.3 | <u>12.7</u> 17.1 | 20.3 21.3 |
| symmetrised | 15.6 * | <u>11.8</u> * | 97.8 * | 98.4 * | <u>12.7</u> ° | <u>12.4</u> ° | <u>14.3</u> ° |

Table 2. FTIMIT - Speaker identification error rates

\bar{M} is the average number of frames for the training sentences across all speakers, and \bar{N} is the average number of frames for the test sentences. The same symmetrisations are applied to $f_{Y/X}^{(A)}$ and $f_{X/Y}^{(B)}$.

5. DATABASE AND SIGNAL ANALYSIS

We use the first 63 speakers of TIMIT [14] and NTIMIT [15] for our experiments (19 females and 44 males)¹. Each of them has read 10 sentences. The signal is sampled at 16 kHz, on 16 bits, on a linear amplitude scale. NTIMIT is a telephone-channel version of TIMIT.

Each sentence is analysed as follows : for each speech token, the speech signal is kept in its integrality; it is decomposed into frames of 31.5 ms at a frame rate of 10 ms, with no pre-emphasis. A Hamming window is applied to each frame. Then the module of a 504 point Fourier Transform is computed, from which 24 Mel-scale triangular filter bank coefficients are extracted. The spectral vectors $\{\mathbf{x}_t\}$ (of dimension $p = 24$) are formed from the logarithm of each filter output. These analysis conditions are identical to those used in [11].

For the TIMIT database, all 24 coefficients of $\{\mathbf{x}_t\}$ are kept. For NTIMIT, 24-dimensional vectors are also extracted, but we keep only the first 17 coefficients, which corresponds to the telephone bandwidth. Experiments are also made on “FTIMIT”, obtained by taking the 17 first coefficients of the vectors $\{\mathbf{x}_t\}$ extracted from TIMIT.

6. EXPERIMENTS

A common training/test protocol is used for all the experiments. It is described in detail in [11] (as protocol “long-short”). Training material consists of 5 sentences (i.e \approx

¹More precisely, we have kept all female and male speakers of “train/dr1” and “test/dr1”, the first female speaker of “train/dr2”, and the first 13 male speakers of “train/dr2”.

14.4 s) which are concatenated into a single reference per speaker. Tests are carried out on 5×1 sentence per speaker (i.e ≈ 3.2 s per sentence) which are tested separately. The total number of independent tests is therefore $63 \times 5 = 315$. The decision rule is the 1-nearest neighbour.

Results of the experiments are given by database (Tables 1 2 and 3). Performances are reported in terms of closed-set speaker identification error rates on the test set for the canonical measures and various combined measures in their asymmetric and their best symmetric form. For the symmetrised measures, a superscript indicates to which symmetrisation (*, ° or •) does the result correspond.

7. DISCUSSION

The following observations can be made :

- Symmetrisation is generally a factor of improvement. However, the appropriate symmetrisation is difficult to predict. It depends on the type of asymmetric measure, and whether the data are in a natural or in a random time order.
- For each database (TIMIT, FTIMIT and NTIMIT), we have underlined the best 10 (or 11) measures. They are (almost) the same ones for all 3 databases. The best one is always obtained with the Gaussian Model.
- With spectral frames in their natural order, VI measures globally outperform IS measures in canonical forms, but the trend is inverted with composed forms.
- With spectral frames in a random order, symmetric composed IS measures outperform all other AR-vector measures, in spite of the loss of the dynamic spectral characteristics.

8. CONCLUSION

In our experiments, we did not succeed in obtaining better speaker identification results with an AR-vector model based measure than with a single Gaussian model classifier.

| function f | a | $\log a$ | g | $\log g$ | $a - \log g - 1$ | $\log(a/g)$ | $a - g$ |
|---|-------------|---------------|-------------|-------------|--------------------|--------------------|---------------|
| AR-vector model - spectral frames in their natural time order | | | | | | | |
| $f_X^{(B/A)} f_Y^{(A/B)}$ | 71.8 54.6 | 71.8 54.6 | 67.3 54.3 | 67.3 54.3 | 78.1 58.4 | 83.8 69.5 | 82.9 67.9 |
| symmetrised | 51.8 • | 52.1 • | 50.5 • | 50.2 • | 57.5 • | 66.0 • | 65.1 • |
| $f_{Y/X}^{(A)} f_{X/Y}^{(B)}$ | 96.8 92.4 | 96.8 92.4 | 97.1 95.6 | 97.1 95.6 | 67.3 88.9 | 66.0 78.7 | 75.2 76.8 |
| symmetrised | 61.9 * | 56.5 * | 68.3 * | 53.0 * | 59.7 ° | 63.2 ° | 66.4 ° |
| AR-vector model - spectral frames in a random time order | | | | | | | |
| $f_{X'}^{(B'/A')} f_{Y'}^{(A'/B')}$ | 64.4 92.1 | 64.1 92.1 | 65.4 91.8 | 65.4 91.8 | 61.9 92.4 | 64.8 93.3 | 64.4 93.0 |
| symmetrised | 65.4 ° | 65.1 ° | 67.9 ° | 68.3 ° | 62.2 ° | 64.4 ° | 64.1 ° |
| $f_{Y'/X'}^{(A')} f_{X'/Y'}^{(B')}$ | 94.0 94.3 | 94.0 94.3 | 98.4 97.5 | 98.4 97.5 | <u>47.0</u> 86.4 | <u>46.0</u> 63.2 | 56.8 77.1 |
| symmetrisation | 61.9 * | 52.4 * | 88.3 * | 72.4 * | 50.2 ° | <u>44.1</u> ° | <u>48.6</u> ° |
| Gaussian model | | | | | | | |
| $f_{Y_o/X_o}^{(I)} f_{X_o/Y_o}^{(I)}$ | 93.0 94.6 | 93.0 94.6 | 98.4 98.4 | 98.4 98.4 | <u>44.1</u> 75.9 | <u>42.5</u> 59.7 | 56.2 73.3 |
| symmetrised | 58.1 * | <u>49.8</u> * | 97.8 * | 98.4 * | <u>47.6</u> ° | <u>44.1</u> ° | <u>49.2</u> ° |

Table 3. NTIMIT - Speaker identification error rates

This observation is in contradiction with results reported in [7], but this divergence may be due to different signal pre-processing and analysis.

Moreover, we globally obtained better performances with the AR-vector model on spectral frames in a random time order rather than when we kept the natural time order. Therefore, the role of dynamic speaker characteristics in the success of the AR-vector model can be questioned, as our results suggest that AR-vector models tend to extract indirectly speaker characteristics of a static nature.

Finally, the influence of symmetrisation can be crucial, but its theoretical basis remains to be understood.

REFERENCES

- [1] Yves Grenier. Utilisation de la prédiction linéaire en reconnaissance et adaptation au locuteur. In *XIèmes Journées d'Etude sur la Parole*, pages 163–171, May 1980. Strasbourg, France.
- [2] T. Artières, Y. Bennani, P. Gallinari, and C. Montacié. Connectionist and conventional models for text-free talker identification tasks. In *Proceedings of NEURONIMES 91*, 1991. Nîmes, France.
- [3] C. Montacié, P. Deléglise, F. Bimbot, and M.-J. Caraty. Cinematic techniques for speech processing: temporal decomposition and multivariate linear prediction. In *Proceedings of ICASSP 92*, volume 1, pages 153–156, March 1992. San Francisco, United-States.
- [4] F. Bimbot, L. Mathan, A. de Lima, and G. Chollet. Standard and target-driven AR-vector models for speech analysis and speaker recognition. In *Proceedings of ICASSP 92*, volume 2, pages II.5–II.8, March 1992. San Francisco, United-States.
- [5] Claude Montacié and Jean-Luc Le Floch. AR-vector models for free-text speaker recognition. In *Proceedings of ICSLP 92*, volume 1, pages 611–614, October 1992. Banff, Canada.
- [6] Chintana Griffin, Tomoko Matsui, and Sadoaki Furui. Distance measures for text-independent speaker recognition based on MAR model. In *Proceedings of ICASSP 94*, volume 1, pages 309–312, April 1994. Adelaide, Australia.
- [7] J.-L. Le Floch, C. Montacié, and M.-J. Caraty. Speaker recognition experiments on the NTIMIT database. In *Proceedings of EUROSPEECH 95*, volume 1, pages 379–382, September 1995. Madrid, Spain.
- [8] Yves Grenier. *Identification du locuteur et adaptation au locuteur d'un système de reconnaissance phonémique*. PhD thesis, ENST, 1977.
- [9] Herbert Gish, Michael Krasner, William Russell, and Jared Wolf. Methods and experiments for text-independent speaker recognition over telephone channels. In *Proceedings of ICASSP 86*, volume 2, pages 865–868, April 1986. Tokyo, Japan.
- [10] Frédéric Bimbot and Luc Mathan. Text-free speaker recognition using an arithmetic-harmonic sphericity measure. In *Proceedings of EUROSPEECH 93*, volume 1, pages 169–172, September 1993. Berlin, Germany.
- [11] Frédéric Bimbot, Ivan Magrin-Chagnolleau, and Luc Mathan. Second-order statistical measures for text-independent speaker identification. *Speech Communication*, 17(1–2):177–192, August 1995.
- [12] P. Whittle. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50(1–2):129–134, 1963.
- [13] Fumitada Itakura. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(1):67–72, February 1975.
- [14] William M. Fisher, George R. Doddington, and Kathleen M. Goudie-Marshall. The DARPA speech recognition research database : specifications and status. In *Proceedings of the DARPA workshop on speech recognition*, pages 93–99, February 1986.
- [15] Charles Jankowski, Ashok Kalyanswamy, Sara Basson, and Judith Spitz. NTIMIT: a phonetically balanced, continuous speech, telephone bandwidth speech database. *Proceedings of ICASSP 90*, April 1990. New Mexico, United-States.

**Time-Frequency Principal Components of Speech :
Application to Speaker Identification(1999)**

Ivan Magrin-Chagnolleau and Geoffrey Durou. **Time-Frequency Principal Components of Speech : Application to Speaker Identification.** *Proceedings of EUROSPEECH 99*, Vol. 2, pp. 759-762, Budapest, Hungary, September 1999.

4 pages

TIME-FREQUENCY PRINCIPAL COMPONENTS OF SPEECH: APPLICATION TO SPEAKER IDENTIFICATION

Ivan Magrin-Chagnolleau* and Geoffrey Durou**

*Rice University, Houston, Texas, USA - **Faculté Polytechnique de Mons, Mons, Belgium

ivan@icee.org - durou@tcts.fpms.ac.be

ABSTRACT

In this paper, we propose a formalism, called vector filtering of spectral trajectories, which allows to integrate under a common formalism a lot of speech parameterization approaches. We then propose a new filtering in this framework, called time-frequency principal components (TFPC) of speech. We apply this new filtering in the framework of speaker identification, using a subset of the POLYCOST database. The results show an improvement of roughly 20 % compared to the use of the classical cepstral coefficients augmented by their Δ coefficients.

1. INTRODUCTION

Cepstral coefficients [8] have been widely used for decades in speech processing. Although they provide a good set of feature vectors with nice properties, like a good decorrelation of the coefficients, or their ability to decorrelate in theory the vocal source and the vocal tract filtering [8], we are convinced that they are not the ultimate solution to represent speech signals in most of the situations.

To find a good alternative to cepstral coefficients, a lot of approaches have been adopted. In particular, the lack of ability of the cepstral coefficients to extract dynamic information from speech suggested the use of Δ and $\Delta\Delta$ coefficients [2]. The Auto-Regressive (AR) vector modeling was another attempt to capture dynamic information of speech [3].

A first aim of this paper is to integrate most of these approaches under a common formalism. Actually, almost every approach assumes spectral vectors as a starting point, and tries, in different ways, to extract some information from these spectral vectors by applying different transformations on them. Most of these approaches can then be seen as a *vector filtering of spectral trajectories*, that is, a function applied to the coefficients of several consecutive spectral vectors.

However, most of the previous approaches apply the filtering function only to one spectral vector, or to several of them but only component by component. We propose a new filtering, based on a principal component calculation, which is applied to all the components of several consecutive spectral vectors. Such a function can be seen as a time-frequency function (or mask), that is, a function applied to all the components of a spectral vector (frequency direction) and to its time context (time direction). Since the coefficients are calculated through a principal component analysis, we call these new coefficients *time-frequency principal components (TFPC)* of speech. We also show how the application of this filtering is an attempt to capture dynamic information from speech. We finally apply this new speech analysis in the framework of speaker identification, which allow us to improve considerably the results

compared to the classical cepstral parameterization augmented by the Δ coefficients.

2. VECTOR FILTERING OF SPECTRAL TRAJECTORIES

2.1. Principle

Let $\{\mathbf{x}_t\}_{1 \leq t \leq M}$ denote a sequence of spectral vectors. The principle of the vector filtering of spectral trajectories is to replace the vector \mathbf{x}_t by a new vector \mathbf{f}_t , whose each component is obtained by the application of a function to the coordinates of vector \mathbf{x}_t and of the preceding and following vectors (context of vector \mathbf{x}_t). This is a convolution product, which can be interpreted as the application of a time-frequency mask to a sequence of spectral vectors (see Figure 1). We can see on this figure that each component of \mathbf{f}_t is obtained by the application of a different function. We can also see that the filtering is applied jointly in the time and the frequency directions. Each function can thus be seen as a time-frequency mask.

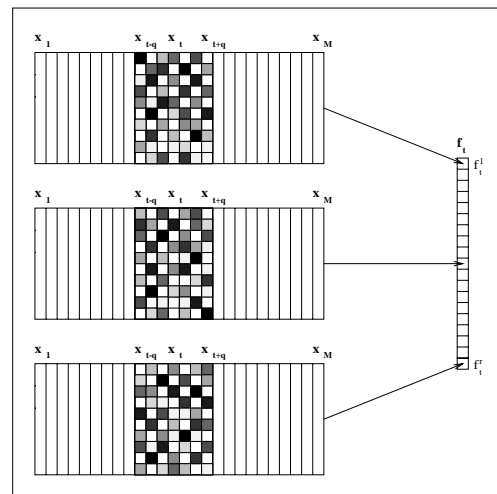


Figure 1: Principle of the vector filtering of spectral trajectories (after [5]).

The approach presented by Milner [7] is quite similar. However, he adopts cepstral coefficients as a starting point, whereas we work directly on spectral coefficients, which eases the interpretation of the new coefficients, and also simplifies the whole process. He also applies the filtering only to the time dimension, and not to the frequency dimension.

Figure 2 illustrates several classical approaches in term of vector

filtering. The first example (a) represents the application of a first derivative approximation function to the first coordinate of several consecutive vectors, which corresponds to the calculation of a Δ coefficient on the first coordinate. The second example (b) is also a Δ coefficient but calculated on the second coordinate. The third example (c) shows the application of a cosine transform to a spectral vector, which corresponds to the calculation of a cepstral coefficient. The last example (d) shows the application of a cosine transform to the first coordinate of several consecutive vectors.

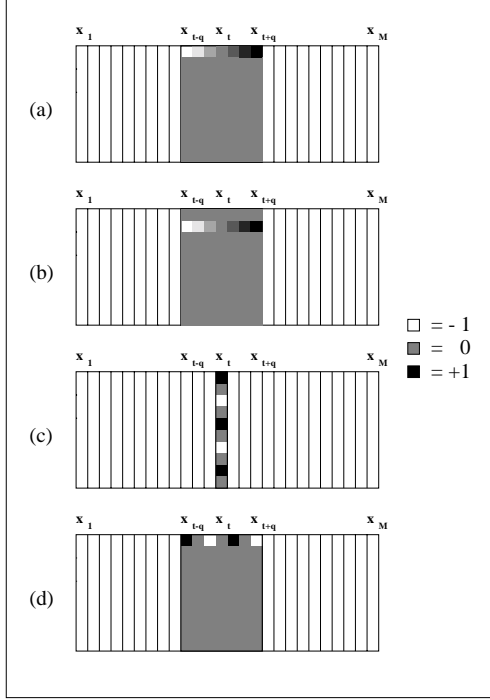


Figure 2: Examples of vector filtering of spectral trajectories (after [5]): (a) Δ parameter on the first coordinate; (b) Δ parameter on the second coordinate; (c) cepstral coefficient; (d) cosine transform of the first coordinate using 7 consecutive vectors.

Once again, we can notice that all these examples show a filtering which operates only in one direction, either the time one or the frequency one, whereas the general approach, as illustrated in Figure 1, operates jointly in both directions.

2.2. Definitions and Notation

Let $\{\mathbf{x}_t^* = \mathbf{x}_t - \bar{\mathbf{x}}\}_{1 \leq t \leq M}$ denote the sequence of centered vectors corresponding to the sequence $\{\mathbf{x}_t\}_{1 \leq t \leq M}$, where $\bar{\mathbf{x}}$ is the corresponding mean vector:

$$\bar{\mathbf{x}} = \frac{1}{M} \sum_{t=1}^M \mathbf{x}_t$$

We also define the sequence of vectors \mathbf{x}_t^* between time $t - q$ and $t + q$:

$$\mathbf{X}_{t-q}^{t+q} = \begin{bmatrix} \mathbf{x}_{t+q}^* \\ \vdots \\ \mathbf{x}_t^* \\ \vdots \\ \mathbf{x}_{t-q}^* \end{bmatrix}$$

By convention, $\mathbf{x}_t^* = 0$ if $t \leq 0$ or $t > M$. The dimension of vector \mathbf{X}_{t-q}^{t+q} is $(2q + 1)p$.

Let then \mathcal{H} be a filtering operating on \mathbf{X}_{t-q}^{t+q} :

$$\mathcal{H} : (\mathbb{R}^p)^{2q+1} \rightarrow \mathbb{R}^r$$

$$\mathbf{X}_{t-q}^{t+q} \mapsto \mathbf{f}_t = \mathcal{H}(\mathbf{X}_{t-q}^{t+q})$$

The dimension of \mathbf{f}_t is r .

In the following, we only consider the case of a linear filtering. In that case, the filtering \mathcal{H} can be expressed in a matricial form:

$$\mathbf{H} = [\mathbf{H}_{-q} \mid \dots \mid \mathbf{H}_0 \mid \dots \mid \mathbf{H}_q]$$

The dimension of \mathbf{H} is $r \times (2q + 1)p$.

And we have:

$$\begin{aligned} \mathbf{f}_t &= \mathbf{H} \cdot \mathbf{X}_{t-q}^{t+q} \\ &= \sum_{k=-q}^{+q} \mathbf{H}_k \cdot \mathbf{x}_{t-k}^* \end{aligned}$$

The dimension of each matricial coefficient \mathbf{H}_k is $r \times p$.

2.3. Examples of Vector Filterings

2.3.1. Cepstral Analysis

The cepstral analysis [8] is one particular filtering of the spectral vectors, where the filtering functions are applied only to the frequency dimension, that is, only to \mathbf{x}_t^* , thus $q = 0$. Each filtering function is a cosine transform. If we apply only the k first cosine transforms to \mathbf{x}_t^* , then $r = k$. And the lines of the matrix \mathbf{H} are the different cosine functions:

$$\mathbf{H} = \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_k^T \end{bmatrix}$$

We finally obtain:

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_t^t = \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_k^T \end{bmatrix} \cdot \mathbf{x}_t^* = \mathbf{c}_t$$

where \mathbf{c}_t is the cepstral vector corresponding to the spectral vector \mathbf{x}_t^* , and containing the cepstral coefficients 1 to k .

2.3.2. Δ and $\Delta\Delta$ Parameters

Δ and $\Delta\Delta$ parameters [2] are also examples of vector filterings. They are applied to the time dimension only. If we use 5 frames to calculate the Δ parameters, then $q = 2$. In fact, these parameters can be calculated for different values of q . We classically use the values $q = 1, 2, 3$. If the filtered vector is composed of the original vector augmented by its Δ coefficients, then the dimension of the new vector is $r = 2p$. If we also add the $\Delta\Delta$ coefficients, then $r = 3p$. In that case, \mathbf{H} can be, for instance:

$$\mathbf{H} = \begin{bmatrix} \mathbf{0}_p & \mathbf{0}_p & \mathbf{I}_p & \mathbf{0}_p & \mathbf{0}_p \\ -2\mathbf{I}_p & -\mathbf{I}_p & \mathbf{0}_p & \mathbf{I}_p & 2\mathbf{I}_p \\ -\mathbf{I}_p & \mathbf{0}_p & 2\mathbf{I}_p & \mathbf{0}_p & -\mathbf{I}_p \end{bmatrix}$$

where $\mathbf{0}_p$ denotes the null matrix of dimension $p \times p$ and \mathbf{I}_p the identity matrix of dimension p .

The corresponding filtered vector is:

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_{t-2}^{t+2} = \begin{bmatrix} \mathbf{x}_t \\ \Delta \mathbf{x}_t \\ \Delta\Delta \mathbf{x}_t \end{bmatrix}$$

2.3.3. Second-Order Auto-Regressive Vector Model

The application of a second-order auto-regressive model to a sequence of vectors [3] can also be interpreted as a vector filtering. In that case, we have $q = 2$, $r = p$, and the filtering matrix is given by:

$$\mathbf{H} = [\mathbf{0}_p \mid \mathbf{0}_p \mid \mathbf{I}_p \mid \mathbf{A}_1 \mid \mathbf{A}_2]$$

where \mathbf{A}_1 and \mathbf{A}_2 are the matricial coefficients of the second-order auto-regressive model.

The filtered vector is the prediction error of the model at time t :

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_{t-2}^{t+2} = \mathbf{x}_t^* + \mathbf{A}_1 \cdot \mathbf{x}_{t-1}^* + \mathbf{A}_2 \cdot \mathbf{x}_{t-2}^* = \mathbf{e}_t$$

This filtering apply to both the time and frequency directions.

2.4. Composition of Filtering

These different filterings can also be composed together, that is, the matrices \mathbf{H} can be multiplied. This is the case for instance when we compute the Δ cepstral parameters. We compose the cepstral filtering with the Δ filtering.

3. TIME-FREQUENCY PRINCIPAL COMPONENTS OF SPEECH

We have defined a formalism which allows to express a lot of approaches in terms of filtering of spectral vectors. However, as it has been seen, most of these classical approaches do not combine the time and the frequency dimensions in the same filtering (except for the second-order AR-vector model). We propose a new filtering operating in both dimensions. This new filtering also assumes a set of training data on which to extract some principal components, which means that it is a data-driven filtering. We call this new filtering *Time-Frequency Principal Components (TFPC)* of Speech.

3.1. Principle of the TFPC Filtering

The idea of the TFPC filtering is to extract time-frequency patterns which are characteristic of the sequence of training vectors, that is, to summarize the evolution of the spectral content by a few spectral sequences extracted from the entire sequence. The original sequence has thus to be long enough, and representative of the class we want to represent with the time-frequency patterns. This strategy can be applied to any pattern recognition problem, as long as we have enough vectors for each class to calculate the time-frequency patterns. Once the patterns have been extracted, they are used to filter the spectral vectors of both the training and the test datasets. And any modeling technique can then be applied on the new vectors, as it is done usually on spectral vectors or cepstral vectors, or any other vector representation of the original signal. As an example, we apply the TFPC filtering in the framework of closed-set text-independent speaker identification, and we extract time-frequency patterns for each speaker of the training database.

3.2. Definition and notation

Let $\{\mathbf{x}_t\}_{1 \leq t \leq M}$ denote again a sequence of spectral vectors, and $\{\mathbf{x}_t^*\}$ the sequence of the corresponding centered vectors.

Let \mathcal{X}_0 denote the covariance matrix of the sequence $\{\mathbf{x}_t\}$:

$$\mathcal{X}_0 = \frac{1}{M} \sum_{t=1}^M (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_t - \bar{\mathbf{x}})^T = \frac{1}{M} \sum_{t=1}^M \mathbf{x}_t^* \cdot \mathbf{x}_t^{*T}$$

and \mathcal{X}_k the lagged covariance matrix at the order k :

$$\mathcal{X}_k = \frac{1}{M} \sum_{t=k+1}^M (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_{t-k} - \bar{\mathbf{x}})^T = \frac{1}{M} \sum_{t=k+1}^M \mathbf{x}_t^* \cdot \mathbf{x}_{t-k}^{*T}$$

The dimension of the covariance matrix and of the lagged covariance matrices is $p \times p$.

We now define a new matrix, \mathbf{X}_{2q+1} , by:

$$\mathbf{X}_{2q+1} = \begin{bmatrix} \mathcal{X}_0 & \mathcal{X}_1 & \dots & \mathcal{X}_{2q} \\ \mathcal{X}_1^T & \mathcal{X}_0 & \dots & \mathcal{X}_{2q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}_{2q}^T & \mathcal{X}_{2q-1}^T & \dots & \mathcal{X}_0 \end{bmatrix}$$

Note that this matrix is block-Toeplitz, and that its dimension is $(2q+1)p \times (2q+1)p$. This matrix can be interpreted as the covariance matrix of the vectors $\{\mathbf{X}_{t-q}^{t+q}\}_{1 \leq t \leq M}$.

We now calculate the principal components of this matrix [4]. It is equivalent to the extraction of eigenvalues and eigenvectors of the matrix. The eigenvector associated with the largest eigenvalue is then the direction of projection which conserves the maximum of the variance; The eigenvector associated to the second largest eigenvalue is the direction of projection which conserves the maximum of the variance uncorrelated (that is orthogonal) to the first one; And so on. We have then:

$$\mathbf{X}_{2q+1} = \mathbf{V}_{2q+1} \cdot \mathbf{\Lambda}_{2q+1} \cdot \mathbf{V}_{2q+1}^T$$

with:

$$\begin{aligned} \mathbf{V}_{2q+1} &= (\mathbf{v}_1, \dots, \mathbf{v}_{2q+1}) \\ \mathbf{\Lambda}_{2q+1} &= \text{diag}(\lambda_1, \dots, \lambda_{2q+1}), \quad \lambda_1 \geq \dots \geq \lambda_{2q+1} \end{aligned}$$

The dimension of the matrix \mathbf{V}_{2q+1} and of the matrix \mathbf{M}_{2q+1} is $(2q+1)p \times (2q+1)p$. The dimension of each vector \mathbf{v}_i , $1 \leq i \leq 2q+1$, is $(2q+1)p$.

3.3. Choice of the Components

Once the principal components have been calculated, we have to decide which ones to keep. Since the decomposition is done according to the maximum of the variance, the first components contain a lot of information, and the last components correspond mainly to noise. The first components are usually kept, the number of them depending on the experiment. Since the eigenvalues correspond to a variance measurement, the criterion for keeping them can be a percentage of the total variance, for instance 80 % [4]. Some other procedures can be used for the choice of the components like the F-ratio or the knock-out procedure. Whatever the method is, the choice of the components can only be done experimentally.

4. APPLICATION TO SPEAKER IDENTIFICATION

4.1. Task

We have tested the TFPC of speech in the framework of closed-set text-independent speaker identification. There is a single reference per speaker. The possibility of rejection is not taken into account: the test speaker is always part of the set of references.

4.2. Database

We use a subset of the POLYCOST database [1], a telephone database, containing 112 speakers (64 females and 48 males). For each speaker, we use in average 90 seconds of speech for the training, and several utterances of 5 seconds in average for the tests. The total number of tests is 560.

4.3. Spectral Analysis

Each utterance is analyzed as followed : the speech signal is decomposed in frames of 30 ms at a frame rate of 10 ms. A Hamming window is applied to each frame. The signal is pre-emphasized with a coefficient 0.95. For each frame, a fast Fourier transform is computed and provides 252 square module values representing the short term power spectrum in the 0-4 kHz band. This Fourier power spectrum is then used to compute 24 filter bank coefficients, using triangular filters placed on a non-uniform frequency scale, similar to the Bark/Mel scale. We finally take the base 10 logarithm of each filter output and multiply the result by 10, to form a 24-dimensional vector of filter bank coefficients in dB.

4.4. TFPC Filtering

Once spectral vectors have been extracted from a training utterance, we calculate the TFPC corresponding to that sequence using $q = 1, 2, 3$, and keep each time all the components. We then filter the spectral vectors by these components to obtain a new set of feature vectors. Thus, for each speaker, we have a set of components for the filtering, and a sequence of filtered vectors by these components. For comparison, we also compute the first 12 cepstral vectors (c_1 to c_{12}) [8], and the corresponding Δ parameters [2] using 3 or 5 frames. Since the TFPC approach can be applied to any kind of vectors, we also apply it to the sequence of cepstral vectors instead of the sequence of spectral vectors.

4.5. Modeling

Each training sequence is then modeled by a Gaussian mixture model [6, 9] using 8 components and diagonal covariance matrices.

4.6. Test Phase

During the test phase, before calculating the log-likelihood of a test sequence of spectral vectors given a training model, we first filter this sequence by the corresponding time-frequency principal components, which can be interpreted as a projection on a particular sub-space, and we then calculate the log-likelihood. The model which gives the highest log-likelihood will determine the identity of the test utterance.

4.7. Results and Discussion

Table 1 presents the results of our experiments.

| Coeff. | static | | static + delta | | TFPC filtering | | |
|---------|---------|---------|----------------|---------|----------------|---------|--|
| | 1 vect. | 3 vect. | 5 vect. | 1 vect. | 3 vect. | 5 vect. | |
| Context | 21.96 | - | - | 9.82 | 9.11 | 9.82 | |
| Spec. | 15.71 | 12.32 | 11.43 | 10.00 | 9.29 | 10.71 | |

Table 1: Percentage of identification error.

The reference identification error rate is the one obtained when using cepstral coefficients augmented by the Δ parameters. In our experiments, the best score is obtained with a Δ calculation over 5 vectors (11.43 %).

In both cases (when starting with spectral coefficients or cepstral coefficients), the error rates obtained after application of the TFPC filtering are better, and are lower when the TFPC filtering is applied using 3 vectors.

Finally, the best error rate is obtained when applying the TFPC filtering to spectral coefficients (9.11 %) and outperforms the reference error rate of 20.3 %.

5. CONCLUSION

We have presented a new formalism, called vector filtering of spectral trajectories, which allows to integrate several speech parameterizations under a common formalism. We have also studied a new speech parameterization called time-frequency principal components. Applied in the framework of closed-set text-independent speaker identification, this new approach shows an improvement in the identification error rate of roughly 20 % when compared to the use of the classical cepstral coefficients augmented by their Δ coefficients.

6. FUTURE DIRECTIONS

In our experiments, we have kept all the TFPC extracted. It may be interesting to study the degradation of the performance if we keep only the first components. We may have a more compact representation with a small loss in performance, or even a gain in performance since the last components correspond mainly to noise.

The TFPC filtering can also be tested in the framework of text-independent speaker verification. In that case, it must also be applied to the sequence of vectors used to train the background models. Then, during the test phase, before calculating the likelihood ratio, the sequence of test vectors will be projected on two different sub-spaces, the sub-space corresponding to the claimed identity model, and the sub-space corresponding to the background model.

Finally, the TFPC filtering can more generally be applied to any pattern recognition problem where the task is to discriminate between classes, as in language recognition or speech recognition for instance.

7. REFERENCES

- [1] *The European COST 250 Action entitled: Speaker Recognition in Telephony*, 1998. (<http://circhp.epfl.ch/polycost>).
- [2] Sadaoki Furui. Comparison of speaker recognition methods using static features and dynamic features. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(3):342–350, June 1981.
- [3] Yves Grenier. Utilisation de la prédiction linéaire en reconnaissance et adaptation au locuteur. In *XIèmes Journées d'Etude sur la Parole*, pages 163–171, May 1980. Strasbourg, France.
- [4] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.
- [5] Ivan Magrin-Chagnolleau. *Approches statistiques et filtrage vectoriel de trajectoires spectrales pour l'identification du locuteur indépendante du texte*. PhD thesis, École Nationale Supérieure des Télécommunications, January 1997.
- [6] Geoffrey J. McLachlan and Kaye E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, 1988.
- [7] Ben Milner. Inclusion of temporal information into features for speech recognition. In *Proceedings of ICSLP 96*, 1996.
- [8] Alan V. Oppenheim and Ronald W. Schaffer. Homomorphic analysis of speech. *IEEE Transactions on Audio and Electroacoustics*, 16(2):221–226, June 1968.
- [9] Douglas A. Reynolds and Richard C. Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83, January 1995.

Application of Time-Frequency Principal Component Analysis to Speaker Verification (2000)

Ivan Magrin-Chagnolleau and Geoffrey Durou. **Application of Time-Frequency Principal Component Analysis to Speaker Verification.** *Digital Signal Processing*, Vol. 10, No. 1-3, pp. 226-236, January/April/July 2000.

11 pages

Application of Time-Frequency Principal Component Analysis to Speaker Verification

Ivan Magrin-Chagnolleau* and Geoffrey Durou†

* *Rice University, Houston, Texas, United States and IRISA/INRIA, Rennes, France.*

† *Faculté Polytechnique de Mons, Mons, Belgium.*

E-mail: ivan@ieee.org - geoffrey.durou@comexis.com

Magrin-Chagnolleau, Ivan, and Durou, Geoffrey, Application of Time-Frequency Principal Component Analysis to Speaker Verification, *Digital Signal Processing* **10** (2000), 226–236.

This article presents the RIMO/ELISA speaker verification system which has been used in the 1999 NIST speaker recognition evaluation. This system is based on a new technique for analyzing speech signals called *time-frequency principal component (TFPC)* analysis. This technique consists in extracting principal components from the contextual covariance matrix, which is the covariance matrix of a sequence of vectors expanded by their temporal context. The database used for the experiments is a subset of the Switchboard corpus. ©2000 Academic Press

Key Words: time-frequency principal components (TFPC); contextual principal components (CPC); contextual covariance matrix; speaker verification; speaker recognition; speaker detection; switchboard.

1. INTRODUCTION

1.1. An Alternative to Cepstral Analysis

Cepstral coefficients [3, 18] have been widely used for decades in speech processing. Although they provide a good set of feature vectors with nice properties, like a good decorrelation of the coefficients, or their ability to decorrelate in theory the vocal source and the vocal tract filtering [18], we are convinced that they are not the ultimate solution in most of the situations.

To find a good alternative to cepstral coefficients, a number of approaches have been adopted. In particular, the lack of ability of the cepstral coefficients to extract dynamic information from speech suggested the use of Δ and $\Delta\Delta$ coefficients [6, 23]. Autoregressive (AR) vector modeling was another attempt to capture dynamic information of speech [1, 7, 16, 17].

Actually, almost every approach assumes spectral vectors as a starting point and tries, in various ways, to extract some information from these spectral vectors by applying transformations on them. Most of these approaches can

then be seen as a *vector filtering of spectral trajectories*, that is, a function applied to the coefficients of several consecutive spectral vectors [11, 12, 13]. It is the case, for instance, of cepstral analysis [3, 18], Δ and $\Delta\Delta$ parameterization [6, 23], autoregressive (AR) vector modeling [1, 7, 16, 17], and temporal principal components [2].

However, most of these approaches applies the filtering function only to one spectral vector, or to several of them but only component by component. In some previous papers, we have proposed a new filtering based on a principal component calculation applied to all the components of several consecutive spectral vectors. Such functions can be seen as time-frequency functions (or masks), that is, functions applied to all the components of a spectral vector (frequency direction) and to its time context (time direction). Since the coefficients are calculated through a principal component analysis, we have called these new coefficients *contextual principal components (CPC)* or *time-frequency principal components (TFPC)* of speech [11, 12, 13]. The latter name is used when the original vectors contain some information about the frequency content of the original signal, as spectral vectors do, for instance. The former name is used otherwise.

In this paper, we apply the TFPC filtering in the framework of speaker verification (or speaker detection). The system described in this article, called RIMO¹/ELISA system, has been used in the 1999 NIST speaker recognition evaluation. The authors of this paper are members of the ELISA consortium, which is a research consortium on the topic of speaker recognition. The ELISA consortium and all the ELISA speaker verification and speaker tracking systems proposed for the NIST 1999 speaker recognition evaluation are presented in [4].

1.2. Outline of the Paper

The outline of the paper is the following. In Section 2, we explain the time-frequency principal component (TFPC) approach and describe its application in the framework of speaker verification. Then experiments are detailed in Section 3. Results are presented and discussed in Section 4. Finally, Section 5 concludes this work and gives some future directions.

2. TIME-FREQUENCY PRINCIPAL COMPONENT ANALYSIS

In this section, we explain in details the time-frequency principal component analysis. We first describe its principle. Then we give its mathematical formulation. We discuss briefly the choice of the components. Finally, we describe the use of this new analysis in the framework of speaker verification.

2.1. Principle of the TFPC Filtering

The idea of the TFPC filtering is to extract time-frequency patterns that are characteristic of a sequence of vectors in order to summarize the evolution of the

¹RIMO stands for **RI**ce University—**F**aculté Polytechnique de **MO**ns.

spectral content by a few spectral sequences extracted from the entire sequence. The original sequence must be long enough to be representative of the class we want to characterize with the time-frequency patterns. This strategy can be applied to any pattern recognition problem, as long as we have enough vectors for each class to calculate the time-frequency patterns. Once the patterns have been extracted, they are used to filter the spectral vectors of both the training and the test dataset. And any modeling technique can then be applied to the new vectors, as it is done usually on spectral vectors or cepstral vectors, or any other vector representation of the original signal. As an example, we apply, in this paper, the TFPC filtering in the framework of text-independent speaker verification, and we extract time-frequency patterns for each speaker in the training dataset and for a population of speakers representative of the world. TFPC filtering has already been applied to closed-set speaker identification in [11, 12, 13].

2.2. Definitions and Notations

Let $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ denote a sequence of spectral vectors, and $\{\mathbf{x}_t^* = \mathbf{x}_t - \bar{\mathbf{x}}\}$ denote the sequence of the corresponding centered vectors. $\bar{\mathbf{x}}$ is the mean vector of the sequence $\{\mathbf{x}_t\}$.

Let $\boldsymbol{\mathcal{X}}_0$ denote the covariance matrix of the sequence $\{\mathbf{x}_t\}$,

$$\boldsymbol{\mathcal{X}}_0 = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_t - \bar{\mathbf{x}})^T = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t^* \cdot \mathbf{x}_t^{*T},$$

and $\boldsymbol{\mathcal{X}}_k$ the lagged covariance matrix at the order k ,

$$\boldsymbol{\mathcal{X}}_k = \frac{1}{T} \sum_{t=k+1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_{t-k} - \bar{\mathbf{x}})^T = \frac{1}{T} \sum_{t=k+1}^T \mathbf{x}_t^* \cdot \mathbf{x}_{t-k}^{*T}.$$

The dimension of the covariance matrix and of the lagged covariance matrices is $p \times p$.

We now define the matrix \mathbf{X}_{2q+1} by

$$\mathbf{X}_{2q+1} = \begin{bmatrix} \boldsymbol{\mathcal{X}}_0 & \boldsymbol{\mathcal{X}}_1 & \dots & \boldsymbol{\mathcal{X}}_{2q} \\ \boldsymbol{\mathcal{X}}_1^T & \boldsymbol{\mathcal{X}}_0 & \dots & \boldsymbol{\mathcal{X}}_{2q-1} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\mathcal{X}}_{2q}^T & \boldsymbol{\mathcal{X}}_{2q-1}^T & \dots & \boldsymbol{\mathcal{X}}_0 \end{bmatrix}.$$

This matrix is block-Toeplitz, and its dimension is $(2q + 1) p \times (2q + 1) p$.

Let define the sequence of vectors \mathbf{x}_t between time $t - q$ and $t + q$ by

$$\mathbf{X}_{t-q}^{t+q} = \begin{bmatrix} \mathbf{x}_{t+q} \\ \vdots \\ \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t-q} \end{bmatrix}.$$

By convention, $\mathbf{x}_t = 0$ if $t \leq 0$ or $t > T$. The dimension of vector \mathbf{X}_{t-q}^{t+q} is $(2q+1)p$.

The matrix \mathbf{X}_{2q+1} can be interpreted as the covariance matrix of the vectors $\{\mathbf{X}_{t-q}^{t+q}\}_{1 \leq t \leq T}$, and it can therefore be called the *contextual covariance matrix*.

We now calculate the principal components of this matrix [8, 22]. This is equivalent to the extraction of eigenvalues and eigenvectors of the matrix. The eigenvector associated with the largest eigenvalue gives the direction of projection that conserves the maximum of the variance, the eigenvector associated with the second largest eigenvalue gives the direction of projection that conserves the maximum of the variance uncorrelated with (that is, orthogonal to) the first one, and so on. We have then

$$\mathbf{X}_{2q+1} = \mathbf{V}_{2q+1} \cdot \mathbf{M}_{2q+1} \cdot \mathbf{V}_{2q+1}^T$$

with:

$$\begin{aligned} \mathbf{V}_{2q+1} &= (\mathbf{v}_1, \dots, \mathbf{v}_{2q+1}) \\ \mathbf{M}_{2q+1} &= \text{diag}(\mu_1, \dots, \mu_{2q+1}), \quad \mu_1 \geq \dots \geq \mu_{2q+1}. \end{aligned}$$

The dimension of the matrix \mathbf{V}_{2q+1} and of the matrix \mathbf{M}_{2q+1} is $(2q+1)p \times (2q+1)p$. The dimension of each vector $\mathbf{v}_i, 1 \leq i \leq 2q+1$, is $(2q+1)p$.

Since the principal components are extracted from the contextual covariance matrix, instead of the covariance matrix itself, we call them *contextual principal components (CPC)*. When the original vectors contain some information about the frequency content of the original signal, as spectral vectors, we can call these components more specifically *time-frequency principal components (TFPC)*.

2.3. Choice of the Components

Once the principal components have been calculated, we have to decide which ones to keep. Since the decomposition is done according to the maximum of the variance, the first components contain a lot of information, and the last components correspond mainly to noise. The first components are usually kept, the number of them depending on the experiment. Since the eigenvalues correspond to a variance measurement, a criterion for keeping them can be a percentage of the total variance, for instance 80% [8, 22]. Some other procedures can be used for the choice of the components such as the F-ratio [25, 9] or the knock-out procedure [21]. Whatever the method is, the choice of the components can only be done experimentally.

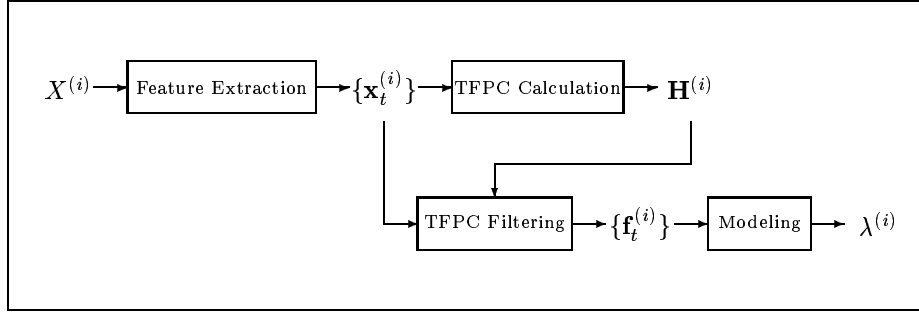


FIG. 1. Application of the time-frequency principal component analysis during the training phase of a speaker verification system.

2.4. Application of the TFPC Analysis to Speaker Verification

In a speaker verification system, there are two phases: the training phase and the test phase. We show how to use the TFPC analysis in each of them.

Training Phase. Figure 1 illustrates the use of the TFPC analysis during the training phase of a speaker verification system. The starting point is a utterance $X^{(i)}$ representing the target speaker $\mathcal{X}^{(i)}$. First, spectral vectors $\{\mathbf{x}_t^{(i)}\}$ are extracted from the utterance. Then, time-frequency principal components are calculated from the spectral vectors and are used to form a filtering matrix $\mathbf{H}^{(i)}$. Then the spectral vectors $\{\mathbf{x}_t^{(i)}\}$ are filtered by this filtering matrix $\mathbf{H}^{(i)}$ to form a new sequence of vectors $\{\mathbf{f}_t^{(i)}\}$, which is a projection of $\{\mathbf{x}_t^{(i)}\}$ on the subspace represented by $\mathbf{H}^{(i)}$. Finally, a modeling technique is applied to the sequence $\{\mathbf{f}_t^{(i)}\}$ to obtain a model $\lambda^{(i)}$ for the speaker $\mathcal{X}^{(i)}$. At the end of the training phase, the speaker $\mathcal{X}^{(i)}$ is represented by a model $\lambda^{(i)}$ and a filtering matrix $\mathbf{H}^{(i)}$. This procedure is repeated for each target speaker and also for the world data, which include a population of female and male speakers representative of the world population.

Test Phase. Figure 2 illustrates the use of the TFPC analysis during the test phase of a speaker verification system. The starting point is, this time, a test

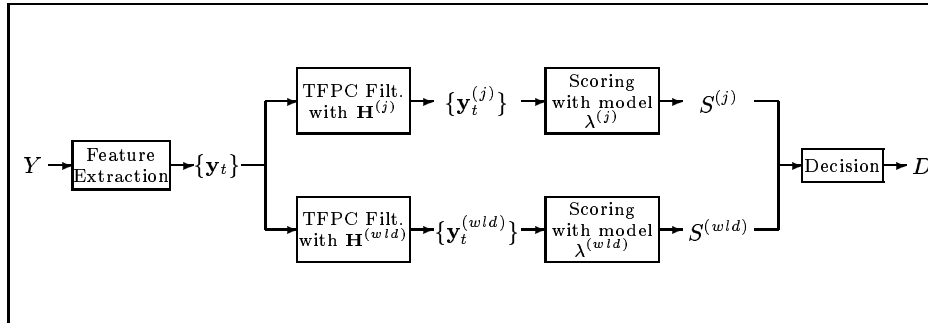


FIG. 2. Application of the time-frequency principal component analysis during the test phase of a speaker verification system.

utterance Y and a claimed identity $\mathcal{X}^{(j)}$. First, spectral vectors $\{\mathbf{y}_t\}$ are extracted from the utterance Y . Then, the sequence $\{\mathbf{y}_t\}$ is filtered by the filtering matrix $\mathbf{H}^{(j)}$, and we obtain the new sequence $\{\mathbf{y}_t^{(j)}\}$. Then, we calculate a score $S^{(j)}$ on this new sequence using the model $\lambda^{(j)}$. On the other hand, we filter the sequence $\{\mathbf{y}_t\}$ by the filtering matrix $\mathbf{H}^{(wld)}$ which has been extracted from the world data. We obtain the sequence $\{\mathbf{y}_t^{(wld)}\}$. Then, we calculate a score $S^{(wld)}$ on these filtered vectors using the world model $\lambda^{(wld)}$. We finally combine the two scores $S^{(j)}$ and $S^{(wld)}$ to make an acceptance or rejection decision.

3. EXPERIMENTS

After having explained the time-frequency principal component analysis and described how to use it in the framework of speaker verification, we detail in this section the experiments we have carried out.

3.1. Task

The task described in this paper is the task called *one-speaker detection* by NIST [19], which is in fact what is classically called speaker verification². The task consists in determining whether a specified target speaker is speaking in a given segment of μ -law encoded telephone speech or not. The hypothesized target speakers are always of the same sex as the segment speaker.

3.2. Database

The data for the NIST 1999 speaker recognition evaluation came from the Switchboard 2–Phase 3 corpus collected by the Linguistic Data Consortium (LDC). The training data for each target speaker consisted of two utterances of 1 min each in average, obtained by concatenating consecutive turns of the speaker. Areas of silence were removed. The total number of target speakers was 539 (309 female speakers and 230 male speakers). The test data for the one-speaker detection task consisted of segments containing speech of only one speaker, obtained by concatenating consecutive turns of this speaker. Areas of silence were also removed. The duration of the test segments was between a few seconds and 1 min. There were 37,620 tests (3,157 of them were target trials and the remaining 34,463 were impostor trials). For a more complete description of the NIST 1999 data, see [19].

3.3. Spectral Analysis

Each utterance was analyzed as follows: the speech signal was converted from a μ -law to a linear representation, then decomposed in frames of 30 ms at a frame rate of 10 ms. A Hamming window was applied to each frame. The signal was preemphasized with a coefficient 0.95. For each frame, a fast Fourier transform was computed and provided 256 square modulus values representing the short-term power spectrum in the 0-4 kHz band. This Fourier power

²The authors of this paper also took part to the NIST speaker tracking task. This work, briefly described in [4], will be reported on in a future communication.

spectrum was then used to compute 24 filter bank coefficients, using triangular filters placed on a nonuniform frequency scale, similar to the Bark/Mel scale. We finally took the base 10 logarithm of each filter output and multiplied the result by 10, to form a 24-dimensional vector of filter bank coefficients in dB.

3.4. TFPC Analysis

Once spectral vectors have been extracted for each speaker of the training dataset, we applied the TFPC analysis as explained in Section 2. For our experiments, we chose to test the value $q = 1$ only (which corresponds to a context of three consecutive vectors) and to keep all the components (which provided us with new vectors of 72 components after filtering). This choice has been done accordingly to some preliminary results in speaker identification [12, 13]. We also tested, for $q = 1$, an other configuration corresponding to the discard of the last principal components, which are usually the most noisy: we discarded the 12 last components, so the new vectors had 60 components.

3.5. Gaussian Mixture Modeling

Once the filtered vectors were calculated, we modeled each target speaker by a Gaussian mixture model of 128 components with diagonal covariance matrices [15, 20, 24]. We used the expectation maximization (EM) algorithm [5] to train the mixtures. The initialization of the EM algorithm has been done with a vector quantization (VQ) algorithm [10].

3.6. World Model

We also needed a world model, that is, a model to represent the world population of speakers. This model was used to normalize the likelihood score during the verification phase. If the test speaker is the same as the target speaker, then the likelihood obtained with the target model should be higher than the likelihood obtained with the world model. If the test speaker is not the same as the target speaker, then the likelihood obtained with the target model should be lower than the likelihood obtained with the world model.

We trained our world model on data provided for the 1998 NIST speaker recognition evaluation. These data also came from the Switchboard corpus. We concatenated 308 utterances of 30 s in average (155 females and 153 males). Some of the utterances have been recorded with an electret microphone, some with a carbon microphone. Our world model was therefore gender-independent and handset-independent.

For the rest of the analysis, the world data were processed the same way as the target speaker data. In particular, we also calculated time-frequency principal components using the spectral vectors extracted from the world data, and we used 128 components and diagonal covariance matrices for the Gaussian mixture modeling.

3.7. Thresholds Setting

In order to determine the thresholds for the verification phase, we needed to simulate target and impostor trials. For each speaker of the training dataset, we

had 2 utterances of 1 min in average. We used each of them as a target trial. We also used 100 utterances of 1 minute in average (50 females and 50 males) taken from the NIST 1998 data as impostor trials. We finally set up the thresholds for verification, using a log-likelihood ratio calculation and a normalization by the length of the utterance, in order to minimize the cost function \mathcal{C}_{Det} given by NIST [19],

$$\mathcal{C}_{Det} = \mathcal{C}_{Miss} \times \mathcal{R}_{Miss} \times P_{Target} + \mathcal{C}_{FalseAlarm} \times \mathcal{R}_{FalseAlarm} \times P_{Impostor},$$

where \mathcal{C}_{Miss} is the cost of a miss (also called false rejection in speaker verification), $\mathcal{C}_{FalseAlarm}$ the cost of a false alarm (also called false acceptance in speaker verification), P_{Target} the a priori probability of a target trial, and $P_{Impostor}$ the a priori probability of an impostor trial. \mathcal{R}_{Miss} and $\mathcal{R}_{FalseAlarm}$ are the miss (or false rejection) rate and the false alarm (or false acceptance) rate, respectively. In the NIST 1999 speaker recognition evaluation, the following values were used:

$$\mathcal{C}_{Miss} = 10; \mathcal{C}_{FalseAlarm} = 1; P_{Target} = 0.01; P_{Impostor} = 1 - P_{Target} = 0.99.$$

During the test phase, we also used a log-likelihood ratio calculation and a normalization by the length of the utterance. Results are reported in the next section.

4. RESULTS AND DISCUSSION

In this section, we show and discuss the results of our speaker verification system.

Figure 3 shows the DET curves [14] for the speaker verification task (called one speaker detection by NIST). This representation is a way to show all of the possible operating points of a system (false alarm rate vs miss rate) in a scale which makes the result curves rather linear. The solid line corresponds to the keeping of all the 72 time-frequency principal components. The dashed line corresponds to the discarding of the last 12 time-frequency principal components. The equal error rates (EER) of the two configurations are very close to each other, and their value is approximately equal to 24%.

The difference between the two configurations is very small. In the higher part of the DET plot, corresponding to a high miss rate and a low false alarm rate, the system including all the time-frequency principal components performs slightly better. In the middle and lower part of the DET plot, the two curves are very close to each other. These results seem to indicate that it is possible to discard the last time-frequency principal components, reducing the number of parameters, without losing too much in performance.

The overall performance of our system is, however, lower than a system which would use cepstral and Δ -cepstral coefficients and Gaussian mixture models [19], although a TFPC-based system performs better than a cepstral-based system in speaker identification [12, 13]. One of the possible explanations

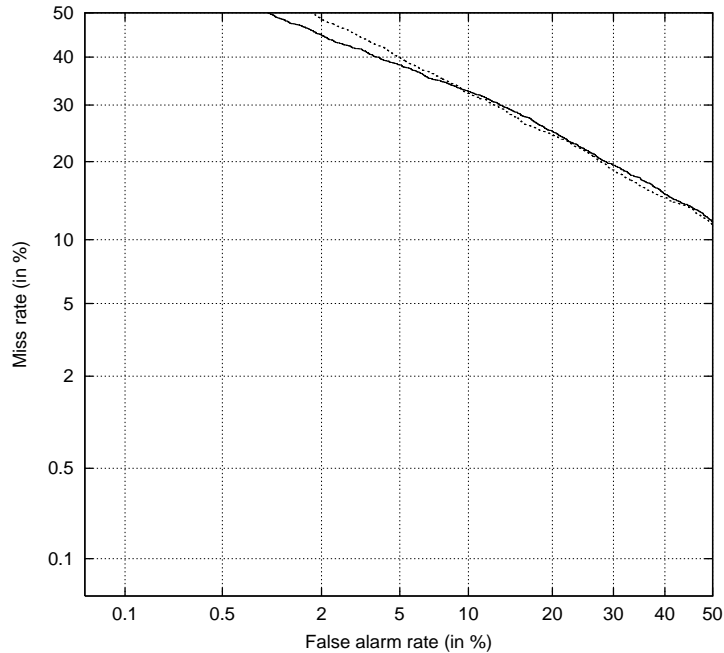


FIG. 3. Results of speaker verification (called one speaker detection by NIST). The solid line corresponds to the keeping of all the 72 time-frequency principal components, the dashed line to the discarding of the last 12 time-frequency principal components.

for the discrepancy between the performance in speaker identification and speaker verification is the normalization issue. When we apply the TFPC scheme to speaker verification, we calculate the ratio of two log-likelihood scores calculated in two different subspaces, one in the subspace of the target speaker, one in the subspace of the world. That may explain why the results obtained so far in speaker verification with a TFPC-based system are not as good as the results obtained by a similar system in speaker identification. The differences in performance between our system and the cepstral-based systems used in the NIST 1999 speaker recognition evaluation could also be explained by other reasons: the training algorithms were different; the world model was trained on different data; the number of Gaussian components was different. In order to compare rigorously the two acoustic analyses, all the other components of the system have to be identical. This comparison will be one of the perspectives of this work.

5. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have presented a new way of analyzing speech signals, called time-frequency principal component (TFPC) analysis. This technique is based on the extraction of time-frequency patterns, which are characteristics of each class that we want to represent. In our case, classes correspond to speakers. We

have applied the TFPC analysis to text-independent speaker verification in the framework of the NIST 1999 speaker recognition evaluation. In spite of very good results with TFPC analysis in speaker identification [12, 13], the TFPC approach did not seem to perform as well in speaker verification.

One of the possible explanations is the normalization issue. Therefore, a theoretical study of the calculation of log-likelihood scores in two different subspaces needs to be conducted.

Other sizes for the context used in the calculation of the TFPC and other configurations of components will also be tested in the future. In particular, the choice of the time-frequency principal components to keep may be done according to a variance criterion.

REFERENCES

1. F. Bimbot, L. Mathan, A. de Lima, and G. Chollet. Standard and target-driven AR-vector models for speech analysis and speaker recognition. In *Proceedings of ICASSP 92*, volume 2, pages II.5–II.8, March 1992. San Francisco, United States.
2. Frédéric Bimbot, Enrico Bocchieri, and Bishnu Atal. Sous-espaces de projection de séquences de trames acoustiques pour l'analyse et la reconnaissance de parole. In *XXIèmes Journées d'Etude sur la Parole*, 1996. Avignon, France.
3. Bruce P. Bogert, M.J.R. Healy, and John W. Tukey. The quefreny alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In Murray Rosenblatt, editor, *Proceedings of the Symposium on Time Series Analysis*, chapter 15, pages 209–243. John Wiley and Sons, 1963.
4. The ELISA Consortium. The ELISA systems for the NIST'99 evaluation in speaker detection and tracking. *Digital Signal Processing*, 10(1), January 2000.
5. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 6(39):1–38, 1977.
6. Sadaoki Furui. Comparison of speaker recognition methods using static features and dynamic features. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(3):342–350, June 1981.
7. Yves Grenier. Utilisation de la prédiction linéaire en reconnaissance et adaptation au locuteur. In *XIèmes Journées d'Etude sur la Parole*, pages 163–171, May 1980. Strasbourg, France.
8. I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.
9. Sir Maurice Kendall and Alan Stuart. *The Advanced Theory of Statistics*, volume 1: Distribution Theory. Charles Griffin, 1977.
10. Yoseph Linde, Andres Buzo, and Robert M. Gray. An algorithm for vector quantization design. *IEEE Transactions on Communications*, 28(1):84–95, January 1980.
11. Ivan Magrin-Chagnolleau. *Approches statistiques et filtrage vectoriel de trajectoires spectrales pour l'identification du locuteur indépendante du texte*. PhD thesis, École Nationale Supérieure des Télécommunications, January 1997.
12. Ivan Magrin-Chagnolleau and Geoffrey Durou. Time-frequency principal components of speech: application to speaker identification. In *Proceedings of EUROSPEECH 99*, pages 759–762, September 1999. Budapest, Hungary.
13. Ivan Magrin-Chagnolleau, Geoffrey Durou, and Frédéric Bimbot. Application of time-frequency principal component analysis to text-independent speaker identification. *Submitted to IEEE Transactions on Speech and Audio Processing*.
14. A. Martin et al. The DET curve in assessment of detection task performance. In *Proceedings of EUROSPEECH 97*, volume 4, pages 1895–1898, September 1997. Rhodes, Greece.
15. Geoffrey J. McLachlan and Kaye E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, 1988.
16. C. Montacié, P. Deléglise, F. Bimbot, and M.-J. Caraty. Cinematic techniques for speech processing: temporal decomposition and multivariate linear prediction. In *Proceedings of ICASSP 92*, volume 1, pages 153–156, March 1992. San Francisco, United States.

17. Claude Montacié and Jean-Luc Le Floch. AR-vector models for free-text speaker recognition. In *Proceedings of ICSLP 92*, volume 1, pages 611–614, October 1992. Banff, Canada.
18. Alan V. Oppenheim and Ronald W. Schafer. Homomorphic analysis of speech. *IEEE Transactions on Audio and Electroacoustics*, 16(2):221–226, June 1968.
19. Mark Przybocki and Alvin Martin. The NIST 1999 speaker recognition evaluation - an overview. *Digital Signal Processing*, 10(1-3), April 2000.
20. Douglas A. Reynolds and Richard C. Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83, January 1995.
21. Marvin R. Sambur. Selection of acoustic features for speaker identification. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(2):176–182, April 1975.
22. Gilbert Saporta. *Probabilités analyse des données et statistique*. Éditions Technip, 1990.
23. Frank K. Soong and Aaron E. Rosenberg. On the use of instantaneous and transitional spectral information in speaker recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(6):871–879, June 1988.
24. D.M. Titterton, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and sons, 1985.
25. Jared J. Wolf. Efficient acoustic parameters for speaker recognition. *Journal of the Acoustical Society of America*, 51(6):2044–2056, 1972. Number 6 (Part 2).

Application of Vector Filtering to Pattern Recognition (2000)

Ivan Magrin-Chagnolleau and Geoffrey Durou. **Application of Vector Filtering to Pattern Recognition.** *Proceedings of ICPR 2000*, Barcelona, Spain, September 2000.

4 pages

Application of Vector Filtering to Pattern Recognition

Ivan Magrin-Chagnollet* and Geoffrey Durou†

*IRISA (CNRS & INRIA) - Campus universitaire de Beaulieu - 35042 Rennes cedex - France - European Union

†Faculté Polytechnique de Mons - Boulevard Dolez, 31 - 7000 Mons - Belgium - European Union

ivan@ieee.org - geoffrey.durou@debis.be

Abstract

In this paper, we present a new formalism, called *vector filtering*, which consists in transforming a sequence of vectors through a *matricial filtering*. This formalism allows us to unify a number of classical approaches. We also show how vector filtering can be integrated in a pattern recognition system. We then propose a new filtering, called *contextual principal components (CPC)*, which consists in calculating principal components on vectors augmented by their context. Then, we apply the new filtering in the framework of text-independent speaker identification, which consists in identifying a speaker by the voice without knowledge about the phonetic content. By using this new filtering, we are able to decrease the identification error rate of roughly 20 % compared to a system using the classical cepstral coefficients augmented by their delta parameters.

1. Introduction

In a pattern recognition system, the training phase consists in learning classes from reference data, and the test phase in recognizing an unknown pattern as belonging to one of these classes. As preliminaries to each of these phases, there is a step of pattern analysis, which leads to a more compact representation of the pattern, most often as a sequence of vectors.

In this paper, we propose a new approach called *vector filtering* [6, 8, 7], which consists in filtering the initial sequence of vectors to extract from it a new sequence containing more characteristic information about the class. This filtering can be interpreted as a projection on a subspace characteristic of the class. This approach is first proposed in a general framework, and we show how a number of classical approaches are particular cases of this filtering. It is the case for instance of cepstral analysis [10], auto-regressive vector models [4], Δ and $\Delta\Delta$ parameters [3], cosine transform of spectral trajectories [2]. Most of these examples come from speech processing, but we can find similar examples in a great number of domains. **The approach proposed in this article can be used in any application where patterns are represented by sequences of vectors.**

Then, we propose a new vector filtering based on a principal component analysis applied to several consecutive vectors. This

new filtering is called *contextual principal components (CPC)*. When the original vectors contain some information about the frequency content of the signal, we can call the new filtering more specifically *time-frequency principal components (TFPC)* [6, 8, 7].

We finally apply this new filtering in the framework of text-independent speaker identification, which consists in recognizing the identity of a speaker by the voice without using any knowledge about the phonetic content. The new filtering allows us to reduce the identification error rate of about 20 % compared to the classical use of cepstral coefficients [10] augmented by the Δ parameters [3].

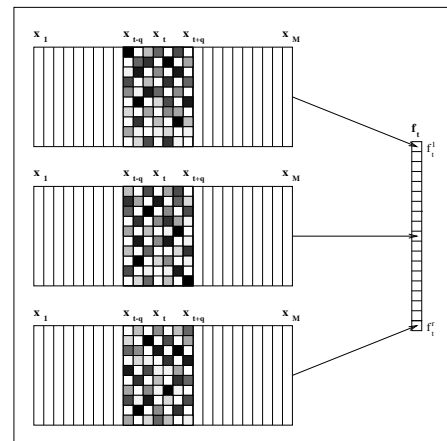


Figure 1. Principle of vector filtering (after [6]).

2. Vector Filtering

2.1. Principle

Let $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ be a sequence of vectors representing a pattern \mathcal{X} . The principle of vector filtering is to replace the vector \mathbf{x}_t by a new vector \mathbf{f}_t , whose each component is obtained by the application of a function to the coordinates of vector \mathbf{x}_t and of the preceding and following vectors (context of vector \mathbf{x}_t). If the function is linear, the operation is similar to a convolution product, which can be interpreted as the application of a mask to the sequence of vectors (see Figure 1). This approach generalizes a number of other filtering approaches presented recently in the literature, but which were operating only in one dimension (see for instance [9]). The great advantage of this new approach is to combine two dimensions in the same filtering.

*This work was started when Ivan Magrin-Chagnollet was with the ECE department of Rice University (Houston, Texas, U.S.A.), and finalized when he joined IRISA, where he is currently working.

†Geoffrey Durou was with the Faculté Polytechnique de Mons (Mons, Belgium) when this work was done. He is currently working at Debis IT (Belgium).

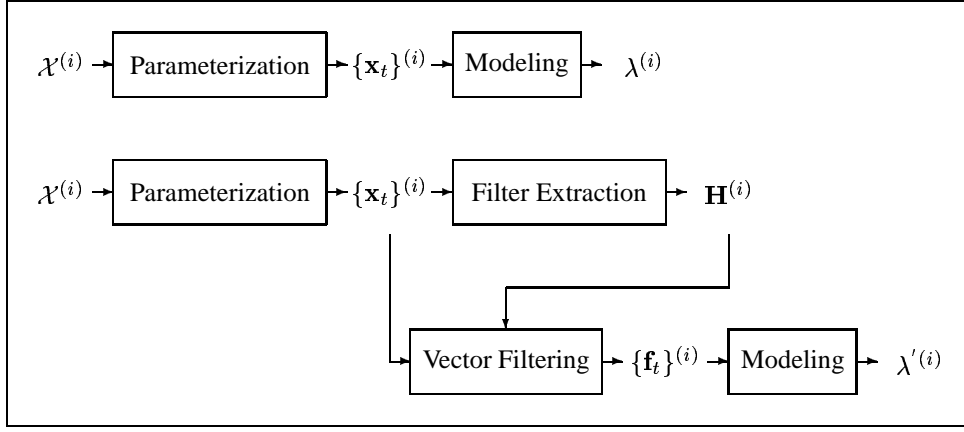


Figure 2. Training phase of a pattern recognition system without (top) or with (bottom) the use of vector filtering.

2.2. Definitions and Notations

General Case Let \mathcal{X} be a pattern represented by a sequence of vectors $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ with dimension p . Let \mathbf{X}_{t-q}^{t+q} denote the sequence of vectors \mathbf{x}_t between time $t-q$ and time $t+q$:

$$\mathbf{X}_{t-q}^{t+q} = \begin{bmatrix} \mathbf{x}_{t+q} \\ \vdots \\ \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t-q} \end{bmatrix}$$

By convention, $\mathbf{x}_t = 0$ for $t \leq 0$ and $t > T$. The dimension of \mathbf{X}_{t-q}^{t+q} is $(2q+1)p$. Let \mathcal{H} be a filtering operating on \mathbf{X}_{t-q}^{t+q} :

$$\mathcal{H} : (\mathbb{R}^p)^{2q+1} \rightarrow \mathbb{R}^r$$

$$\mathbf{X}_{t-q}^{t+q} \mapsto \mathbf{f}_t = \mathcal{H}(\mathbf{X}_{t-q}^{t+q})$$

The dimension of \mathbf{f}_t is noted r .

Case of a Linear Filtering In the following, we suppose that the filtering is linear. The filtering \mathcal{H} can then be expressed in a matricial form:

$$\mathbf{H} = [\mathbf{H}_{-q} \mid \dots \mid \mathbf{H}_0 \mid \dots \mid \mathbf{H}_q]$$

The dimension of \mathbf{H} is $r \times (2q+1)p$. We have then:

$$\begin{aligned} \mathbf{f}_t &= \mathbf{H} \cdot \mathbf{X}_{t-q}^{t+q} \\ &= \sum_{k=-q}^{+q} \mathbf{H}_k \cdot \mathbf{x}_{t-k} \end{aligned}$$

The dimension of each matricial coefficient \mathbf{H}_k is $r \times p$.

2.3. Examples of Vector Filtering

Cepstral Analysis The cepstral analysis [10] is one particular filtering, where the filtering functions are applied only to \mathbf{x}_t , therefore $q = 0$. Each filtering function is a cosine transform. If we apply only the k first cosine transforms to \mathbf{x}_t , then $r = k$. And the lines of the matrix \mathbf{H} are the different cosine functions:

$$\mathbf{H} = \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_k^T \end{bmatrix}$$

We finally obtain:

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_t^t = \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_k^T \end{bmatrix} \cdot \mathbf{x}_t = \mathbf{c}_t$$

where \mathbf{c}_t is the cepstral vector corresponding to the spectral vector \mathbf{x}_t , and containing the cepstral coefficients 1 to k .

Second-Order Auto-Regressive Vector Model The application of a second-order auto-regressive model [4] to a sequence of vectors can also be interpreted as a vector filtering. In that case, we have $q = 2$, $r = p$, and the filtering matrix is given by:

$$\mathbf{H} = [\mathbf{0}_p \mid \mathbf{0}_p \mid \mathbf{I}_p \mid \mathbf{A}_1 \mid \mathbf{A}_2]$$

where \mathbf{A}_1 and \mathbf{A}_2 are the matricial coefficients of the second-order auto-regressive model. The filtered vector is the prediction error of the model at time t :

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_{t-2}^{t+2} = \mathbf{x}_t + \mathbf{A}_1 \cdot \mathbf{x}_{t-1} + \mathbf{A}_2 \cdot \mathbf{x}_{t-2} = \mathbf{e}_t$$

2.4. Application to Pattern Recognition

We can use the vector filtering in the framework of a pattern recognition system. Figure 2 shows how to integrate the vector filtering approach in the training phase of a pattern recognition system. The top part corresponds to a system which does not use vector filtering. The bottom part corresponds to a system using vector filtering. Figure 3 shows how to integrate the vector filtering in the test phase of a pattern recognition system. The top part corresponds to a system which does not use vector filtering. The bottom part corresponds to a system using vector filtering.

3. Contextual Principal Components

Now that we have defined the vector filtering, and shown how to integrate it in a pattern recognition system, we propose a new filtering called *contextual principal components (CPC)* in the general case, and *time-frequency principal components (TFPC)* in the case where the original vectors contain some information about the frequency content of the signal.

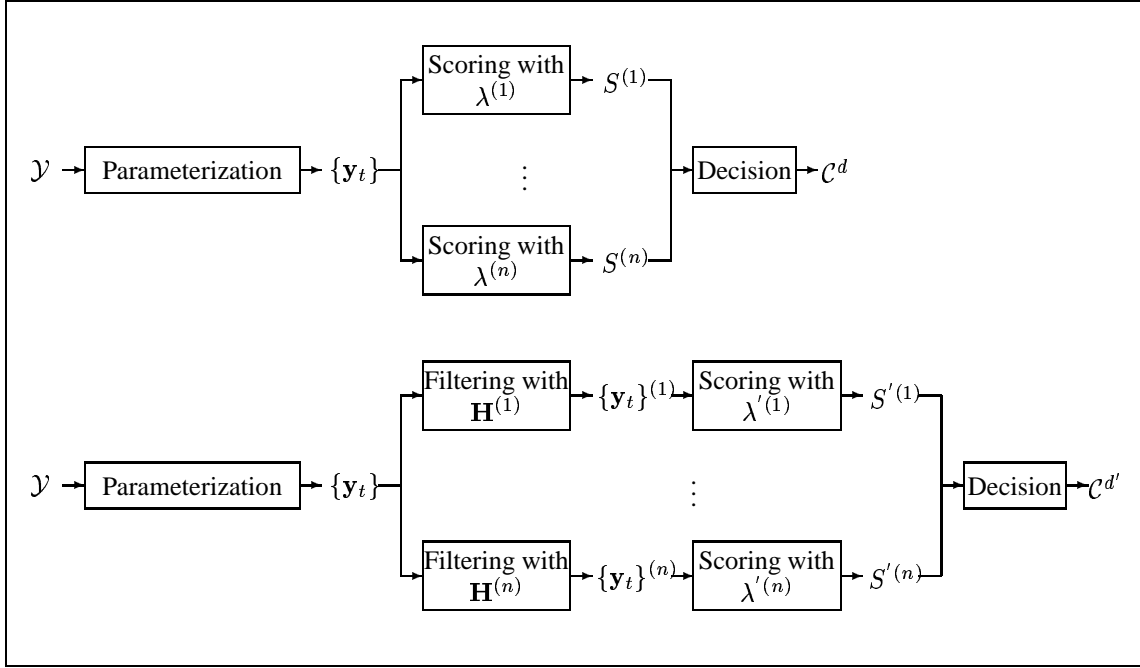


Figure 3. Test phase of a pattern recognition system without (top) or with (bottom) vector filtering.

3.1. Definitions et Notations

Let $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ denote again a sequence of spectral vectors, and $\{\mathbf{x}_t^*\}$ the sequence of the corresponding centered vectors. Let \mathbf{X}_0 denote the covariance matrix of the sequence $\{\mathbf{x}_t\}$:

$$\mathbf{X}_0 = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_t - \bar{\mathbf{x}})^T = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t^* \cdot \mathbf{x}_t^{*T}$$

and \mathbf{X}_k the lagged covariance matrix at the order k :

$$\mathbf{X}_k = \frac{1}{T} \sum_{t=k+1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_{t-k} - \bar{\mathbf{x}})^T = \frac{1}{T} \sum_{t=k+1}^T \mathbf{x}_t^* \cdot \mathbf{x}_{t-k}^{*T}$$

The dimension of the covariance matrix and of the lagged covariance matrices is $p \times p$. We now define a new matrix, \mathbf{X}_{2q+1} , by:

$$\mathbf{X}_{2q+1} = \begin{bmatrix} \mathbf{X}_0 & \mathbf{X}_1 & \dots & \mathbf{X}_{2q} \\ \mathbf{X}_1^T & \mathbf{X}_0 & \dots & \mathbf{X}_{2q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{2q}^T & \mathbf{X}_{2q-1}^T & \dots & \mathbf{X}_0 \end{bmatrix}$$

Note that this matrix is block-Toeplitz, and that its dimension is $(2q+1)p \times (2q+1)p$. This matrix can be interpreted as the covariance matrix of the vectors $\{\mathbf{X}_{t-q}^{t+q}\}_{1 \leq t \leq T}$. We now calculate the principal components of this matrix [5]. It is equivalent to the extraction of eigenvalues and eigenvectors of the matrix. The eigenvector associated with the largest eigenvalue is the direction of projection which conserves the maximum of the variance; The eigenvector associated to the second largest eigenvalue is the direction of projection which conserves the maximum of the variance uncorrelated (that is, orthogonal) to the first one; And so on. We have then:

$$\mathbf{X}_{2q+1} = \mathbf{V}_{2q+1} \cdot \mathbf{\Lambda}_{2q+1} \cdot \mathbf{V}_{2q+1}^T$$

with:

$$\mathbf{V}_{2q+1} = (\mathbf{v}_1, \dots, \mathbf{v}_{2q+1})$$

$$\mathbf{\Lambda}_{2q+1} = \text{diag}(\lambda_1, \dots, \lambda_{2q+1}), \quad \lambda_1 \geq \dots \geq \lambda_{2q+1}$$

The dimension of the matrix \mathbf{V}_{2q+1} and of the matrix $\mathbf{\Lambda}_{2q+1}$ is $(2q+1)p \times (2q+1)p$. The dimension of each vector $\mathbf{v}_i, 1 \leq i \leq 2q+1$, is $(2q+1)p$. Finally, the filtering matrix \mathbf{H} is composed of some of the eigenvectors \mathbf{v}_i , for instance $\mathbf{H} = [\mathbf{v}_2; \dots; \mathbf{v}_7; \mathbf{v}_{10}; \dots; \mathbf{v}_{14}]^T$.

4. Application to Speaker Identification

Task We tested the CPC in the framework of closed-set text-independent speaker identification. Closed-set means that the possibility of rejection is not taken into account: the test speaker is always part of the set of references. Text-independent means that we do not use any phonetic knowledge to do the identification.

Database We used a subset of the POLYCOST database [1], a telephone database, containing 112 speakers (64 females and 48 males). For each speaker, we used in average 90 seconds of speech for the training, and several utterances of 5 seconds in average for the tests. The total number of tests was 560.

Spectral and Cepstral Analysis Each utterance was analyzed as follows: the speech signal was decomposed in frames of 30 ms at a frame rate of 10 ms. A Hamming window was applied to each frame. The signal was pre-emphasized with a coefficient 0.95. For each frame, a fast Fourier transform was computed and provided 256 square module values representing the short term power spectrum in the 0-4 kHz band. This Fourier power spectrum was then used to compute 13 filter bank coefficients, using triangular filters placed on a non-uniform frequency scale, similar to the Bark/Mel scale. We finally took the base 10 logarithm of each filter output and multiplied the result by 10, to form a 13-dimensional vector of filter bank coefficients in dB. As an alternative to spectral analysis, we also extracted the cepstral coefficients c_1 to c_{12} from the spectral coefficients.

| Coeff. | static | | | static + Δ | | | One CPC filt. for all speakers | | | One CPC filt. for each speaker | | |
|--------|----------|-----------|--------------|-------------------|--------------|-----------|--------------------------------|-------------|-----------|--------------------------------|-----------|-----------|
| | 1 vector | 3 vectors | 5 vectors | 1 vector | 3 vectors | 5 vectors | 1 vector | 3 vectors | 5 vectors | 1 vector | 3 vectors | 5 vectors |
| Spec. | 21.96 | 19.64 | 18.93 | 11.96 | 11.35 | 11.61 | 9.82 | 9.11 | 9.82 | | | |
| Ceps. | 15.71 | 12.32 | 11.43 | 11.43 | 12.50 | 16.25 | 10.00 | 9.29 | 10.71 | | | |

Table 1. Speaker identification results when using the static vectors only, the static vectors and the Δ parameters, or the CPC filtering. Results are given in percentage of identification error.

CPC and Δ Parameters Once spectral or cepstral vectors have been extracted from a training utterance, we calculated the corresponding CPC using $q = 1, 2, 3$, and kept each time all the components. We then filtered the original vectors by these components to obtain a new set of feature vectors. Thus, for each speaker, we had a set of components for the filtering, and a sequence of filtered vectors by these components. We also computed the CPC corresponding to all the training data pooled together, which provided us with one CPC filtering for all the speakers instead of one CPC filtering for each speaker as previously. As an alternative to the CPC filtering, we also used the Δ parameters calculated over 3 or 5 vectors [3].

Modeling Each training sequence was then modeled by a Gaussian mixture model [11] using 8 components and diagonal covariance matrices.

5. Results

Results are reported in Table 1. The score of reference, 11.43 %, was obtained using cepstral coefficients augmented by the Δ parameters, with a Δ calculation over 5 vectors. We can observe that cepstral coefficients perform better than spectral coefficients when using the Δ parameters, which is a classical result in speaker identification. The error rates obtained after application of one CPC filtering for all speakers using the cepstral vectors as original vectors are not better than the score of reference. The error rates even degrade if we use 3 or 5 vectors. This is probably due to the loss of information. The cepstral parameterization is already a filtering of the spectral vectors, and then we add another filtering, which reduces too much the quantity of information contained in the new parameters. However, when we applied one CPC filtering for all speakers using the spectral vectors as original vectors, we performed better than the spectral vectors augmented by the Δ parameters, and even slightly better than the score of reference. When we used 3 vectors for the computation of the CPC, we obtained an error rate of 11.35 %, which allowed us to outperform the score of reference by 7.0 %. But our best error rates were obtained when we applied one CPC filtering for each speaker. When using cepstral vectors as original vectors, our best error rate, 9.29 %, was obtained with a calculation of the CPC over 3 vectors, and gave an improvement of 18.7 % over the score of reference. When using spectral vectors as original vectors, our best error rate, 9.11 %, was obtained with a calculation of the CPC over 3 vectors, and gave an improvement of 20.3 % over the score of reference.

6. Conclusions and Perspectives

In this paper, we proposed a new formalism called *vector filtering*, which consists in transforming a sequence of vectors through a matricial filtering. This formalism unifies a number of classical approaches like cepstral analysis or Δ parameters. We also showed how to integrate vector filtering in a pattern

recognition system. We then proposed a new filtering called *contextual principal components (CPC)* in the general case, or *time-frequency principal components (TFPC)* when the original vectors represent some information about the frequency content of the signal. We finally applied this new filtering to closed-set text-independent speaker identification, and improved the performance by roughly 20 % compared to the use of the classical cepstral coefficients augmented by their Δ parameters.

This new approach is therefore very promising, and this work can be prolonged in several directions. We intend to study the evolution of the performance when reducing the dimension of the new vectors by not keeping all the components. It is also possible to apply this new technique to other applications like language recognition or speech recognition. Finally, vector filtering provides a very interesting framework to study new signal parameterizations and new filterings.

References

- [1] *The European COST 250 Action entitled: Speaker Recognition in Telephony*, 1998. (<http://circhp.epfl.ch/polycost>).
- [2] F. Bimbot, E. Bocchieri, and B. Atal. Sous-espaces de projection de séquences de trames acoustiques pour l'analyse et la reconnaissance de parole. In *XXIèmes Journées d'Etude sur la Parole*, 1996. Avignon, France.
- [3] S. Furui. Comparison of speaker recognition methods using static features and dynamic features. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(3):342–350, June 1981.
- [4] Y. Grenier. Utilisation de la prédiction linéaire en reconnaissance et adaptation au locuteur. In *XIèmes Journées d'Etude sur la Parole*, pages 163–171, May 1980. Strasbourg, France.
- [5] I. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.
- [6] I. Magrin-Chagnolleau. *Approches statistiques et filtrage vectoriel de trajectoires spectrales pour l'identification du locuteur indépendante du texte*. PhD thesis, École Nationale Supérieure des Télécommunications, January 1997.
- [7] I. Magrin-Chagnolleau. Utilisation du filtrage vectoriel en reconnaissance des formes. In *Proceedings of RFIA 2000*, volume II, pages 157–166, February 2000. Paris, France.
- [8] I. Magrin-Chagnolleau and G. Durou. Time-frequency principal components of speech: application to speaker identification. In *Proceedings of EUROSPEECH 99*, pages 759–762, September 1999. Budapest, Hungary.
- [9] B. Milner. Inclusion of temporal information into features for speech recognition. In *Proceedings of ICSLP 96*, 1996.
- [10] A. V. Oppenheim and R. W. Schaffer. Homomorphic analysis of speech. *IEEE Transactions on Audio and Electroacoustics*, 16(2):221–226, June 1968.
- [11] D. Titterton, A. Smith, and U. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and sons, 1985.

A Further Investigation on Speech Features for Speaker Characterization (2000)

Ivan Magrin-Chagnolleau, Guillaume Gravier, Mouhamadou Seck, Olivier Boeffard, Raphaël Blouet, and Frédéric Bimbot. **A Further Investigation on Speech Features for Speaker Characterization.** *Proceedings of ICSLP 2000*, Beijing, China, October 2000.

4 pages

A FURTHER INVESTIGATION ON SPEECH FEATURES FOR SPEAKER CHARACTERIZATION

I. Magrin-Chagnolleau, G. Gravier, M. Seck, O. Boeffard, R. Blouet, F. Bimbot.*

IRISA (CNRS & INRIA), Campus universitaire de Beaulieu, 35042 Rennes cedex, France – European Union.

* also with ENSSAT, 6 rue de Kerampont, BP447, 22305 Lannion cedex, France – European Union.

ivan@ieee.org - ggravier@irisa.fr - mseck@irisa.fr
olivier.boeffard@enssat.fr - rblouet@irisa.fr - bimbot@irisa.fr

ABSTRACT

In this article, we investigate on alternative speech features for speaker characterization. We study Line Spectrum Pairs features, Time-Frequency Principal Components and Discriminant Components of the Spectrum. These alternative features are tested and compared on a task of speaker verification. This task consists in verifying a claimed identity from a speech segment. Systems are evaluated on a subset of the evaluation data of the NIST 1999 speaker recognition campaign. The new speech features are also compared to the classical cepstral coefficients, which remain, in our experiments, the best performing features.

1. INTRODUCTION

Cepstral coefficients and their delta parameters have been intensively used in speech processing. Although they have been developed for speech recognition, they are also used for speaker recognition mainly because of their nice statistical properties and the possibility of doing channel compensation in the cepstral domain. In this paper, we investigate on alternative speech features, namely Line Spectrum Pairs (LSP), Time-Frequency Principal Components (TFPC) and Discriminant Components of the Spectrum (DCS).

LSP features have been used for speech coding [9] but rarely for speaker characterization. Liu *et al.* studied LSP derived parameters in a VQ based text-dependent speaker verification system and concluded towards a better performance of the LSP frequencies over the cepstral coefficients [1]. In this paper, we use these parameters for text-independent speaker verification using Gaussian mixture models (GMM).

The TFPC analysis [2] is a way of capturing dynamic information in the speech signal by a principal component

A part of this work has been done in the framework of the first phase of the AGIR project, funded by the French RNRT Programme.

analysis of the contextual matrix, which is the covariance matrix of feature vectors augmented by their time context. In former work, TFPC has shown interesting properties in enhancing some speaker-specific information.

DCS features are obtained by a Linear Discriminant Analysis on the log-magnitude spectrum. LDA has already been used for speech analysis in speaker verification. However, in our approach, we take into account a priori knowledge, such as the gender of the speaker and the handset type of the call, for estimating the discriminant components.

These three speech features are tested and compared on a speaker verification task. This task consists in verifying a claimed identity from a speech segment. The experiments are conducted on a subset of the evaluation data of the NIST 99 speaker recognition campaign. This corpus was extracted from the Switchboard corpus and is composed of conversations recorded over the telephone.

2. FEATURES

2.1. Line spectrum Pairs

LSP frequencies are related to linear predictive analysis. Considering the p order LPC inverse filter $A_p(z)$, the LSP frequencies are defined as the zeros of the two polynomials defined by $A_p(z) \pm z^{-(p+1)}A_p(z^{-1})$. It can be shown that these zeros are on the unit circle and correspond to frequencies related to the formant frequencies of the LPC filter [9].

The p LSP frequencies are ordered to form a feature vector. The main difficulty with the LSP comes from the fact that the features are ordered and take their values in a bounded interval, namely $]0, 0.5[$. These two properties are not very suited for Gaussian modeling. In order to overcome the ordering property, the LSP frequencies are centered by subtracting the long term mean of each feature. In this pa-

per, 16th order linear prediction is considered. The frame length is set to 20 ms with a 50% overlap.

2.2. Time-Frequency Principal Components

A potential way to characterize a speaker is to extract time-frequency patterns that are characteristic of the acoustic vector trajectory of that speaker. In the case of TFPC, these time-frequency patterns are obtained by calculating the principal components of the contextual feature vector, which is the feature vector augmented by its time context. The original sequence must be long enough to be representative of the speaker we want to characterize with the time-frequency patterns.

Once the patterns have been extracted, they are used to filter the acoustic trajectory of the training material and of the test material. Any modeling technique can then be applied to the new vectors, as it is done usually on cepstral vectors. For a detailed presentation of the TFPC analysis, see [2, 3].

2.3. Discriminant Components of the Spectrum

In the case of DCS, we consider the set of spectral features (primary vectors) of different speakers as belonging to different classes. Linear Discriminant Analysis (LDA) aims at finding a new feature vector, obtained by linear transformation of the primary vectors, with lower dimension and which maximally separates the different classes. Here, a class corresponds to a speaker, and primary vectors are the log-module of the spectrum magnitude, obtained by Fourier Transform.

We learn the LDA on a variety of speakers, in order to obtain features which are speaker-discriminant on the learning set. Coefficients of the linear discriminant transformation are the eigenvectors of the matrix $V^{-1}B$, where V is the total covariance matrix and B is the covariance matrix between classes. Discriminant components are selected according to their corresponding eigenvalues (in decreasing order) and form the Discriminant Components of the Spectrum (DCS).

The result of the LDA depends on the composition of analysis data, between the different speakers and the two handset types¹. For taking care of the gender variety, we use the same number of speakers for each gender. For handling the different type of handsets, we weight the electret and carbon observations, in order to have equal total weights for the two types of data. As a consequence, observation from speaker s and handset type h , have the weight $\frac{1}{2 * S * N_{sh}}$, where S is the number of learning speakers, and N_{sh} is the the number

¹In the USA, two types of handset are used : electret and carbon.

of observations from speaker s and handset type h . In this experiment, $S = 50$.

Note that an important limitation of DCS, is the impossibility of applying mean subtraction (which is however a well-known technique for channel mismatch compensation). In fact, if mean subtraction were applied in the spectral domain, class means would all be equal to zero, leading to a null covariance matrix between classes B .

In this paper, each speaker have at least 30 seconds of data for each handset type. For robustness of LDA to channel effects, different recordings are also used for a speaker. Frames of 16 ms duration with 50% overlap are used. Spectral analysis is restricted to the telephone band [340 Hz, 3400 Hz]. Features are 16 DCS, augmented by their delta and the delta log-energy.

3. EXPERIMENTS

3.1. Database

The evaluation data for the NIST 1999 speaker recognition campaign [6] comes from the Switchboard 2–Phase 3 corpus collected by the Linguistic Data Consortium (LDC). The training data for each target speaker consist of two utterances of 1 minute each in average, obtained by concatenating consecutive turns of the speaker. Segments of silence were removed. We use a subset of this corpus for our experiments. The total number of target speakers is 119 (79 female speakers and 40 male speakers).

The test data consist of segments containing speech of only one speaker, obtained by concatenating consecutive turns of this speaker. Here also, segments of silence were removed. The duration of the test segments is between a few seconds and 1 min. There are 857 test segments, tested against several claimed identities, totaling 9427 tests. Claimed speakers have always the same gender than the test segment speaker.

Different test conditions are considered by NIST. Test conditions are defined according to the call phone number and handset type mismatches, between the test segment and train data (when the claimed identity is the good one). The four test conditions are : Same Number / Same Type (SNST), Same Number / Different Type (SNDT), Different Number / Same Type (DNST) and Different Number / Different Type (DNST).

3.2. Verification system

The verification system is a variant of the ELISA Text-Independent Speaker Verification Platform [7]. It proceeds by

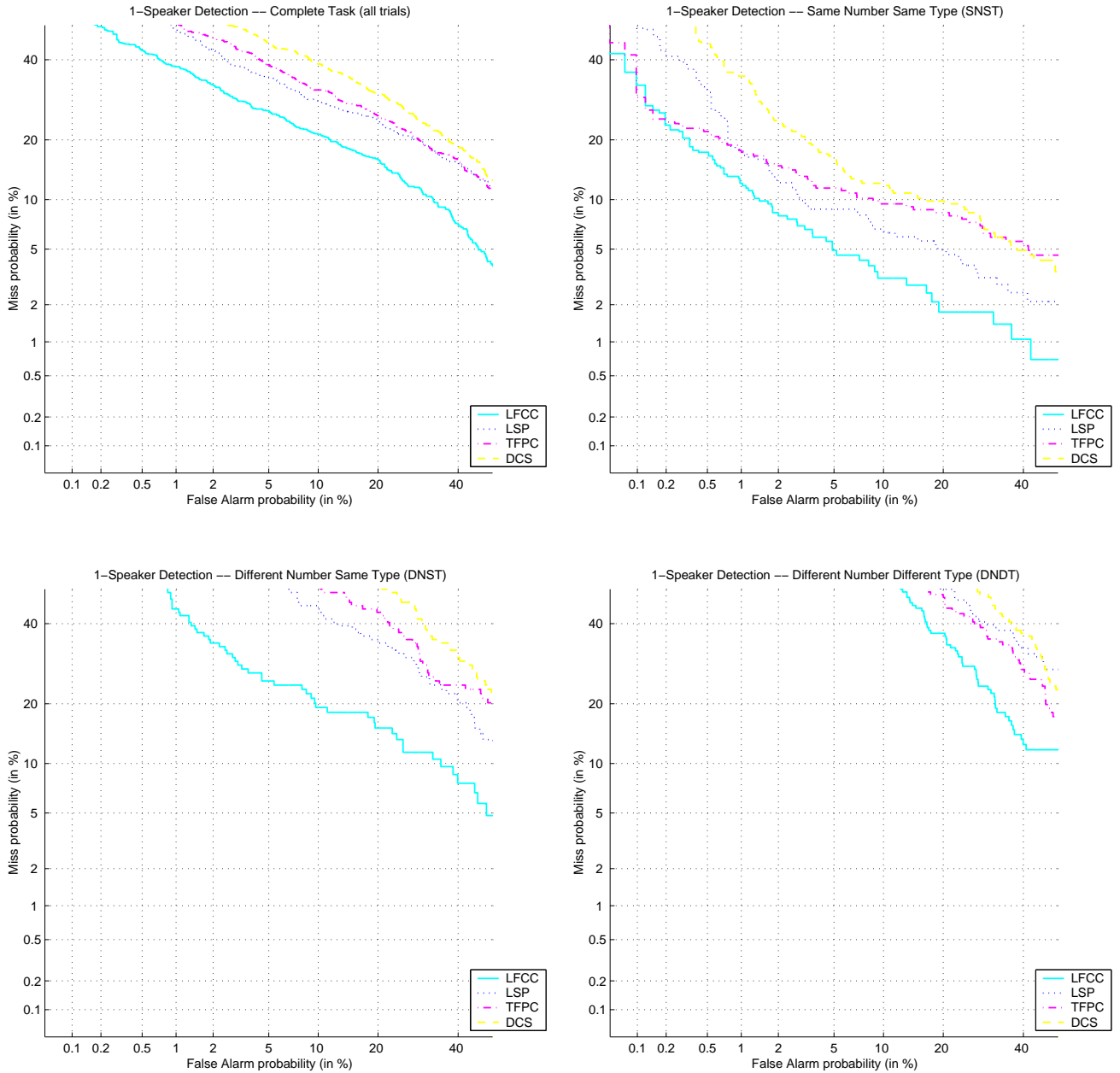


Figure 1: Performances of LFCC, LSP, TFPC and DCS, on all trials and for the three following NIST test conditions : SNST, DNST, DNDT

training a target speaker model for each client-speaker, and a speaker-independent world-model for each gender. Both speaker models and world models are mixtures of 128 Gaussians with diagonal covariance matrices [8]. For each frame in a test segment, the log-likelihood ratio (LLR) between the target speaker model and the world-model is calculated. The utterance score is obtained by averaging the frame-

based LLR. This score is compared to a threshold to make the final decision.

We compare the features defined in section 2 with a reference configuration based on Linear Frequency Cepstral Coefficients (LFCC) plus their first order deltas augmented with the delta log energy. For the reference feature set, a

24 channel triangular filter-bank is applied to 20 ms frames every 10 ms. The 24 filter central frequencies are regularly spread along the frequency range [340 Hz - 3400 Hz]. Finally, cepstral coefficients c_1 to c_{16} are computed from the filter bank output.

3.3. Evaluation

The evaluation of the systems is done using Detection Error Trade off (DET) curves [4]. This representation is a way to show all of the possible operating points of a system (false alarm rate vs miss rate) in a scale which makes the result curves rather linear and easier to compare.

4. RESULTS

Figure 1 represents the performance of the different features, when evaluated on all trials and when restricted to the three following test conditions : SNST, DNST, DNDT.

As usually observed, performance decreases when the call phone number is different (from SNST to DNST) and drops further when the type of handset changes (DNST to DNDT). All four feature sets seem very sensitive to phone number (channel) mismatch effects.

The LSP and TFPC alternate in second and third position, according to the handset-type condition. This suggests that LSP may be more sensitive than TFPC to handset type mismatch.

DCS perform systematically worse than the other feature sets. This result may be explained by the fact that the population of learning speakers is rather limited in this experiment. Note also that, as mentioned earlier, no channel compensation was carried out.

Finally, all tested parameters are clearly outperformed by the conventional LFCC, which shows that, even though these acoustic features are not specifically designed for speaker characterization, they possess good properties for this task.

5. SUMMARY AND CONCLUSIONS

Our experiments have not succeeded in evidencing a benefit of alternate acoustic features over cepstral coefficients, for speaker characterization. This may be partly owed to the lack in implementing equivalent pre- and post-processing for computing the alternate parameters. Future efforts should focus on attempts in unbounding the LSP, normalizing the likelihood ratio according to the speaker-dependent transformation for TFPC, better addressing channel compensa-

tion issues for the DCS, etc...

However, our experiments confirm the difficulty in finding, for speaker verification, better acoustic features than the conventional cepstrum coefficients, for which the theory and the know-how are well mastered.

6. REFERENCES

- [1] Chi-Shi Liu, Min-Tau Lin, Wern-Jun Wang and Hsiao-Chuan Wang Study of Line Spectrum Pair frequencies for speaker recognition. *ICASSP*, 1990.
- [2] Ivan Magrin-Chagnolleau. *Approches statistiques et filtrage vectoriel de trajectoires spectrales pour l'identification du locuteur indépendante du texte*. PhD thesis, École Nationale Supérieure des Télécommunications, January 1997.
- [3] Ivan Magrin-Chagnolleau and Geoffrey Durou. Application of time-frequency principal component analysis to speaker verification. *Digital Signal Processing*, 10(1-3), April 2000.
- [4] A. Martin et al. The DET curve in assessment of detection task performance. In *Proceedings of EURO-SPEECH 97*, volume 4, pages 1895–1898, September 1997. Rhodes, Greece.
- [5] Geoffrey J. McLachlan and Kaye E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, 1988.
- [6] Mark Przybocki and Alvin Martin. The NIST 1999 speaker recognition evaluation - an overview. *Digital Signal Processing*, 10(1-3), April 2000.
- [7] The ELISA Consortium. The ELISA Systems for the NIST'99 Evaluation in Speaker Detection and Tracking. *Digital Signal Processing*, 10(1-3), April 2000.
- [8] Douglas A. Reynolds. Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication*, 17(1-2):91–108, August 1995.
- [9] Frank K. Soong and Biing-Hwang Juang Line Spectrum Pair (LSP) and Speech data compression. *ICASSP*, 1984.
- [10] D. M. Titterton, A. F. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and Sons, 1985.

**Overview of the 2000-2001 ELISA Consortium
Research Activities (2001)**

Ivan Magrin-Chagnolleau, Guillaume Gravier, and Raphaël Blouet for the ELISA consortium. **Overview of the 2000-2001 ELISA Consortium Research Activities.** *Proceedings of the Workshop 2001 : A Speaker Odyssey*, Chania, Crete, Greece, June 2001.

6 pages

OVERVIEW OF THE 2000-2001 ELISA CONSORTIUM RESEARCH ACTIVITIES

Ivan Magrin-Chagnolleau, Guillaume Gravier, and Raphaël Blouet

for the ELISA consortium.

elisa@listes.univ-avignon.fr

ABSTRACT

This paper summarizes the research activities in speaker recognition in the framework of the ELISA consortium. The ELISA speaker recognition common platform is first presented, including the common evaluation protocol and the functioning of the consortium. Then experiments with this platform on the development data of the NIST 2001 speaker recognition campaign are reported. Finally, a survey of the research directions in the various ELISA laboratories is given.

1. INTRODUCTION

The ELISA consortium was originally created by ENST, EPFL, IDIAP, IRISA and LIA in 1998 with the aim of developing a common state-of-the-art speaker verification system and participating in the yearly NIST speaker recognition evaluation campaigns. Along the years, the composition of the consortium has changed and today ENST, IDIAP, IRISA and LIA are members of the consortium. Since 1998, the members of the Consortium have been participating in the NIST evaluation campaigns in speaker recognition and a comparative study of the various systems presented in the 1999 campaign can be found in [2].

The aim of the Consortium is to promote scientific exchanges between members. To reach this goal, a common baseline reference platform is maintained by all the members. The reference platform is modular in order to be easily modified and reflects state-of-the-art performance achieved with gaussian mixture models (GMMs). Modules are provided for the various tasks of the NIST evaluations, namely speaker verification, detection, tracking, and segmentation. The possibility of the platform to deal with segmental approaches [4] at the score computation level enables an easy integration of the speaker detection, tracking, and segmentation tasks in the same platform. A common evaluation protocol,

The list of the current members of the consortium is, in alphabetical order: F. Bimbot, R. Blouet, J.F. Bonastre, G. Chollet, C. Fredouille, G. Gravier, J. Kharroubi, I. Magrin-Chagnolleau, J. Mariethoz, S. Meignier, T. Merlin, and M. Seck.

derived from the NIST evaluation rules, is shared by all the Consortium members to allow fair comparisons between the variants of the baseline system.

In this paper, we describe in Section 2 the common resources of the ELISA consortium, including the architecture of the platform, the common evaluation protocol, and the functioning of the consortium. In Section 3, we report on some experiments that were carried out to bring the platform to state-of-the-art performance in speaker verification. In Section 4, we point out the various research directions of the laboratories member of the Consortium.

2. THE ELISA COMMON RESOURCES

2.1. Platform Architecture

The ELISA platform is composed of the following main modules: *speech parameterization*, *modeling*, *likelihood calculation*, *normalization*, and *decision / scoring*. The overall architecture of the platform is illustrated on Fig. 1. The *speech parameterization* module implements classical speech analyses such as filterbank analysis or cepstral analysis plus frame selection methods. The *modeling* module is based on gaussian mixture models (GMMs) with maximum likelihood (ML) parameter estimation and/or maximum a posteriori (MAP) adaptation of a speaker independent model. Various MAP adaptation techniques have been implemented. The *likelihood* module is in charge of log-likelihood computation and of mixture component selection for scoring. Several likelihood ratio normalization procedures are possible in the *normalization* module, such as z-norm [8], h-norm [13], t-norm [1], and World + MAP [5]. The *decision / scoring* module makes the decision by comparing a normalized likelihood ratio to a threshold and plots the DET curves [10].

2.2. Common Evaluation Protocol

The main goal of the ELISA Common Evaluation Protocol (CEP) is to make the results comparable within the Consortium while enabling the preparation of the next NIST

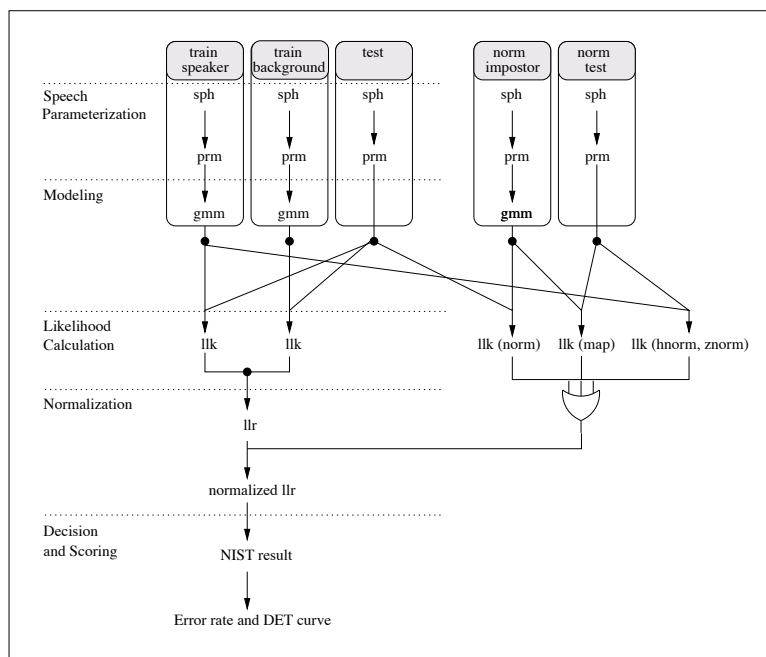


Fig. 1. Architecture of the ELISA reference platform.

evaluation campaign. The protocol is therefore redefined every year on a subset of the development data of the upcoming NIST evaluation campaign. The CEP for the 2001 campaign was defined as follows. 4 subsets of the NIST development data were defined. The first subset, called the *world* subset, is composed of 316 speakers (186 females using electret handsets, 10 females using carbon handsets, 102 males using electret handsets, and 18 males using carbon handsets), and is used to train gender and handset dependent world models. The second subset, called the *dev* subset, is composed of 100 speakers (50 females and 50 males), and is used to do some development experiments. The third subset, called the *eva* subset, is also composed of 100 speakers (50 females and 50 males), and is used to cross-validate the experiments done on the *dev* subset. Finally, the fourth subset, called the *norm* subset, is composed of 159 speakers (50 females using electret handsets, 24 females using carbon handsets, 50 males using electret handsets, and 35 males using carbon handsets), and is used for various normalizations of the log-likelihood ratios. These four subsets are used by all the members of the Consortium in order to have comparable results.

2.3. Functioning of the Consortium

Members of the consortium have regular meetings during the year. In average, they meet once every trimester. These meetings are the occasion to discuss current development and research issues, to compare the last results obtained by

each laboratory, and to set up new goals until the next meeting. Each laboratory decides to focus on a particular aspect in order to avoid redundant work.

Each year, a standard configuration is defined as being the baseline configuration. A new configuration is integrated only if it has been proven to provide better performance than the baseline configuration. In that case, the new configuration becomes the new baseline system.

Additionally, results are regularly shared on the web site of the consortium, and a mailing list allows the consortium members to regularly discuss encountered problems, or any topic of interest.

3. EXPERIMENTS

After a presentation of the common resources of the ELISA consortium and its functioning, we present in this section experiments carried out to bring the platform to state-of-the-art performance.

3.1. Task

Although the ELISA platform can be used to deal with any of the tasks proposed in the NIST speaker recognition evaluation, namely speaker verification, speaker detection, speaker tracking, or speaker segmentation, we focus in this section on text-independent speaker verification (called *one speaker*

detection by NIST), which consists in verifying a claimed identity from a recorded speech utterance, without using any prior phonetic knowledge.

3.2. Database

Experiments are reported on the *dev* subset described in the common evaluation protocol.

3.3. Speech Analysis

Each speech utterance is converted from a μ -law into a linear representation with a sampling frequency of 8 kHz. Each utterance is decomposed into frames of 20 ms extracted every 10 ms. A Hamming window is applied to each frame. The signal is not pre-emphasized. For each frame, a fast Fourier transform is computed and provides square modulus values representing the short term power spectrum in the 0-4000 Hz band. This Fourier power spectrum is then used to compute 24 filterbank coefficients, using triangular filters placed on a linear frequency scale in the 300-3400 Hz band. The base 10 logarithm of each filter output is taken and multiplied by 10, to form a 24-dimensional vector of filterbank coefficients in dB. Then, cepstral coefficients [12] c_1 to c_{16} , augmented by their Δ coefficients [6] (calculated over 5 vectors) are calculated, and a cepstral mean subtraction (CMS) is applied. We finally obtain 32-dimensional feature vectors.

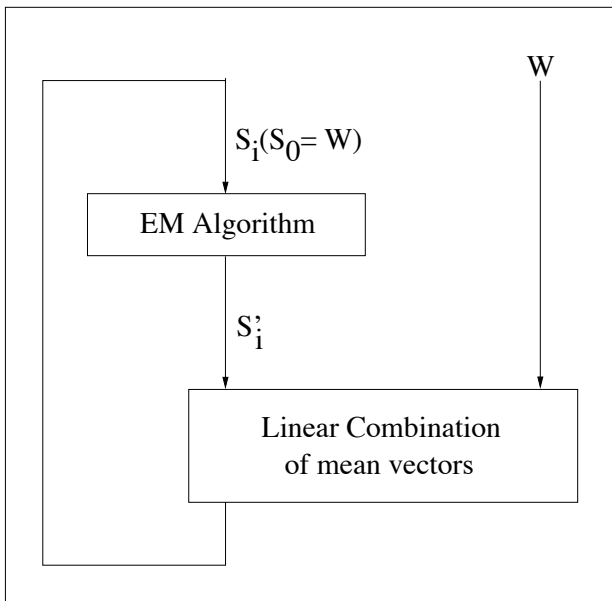


Fig. 2. MAP procedure used for the training of a speaker model.

3.4. World Models

For the world models, gaussian mixture models (GMMs) [14] with 128 components and diagonal covariance matrices are trained on speech utterances from the various *world* subsets of the common evaluation protocol using the EM algorithm [3]. The world models are gender and handset dependent.

3.5. Speaker Models

Speaker models are trained using a MAP procedure described in Fig. 2. The corresponding world model, denoted by W , is used as an initialization. Then, one iteration of the EM algorithm is applied, leading to the model S' . Then a new model S_1 is built with the mean vectors being a linear combination of the mean vectors of the two models W and S' :

$$\mu_{S_1}^{(i)} = \frac{0.25 \cdot \pi_W^{(i)} \cdot \mu_W^{(i)} + 0.75 \cdot \pi_{S'}^{(i)} \cdot \mu_{S'}^{(i)}}{0.25 \cdot \pi_W^{(i)} + 0.75 \cdot \pi_{S'}^{(i)}}$$

with $\pi_W^{(i)}$ and $\pi_{S'}^{(i)}$ being the weights of the i -th Gaussian component of models W and S' respectively.

Finally, S_1 is used as the initialization for the second iteration of the EM algorithm. 20 iterations are done that way.

3.6. Evaluation

Results of the various systems are measured by a DET curve [10]. For the speaker verification task, the false alarm rate and the miss rate are defined as follows:

$$\mathcal{R}_{FA} = \frac{\text{Number of impostor utterances wrongly accepted}}{\text{Total number of impostor utterances}}$$

$$\mathcal{R}_{MI} = \frac{\text{Number of client utterances wrongly rejected}}{\text{Total number of client utterances}}$$

3.7. Experiments on the Energy

The first set of experiments concerns the integration of the Δ -log-energy and/or the log-energy in the parameter vectors. Fig. 3 shows the results without any normalization technique. The score used is simply the averaged log-likelihood ratio.

The addition of the Δ -log-energy in the feature vectors gives very similar results to a system using only the cepstral and the Δ -cepstral coefficients. Adding the log-energy to the feature vectors clearly degrades the performance. This latter result is surprising because the energy should be used by the GMMs to make a pre-classification between high and low energy frames, thus enabling a better modeling. One possible explanation is the lack of normalization of the

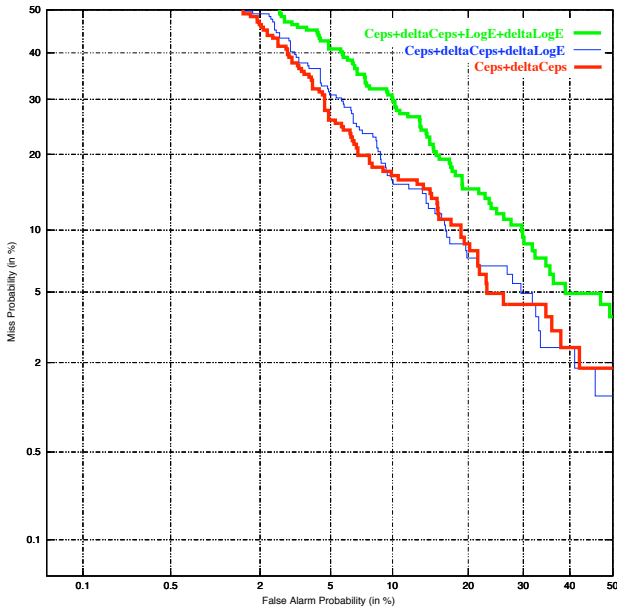


Fig. 3. Influence of the log-energy and the Δ -log-energy on the performance.

log-energy. Some kind of normalization of the log-energy should be used to limit the differences between two utterances.

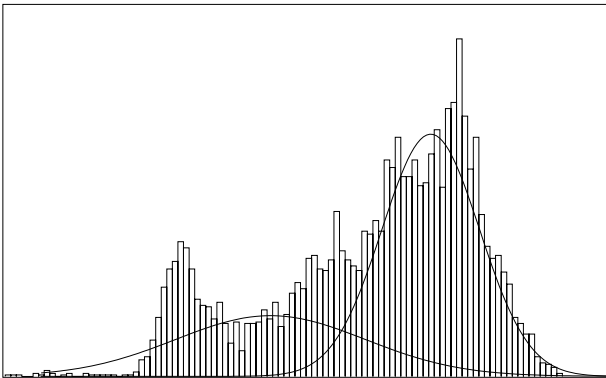


Fig. 4. Bi-gaussian modeling of the energy distribution.

3.8. Experiments on Frame Removal

The second set of experiments concerns the study of the influence of frame removal on performance. The method used is based on a bi-gaussian modeling of the energy distribution. First, each frame with a zero energy is automatically discarded. These frames correspond generally to the beginning or the end of a recorded segment, and there are frequently such frames in the Switchboard data. Then, the energy distribution of the remaining frames is calculated, and

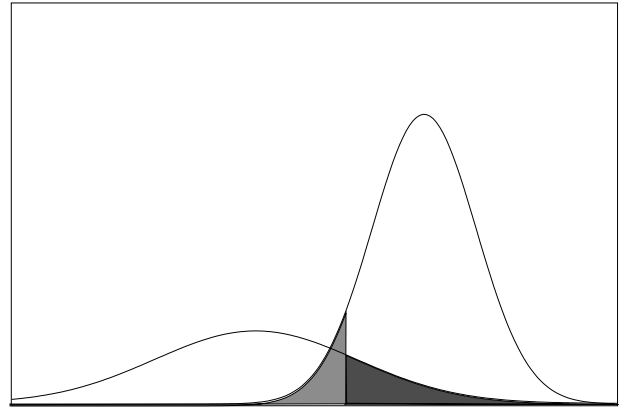


Fig. 5. Calculation of the threshold on the energy.

a bi-gaussian model is learned from this distribution using the expectation-maximization (EM) algorithm [3]. Fig. 4 shows an example of energy distribution for a typical test utterance of the NIST speaker recognition evaluation. The bi-gaussian model is also represented on the figure. Once the bi-gaussian model has been estimated, a threshold on the energy is calculated such that the residual surfaces under each gaussian are equal (see Fig. 5). Finally, each frame with an energy below the threshold is discarded.

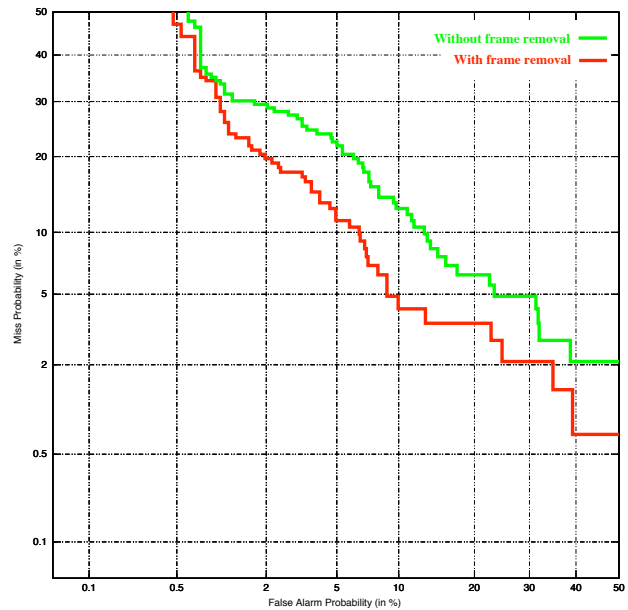


Fig. 6. Influence of frame removal on the performance.

Fig. 6 summarizes these experiments and shows results after the application of the h-norm. The application of a frame removal technique improves the performance considerably.

This suggests that the information extracted from low energy frames is not reliable and/or cannot be modeled accurately with GMMs. It therefore helps to simply discard such frames. However, it may be interesting in the future to investigate which, if any, information can be used from those low energy frames.

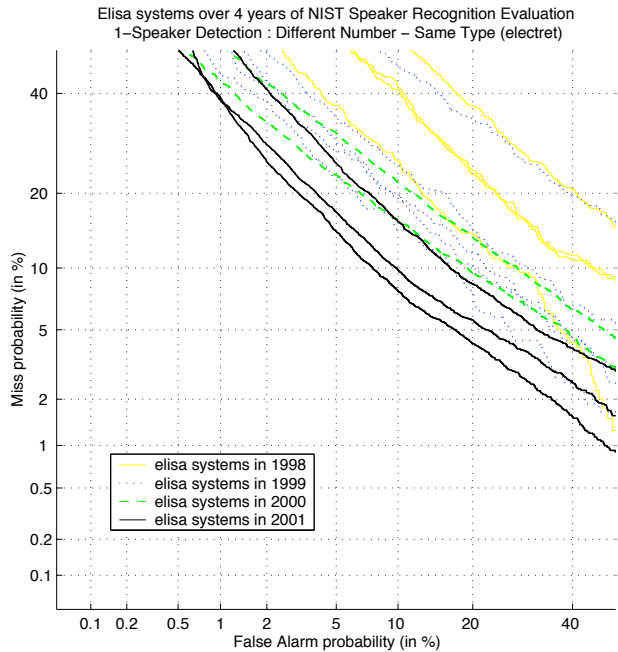


Fig. 7. ELISA systems over 4 years of NIST speaker recognition evaluation.

4. RESEARCH DIRECTIONS

Being at the level of the state-of-the-art speaker verification systems allows the Consortium members to investigate on more original research directions. We present the main ones in this section. Some of them are described in details in other papers, other approaches are still currently under investigation.

- Contextual principal components and contextual independent components as an alternative to cepstral analysis for the speech representation module [9];
- Evaluation of the intrinsic quality of a new parameterization using a mutual information criterion;
- Alternative initialization procedures of the world models (vector quantization, mixture of models learned on various random subsets of data, etc.);

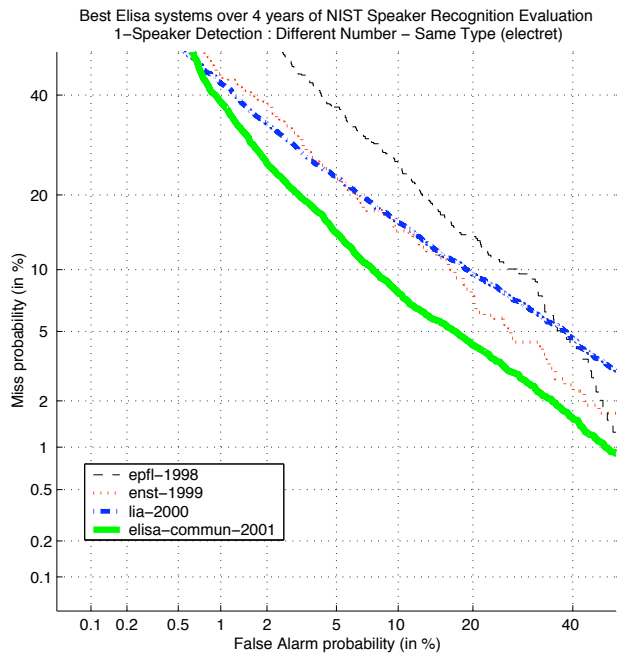


Fig. 8. Best ELISA systems over 4 years of NIST speaker recognition evaluation.

- Modeling by a mixture of models, one being trained with an ML procedure, the other being trained with a MAP procedure;
- Alternative MAP procedures to train the speaker models;
- Evaluation of the intrinsic quality of the world and speaker models;
- The use of support vector machines (SVM) as an alternative to log-likelihood-based scoring [7];
- The study of new approaches for score normalization [5];
- Evolutive hidden Markov models for speaker indexing [11].

5. CONCLUSIONS

The ELISA consortium has been created 4 years ago, and a lot of progress has been made since then (see Fig. 7¹ and Fig. 8). For the first time this year, an ELISA platform has been able to provide state-of-the-art performance

¹The training material and the conditions of the evaluation have changed over the years. Read the evaluation plans provided by NIST on <http://www.nist.gov/speech/tests/spk/index.htm> for details.

in speaker verification, allowing the members of the consortium to have of a good platform for experimentation in speaker recognition, and leaving time for original research on this topic. One of the main goals of the consortium has been reached thanks to common work between several researchers belonging to various research laboratories. This is an excellent example of collaborative research, and we hope that this year accomplishment will stimulate creative research inside (and outside) the consortium.

6. REFERENCES

- [1] Roland Auckenthaler, Michael Carey, and Harvey Lloyd-Thomas. Score normalization for text-independent speaker verification systems. *Digital Signal Processing*, 10(1-3):42-54, January/April/July 2000.
- [2] The ELISA Consortium. The ELISA 99 speaker recognition and tracking systems. *Digital Signal Processing*, 10(1-3), January/April/July 2000.
- [3] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 6(39):1-38, 1977.
- [4] Corinne Fredouille, Jean-François Bonastre, and Teva Merlin. AMIRAL: A block-segmental multirecognizer architecture for automatic speaker recognition. *Digital Signal Processing*, 10(1-3):172-197, January/April/July 2000.
- [5] Corinne Fredouille, Jean-François Bonastre, and Teva Merlin. Bayesian approach based-decision in speaker verification. In *Proceedings of 2001: A Speaker Odyssey*, June 2001. Crete, Greece.
- [6] Sadaoki Furui. Comparison of speaker recognition methods using static features and dynamic features. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(3):342-350, June 1981.
- [7] Jamal Kharroubi and Gérard Chollet. Text-independent speaker verification using support vector machines. In *ICASSP Student Forum*, May 2001.
- [8] Kung-Pu Li and Jack E. Porter. Normalizations and selection of speech segments for speaker recognition scoring. In *Proceedings of ICASSP 88*, pages 1895-1998, 1988.
- [9] Ivan Magrin-Chagnolleau, Geoffrey Durou, and Frédéric Bimbot. Application of time-frequency principal component analysis to text-independent speaker identification. *Accepted for publication in IEEE Transactions on Speech and Audio Processing*.
- [10] Alvin Martin et al. The DET curve in assessment of detection task performance. In *Proceedings of EUROSPEECH 97*, volume 4, pages 1895-1898, September 1997. Rhodes, Greece.
- [11] Sylvain Meignier, Jean-François Bonastre, and Stéphane Igounet. E-HMM approach for learning and adapting sound models for speaker indexing. In *Proceedings of 2001: A Speaker Odyssey*, June 2001. Crete, Greece.
- [12] Alan V. Oppenheim and Ronald W. Schafer. Homomorphic analysis of speech. *IEEE Transactions on Audio and Electroacoustics*, 16(2):221-226, June 1968.
- [13] Douglas Reynolds. Comparison of background normalization methods for text-independent speaker verification. In *Proceedings of EUROSPEECH 97*, pages 963-966, 1997. Rhodes, Greece.
- [14] Douglas A. Reynolds. Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication*, 17(1-2):91-108, August 1995.

**SPeaker and Language Characterization (SpLC) : A
Special Interest Group (SIG) of ISCA (2001)**

Jean-François Bonastre, Ivan Magrin-Chagnolleau, Stefan Euler, François Pellegrino, Régine André-Obrecht, John Mason, and Frédéric Bimbot.
SPeaker and Language Characterization (SpLC) : A Special Interest Group (SIG) of ISCA. *Proceedings of EUROSPEECH 2001*, Aalborg, Denmark, September 2001.

4 pages

SpLC: Speaker and Language Characterization (SpLC): A Special Interest Group (SIG) of ISCA

*Jean-François Bonastre (1), Ivan Magrin-Chagnolleau (1), Stephan Euler (2)
François Pellegrino (3), Régine André-Obrecht (4)
John Mason (5), Frédéric Bimbot (6)*

- (1) LIA University of Avignon, Avignon (France)
- (2) Robert-Bosch GmbH, Stuttgart (Germany)
- (3) DDL, University of Lyon II, Lyon (France)
- (4) IRIT, University Paul Sabatier, Toulouse (France)
- (5) University of Swansea, Swansea (UK)
- (6) IRISA/INRIA, Rennes (France)

Abstract

Last year, SpLC - an ISCA Special Interest Group (SIG) centered around Speaker and Language Characterization was born. The aims of this paper are to present the SpLC SIG, its objectives, and the work done during the first year.

1. Introduction

The SpLC ISCA-SIG was launched during the RLA2C workshop, in 1998, with the objectives of providing an entry point and a network for all the persons involved in the field of Speaker and Language Characterization, in order to facilitate the exchange of information, to stimulate discussions between the different communities and countries, to compare viewpoints and approaches, etc.

People interested in speaker and language characterization have various origins: computer science, computer engineering, electrical engineering, speech science, phonetics, linguistics, etc. This characteristic highlights the need of collaboration and exchange of information.

The aims of SpLC are:

- to facilitate exchanges between members;
- to provide information on the area;
- to promote SpLC activities;
- to provide members of ISCA with a special interest in SpLC with a means of exchanging news of recent research developments and other matters of interest in SpLC;
- to sponsor meetings and workshops in SpLC that appear to be timely and worthwhile, operating within the framework of ISCA's by-laws for SIGs;
- to provide and make available resources relevant to SpLC, including text and speech corpora, analysis

tools, analysis and generation software, research papers and generated data.

2. SPLC Internal Organization

SpLC organizes its activities around four main areas:

- Academic research on Speaker Characterization and Recognition;
- Forensic aspects of Speaker Characterization and Recognition;
- Language Characterization and Recognition;
- Commercial Applications for Speaker and Language Recognition.

In addition, several transversal activities will be proposed. For the moment, three are already defined:

- *Evaluation group*: this group works on the evaluation paradigm, for all the various axes of the SpLC domains. Particularly, the group will highlight evaluation tasks for commercial applications, for speaker tracking and indexing, and for language identification.
- *Meetings and Workshops group*: this group concerns the organization of SpLC-events, like workshops, meetings, summer schools, exhibitions, etc.
- *Speaker Indexing group*: this group proposes to work on algorithms to do speaker indexing, which is a recent task in the speaker recognition community.

3. The main areas

As written before, SPLC is organized around four main areas. Each main activity has (or will have) its own group leader(s) and its electronic communication tools (at the beginning, one electronic mailing list). Each activity organizes the work, the meetings, etc.

3.1. Academic research on Speaker Characterization and Recognition

Automatic speaker recognition is just one kind of biometric security mode. It is deemed not to be particularly accurate in relation to some other modes, but is very convenient. Thus increasing the accuracy of the speech mode has been the goal of many researchers for many years. The recent NIST open trials has focused the work of several leading sites, with controlled, inherently competitive evaluations. This has unquestionably helped towards achieving the goal of greater accuracy, largely it would seem through the introduction of new normalization techniques.

Speaker recognition accuracy in practice is constrained by natural variations in an individual's speech from one occasion to another and the recording conditions. Both cases demand capturing sufficient representative training data, since strategies for generalization are currently inadequate. Strategies to overcome or circumvent these limitations are topics for today's research.

One potentially productive avenue for improved accuracy is multi-modal biometrics. The suitability of different modes is likely to be very application dependent, though obvious candidates to complement speech (audio) are visual speech based on lips and the face generally [1],[2],[3].

Possible roles for SpLC here are clear:

- to foster multi-modal work in biometric security where speech is an integral part.
- to evolve new groupings that identify and publicise on-going research in multi-modal biometrics.

A good example of the latter is AVBPA (Audio- and Video-based Biometric Person Authentication). In the Association's own words:



“AVBPA has an ambition to bring together the leading biometric signal analysis researchers of image and voice modalities, in an attempt to contribute with robust solutions to efficient and secure communication. The conference, that has been organized biannually, is an official event of the International Association of Pattern Recognition”

The third biannual conference is to be held in Halmstad, Sweden in June 2001.

As mentioned above, speech is deemed not to be one of the more accurate of modalities. However it does have one distinct advantage over most other biometric modes, namely the ease with which an identity claim can be made; for this can be a fully integrated part of the authentication process. Thus

speech is an ideal partner to other modalities. Consequently, SpLC activities can look towards the example of AVBPA and extend current activities in multi-modal work.

In so doing it might well be worth considering the different modes under the biometric classification of Roethenbaugh and Mansfield [4]. In their revised glossary of terminology emphasis is given to *physiological* and *behavioural* biometrics. Essentially, physiological biometric is what you *are*, while a behavioural biometric is what you *do*. Some examples are given in the table below.

| | Physiological What you <i>are</i> | Behavioural What you <i>do</i> |
|---|--------------------------------------|-----------------------------------|
| Eye-related scan | are | |
| Fingerprints | are | |
| Speech(audio) | | do |
| Handwriting | | do |
| Face  | are | do |
| Lips  | are | do |

Thus speech provides an interesting dual biometric

3.2. Forensic aspects of Speaker Characterization and Recognition

Forensic applications of speaker characterization and recognition occupy a special position in the field of speech processing.

Firstly, because forensic applications do not have only a financial impact, but also a social one, given their potential consequences on the life of a person. Secondly, because the forensic field involves a wide range of people, i.e. researchers, industrialists, lawyers, judges, investigators, policemen, and sometimes even jury members, that is, anybody. This requires efforts to improve communication between this variety of people.

A lot of issues related to forensic speaker recognition remain open. Is it scientifically possible to identify someone using his/her voice? Is it possible to provide some robust information in order to help investigations? How can the reliability of a forensic technique be assessed?

The SpLC is composed of several groups, among which the “forensic group”. An important activity of the forensic group could be to make a description of the judicial systems in various countries all over the world, and to compare the place of voice analysis and speaker recognition in the different systems.

3.3. Language Characterization and Recognition

Automatic Language Identification appeared in the United States 25 years ago; the 90's saw the expansion of this theme, when the first multilingual corpora were made available. Beyond the interesting linguistic point of view, significant stakes lie in the Multilingual Man-Computer Interfaces and Computer-Assisted Communication areas; Automatic Language Identification is consequently a challenge for the user-friendly society of the XXIst century.

The state of the art approach is based on phonotactics modeling. Nowadays, best systems identify quite well (90% of correct identification) one language among eleven, with 45s duration records [5,6,7]. Even if the results are good, only marginal improvements have been performed since '96, and it is now crucial to consider alternative features to overcome these limits [8]. Especially, it is necessary to reduce the record duration, and to reach these purposes, the classical ALI system must be improved, and the wide range of distinctive features available to characterize a language (phonetic [5], phonotactic [9], prosodic [10], lexical, morpho-syntactic features) must certainly be exploited.

European researches on this topic are still rare [11,12,8], while Europe is deeply a multilingual area. A few laboratories are already involved in ALI projects across the Europe, but the SPLC SIG is the ideal framework for the rise of a world network consisting of the different researchers (linguists, neuroscientists, speech processing engineers, etc.).

3.4. Commercial Applications for Speaker and Language Recognition

Recently, an increasing interest in commercial applications of techniques such as speaker verification, speaker identification and language recognition can be observed. In a world of mobile, networked users, these techniques offer novel solutions for increased security and user friendliness.

SpLC can help in the process of bringing these techniques into products by:

- Enhancing the communication between researchers and technology providers;
- Giving an overview on ongoing research efforts as well as existing prototype implementations in these areas;
- Providing a link to ongoing standardization organizations and projects (e.g. COST).

In general, the described technologies are only part of products. Successful integration of speaker verification requires a thorough understanding of the technical possibilities and limitations as well as the application.

Speaker verification inherently can not provide 100% accuracy. Therefore the combination with other security techniques, in particular other biometric methods, is crucial. Furthermore, such novel methods raise a number of legal questions. SpLC will provide a forum to discuss these interdisciplinary questions by bringing together experts from the various fields. The discussion during the workshop in RLA2C workshop in 1998 was an important step in this direction. The workshop, however, also showed clearly that further discussions are needed to develop a common language and understanding between the various fields.

4. Transversal/special activities

4.1. Evaluation group

The aim of the evaluation group is to work on the evaluation paradigm, for all the various axes of the SpLC domains. Particularly, the group wants to highlight evaluation tasks for commercial applications, for speaker tracking and indexing, and for language identification.

For each task, the evaluation group intends to propose an evaluation scheme, a format for the representation of the results, and when possible some programs to do the scoring.

Especially for language identification, the task will be perhaps more difficult because there is no precedent, and the evaluation scheme cannot be an improvement of any previous one; databases will be one of the key point [13].

The evaluation group also intends to work if possible in collaboration with other evaluation projects (like NIST/NSA). A web page focused on these issues will be available very soon on the web site of SpLC.

4.2. Meetings and Workshops group

As said in the introduction, SpLC SIG was launched during the RLA2C workshop, in 1998. This point illustrates the necessity of thematic workshops to define the state-of-the-art and to increase collaboration and exchanges. Workshops as RLA2C (Avignon, 98) or Martigny ETRW (94) are two good examples. The main role of SpLC is to organize next workshops with the objectives to guaranty the independence of the workshops and the continuity of the editorial policy.

The next workshop, the first SpLC event (2001: A Speaker Odyssey), takes place in Crete (Greece) during June 2001 (18-22). Extended information is available at <http://www.odyssey.westhost.com/>.

During this workshop, SpLC also organizes its first plenary meeting.

4.3. Speaker Indexing

Speaker indexing consists in extracting information about speakers from audio documents. This is still a novel task in speaker recognition. The first papers on the topic were presented only about 5-10 years ago [14,15,16]. It has been proposed by NIST for the speaker recognition evaluation workshop only 2 years ago. And this topic is regularly presented in conferences only for 1 or 2 years now.

New algorithms need to be developed, evaluation procedures need to be defined, and databases for evaluation need to be designed, built, and made available. The speaker indexing group of SpLC proposes to be a place to discuss all these issues.

5. SpLC Web Site

A new web site has been designed in order to facilitate the circulation of information between SpLC members. The web site is composed of several pages. The Home page lists the four main areas of SpLC, the various specialized groups, and the recent news of SpLC. The other pages of the site include a list of events organized by SpLC, the list of the board members, a page to subscribe automatically to SpLC, the constitution of SpLC, a way to contact the SpLC board, a few interesting links including some links to related workshops and conferences or to laboratories working on speaker and language characterization, and a bibliography of the domain with the possibility to add new references.

The web site is currently hosted by the Laboratoire Informatique d'Avignon at <http://www.lia.univ-avignon.fr/heberges/SPLC>. The site can also be reached from the web page of ISCA dedicated to the SIGs at <http://www.isca-speech.org/sig.html>. Questions and suggestions on the web site are very welcomed and can be sent directly to the webmaster of the site. Contributions from the specialized groups and from all the members of SpLC are also waited.

6. Conclusion

During the first year of SpLC SIG, a lot of work has already been accomplished. A workshop has been organized (2001: A Speaker Odyssey), a web site has been designed and is going to present regularly the work done in the framework of SpLC, and several discussion groups have been initiated.

However, there are still a lot of things to do. Every main area should have its own leader and its own electronic

means, as for the specialized groups. New propositions need also to be done in order to keep SpLC very active in the domain and up-to-date. If you are interested in speaker and language characterization, do not hesitate to visit the web site, to subscribe, or to send your comments or questions. SpLC faces a very interesting challenge: be the place for stimulating exchange and innovative research thanks to the contribution and participation of all its members.

7. Reference

- [1] J. Luetttin, N.A. Thacker, and S.W. Beet. Speaker identification by lipreading. *Proceedings of ICSLP 96*, vol. 1, pp. 62-65, 1996.
- [2] Claude C. Chibelushi, Farzin Deravi, and John S.D. Mason. A review of speech-based bimodal recognition. Accepted for publication in *IEEE Transactions on Multimedia*.
- [3] J.D.Brand, J.S.Mason, and S.Colomb. Visual speech: a physiological or behavioural biometric? *Proceedings of AVBPA 2001*, Halmstad, Sweden, 2001.
- [4] G. Roethenbaugh. "Biometrics Explained", <http://www.icsa.net/services/consortia/cbdc/explained.shtml>, 1999.
- [5] T. J. Hazen and V. W. Zue. Segment-based automatic language identification. *Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2323-2331, 1997.
- [6] J. Hieronymus and S. Kadambe. Robust spoken language identification using large vocabulary speech recognition. *Proceedings of ICASSP '97*, vol. 2, pp. 1111-1114, Munich, Germany, 1997.
- [7] M. A. Zissman. Comparison of four approaches to automatic language identification of telephone speech. *IEEE Trans. on Speech and Audio Processing*, vol. 4, no. 1, Jan 1996.
- [8] Michel Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot. Language recognition using time-frequency principal component analysis and acoustic modeling. *Proceedings of ICSLP 2000*, Beijing, China, October 2000.
- [9] H. Kwan and K. Hirose. Recognized phoneme-based N-gram modeling in automatic language identification. *Proceedings of EUROSPEECH '95*, pp. 1367-1370, Madrid, Spain, September 1995
- [10] S. Itahashi and L. Du. Language identification based on speech fundamental frequency. *Proceedings of EUROSPEECH '95*, pp. 1359-1362, Madrid, Spain, September 1995.
- [11] L. Lamel and J.L. Gauvain. Language identification using phone-based acoustic likelihood. *Proceedings of ICASSP '94*, pp. 293-296, Adelaide, Australia, 1994.
- [12] F. Pellegrino and R. André-Obrecht. From vocalic detection to automatic emergence of vowel systems. *Proceedings of ICASSP '97*, Munich, Germany, April 1997
- [13] T. L. Lander et al. The OGI 22 language telephone speech corpus. *Proceedings of EUROSPEECH '95*, pp. 817-820, Madrid, Spain, September 1995.
- [14] Man-Hung Siu, George Yu, and Herbert Gish. An unsupervised, sequential learning algorithm for the segmentation of speech waveforms with multiple speakers. *Proceedings of ICASSP '92*, vol. 2, pp. 189-192, San Francisco, USA, March 1992.
- [15] Lynn Wilcox, Francine Chen, Don Kimber, and Vijay Balasubramanian. Segmentation of speech using speaker identification. *Proceedings of ICASSP '94*, pp. I.161-I.164, 1994.
- [16] Aaron E. Rosenberg, Ivan Magrin-Chagnolleau, S.Parthasarathy, and Qian Huang. Speaker detection in broadcast speech databases. *Proceedings of ICSLP '98*, Sydney, Australia, December 1998.

Application of Time-Frequency Principal Component Analysis to Text-Independent Speaker Identification (2002)

Ivan Magrin-Chagnolleau, Geoffrey Durou, and Frédéric Bimbot. **Application of Time-Frequency Principal Component Analysis to Text-Independent Speaker Identification.** *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 5, July 2002.

8 pages

Application of Time-Frequency Principal Component Analysis to Text-Independent Speaker Identification

Ivan Magrin-Chagnolleau, *Senior Member, IEEE*, Geoffrey Durou, and Frédéric Bimbot

Abstract—We propose a formalism, called **vector filtering of spectral trajectories**, that allows the integration of a number of speech parameterization approaches (cepstral analysis, Δ and $\Delta\Delta$ parameterizations, auto-regressive vector modeling, ...) under a common formalism. We then propose a new filtering, called **contextual principal components (CPC) or time-frequency principal components (TFPC)**. This filtering consists in extracting the principal components of the contextual covariance matrix, which is the covariance matrix of a sequence of vectors expanded by their context. We apply this new filtering in the framework of closed-set speaker identification, using a subset of the POLYCOST database. When using speaker-dependent TFPC filters, our results show a relative improvement of approximately 20% compared to the use of the classical cepstral coefficients augmented by their Δ -coefficients, which is significantly better with a 90% confidence level.

Index Terms—Closed-set speaker identification, contextual covariance matrix, contextual principal components (CPC), POLYCOST database, speaker recognition, speech analysis, speech representation, time-frequency principal components (TFPC), vector filtering of spectral trajectories.

I. INTRODUCTION

CEPSTRAL coefficients [1], [2] have been widely used for decades in speech processing. Although they provide a good set of feature vectors with nice properties, like a good decorrelation of the coefficients, or a good ability to deconvolve in theory the vocal source and the vocal tract filtering [2], they may not be the best solution in all situations. This is a commonly-shared point of view as proved by the abundant literature on the topic of feature parameters (see for instance [3]–[12]).

To find a good alternative to cepstral coefficients, a number of approaches have been adopted. In particular, the inability of the cepstral coefficients to extract dynamic information from speech suggested the use of Δ and $\Delta\Delta$ coefficients [13], [14]. The autoregressive (AR) vector modeling was another attempt to capture some of the dynamic information in speech signals [15]–[18].

A first aim of this paper is to propose a formalism that integrates most of these approaches under a common framework.

Manuscript received January 17, 2000; revised March 11, 2002. This work was begun while I. Magrin-Chagnolleau was with the Digital Signal Processing Group, Rice University, Houston, TX 77001 USA and finished while he was with IRISA (CNRS & INRIA), Rennes, France. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Peter Kabal.

I. Magrin-Chagnolleau is with CNRS, Lyon, France (e-mail: ivan@ieee.org).

G. Durou was with the Faculté Polytechnique de Mons, Mons, Belgium, when he collaborated on this work (e-mail: geoffrey.durou@debis.be).

F. Bimbot is with IRISA (CNRS & INRIA), Rennes, France (e-mail: bimbot@irisa.fr).

Digital Object Identifier 10.1109/TSA.2002.800557.

Actually, almost every approach assumes spectral vectors as a starting point, and tries, in various ways, to extract some information from these spectral vectors by applying transformations to them. Most of these approaches can then be seen as a *vector filtering of spectral trajectories*, that is, a function applied to the coefficients of several consecutive spectral vectors [19], [20]. It is the case, for instance, of cepstral analysis [1], [2], AR vector modeling [15]–[18], Δ and $\Delta\Delta$ parameterizations [13], [14], and temporal principal components [21].

However, most of these approaches apply the filtering function only to one spectral vector, or to several of them but only component by component. We propose an other approach, based on a principal component calculation, that is applied to all the components of several consecutive spectral vectors. Such a function can be seen as a time-frequency function (or mask), that is, a function applied to all the components of a spectral vector (frequency direction) and to its time context (time direction). Since the coefficients are calculated through a principal component analysis, we call these new coefficients *contextual principal components (CPC) or time-frequency principal components (TFPC)* of speech [19], [20]. The latter name will be used when the original vectors contain some information about the frequency content of the signal, as spectral vectors for instance. The former name will be used otherwise.

We finally apply this new speech analysis technique in the framework of closed-set speaker identification. It leads to a significant improvement of the results compared to the classical cepstral parameterization augmented by the Δ coefficients. We obtain a relative reduction of the identification error rate of approximately 20%.

The outline of this paper is the following. Section II illustrates the principle of the vector filtering of spectral trajectories, gives a mathematical formulation, provides an interpretation, and finally shows a couple of examples. The TFPC approach is presented in Section III, with a discussion about how to choose the components, an interpretation of the TFPC filters in terms of time-frequency masks, and a description of the use of the TFPC filtering in a pattern recognition system. An application to closed-set speaker identification is proposed in Section IV. Section V concludes this work and gives some future directions.

II. VECTOR FILTERING OF SPECTRAL TRAJECTORIES

A. Principle of the Vector Filtering

Let $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ denote a sequence of spectral vectors. The principle of the vector filtering of spectral trajectories is to replace the vector \mathbf{x}_t by a new vector \mathbf{f}_t , each coordinate of which is obtained by the application of a function to the coordinates

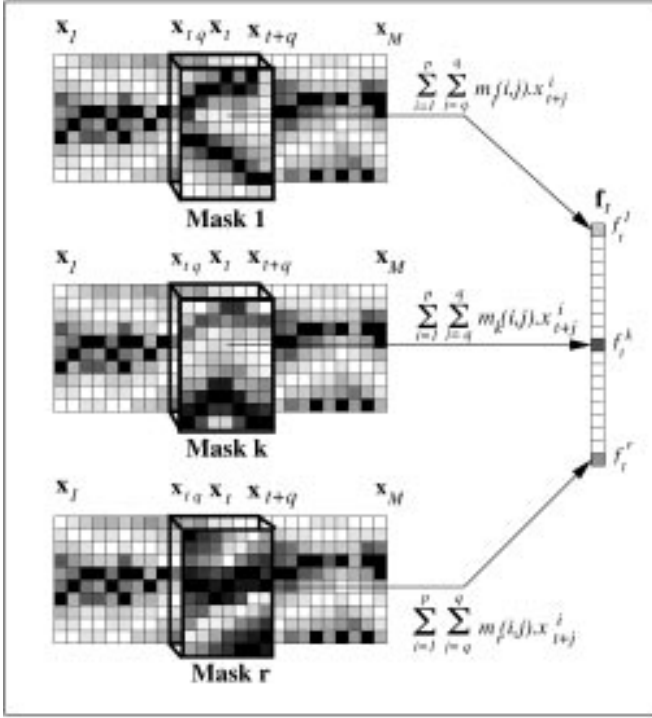


Fig. 1. Principle of the vector filtering of spectral trajectories (after [19]).

of vector \mathbf{x}_t and of the preceding and the following vectors (context of vector \mathbf{x}_t). If the filtering is linear, this is a convolution product, which can be interpreted as the application of a time-frequency mask to a sequence of spectral vectors (see Fig. 1).

Each component of \mathbf{f}_t is obtained by the application of a different function. The filtering is applied jointly in the time and the frequency directions. Each function can therefore be seen as a time-frequency mask.

This approach generalizes other filtering approaches recently presented, but that were applied only to one dimension, either the time dimension [22]–[24] or the frequency dimension [23], [25].

Our work has some similarities with the work by Milner [26]–[28]. However, we work directly on spectral coefficients instead of cepstral coefficients. This makes the interpretation of the new coefficients easier.

B. Definitions and Notations

Let \mathbf{X}_{t-q}^{t+q} denote the sequence of vectors \mathbf{x}_t between time $t-q$ and $t+q$

$$\mathbf{X}_{t-q}^{t+q} = \begin{bmatrix} \mathbf{x}_{t+q} \\ \vdots \\ \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t-q} \end{bmatrix}.$$

By convention, $\mathbf{x}_t = 0$ if $t \leq 0$ or $t > T$. The dimension of vector \mathbf{X}_{t-q}^{t+q} is $(2q+1)p$.

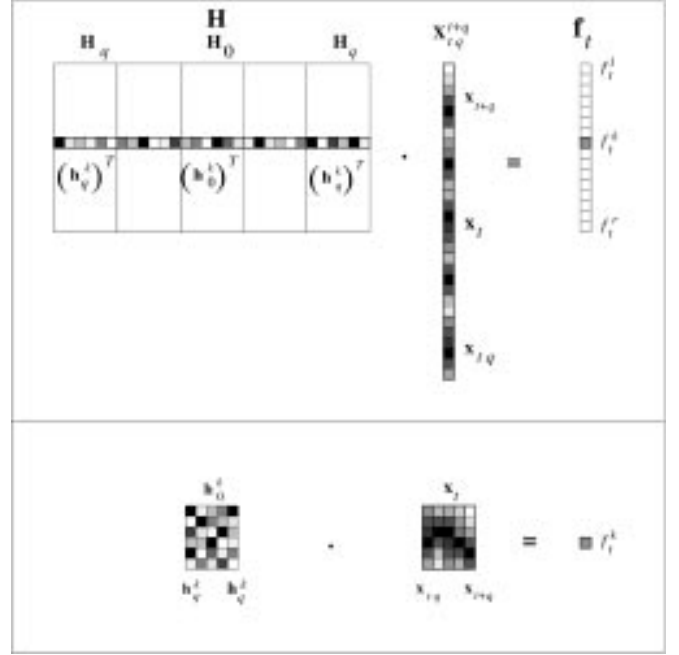


Fig. 2. Interpretation of the filtering matrix \mathbf{H} in terms of time-frequency masks (after [19]).

Let \mathcal{H} be a filtering operating on \mathbf{X}_{t-q}^{t+q}

$$\mathcal{H}: (\mathbb{R}^p)^{2q+1} \longrightarrow \mathbb{R}^r$$

$$\mathbf{X}_{t-q}^{t+q} \longmapsto \mathbf{f}_t = \mathcal{H}(\mathbf{X}_{t-q}^{t+q}).$$

The dimension of \mathbf{f}_t is r , which can be related or not to the dimension p of the original vectors.

In the following, we only consider linear filtering. The filtering \mathcal{H} can then be expressed in matrix form

$$\mathbf{H} = [\mathbf{H}_{-q} \mid \cdots \mid \mathbf{H}_0 \mid \cdots \mid \mathbf{H}_q].$$

The dimension of \mathbf{H} is $r \times (2q+1)p$.

We also have

$$\begin{aligned} \mathbf{f}_t &= \mathbf{H} \cdot \mathbf{X}_{t-q}^{t+q} \\ &= [\mathbf{H}_{-q} \mid \cdots \mid \mathbf{H}_0 \mid \cdots \mid \mathbf{H}_q] \cdot \begin{bmatrix} \mathbf{x}_{t+q} \\ \vdots \\ \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t-q} \end{bmatrix} \\ &= \sum_{k=-q}^{+q} \mathbf{H}_k \cdot \mathbf{x}_{t-k}. \end{aligned}$$

The dimension of each matrix block \mathbf{H}_k is $r \times p$.

C. Interpretation of the Vector Filtering

Fig. 2 proposes an interpretation of the filtering matrix \mathbf{H} in terms of time-frequency masks.

The top part of the figure represents the vector filtering as it is mathematically formulated. The vector filtering is the product of a filtering matrix \mathbf{H} by the expanded vector \mathbf{X}_{t-q}^{t+q} , and produces the filtered vector \mathbf{f}_t . The bottom part of the figure rep-

resents the vector filtering in terms of time-frequency masks. The vector filtering can be seen as the application of a series of time-frequency masks (a reorganization of the lines of the filtering matrix \mathbf{H} by groups of p coordinates) to a sequence of spectral vectors (a reorganization of the expanded vector \mathbf{X}_{t-q}^{t+q} in a sequence of consecutive spectral vectors). The dot product of the top part denotes the classical matrix product. The dot product of the bottom part is not a matrix product, but denotes a convolution between one time-frequency mask (on the figure, the k th line of the filtering matrix \mathbf{H}) and the sequence of spectral vectors: each element of the sequence of spectral vectors is multiplied by the corresponding element of the time-frequency mask and all the products are then added.

D. Examples of Vector Filterings

1) *Cepstral Analysis*: The cepstral analysis [1], [2] is one particular filtering of the spectral vectors, for which the filtering functions apply only to the frequency dimension, that is, only to \mathbf{x}_t . Therefore, $q = 0$. Each filtering function is a cosine transform. If the k first cosine functions are applied to \mathbf{x}_t , then $r = k$. And the lines of the matrix \mathbf{H} are the cosine functions

$$\mathbf{H} = \begin{bmatrix} \gamma_1^T \\ \vdots \\ \gamma_k^T \end{bmatrix}.$$

We finally obtain

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_t^t = \begin{bmatrix} \gamma_1^T \\ \vdots \\ \gamma_k^T \end{bmatrix} \cdot \mathbf{x}_t = \mathbf{c}_t$$

where \mathbf{c}_t is the cepstral vector corresponding to the spectral vector \mathbf{x}_t , and containing the cepstral coefficients 1 to k .

2) Δ and $\Delta\Delta$ Parameters: Δ and $\Delta\Delta$ parameters [13] are also examples of vector filterings. They are applied to the time dimension only. If five frames are used to calculate the Δ parameters, then $q = 2$. In fact, these parameters can be calculated for different values of q . The values $q = 1, 2, 3$ are classically chosen. If the filtered vector is composed of the original vector augmented by its Δ coefficients, then the dimension of the new vector is $r = 2p$. If the $\Delta\Delta$ coefficients are also added, then $r = 3p$. In that case, \mathbf{H} can be, for instance

$$\mathbf{H} = \begin{bmatrix} \mathbf{0}_p & \mathbf{0}_p & \mathbf{I}_p & \mathbf{0}_p & \mathbf{0}_p \\ -2\mathbf{I}_p & -\mathbf{I}_p & \mathbf{0}_p & \mathbf{I}_p & 2\mathbf{I}_p \\ -\mathbf{I}_p & \mathbf{0}_p & 2\mathbf{I}_p & \mathbf{0}_p & -\mathbf{I}_p \end{bmatrix}$$

where $\mathbf{0}_p$ denotes the null matrix of dimension $p \times p$ and \mathbf{I}_p the identity matrix of dimension p .

The corresponding filtered vector is

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_{t-2}^{t+2} = \begin{bmatrix} \mathbf{x}_t \\ \Delta\mathbf{x}_t \\ \Delta\Delta\mathbf{x}_t \end{bmatrix}.$$

3) *Second-Order Auto-Regressive Vector Model*: The application of a second-order auto-regressive model to a sequence of

vectors [15]–[18] can also be interpreted as a vector filtering. $q = 2$, $r = p$, and the filtering matrix is given by

$$\mathbf{H} = [\mathbf{0}_p \mid \mathbf{0}_p \mid \mathbf{I}_p \mid \mathbf{A}_1 \mid \mathbf{A}_2]$$

where \mathbf{A}_1 and \mathbf{A}_2 are the matrix coefficients of the second-order auto-regressive model. The filtered vector is the prediction error of the model at time t

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_{t-2}^{t+2} = \mathbf{x}_t + \mathbf{A}_1 \cdot \mathbf{x}_{t-1} + \mathbf{A}_2 \cdot \mathbf{x}_{t-2} = \mathbf{e}_t.$$

This filtering applies to both the time and frequency directions.

E. Composition of Filterings

Several filterings can also be composed together. In that case, the filtering matrices \mathbf{H} are multiplied. For instance, Δ -cepstral parameters correspond to the composition of the cepstral filtering with the Δ filtering.

III. TIME-FREQUENCY PRINCIPAL COMPONENTS OF SPEECH

We have defined a formalism that makes it possible to express a number of approaches in terms of filtering of spectral vectors. We now propose a new filtering operating in both the time and the frequency dimensions. This approach also assumes a set of training data on which to extract some principal components: it is a data-driven filtering. We call this new filtering *CPC* or *TFPC* of speech.

A. Principle of the TFPC Filtering

The idea of the TFPC filtering is to extract time-frequency patterns that are characteristic of a sequence of vectors in order to summarize the evolution of the spectral content by a few spectral sequences extracted from the entire sequence. The original sequence must be long enough to be representative of the class that we want to represent with the time-frequency patterns. This strategy can be applied to any pattern recognition problem, as long as there are enough vectors for each class to calculate the time-frequency patterns. Once the patterns have been extracted, they are used to filter the spectral vectors of both the training and the test data. And any modeling technique can then be applied to the new vectors, as it is done usually on spectral vectors or cepstral vectors, or any other vector representation of the original signal. For example, in Section IV, we apply the TFPC filtering in the framework of closed-set speaker identification, and we extract time-frequency patterns for each speaker of the training dataset.

B. Definitions and Notations

Let $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ denote again a sequence of spectral vectors, and $\{\mathbf{x}_t^* = \mathbf{x}_t - \bar{\mathbf{x}}\}$ the sequence of the corresponding centered vectors. $\bar{\mathbf{x}}$ is the mean vector of the sequence $\{\mathbf{x}_t\}$.

Let \mathcal{X}_0 denote the covariance matrix of the sequence $\{\mathbf{x}_t\}$

$$\mathcal{X}_0 = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_t - \bar{\mathbf{x}})^T = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t^* \cdot \mathbf{x}_t^{*T}$$

and \mathcal{X}_k denote the k -lagged covariance matrix

$$\mathcal{X}_k = \frac{1}{T} \sum_{t=k+1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_{t-k} - \bar{\mathbf{x}})^T = \frac{1}{T} \sum_{t=k+1}^T \mathbf{x}_t^* \cdot \mathbf{x}_{t-k}^{*T}.$$

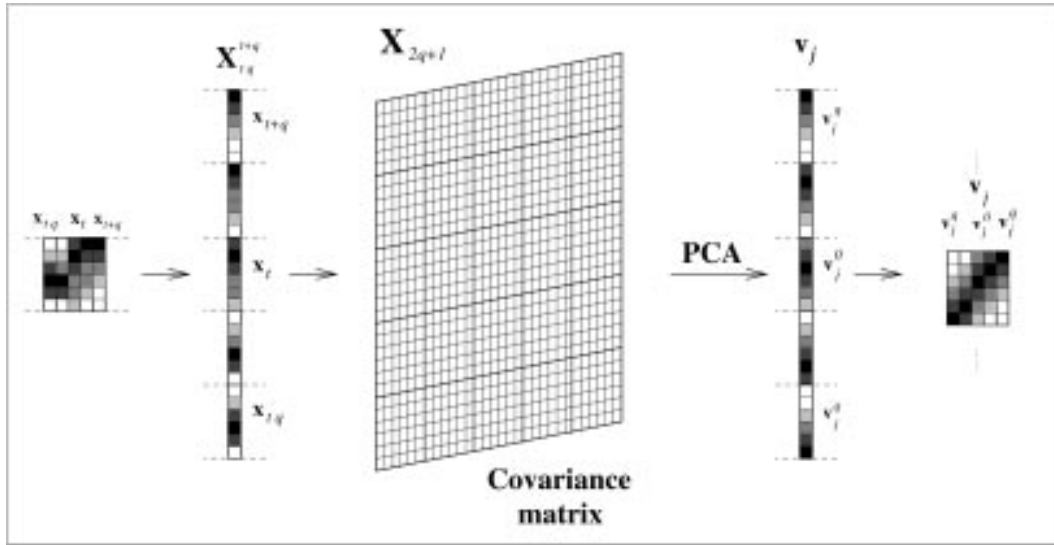


Fig. 3. Interpretation of a principal component in terms of a time-frequency mask (after [19]).

The dimension of the covariance matrix and of the lagged covariance matrices is $p \times p$.

We now define a new matrix, \mathbf{X}_{2q+1} , by

$$\mathbf{X}_{2q+1} = \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \cdots & \mathbf{x}_{2q} \\ \mathbf{x}_1^T & \mathbf{x}_0 & \cdots & \mathbf{x}_{2q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{2q}^T & \mathbf{x}_{2q-1}^T & \cdots & \mathbf{x}_0 \end{bmatrix}.$$

This matrix is block-Toeplitz, and its dimension is $(2q+1)p \times (2q+1)p$.¹ This matrix can be interpreted as the covariance matrix of the vectors $\{\mathbf{X}_{t-q}^{t+q}\}_{1 \leq t \leq T}$, and can therefore be called *contextual covariance matrix*.

Then the principal components of this matrix [29], [30] are calculated. It is equivalent to the extraction of eigenvalues and eigenvectors of the matrix. The eigenvector associated with the largest eigenvalue is the direction of projection that preserves the maximum variance, the eigenvector associated with the second largest eigenvalue is the direction of projection that preserves the maximum variance uncorrelated (that is, orthogonal) to the first one, and so on. We finally have

$$\mathbf{X}_{2q+1} = \mathbf{V}_{2q+1} \cdot \mathbf{\Lambda}_{2q+1} \cdot \mathbf{V}_{2q+1}^T$$

with

$$\begin{aligned} \mathbf{V}_{2q+1} &= (\mathbf{v}_1, \dots, \mathbf{v}_{2q+1}) \\ \mathbf{V}_{2q+1}^T \cdot \mathbf{V}_{2q+1} &= \mathbf{I}_{2q+1} \\ \mathbf{\Lambda}_{2q+1} &= \text{diag}(\lambda_1, \dots, \lambda_{2q+1}), \\ \lambda_1 &\geq \dots \geq \lambda_{2q+1}. \end{aligned}$$

The dimension of matrices \mathbf{V}_{2q+1} and $\mathbf{\Lambda}_{2q+1}$ is $(2q+1)p \times (2q+1)p$. The dimension of each vector \mathbf{v}_i , $1 \leq i \leq 2q+1$, is $(2q+1)p$.

¹Although this matrix is $(2q+1)p \times (2q+1)p$, we adopt the notation \mathbf{X}_{2q+1} in order to have a simpler notation. This is also the case for the matrices \mathbf{V}_{2q+1} and $\mathbf{\Lambda}_{2q+1}$.

Since the principal components are extracted from the contextual covariance matrix, instead of the covariance matrix itself, we call them *CPC*. When the original vectors contain some information about the frequency content of the signal, as spectral vectors, we can call these components more specifically *TFPC*.

C. Choice of the Components

Once the principal components have been calculated, it is necessary to select which ones to keep. It is common to keep the first components, their number depending on the experiment. Since the eigenvalues correspond to a variance measurement, a criterion for keeping them can be a percentage of the total variance, for instance 80% [29], [30]. Some other procedures can be used for the choice of the components like the *F*-ratio [31], [32] or the knock-out procedure [33]. Here is an example of the filtering matrix \mathbf{H} , corresponding to the selection of the first eight components

$$\mathbf{H} = [\mathbf{v}_1, \dots, \mathbf{v}_8]^T.$$

If all the components are kept, then $\mathbf{H} = \mathbf{V}_{2q+1}^T$.

D. Interpretation

Each principal component can be interpreted in terms of a time-frequency mask, as shown in Fig. 3.

On the left part of the figure, a sequence of spectral vectors $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ is represented. Then, the corresponding expanded vectors $\{\mathbf{X}_{t-q}^{t+q}\}_{1 \leq t \leq T}$ are formed. The covariance matrix of these expanded vectors, \mathbf{X}_{2q+1} , is calculated. It is a block-Toeplitz matrix. Then the principal component analysis is applied and the eigenvectors of this matrix are obtained. The dimension of each eigenvector \mathbf{v}_i , $1 \leq i \leq 2q+1$, is $(2q+1)p$. Each of them can be reordered as a sequence of p -dimensional vectors to form a time-frequency mask. Finally, for each eigenvector, a time-frequency mask is obtained, as shown on the right part of the figure.

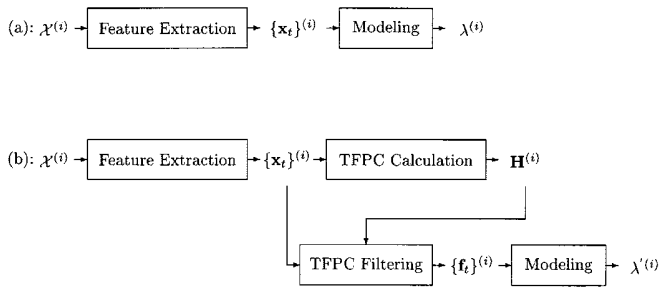


Fig. 4. Training phase of a pattern recognition system: (a) without the TFPC filtering and (b) with the TFPC filtering.

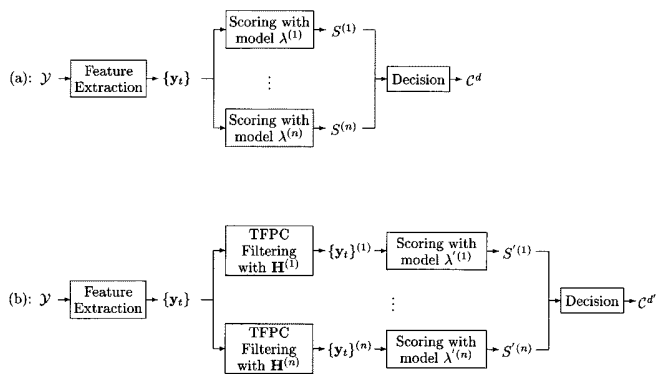


Fig. 5. Test phase of a pattern recognition system: (a) without the TFPC filtering and (b) with the TFPC filtering.

E. TFPC Filtering and Pattern Recognition

The next paragraphs explain how to integrate the TFPC filtering in a speaker recognition system.²

1) *Training Phase*: Fig. 4 shows the training phase of the recognition system: (a) without the TFPC filtering; and (b) with the TFPC filtering. In both cases, we dispose of a training pattern $\mathcal{X}^{(i)}$ representing the class $\mathcal{C}^{(i)}$. Feature parameters are extracted from this training pattern (for instance spectral vectors) and the sequence of vectors $\{\mathbf{x}_t\}^{(i)}$ is formed.

In the first case, this sequence of vectors is then modeled (for instance using a Gaussian mixture model [34]–[36]) and a model $\lambda^{(i)}$ is obtained for class $\mathcal{C}^{(i)}$. At the end, the class $\mathcal{C}^{(i)}$ is represented by the model $\lambda^{(i)}$.

In the second case, some time-frequency principal components are first extracted from the feature vectors $\{\mathbf{x}_t\}^{(i)}$ and a filtering matrix $\mathbf{H}^{(i)}$ is obtained for the class $\mathcal{C}^{(i)}$. This matrix is then used to filter the sequence $\{\mathbf{x}_t\}^{(i)}$ and a sequence of new vectors $\{\mathbf{f}_t\}^{(i)}$ is formed. Finally, this new sequence of vectors is modeled and a model $\lambda'^{(i)}$ is obtained for class $\mathcal{C}^{(i)}$. At the end, the class $\mathcal{C}^{(i)}$ is represented by a model $\lambda'^{(i)}$ and a filtering matrix $\mathbf{H}^{(i)}$.

2) *Test Phase*: Fig. 5 represents the test phase of the recognition system: (a) without the TFPC filtering and (b) with the TFPC filtering. In both cases, the test pattern \mathcal{Y} has to be classified in one of the reference classes \mathcal{C}^1 to \mathcal{C}^n , where n is the number of classes. First, parameter vectors are extracted from the pattern \mathcal{Y} , and the sequence $\{\mathbf{y}_t\}$ is formed.

In the first case, a score is calculated on this sequence with each reference model, $\lambda^{(1)}$ to $\lambda^{(n)}$, and the n scores S^1 to S^n are obtained. Then, a decision algorithm is applied to decide to which class \mathcal{C}^d the test pattern \mathcal{Y} belongs.

In the second case, the TFPC filtering of each reference class, $\mathbf{H}^{(1)}$ to $\mathbf{H}^{(n)}$, is applied to the sequence $\{\mathbf{y}_t\}$. This leads to the new sequences $\{\mathbf{y}_t\}^{(1)}$ to $\{\mathbf{y}_t\}^{(n)}$. Then, a score is calculated on these sequences with each corresponding reference model, $\lambda'^{(1)}$ to $\lambda'^{(n)}$, and the n new scores S'^1 to S'^n are obtained. Finally, a decision algorithm is applied to decide to which class \mathcal{C}^d the test pattern \mathcal{Y} belongs.

IV. APPLICATION OF THE TFPC FILTERING TO CLOSED-SET SPEAKER IDENTIFICATION

We have defined the TFPC filtering and explained how to integrate it in a speaker recognition system. To evaluate the effectiveness of this approach, we now apply it to closed-set speaker identification [20].

A. Experimental Protocol

1) *Task*: The task considered in our experiments is closed-set text-independent speaker identification, which consists in recognizing the identity of the speaker of a test utterance, without using any phonetic information about the utterance. Closed-set means that the system always gives the identity of one of the speakers of the reference set as the answer.

2) *Database*: In our experiments, a subset of the POLY-COST database [37] was used. This subset is composed of 112 speakers (64 females and 48 males), who have pronounced some free-text utterances and some guided utterances in English through the telephone. Speakers were originating from various European countries, and most of them were not English native speakers. Free-text utterances were used for training, guided utterances were used for testing. Both the training and the test data were recorded during 10 sessions in average, spread over three months. We used about 90 s of speech for the training, and several utterances of 5 s in average for the test. The total number of tests was 560.

3) *Spectral Analysis*: Each utterance was analyzed as follows: the speech signal was decomposed in frames of 30 ms at a frame rate of 10 ms. A Hamming window was applied to each frame. The signal was pre-emphasized with a coefficient equal to 0.95. For each frame, a fast Fourier transform was computed and provided 256 square module values representing the short term power spectrum in the 0–4 kHz band. This Fourier power spectrum was then used to compute 13 filter bank coefficients. Each filter was triangular. The filters were placed on a nonuniform frequency scale, similar to the Bark/Mel scale. The central frequency of the 13 filters were, in Hz: 102, 219, 353, 506, 682, 883, 1114, 1378, 1681, 2028, 2425, 2881, and 3402. Each filter covered a spectral range from the central frequency of the previous filter to the central frequency of the next filter, with a maximum value of one for its own central frequency. We finally took the base 10 logarithm of each filter output and multiplied the result by 10, to form a 13-dimensional vector of filter bank coefficients in decibels.

²This approach can easily be reused for other pattern recognition problems.

4) *Application of the TFPC Filtering to Spectral Vectors*: We then calculated for each speaker the TFPC for the values $q = 0, 1, 2$, which corresponds to one frame (the classical principal components analysis), three frames, and five frames, respectively. For each value of q , all the components were kept. Then, the extracted speaker-dependent TFPC filtering was applied to the initial spectral vectors in order to obtain the new vectors. The filtering matrices were kept to be used also during the test phase. As an additional experiment, TFPC were also extracted from all the training data pooled together (see [19]). A speaker-independent TFPC filtering, calculated on multispeaker speech, was obtained. Various values of q (0, 1, and 2) were also tested, and all the components were kept for each value of q . The last filtering that was tested was a classical Δ calculation [13], [14], using three or five frames, which provided us with 26-dimensional parameter vectors.

5) *Baseline: Cepstrum and Δ -Cepstrum*: For comparison, cepstral coefficients [1], [2] were also extracted from the spectral vectors. The first coefficient c_0 , which is highly correlated with the energy of the frame, was not kept. Only c_1 to c_{12} were used. The corresponding Δ parameters [13], [14] were calculated using three or five frames, and appended to the cepstral coefficients. Finally, 24-dimensional parameter vectors were obtained. To complete our experiments, both speaker-dependent and speaker-independent TFPC filterings were also tested on the cepstral vectors, which is equivalent to composing a cepstral filtering and a TFPC filtering, and applying the composition directly to the spectral vectors.

6) *Modeling*: A Gaussian mixture model (GMM) [34]–[36] was used to represent each reference speaker. Each mixture was composed of eight components, and diagonal covariance matrices were chosen. The Gaussian mixture models were trained using the expectation-maximization (EM) algorithm [34], [35], [38]. The initialization of the EM algorithm was done with a vector quantization (VQ) algorithm [39].

B. Results and Discussion

Identification error rates are reported in Table I. We also give, in the discussion, the 95% and the 90% confidence intervals of the baseline score assuming that the probability distribution of the error rate is binomial. If R is the identification error rate of one experiment and N the total number of tests, then the confidence interval is given by [30], [40]

$$R \pm u \cdot \sqrt{\frac{R(1-R)}{N}}$$

with $u = 1.96$ for the 95% confidence interval and $u = 1.65$ for the 90% confidence interval.

The top part of Table I presents classical results using spectral or cepstral coefficients and Δ parameters. The baseline score, 11.43%, was obtained using cepstral coefficients augmented by the Δ parameters, with a Δ calculation over five vectors. Cepstral coefficients perform better than spectral coefficients when using the Δ parameters, which is a very classical result in speaker identification.

TABLE I
SPEAKER IDENTIFICATION RESULTS WHEN USING THE Δ PARAMETERS, ONE TFPC FILTER FOR ALL SPEAKERS, OR ONE TFPC FILTER FOR EACH SPEAKER. SEVERAL SIZES OF CONTEXT (1, 3, AND 5 VECTORS) WERE TESTED. RESULTS ARE GIVEN AS A PERCENTAGE OF IDENTIFICATION ERROR

| Coefficients | static + Δ | | | |
|--------------|-------------------|----------|--------------|-----------|
| | static | 1 vector | 3 vectors | 5 vectors |
| Context | | | | |
| Spectral | 21.96 | 19.64 | 18.93 | |
| Cepstral | 15.71 | 12.32 | 11.43 | |

| Coefficients | Speaker-independent TFPC filtering | | |
|--------------|------------------------------------|--------------|-----------|
| | 1 vector | 3 vectors | 5 vectors |
| Context | | | |
| Spectral | 11.96 | 11.35 | 11.61 |
| Cepstral | 11.43 | 12.50 | 16.25 |

| Coefficients | Speaker-dependent TFPC filtering | | |
|--------------|----------------------------------|-------------|-----------|
| | 1 vector | 3 vectors | 5 vectors |
| Context | | | |
| Spectral | 9.82 | 9.11 | 9.82 |
| Cepstral | 10.00 | 9.29 | 10.71 |

The middle part of Table I presents results using a speaker-independent TFPC filtering. When cepstral vectors were used as original vectors, error rates were not better than the baseline score. The error rates even degrade if we use three or five vectors. When spectral vectors were used as original vectors, the best error rate obtained was 11.35%, which is not significantly better than the baseline score given its 95% confidence interval (8.79%, 14.07%).

More interesting results were obtained using speaker-dependent TFPC filterings as reported in the bottom part of Table I. The best error rate, 9.11%, outperformed the baseline score of about 20% relatively. The figure of 9.11% falls near the border of the 95% confidence interval of the baseline score (8.79%, 14.07%). However, it falls outside the 90% confidence interval (9.21%, 13.65%). Therefore, the difference observed is significant with a 90% confidence level.

V. CONCLUSIONS AND PERSPECTIVES

In this paper, we have presented a new formalism, called *vector filtering of spectral trajectories*, that integrates under a common formalism several approaches such as cepstral analysis, Δ and $\Delta\Delta$ parameters, or auto-regressive vector models. In the framework of this formalism, we proposed a new filtering called *time-frequency principal components (TFPC)* or, more generally, *contextual principal components (CPC)*, which consists of calculating the principal components of the contextual covariance matrix, i.e., the covariance matrix of a sequence of vectors augmented by their time context. We also explained how to integrate the new filtering in a speaker recognition system. We applied the new filtering to closed-set speaker identification, with several sizes of context, and using spectral or cepstral vectors as original vectors. Our best configuration, based on speaker-dependent TFPC filterings, outperforms the classical parameterization using cepstral and Δ -cepstral coefficients by approximately 20%, which is significantly better with a 90% confidence level.

A number of issues are still pending. One of them is the selection of the components. In our experiments, we systematically kept all the components. An interesting experiment would

be the removal of the last components, which should not decrease noticeably the performance, and would provide a more compact representation of the signal. However, in the case of speaker-dependent TFPC filterings, the choice of the components to keep must be done more carefully, as some parameters having a small variance may still be a very good indicator of a particular speaker.

Another extension of this work is the application of the TFPC filtering to speaker verification and speaker tracking rather than speaker identification. However, in these cases, a score normalization is needed, and the most common approach is the calculation of a log-likelihood ratio using a nonspeaker model. Therefore, a theoretical study of the normalization problem needs to be done.

The TFPC filtering can also be applied to speech recognition. It may be useful to neutralize some of the effect of the speaker variability and/or the transmission channel variability.

More generally, the TFPC approach can be investigated on for a number of pattern recognition problems, and we hope that this preliminary work will encourage some of our colleagues to experiment with this approach.

REFERENCES

- [1] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking," in *Proceedings of the Symposium on Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, 1963, ch. 15, pp. 209–243.
- [2] A. V. Oppenheim and R. W. Schaffer, "Homomorphic analysis of speech," *IEEE Tran. Audio Electroacoust.*, vol. AE-16, pp. 221–226, June 1968.
- [3] L. Xu, "Perceptually-based features for speaker identification," Ph.D. dissertation, Univ. College Swansea, Univ. Wales, Wales, U.K., Feb. 1992.
- [4] Z. P. Sun and J. S. Mason, "Combining features via LDA in speaker recognition," in *Proc. EUROSPEECH 93*, vol. 3, Berlin, Germany, Sept. 1993, pp. 2287–2290.
- [5] J. Thompson and J. S. Mason, "Within class optimization of cepstra for speaker recognition," in *Proc. EUROSPEECH 93*, vol. 1, Berlin, Germany, Sept. 1993, pp. 165–168.
- [6] X. Wang and G. Zhao, "Text-dependent speaker verification using recurrent time delay neural networks for feature extraction," in *Proc. ICSP 93*, vol. 1, Beijing, China, Oct. 1993, pp. 674–677.
- [7] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 578–589, Oct. 1994.
- [8] N. Higuchi and M. Hashimoto, "Analysis of acoustic features affecting speaker identification," in *Proc. EUROSPEECH 95*, vol. 1, 1995, pp. 435–438.
- [9] C. R. Jankowski Jr., T. F. Quatieri, and D. A. Reynolds, "Fine structure features for speaker identification," in *Proc. ICASSP 96*, vol. 2, Atlanta, GA, May 1996, pp. 689–692.
- [10] T. Kitamura and S. Takei, "Speaker recognition model using two-dimensional mel-cepstrum and predictive neural network," in *Proc. ICSLP 96*, 1996.
- [11] C.-S. Liu, "A general framework of feature extraction: Application to speaker recognition," in *Proc. ICASSP 96*, vol. 2, Atlanta, GA, May 1996, pp. 669–672.
- [12] L. Besacier and J.-F. Bonastre, "Subband approach for automatic speaker recognition: Optimal division of the frequency domain," in *Proc. Workshop Audio and Video Biometric Person Authentication*, Craus-Montana, Switzerland, 1997, pp. 195–202.
- [13] S. Furui, "Comparison of speaker recognition methods using static features and dynamic features," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, no. 3, pp. 342–350, June 1981.
- [14] F. K. Soong and A. E. Rosenberg, "On the use of instantaneous and transitional spectral information in speaker recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 871–879, June 1988.
- [15] Y. Grenier, "Utilization de la prédiction linéaire en reconnaissance et adaptation au locuteur," in *XIèmes Journées d'Etude sur la Parole*, Strasbourg, France, May 1980, pp. 163–171.
- [16] F. Bimbot, L. Mathan, A. de Lima, and G. Chollet, "Standard and target-driven AR-vector models for speech analysis and speaker recognition," in *Proc. ICASSP 92*, vol. 2, San Francisco, CA, Mar. 1992, pp. II.5–II.8.
- [17] C. Montacié, P. Deléglise, F. Bimbot, and M.-J. Caraty, "Cinematic techniques for speech processing: Temporal decomposition and multivariate linear prediction," in *Proc. ICASSP 92*, vol. 1, San Francisco, CA, Mar. 1992, pp. 153–156.
- [18] C. Montacié and J.-L. Le Floch, "AR-vector models for free-text speaker recognition," in *Proc. ICSLP 92*, vol. 1, Banff, AB, Canada, Oct. 1992, pp. 611–614.
- [19] I. Magrin-Chagnolleau, "Approches statistiques et filtrage vectoriel de trajectoires spectrales pour l'identification du locuteur indépendante du texte," Ph.D. dissertation, École Nat. Supérieure Télécommun., Jan. 1997.
- [20] I. Magrin-Chagnolleau and G. Durou, "Time-frequency principal components of speech: Application to speaker identification," in *Proc. EUROSPEECH 99*, Budapest, Hungary, Sept. 1999, pp. 759–762.
- [21] F. Bimbot, E. Bocchieri, and B. Atal, "Sous-espaces de projection de séquences de trames acoustiques pour l'analyse et la reconnaissance de parole," in *XXIèmes Journées d'Etude sur la Parole*, Avignon, France, 1996.
- [22] J.-L. Shen, W.-L. Hwang, and L.-S. Lee, "Robust speech recognition features based on temporal trajectory filtering of frequency band spectrum," in *Proc. ICSLP 96*, 1996.
- [23] C. Nadeu, J. B. Marino, J. Hernando, and A. Nogueiras, "Frequency and time filtering of filter-bank energies for HMM speech recognition," in *Proc. ICSLP 96*, 1996.
- [24] J. P. Openshaw and J. S. Mason, "Noise robust estimate of speech dynamics for speaker recognition," in *Proc. ICSLP 96*, 1996.
- [25] D. J. Darlington and D. R. Campbell, "Sub-band adaptive filtering applied to speech enhancement," in *Proc. ICSLP 96*, 1996.
- [26] B. P. Milner and S. V. Vaseghi, "An analysis of cepstral-time matrices for noise and channel robust speech recognition," in *Proc. EUROSPEECH 95*, vol. 1, 1995, pp. 519–522.
- [27] B. Milner, "Inclusion of temporal information into features for speech recognition," in *Proc. ICSLP 96*, 1996.
- [28] N. Harte, S. Vaseghi, and B. Milner, "Dynamic features for segmental speech recognition," in *Proc. ICSLP 96*, 1996.
- [29] I. T. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer-Verlag, 1986.
- [30] G. Saporta, *Probabilités, analyse des données et statistique*. Paris, France: Éditions Technip, 1990.
- [31] J. J. Wolf, "Efficient acoustic parameters for speaker recognition," *J. Acoust. Soc. Amer.*, pt. 2, vol. 51, no. 6, pp. 2044–2056, 1972.
- [32] S. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*. London, U.K.: Griffin, 1977.
- [33] M. R. Sambur, "Selection of acoustic features for speaker identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 176–182, Apr. 1975.
- [34] D. M. Titterton, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.
- [35] G. J. McLachlan and K. E. Basford, *Mixture Models: Inference and Applications to Clustering*. London, U.K.: Marcel Dekker, 1988.
- [36] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 72–83, Jan. 1995.
- [37] (1998) *Speaker Recognition in Telephony*. Eur. COST 250 Action. [Online]. Available: <http://circhp.epfl.ch/polycost>.
- [38] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 6, no. 39, pp. 1–38, 1977.
- [39] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantization design," *IEEE Trans. Commun.*, vol. 28, pp. 84–95, Jan. 1980.
- [40] T. H. Wonnacott and R. J. Wonnacott, *Introductory Statistics for Business and Economics*. New York: Wiley, 1990.



Ivan Magrin-Chagnolleau (S'94-M'97-SM'02) was born in Paris, France, in 1969. He received the engineer diploma in electrical engineering from the ENSEA, Cergy-Pontoise, France, in 1992, the M.S. degree in electrical engineering from Paris XI University, Orsay, France, in 1993, the M.A. degree in phonetics from Paris III University in 1996, and the Ph.D. degree in electrical engineering from the ENST, Paris, in 1997.

In February 1997, he joined the Speech and Image Processing Services Laboratory, AT&T Labs Research, Florham Park, NJ. In October 1998, he visited the Digital Signal Processing Group, Electrical and Computer Engineering Department, Rice University, Houston, TX. In October 1999, he joined IRISA, Rennes, France. From October 2000 to August 2001, he was an Assistant Professor with the Computer Science Laboratory, University of Avignon, Avignon, France. In October 2001, he became a permanent Researcher with the French National Center for Scientific Research and is currently working at the "Laboratoire Dynamique Du Langage," one of the CNRS associated laboratories. His research interests include audio indexing, speaker recognition, language identification, speech recognition, statistical pattern recognition, signal representations and decompositions, and data analysis.

Dr. Magrin-Chagnolleau is a member of the IEEE Signal Processing Society, the IEEE Computer Society, and the International Speech Communication Association (ISCA).



Geoffrey Durou was born in Tournai, Belgium, in 1973. He received the engineer diploma from the Faculté Polytechnique de Mons (FPM), Mons, Belgium, in June 1996.

In September 1996, he joined the Speech and Image Processing Laboratory, FPM. His research interests include vector quantization, speaker recognition, signal representations and decompositions.



Frédéric Bimbot was born in Neuilly-sur-Seine, France, in 1963. He received the telecommunication engineer degree and the Ph.D. degree in signal processing on the topic of speech synthesis using temporal decomposition from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1985 and 1988, respectively. He also received the B.A. degree in linguistics from the Sorbonne Nouvelle University, Paris III, in 1987.

In 1990, he joined CNRS (the French National Center for Scientific Research) and became a permanent Researcher. He continued working for ENST (one of the CNRS-associated laboratories) for 7 years and then moved to IRISA (another CNRS-associated laboratory, also part of INRIA), where he has been working since 1997. He visited AT&T Bell Laboratories several times in the past few years (for a total of 15 months). He has been involved in many national and European projects, in particular Esprit projects SPRINT on speech recognition using neural networks (1989-1990), SAM-A on assessment methodology (1992-1993), and DiVAN on audio data indexing (1997-2000). He was also the Research Work-Package Manager in the Telematics projects CAVE (1995-1997) and PICASSO (1998-2000), both on speaker verification. His research interests are focused on speech and audio signal analysis and modeling, speaker characterization and verification, speech recognition and assessment methodology of speech systems. Since 2001, he is heading at IRISA a research group called METISS, centered on these topics.

Dr. Bimbot was the Chairman of the GFCP (a branch of the French Acoustic Society dedicated to spoken communication) from 1996 to 2000 and since 1998, he has been a member of the ISCA Board (International Speech Communication Association, formerly known as ESCA).

A Tutorial on Text-Independent Speaker Verification (2004)

Frédéric Bimbot et al. **A Tutorial on Text-Independent Speaker Verification**. *EURASIP Journal on Applied Signal Processing*. 2004 :4, 430–451.

22 pages

A Tutorial on Text-Independent Speaker Verification

Frédéric Bimbot,¹ Jean-François Bonastre,² Corinne Fredouille,² Guillaume Gravier,¹
Ivan Magrin-Chagnolleau,³ Sylvain Meignier,² Teva Merlin,² Javier Ortega-García,⁴
Dijana Petrovska-Delacrétaz,⁵ and Douglas A. Reynolds⁶

¹IRISA, INRIA & CNRS, 35042 Rennes Cedex, France
Emails: bimbot@irisa.fr; ggravier@irisa.fr

²LIA, University of Avignon, 84911 Avignon Cedex 9, France
Emails: jean-francois.bonastre@lia.univ-avignon.fr; corinne.fredouille@lia.univ-avignon.fr;
sylvain.meignier@lia.univ-avignon.fr; teva.merlin@lia.univ-avignon.fr

³Laboratoire Dynamique du Langage, CNRS, 69369 Lyon Cedex 07, France
Email: ivan@ieee.org

⁴ATVS, Universidad Politécnica de Madrid, 28040 Madrid, Spain
Email: jortega@diac.upm.es

⁵DIVA Laboratory, Informatics Department, Fribourg University, CH-1700 Fribourg, Switzerland
Email: dijana.petrovski@unifr.ch

⁶Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02420-9108, USA
Email: dar@ll.mit.edu

Received 2 December 2002; Revised 8 August 2003

This paper presents an overview of a state-of-the-art text-independent speaker verification system. First, an introduction proposes a modular scheme of the training and test phases of a speaker verification system. Then, the most commonly speech parameterization used in speaker verification, namely, cepstral analysis, is detailed. Gaussian mixture modeling, which is the speaker modeling technique used in most systems, is then explained. A few speaker modeling alternatives, namely, neural networks and support vector machines, are mentioned. Normalization of scores is then explained, as this is a very important step to deal with real-world data. The evaluation of a speaker verification system is then detailed, and the detection error trade-off (DET) curve is explained. Several extensions of speaker verification are then enumerated, including speaker tracking and segmentation by speakers. Then, some applications of speaker verification are proposed, including on-site applications, remote applications, applications relative to structuring audio information, and games. Issues concerning the forensic area are then recalled, as we believe it is very important to inform people about the actual performance and limitations of speaker verification systems. This paper concludes by giving a few research trends in speaker verification for the next couple of years.

Keywords and phrases: speaker verification, text-independent, cepstral analysis, Gaussian mixture modeling.

1. INTRODUCTION

Numerous measurements and signals have been proposed and investigated for use in biometric recognition systems. Among the most popular measurements are fingerprint, face, and voice. While each has pros and cons relative to accuracy and deployment, there are two main factors that have made voice a compelling biometric. First, speech is a natural signal to produce that is not considered threatening by users to provide. In many applications, speech may be the main (or only, e.g., telephone transactions) modality, so users do not consider providing a speech sample for authentication

as a separate or intrusive step. Second, the telephone system provides a ubiquitous, familiar network of sensors for obtaining and delivering the speech signal. For telephone-based applications, there is no need for special signal transducers or networks to be installed at application access points since a cell phone gives one access almost anywhere. Even for non-telephone applications, sound cards and microphones are low-cost and readily available. Additionally, the speaker recognition area has a long and rich scientific basis with over 30 years of research, development, and evaluations.

Over the last decade, speaker recognition technology has made its debut in several commercial products. The specific

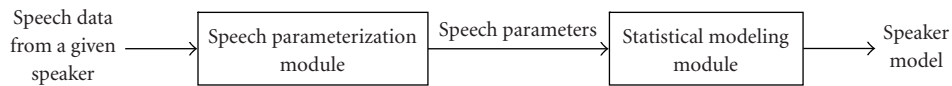


FIGURE 1: Modular representation of the training phase of a speaker verification system.

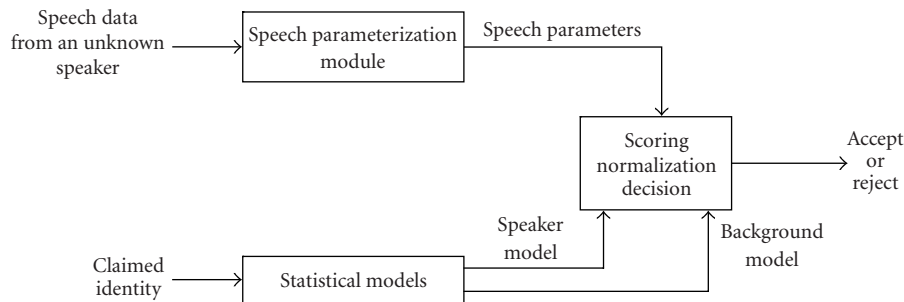


FIGURE 2: Modular representation of the test phase of a speaker verification system.

recognition task addressed in commercial systems is that of verification or detection (determining whether an unknown voice is from a particular enrolled speaker) rather than identification (associating an unknown voice with one from a set of enrolled speakers). Most deployed applications are based on scenarios with cooperative users speaking fixed digit string passwords or repeating prompted phrases from a small vocabulary. These generally employ what is known as text-dependent or text-constrained systems. Such constraints are quite reasonable and can greatly improve the accuracy of a system; however, there are cases when such constraints can be cumbersome or impossible to enforce. An example of this is background verification where a speaker is verified behind the scene as he/she conducts some other speech interactions. For cases like this, a more flexible recognition system able to operate without explicit user cooperation and independent of the spoken utterance (called text-independent mode) is needed. This paper focuses on the technologies behind these text-independent speaker verification systems.

A speaker verification system is composed of two distinct phases, a training phase and a test phase. Each of them can be seen as a succession of independent modules. Figure 1 shows a modular representation of the training phase of a speaker verification system. The first step consists in extracting parameters from the speech signal to obtain a representation suitable for statistical modeling as such models are extensively used in most state-of-the-art speaker verification systems. This step is described in Section 2. The second step consists in obtaining a statistical model from the parameters. This step is described in Section 3. This training scheme is also applied to the training of a background model (see Section 3).

Figure 2 shows a modular representation of the test phase of a speaker verification system. The entries of the system are a claimed identity and the speech samples pronounced by an unknown speaker. The purpose of a speaker verification

system is to verify if the speech samples correspond to the claimed identity. First, speech parameters are extracted from the speech signal using exactly the same module as for the training phase (see Section 2). Then, the speaker model corresponding to the claimed identity and a background model are extracted from the set of statistical models calculated during the training phase. Finally, using the speech parameters extracted and the two statistical models, the last module computes some scores, normalizes them, and makes an acceptance or a rejection decision (see Section 4). The normalization step requires some score distributions to be estimated during the training phase or/and the test phase (see the details in Section 4).

Finally, a speaker verification system can be text-dependent or text-independent. In the former case, there is some constraint on the type of utterance that users of the system can pronounce (for instance, a fixed password or certain words in any order, etc.). In the latter case, users can say whatever they want. This paper describes state-of-the-art text-independent speaker verification systems.

The outline of the paper is the following. Section 2 presents the most commonly used speech parameterization techniques in speaker verification systems, namely, cepstral analysis. Statistical modeling is detailed in Section 3, including an extensive presentation of Gaussian mixture modeling (GMM) and the mention of several speaker modeling alternatives like neural networks and support vector machines (SVMs). Section 4 explains how normalization is used. Section 5 shows how to evaluate a speaker verification system. In Section 6, several extensions of speaker verification are presented, namely, speaker tracking and speaker segmentation. Section 7 gives a few applications of speaker verification. Section 8 details specific problems relative to the use of speaker verification in the forensic area. Finally, Section 9 concludes this work and gives some future research directions.

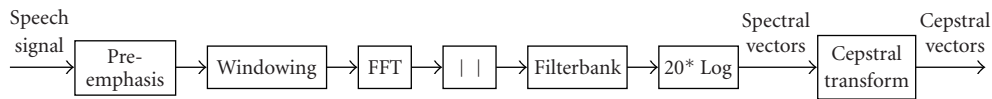


FIGURE 3: Modular representation of a filterbank-based cepstral parameterization.

2. SPEECH PARAMETERIZATION

Speech parameterization consists in transforming the speech signal to a set of feature vectors. The aim of this transformation is to obtain a new representation which is more compact, less redundant, and more suitable for statistical modeling and the calculation of a distance or any other kind of score. Most of the speech parameterizations used in speaker verification systems relies on a cepstral representation of speech.

2.1. Filterbank-based cepstral parameters

Figure 3 shows a modular representation of a filterbank-based cepstral representation.

The speech signal is first preemphasized, that is, a filter is applied to it. The goal of this filter is to enhance the high frequencies of the spectrum, which are generally reduced by the speech production process. The preemphasized signal is obtained by applying the following filter:

$$x_p(t) = x(t) - a \cdot x(t-1). \quad (1)$$

Values of a are generally taken in the interval $[0.95, 0.98]$. This filter is not always applied, and some people prefer not to preemphasize the signal before processing it. There is no definitive answer to this topic but empirical experimentation.

The analysis of the speech signal is done locally by the application of a window whose duration in time is shorter than the whole signal. This window is first applied to the beginning of the signal, then moved further and so on until the end of the signal is reached. Each application of the window to a portion of the speech signal provides a spectral vector (after the application of an FFT—see below). Two quantities have to be set: the length of the window and the shift between two consecutive windows. For the length of the window, two values are most often used: 20 milliseconds and 30 milliseconds. These values correspond to the average duration which allows the stationary assumption to be true. For the delay, the value is chosen in order to have an overlap between two consecutive windows; 10 milliseconds is very often used. Once these two quantities have been chosen, one can decide which window to use. The Hamming and the Hanning windows are the most used in speaker recognition. One usually uses a Hamming window or a Hanning window rather than a rectangular window to taper the original signal on the sides and thus reduce the side effects. In the Fourier domain, there is a convolution between the Fourier transform of the portion of the signal under consideration and the Fourier transform of the window. The Hamming window and the Han-

ning window are much more selective than the rectangular window.

Once the speech signal has been windowed, and possibly preemphasized, its fast Fourier transform (FFT) is calculated. There are numerous algorithms of FFT (see, for instance, [1, 2]).

Once an FFT algorithm has been chosen, the only parameter to fix for the FFT calculation is the number of points for the calculation itself. This number N is usually a power of 2 which is greater than the number of points in the window, classically 512.

Finally, the modulus of the FFT is extracted and a power spectrum is obtained, sampled over 512 points. The spectrum is symmetric and only half of these points are really useful. Therefore, only the first half of it is kept, resulting in a spectrum composed of 256 points.

The spectrum presents a lot of fluctuations, and we are usually not interested in all the details of them. Only the envelope of the spectrum is of interest. Another reason for the smoothing of the spectrum is the reduction of the size of the spectral vectors. To realize this smoothing and get the envelope of the spectrum, we multiply the spectrum previously obtained by a filterbank. A filterbank is a series of band-pass frequency filters which are multiplied one by one with the spectrum in order to get an average value in a particular frequency band. The filterbank is defined by the shape of the filters and by their frequency localization (left frequency, central frequency, and right frequency). Filters can be triangular, or have other shapes, and they can be differently located on the frequency scale. In particular, some authors use the Bark/Mel scale for the frequency localization of the filters. This scale is an auditory scale which is similar to the frequency scale of the human ear. The localization of the central frequencies of the filters is given by

$$f_{\text{MEL}} = 1000 \cdot \frac{\log(1 + f_{\text{LIN}}/1000)}{\log 2}. \quad (2)$$

Finally, we take the log of this spectral envelope and multiply each coefficient by 20 in order to obtain the spectral envelope in dB. At the stage of the processing, we obtain spectral vectors.

An additional transform, called the cosine discrete transform, is usually applied to the spectral vectors in speech processing and yields cepstral coefficients [2, 3, 4]:

$$c_n = \sum_{k=1}^K S_k \cdot \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right], \quad n = 1, 2, \dots, L, \quad (3)$$

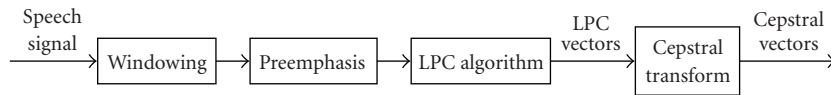


FIGURE 4: Modular representation of an LPC-based cepstral parameterization.

where K is the number of log-spectral coefficients calculated previously, S_k are the log-spectral coefficients, and L is the number of cepstral coefficients that we want to calculate ($L \leq K$). We finally obtain cepstral vectors for each analysis window.

2.2. LPC-based cepstral parameters

Figure 4 shows a modular representation of an LPC-based cepstral representation.

The LPC analysis is based on a linear model of speech production. The model usually used is an auto regressive moving average (ARMA) model, simplified in an auto regressive (AR) model. This modeling is detailed in particular in [5].

The speech production apparatus is usually described as a combination of four modules: (1) the glottal source, which can be seen as a train of impulses (for voiced sounds) or a white noise (for unvoiced sounds); (2) the vocal tract; (3) the nasal tract; and (4) the lips. Each of them can be represented by a filter: a lowpass filter for the glottal source, an AR filter for the vocal tract, an ARMA filter for the nasal tract, and an MA filter for the lips. Globally, the speech production apparatus can therefore be represented by an ARMA filter. Characterizing the speech signal (usually a windowed portion of it) is equivalent to determining the coefficients of the global filter. To simplify the resolution of this problem, the ARMA filter is often simplified in an AR filter.

The principle of LPC analysis is to estimate the parameters of an AR filter on a windowed (preemphasized or not) portion of a speech signal. Then, the window is moved and a new estimation is calculated. For each window, a set of coefficients (called predictive coefficients or LPC coefficients) is estimated (see [2, 6] for the details of the various algorithms that can be used to estimate the LPC coefficients) and can be used as a parameter vector. Finally, a spectrum envelope can be estimated for the current window from the predictive coefficients. But it is also possible to calculate cepstral coefficients directly from the LPC coefficients (see [6]):

$$\begin{aligned}
 c_0 &= \ln \sigma^2, \\
 c_m &= a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad 1 \leq m \leq p, \\
 c_m &= \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad p < m,
 \end{aligned} \tag{4}$$

where σ^2 is the gain term in the LPC model, a_m are the LPC

coefficients, and p is the number of LPC coefficients calculated.

2.3. Centered and reduced vectors

Once the cepstral coefficients have been calculated, they can be centered, that is, the cepstral mean vector is subtracted from each cepstral vector. This operation is called cepstral mean subtraction (CMS) and is often used in speaker verification. The motivation for CMS is to remove from the cepstrum the contribution of slowly varying convolutive noises.

The cepstral vectors can also be reduced, that is, the variance is normalized to one component by component.

2.4. Dynamic information

After the cepstral coefficients have been calculated, and possibly centered and reduced, we also incorporate in the vectors some dynamic information, that is, some information about the way these vectors vary in time. This is classically done by using the Δ and $\Delta\Delta$ parameters, which are polynomial approximations of the first and second derivatives [7]:

$$\begin{aligned}
 \Delta c_m &= \frac{\sum_{k=-l}^l k \cdot c_{m+k}}{\sum_{k=-l}^l |k|}, \\
 \Delta\Delta c_m &= \frac{\sum_{k=-l}^l k^2 \cdot c_{m+k}}{\sum_{k=-l}^l k^2}.
 \end{aligned} \tag{5}$$

2.5. Log energy and Δ log energy

At this step, one can choose whether to incorporate the log energy and the Δ log energy in the feature vectors or not. In practice, the former one is often discarded and the latter one is kept.

2.6. Discarding useless information

Once all the feature vectors have been calculated, a very important last step is to decide which vectors are useful and which are not. One way of looking at the problem is to determine vectors corresponding to speech portions of the signal versus those corresponding to silence or background noise. A way of doing it is to compute a bi-Gaussian model of the feature vector distribution. In that case, the Gaussian with the “lowest” mean corresponds to silence and background noise, and the Gaussian with the “highest” mean corresponds to speech portions. Then vectors having a higher likelihood with the silence and background noise Gaussian are discarded. A similar approach is to compute a bi-Gaussian model of the log energy distribution of each speech segment and to apply the same principle.

3. STATISTICAL MODELING

3.1. Speaker verification via likelihood ratio detection

Given a segment of speech Y and a hypothesized speaker S , the task of speaker verification, also referred to as detection, is to determine if Y was spoken by S . An implicit assumption often used is that Y contains speech from only one speaker. Thus, the task is better termed singlespeaker verification. If there is no prior information that Y contains speech from a single speaker, the task becomes multispeaker detection. This paper is primarily concerned with the single-speaker verification task. Discussion of systems that handle the multispeaker detection task is presented in other papers [8].

The single-speaker detection task can be stated as a basic hypothesis test between two hypotheses:

H0: Y is from the hypothesized speaker S ,

H1: Y is *not* from the hypothesized speaker S .

The optimum test to decide between these two hypotheses is a likelihood ratio (LR) test¹ given by

$$\frac{p(Y|H0)}{p(Y|H1)} \begin{cases} > \theta, & \text{accept H0,} \\ < \theta, & \text{accept H1,} \end{cases} \quad (6)$$

where $p(Y|H0)$ is the probability density function for the hypothesis H0 evaluated for the observed speech segment Y , also referred to as the “likelihood” of the hypothesis H0 given the speech segment.² The likelihood function for H1 is likewise $p(Y|H1)$. The decision threshold for accepting or rejecting H0 is θ . One main goal in designing a speaker detection system is to determine techniques to compute values for the two likelihoods $p(Y|H0)$ and $p(Y|H1)$.

Figure 5 shows the basic components found in speaker detection systems based on LRs. As discussed in Section 2, the role of the front-end processing is to extract from the speech signal features that convey speaker-dependent information. In addition, techniques to minimize confounding effects from these features, such as linear filtering or noise, may be employed in the front-end processing. The output of this stage is typically a sequence of feature vectors representing the test segment $X = \{\vec{x}_1, \dots, \vec{x}_T\}$, where \vec{x}_t is a feature vector indexed at discrete time $t \in [1, 2, \dots, T]$. There is no inherent constraint that features extracted at synchronous time instants be used; as an example, the overall speaking rate of an utterance could be used as a feature. These feature vectors are then used to compute the likelihoods of H0 and H1. Mathematically, a model denoted by λ_{hyp} represents H0, which characterizes the hypothesized speaker S in the feature space of \vec{x} . For example, one could assume that a Gaussian distribution best represents the distribution of feature vectors for H0 so that λ_{hyp} would contain the mean vector and covariance matrix parameters of the Gaussian distribution. The model

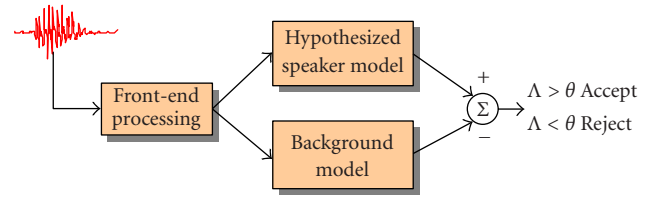


FIGURE 5: Likelihood-ratio-based speaker verification system.

λ_{hyp} represents the alternative hypothesis, H1. The likelihood ratio statistic is then $p(X|\lambda_{\text{hyp}})/p(X|\lambda_{\text{hyp}})$. Often, the logarithm of this statistic is used giving the log LR

$$\Lambda(X) = \log p(X|\lambda_{\text{hyp}}) - \log p(X|\lambda_{\text{hyp}}). \quad (7)$$

While the model for H0 is well defined and can be estimated using training speech from S , the model for λ_{hyp} is less well defined since it potentially must represent the entire space of possible alternatives to the hypothesized speaker. Two main approaches have been taken for this alternative hypothesis modeling. The first approach is to use a set of other speaker models to cover the space of the alternative hypothesis. In various contexts, this set of other speakers has been called likelihood ratio sets [9], cohorts [9, 10], and background speakers [9, 11]. Given a set of N background speaker models $\{\lambda_1, \dots, \lambda_N\}$, the alternative hypothesis model is represented by

$$p(X|\lambda_{\text{hyp}}) = f(p(X|\lambda_1), \dots, p(X|\lambda_N)), \quad (8)$$

where $f(\cdot)$ is some function, such as average or maximum, of the likelihood values from the background speaker set. The selection, size, and combination of the background speakers have been the subject of much research [9, 10, 11, 12]. In general, it has been found that to obtain the best performance with this approach requires the use of speaker-specific background speaker sets. This can be a drawback in applications using a large number of hypothesized speakers, each requiring their own background speaker set.

The second major approach to the alternative hypothesis modeling is to pool speech from several speakers and train a single model. Various terms for this single model are a general model [13], a world model, and a universal background model (UBM) [14]. Given a collection of speech samples from a large number of speakers representative of the population of speakers expected during verification, a single model λ_{bkg} is trained to represent the alternative hypothesis. Research on this approach has focused on selection and composition of the speakers and speech used to train the single model [15, 16]. The main advantage of this approach is that a single speaker-independent model can be trained once for a particular task and then used for all hypothesized speakers in that task. It is also possible to use multiple background models tailored to specific sets of speakers [16, 17]. The use of a single background model has become the predominate approach used in speaker verification systems.

¹Strictly speaking, the likelihood ratio test is only optimal when the likelihood functions are known exactly. In practice, this is rarely the case.

² $p(A|B)$ is referred to as a likelihood when B is considered the independent variable in the function.

3.2. Gaussian mixture models

An important step in the implementation of the above likelihood ratio detector is the selection of the actual likelihood function $p(X|\lambda)$. The choice of this function is largely dependent on the features being used as well as specifics of the application. For text-independent speaker recognition, where there is no prior knowledge of what the speaker will say, the most successful likelihood function has been GMMs. In text-dependent applications, where there is a strong prior knowledge of the spoken text, additional temporal knowledge can be incorporated by using hidden Markov models (HMMs) for the likelihood functions. To date, however, the use of more complicated likelihood functions, such as those based on HMMs, have shown no advantage over GMMs for text-independent speaker detection tasks like in the NIST speaker recognition evaluations (SREs).

For a D -dimensional feature vector \vec{x} , the mixture density used for the likelihood function is defined as follows:

$$p(\vec{x}|\lambda) = \sum_{i=1}^M w_i p_i(\vec{x}). \quad (9)$$

The density is a weighted linear combination of M unimodal Gaussian densities $p_i(\vec{x})$, each parameterized by a $D \times 1$ mean vector $\vec{\mu}_i$ and a $D \times D$ covariance matrix Σ_i :

$$p_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} e^{-(1/2)(\vec{x}-\vec{\mu}_i)' \Sigma_i^{-1} (\vec{x}-\vec{\mu}_i)}. \quad (10)$$

The mixture weights w_i further satisfy the constraint $\sum_{i=1}^M w_i = 1$. Collectively, the parameters of the density model are denoted as $\lambda = (w_i, \vec{\mu}_i, \Sigma_i), i = (1, \dots, M)$.

While the general model form supports full covariance matrices, that is, a covariance matrix with all its elements, typically only diagonal covariance matrices are used. This is done for three reasons. First, the density modeling of an M th-order full covariance GMM can equally well be achieved using a larger-order diagonal covariance GMM.³ Second, diagonal-matrix GMMs are more computationally efficient than full covariance GMMs for training since repeated inversions of a $D \times D$ matrix are not required. Third, empirically, it has been observed that diagonal-matrix GMMs outperform full-matrix GMMs.

Given a collection of training vectors, maximum likelihood model parameters are estimated using the iterative expectation-maximization (EM) algorithm [18]. The EM algorithm iteratively refines the GMM parameters to monotonically increase the likelihood of the estimated model for the observed feature vectors, that is, for iterations k and $k+1$, $p(X|\lambda^{(k+1)}) \geq p(X|\lambda^{(k)})$. Generally, five–ten iterations are sufficient for parameter convergence. The EM equations for training a GMM can be found in the literature [18, 19, 20].

³GMMs with $M > 1$ using diagonal covariance matrices can model distributions of feature vectors with correlated elements. Only in the degenerate case of $M = 1$ is the use of a diagonal covariance matrix incorrect for feature vectors with correlated elements.

Under the assumption of independent feature vectors, the log-likelihood of a model λ for a sequence of feature vectors $X = \{\vec{x}_1, \dots, \vec{x}_T\}$ is computed as follows:

$$\log p(X|\lambda) = \frac{1}{T} \sum_t \log p(\vec{x}_t|\lambda), \quad (11)$$

where $p(\vec{x}_t|\lambda)$ is computed as in equation (9). Note that the average log-likelihood value is used so as to normalize out duration effects from the log-likelihood value. Also, since the incorrect assumption of independence is underestimating the actual likelihood value with dependencies, scaling by T can be considered a rough compensation factor.

The GMM can be viewed as a hybrid between parametric and nonparametric density models. Like a parametric model, it has structure and parameters that control the behavior of the density in known ways, but without constraints that the data must be of a specific distribution type, such as Gaussian or Laplacian. Like a nonparametric model, the GMM has many degrees of freedom to allow arbitrary density modeling, without undue computation and storage demands. It can also be thought of as a single-state HMM with a Gaussian mixture observation density, or an ergodic Gaussian observation HMM with fixed, equal transition probabilities. Here, the Gaussian components can be considered to be modeling the underlying broad phonetic sounds that characterize a person's voice. A more detailed discussion of how GMMs apply to speaker modeling can be found elsewhere [21].

The advantages of using a GMM as the likelihood function are that it is computationally inexpensive, is based on a well-understood statistical model, and, for text-independent tasks, is insensitive to the temporal aspects of the speech, modeling only the underlying distribution of acoustic observations from a speaker. The latter is also a disadvantage in that higher-levels of information about the speaker conveyed in the temporal speech signal are not used. The modeling and exploitation of these higher-levels of information may be where approaches based on speech recognition [22] produce benefits in the future. To date, however, these approaches (e.g., large vocabulary or phoneme recognizers) have basically been used only as means to compute likelihood values, without explicit use of any higher-level information, such as speaker-dependent word usage or speaking style. Some recent work, however, has shown that high-level information can be successfully extracted and combined with acoustic scores from a GMM system for improved speaker verification performance [23, 24].

3.3. Adapted GMM system

As discussed earlier, the dominant approach to background modeling is to use a single, speaker-independent background model to represent $p(X|\lambda_{\text{hyp}})$. Using a GMM as the likelihood function, the background model is typically a large GMM trained to represent the speaker-independent distribution of features. Specifically, speech should be selected that reflects the expected alternative speech to be encountered during recognition. This applies to the type and quality of speech as well as the composition of speakers. For

example, in the NIST SRE single-speaker detection tests, it is known a priori that the speech comes from local and long-distance telephone calls, and that male hypothesized speakers will only be tested against male speech. In this case, we would train the UBM used for male tests using only male telephone speech. In the case where there is no prior knowledge of the gender composition of the alternative speakers, we would train using gender-independent speech. The GMM order for the background model is usually set between 512–2048 mixtures depending on the data. Lower-order mixtures are often used when working with constrained speech (such as digits or fixed vocabulary), while 2048 mixtures are used when dealing with unconstrained speech (such as conversational speech).

Other than these general guidelines and experimentation, there is no objective measure to determine the right number of speakers or amount of speech to use in training a background model. Empirically, from the NIST SRE, we have observed no performance loss using a background model trained with one hour of speech compared to a one trained using six hours of speech. In both cases, the training speech was extracted from the same speaker population.

For the speaker model, a single GMM can be trained using the EM algorithm on the speaker's enrollment data. The order of the speaker's GMM will be highly dependent on the amount of enrollment speech, typically 64–256 mixtures. In another more successful approach, the speaker model is derived by adapting the parameters of the background model using the speaker's training speech and a form of Bayesian adaptation or maximum a posteriori (MAP) estimation [25]. Unlike the standard approach of maximum likelihood training of a model for the speaker, independently of the background model, the basic idea in the adaptation approach is to derive the speaker's model by updating the well-trained parameters in the background model via adaptation. This provides a tighter coupling between the speaker's model and background model that not only produces better performance than decoupled models, but, as discussed later in this section, also allows for a fast-scoring technique. Like the EM algorithm, the adaptation is a two-step estimation process. The first step is identical to the "expectation" step of the EM algorithm, where estimates of the sufficient statistics⁴ of the speaker's training data are computed for each mixture in the UBM. Unlike the second step of the EM algorithm, for adaptation, these "new" sufficient statistic estimates are then combined with the "old" sufficient statistics from the background model mixture parameters using a data-dependent mixing coefficient. The data-dependent mixing coefficient is designed so that mixtures with high counts of data from the speaker rely more on the new sufficient statistics for final parameter estimation, and mixtures with low counts of data from the speaker rely more on the old sufficient statistics for final parameter estimation.

⁴These are the basic statistics required to compute the desired parameters. For a GMM mixture, these are the count, and the first and second moments required to compute the mixture weight, mean and variance.

The specifics of the adaptation are as follows. Given a background model and training vectors from the hypothesized speaker, we first determine the probabilistic alignment of the training vectors into the background model mixture components. That is, for mixture i in the background model, we compute

$$\Pr(i|\vec{x}_t) = \frac{w_i p_i(\vec{x}_t)}{\sum_{j=1}^M w_j p_j(\vec{x}_t)}. \quad (12)$$

We then use $\Pr(i|\vec{x}_t)$ and \vec{x}_t to compute the sufficient statistics for the weight, mean, and variance parameters:⁵

$$\begin{aligned} n_i &= \sum_{t=1}^T \Pr(i|\vec{x}_t), \\ E_i(\vec{x}) &= \frac{1}{n_i} \sum_{t=1}^T \Pr(i|\vec{x}_t) \vec{x}_t, \\ E_i(\vec{x}^2) &= \frac{1}{n_i} \sum_{t=1}^T \Pr(i|\vec{x}_t) \vec{x}_t^2. \end{aligned} \quad (13)$$

This is the same as the expectation step in the EM algorithm.

Lastly, these new sufficient statistics from the training data are used to update the old background model sufficient statistics for mixture i to create the adapted parameters for mixture i with the equations

$$\begin{aligned} \hat{w}_i &= [\alpha_i n_i / T + (1 - \alpha_i) w_i] \gamma, \\ \hat{\mu}_i &= \alpha_i E_i(\vec{x}) + (1 - \alpha_i) \vec{\mu}_i, \\ \hat{\sigma}_i^2 &= \alpha_i E_i(\vec{x}^2) + (1 - \alpha_i) (\vec{\sigma}_i^2 + \vec{\mu}_i^2) - \hat{\mu}_i^2. \end{aligned} \quad (14)$$

The scale factor γ is computed over all adapted mixture weights to ensure they sum to unity. The adaptation coefficient controlling the balance between old and new estimates is α_i and is defined as follows:

$$\alpha_i = \frac{n_i}{n_i + r}, \quad (15)$$

where r is a fixed "relevance" factor.

The parameter updating can be derived from the general MAP estimation equations for a GMM using constraints on the prior distribution described in Gauvain and Lee's paper [25, Section V, equations (47) and (48)]. The parameter updating equation for the weight parameter, however, does not follow from the general MAP estimation equations.

Using a data-dependent adaptation coefficient allows mixture-dependent adaptation of parameters. If a mixture component has a low probabilistic count n_i of new data, then $\alpha_i \rightarrow 0$ causing the deemphasis of the new (potentially under-trained) parameters and the emphasis of the old (better trained) parameters. For mixture components with high probabilistic counts, $\alpha_i \rightarrow 1$ causing the use of the new speaker-dependent parameters. The relevance factor is a way

⁵ \vec{x}^2 is shorthand for $\text{diag}(\vec{x}\vec{x}^T)$.

of controlling how much new data should be observed in a mixture before the new parameters begin replacing the old parameters. This approach should thus be robust to limited training data. This factor can also be made parameter dependent, but experiments have found that this provides little benefit. Empirically, it has been found that only adapting the mean vectors provides the best performance.

Published results [14] and NIST evaluation results from several sites strongly indicate that the GMM adaptation approach provides superior performance over a decoupled system, where the speaker model is trained independently of the background model. One possible explanation for the improved performance is that the use of adapted models in the likelihood ratio is not affected by “unseen” acoustic events in recognition speech. Loosely speaking, if one considers the background model as covering the space of speaker-independent, broad acoustic classes of speech sounds, then adaptation is the speaker-dependent “tuning” of those acoustic classes observed in the speaker’s training speech. Mixture parameters for those acoustic classes not observed in the training speech are merely copied from the background model. This means that during recognition, data from acoustic classes unseen in the speaker’s training speech produce approximately zero log LR values that contribute evidence neither towards nor against the hypothesized speaker. Speaker models trained using only the speaker’s training speech will have low likelihood values for data from classes not observed in the training data thus producing low likelihood ratio values. While this is appropriate for speech not for the speaker, it clearly can cause incorrect values when the unseen data occurs in test speech from the speaker.

The adapted GMM approach also leads to a fast-scoring technique. Computing the log LR requires computing the likelihood for the speaker and background model for each feature vector, which can be computationally expensive for large mixture orders. However, the fact that the hypothesized speaker model was adapted from the background model allows a faster scoring method. This fast-scoring approach is based on two observed effects. The first is that when a large GMM is evaluated for a feature vector, only a few of the mixtures contribute significantly to the likelihood value. This is because the GMM represents a distribution over a large space but a single vector will be near only a few components of the GMM. Thus likelihood values can be approximated very well using only the top C best scoring mixture components. The second observed effect is that the components of the adapted GMM retain a correspondence with the mixtures of the background model so that vectors close to a particular mixture in the background model will also be close to the corresponding mixture in the speaker model.

Using these two effects, a fast-scoring procedure operates as follows. For each feature vector, determine the top C scoring mixtures in the background model and compute background model likelihood using only these top C mixtures. Next, score the vector against only the corresponding C components in the adapted speaker model to evaluate the speaker’s likelihood.

For a background model with M mixtures, this requires only $M + C$ Gaussian computations per feature vector compared to $2M$ Gaussian computations for normal likelihood ratio evaluation. When there are multiple hypothesized speaker models for each test segment, the savings become even greater. Typically, a value of $C = 5$ is used.

3.4. Alternative speaker modeling techniques

Another way to solve the classification problem for speaker verification systems is to use discrimination-based learning procedures such as artificial neural networks (ANN) [26, 27] or SVMs [28]. As explained in [29, 30], the main advantages of ANN include their discriminant-training power, a flexible architecture that permits easy use of contextual information, and weaker hypothesis about the statistical distributions. The main disadvantages are that their optimal structure has to be selected by trial-and-error procedures, the need to split the available train data in training and cross-validation sets, and the fact that the temporal structure of speech signals remains difficult to handle. They can be used as binary classifiers for speaker verification systems to separate the speaker and the nonspeaker classes as well as multiclassifiers for speaker identification purposes. ANN have been used for speaker verification [31, 32, 33]. Among the different ANN architectures, multilayer perceptrons (MLP) are often used [6, 34].

SVMs are an increasingly popular method used in speaker verification systems. SVM classifiers are well suited to separate rather complex regions between two classes through an optimal, nonlinear decision boundary. The main problems are the search for the appropriate kernel function for a particular application and their inappropriateness to handle the temporal structure of the speech signals. There are also some recent studies [35] in order to adapt the SVM to the multiclassification problem. The SVM were already applied for speaker verification. In [23, 36], the widely used speech feature vectors were used as the input training material for the SVM.

Generally speaking, the performance of speaker verification systems based on discrimination-based learning techniques can be tuned to obtain comparable performance to the state-of-the-art GMM, and in some special experimental conditions, they could be tuned to outperform the GMM. It should be noted that, as explained earlier in this section, the tuning of a GMM baseline systems is not straightforward, and different parameters such as the training method, the number of mixtures, and the amount of speech to use in training a background model have to be adjusted to the experimental conditions. Therefore, when comparing a new system to the classical GMM system, it is difficult to be sure that the baseline GMM used are comparable to the best performing ones.

Another recent alternative to solve the speaker verification problem is to combine GMM with SVMs. We are not going to give here an extensive study of all the experiments done [37, 38, 39], but we are rather going to illustrate the problem with one example meant to exploit together the GMM and SVM for speaker verification purposes. One of the

problems with the speaker verification is the score normalization (see Section 4). Because SVM are well suited to determine an optimal hyperplan separating data belonging to two classes, one way to use them for speaker verification is to separate the likelihood client and nonclient values with an SVM. That was the idea implemented in [37], and an SVM was constructed to separate two classes, the clients from the impostors. The GMM technique was used to construct the input feature representation for the SVM classifier. The speaker GMM models were built by adaptation of the background model. The GMM likelihood values for each frame and each Gaussian mixture were used as the input feature vector for the SVM. This combined GMM-SVM method gave slightly better results than the GMM method alone. Several points should be emphasized: the results were obtained on a subset of NIST'1999 speaker verification data, only the Znorm was tested, and neither the GMM nor the SVM parameters were thoroughly adjusted. The conclusion is that the results demonstrate the feasibility of this technique, but in order to fully exploit these two techniques, more work should be done.

4. NORMALIZATION

4.1. Aims of score normalization

The last step in speaker verification is the decision making. This process consists in comparing the likelihood resulting from the comparison between the claimed speaker model and the incoming speech signal with a decision threshold. If the likelihood is higher than the threshold, the claimed speaker will be accepted, else rejected.

The tuning of decision thresholds is very troublesome in speaker verification. If the choice of its numerical value remains an open issue in the domain (usually fixed empirically), its reliability cannot be ensured while the system is running. This uncertainty is mainly due to the score variability between trials, a fact well known in the domain.

This score variability comes from different sources. First, the nature of the enrollment material can vary between the speakers. The differences can also come from the phonetic content, the duration, the environment noise, as well as the quality of the speaker model training. Secondly, the possible mismatch between enrollment data (used for speaker modeling) and test data is the main remaining problem in speaker recognition. Two main factors may contribute to this mismatch: the speaker him-/herself through the intraspeaker variability (variation in speaker voice due to emotion, health state, and age) and some environment condition changes in transmission channel, recording material, or acoustical environment. On the other hand, the interspeaker variability (variation in voices between speakers), which is a particular issue in the case of speaker-independent threshold-based system, has to be also considered as a potential factor affecting the reliability of decision boundaries. Indeed, as this interspeaker variability is not directly measurable, it is not straightforward to protect the speaker verification system (through the decision making process) against all potential impostor

attacks. Lastly, as for the training material, the nature and the quality of test segments influence the value of the scores for client and impostor trials.

Score normalization has been introduced explicitly to cope with score variability and to make speaker-independent decision threshold tuning easier.

4.2. Expected behavior of score normalization

Score normalization techniques have been mainly derived from the study of Li and Porter [40]. In this paper, large variances had been observed from both distributions of client scores (intraspeaker scores) and impostor scores (interspeaker scores) during speaker verification tests. Based on these observations, the authors proposed solutions based on impostor score distribution normalization in order to reduce the overall score distribution variance (both client and impostor distributions) of the speaker verification system. The basic of the normalization technique is to center the impostor score distribution by applying on each score generated by the speaker verification system the following normalization. Let $L_\lambda(X)$ denote the score for speech signal X and speaker model λ . The normalized score $\tilde{L}_\lambda(X)$ is then given as follows:

$$\tilde{L}_\lambda(X) = \frac{L_\lambda(X) - \mu_\lambda}{\sigma_\lambda}, \quad (16)$$

where μ_λ and σ_λ are the normalization parameters for speaker λ . Those parameters need to be estimated.

The choice of normalizing the impostor score distribution (as opposed to the client score distribution) was initially guided by two facts. First, in real applications and for text-independent systems, it is easy to compute impostor score distributions using pseudo-impostors, but client distributions are rarely available. Secondly, impostor distribution represents the largest part of the score distribution variance. However, it would be interesting to study client score distribution (and normalization), for example, in order to determine theoretically the decision threshold. Nevertheless, as seen previously, it is difficult to obtain the necessary data for real systems and only few current databases contain enough data to allow an accurate estimate of client score distribution.

4.3. Normalization techniques

Since the study of Li and Porter [40], various kinds of score normalization techniques have been proposed in the literature. Some of them are briefly described in the following section.

World-model and cohort-based normalizations

This class of normalization techniques is a particular case: it relies more on the estimation of antispeaker hypothesis ("the target speaker does not pronounce the record") in the Bayesian hypothesis test than on a normalization scheme. However, the effects of this kind of techniques on the different score distributions are so close to the normalization method ones that we have to present here.

The first proposal came from Higgins et al. in 1991 [9], followed by Matsui and Furui in 1993 [41], for which the normalized scores take the form of a ratio of likelihoods as follows:

$$\tilde{L}_\lambda(X) = \frac{L_\lambda(X)}{L_{\bar{\lambda}}(X)}. \quad (17)$$

For both approaches, the likelihood $L_{\bar{\lambda}}(y)$ was estimated from a cohort of speaker models. In [9], the cohort of speakers (also denoted as a cohort of impostors) was chosen to be close to speaker λ . Conversely, in [41], the cohort of speakers included speaker λ . Nevertheless, both normalization schemes equally improve speaker verification performance.

In order to reduce the amount of computation, the cohort of impostor models was replaced later with a unique model learned using the same data as the first ones. This idea is the basic of world-model normalization (the world model is also named “background model”) firstly introduced by Carey et al. [13]. Several works showed the interest in world-model-based normalization [14, 17, 42].

All the other normalizations discussed in this paper are applied on world-model normalized scores (commonly named likelihood ratio in the way of statistical approaches), that is, $\tilde{L}_\lambda(X) = \Lambda_\lambda(X)$.

Centered/reduced impostor distribution

This family of normalization techniques is the most used. It is directly derived from (16), where the scores are normalized by subtracting the mean and then dividing by the standard deviation, both estimated from the (pseudo)impostor score distribution. Different possibilities are available to compute the impostor score distribution.

Znorm

The zero normalization (Znorm) technique is directly derived from the work done in [40]. It has been massively used in speaker verification in the middle of the nineties. In practice, a speaker model is tested against a set of speech signals produced by some impostor, resulting in an impostor similarity score distribution. Speaker-dependent mean and variance—normalization parameters—are estimated from this distribution and applied (see (16) on similarity scores yielded by the speaker verification system when running. One of the advantages of Znorm is that the estimation of the normalization parameters can be performed offline during speaker model training.

Hnorm

By observing that, for telephone speech, most of the client speaker models respond differently according to the handset type used during testing data recording, Reynolds [43] had proposed a variant of Znorm technique, named handset normalization (Hnorm), to deal with handset mismatch between training and testing.

Here, handset-dependent normalization parameters are estimated by testing each speaker model against handset-

dependent speech signals produced by impostors. During testing, the type of handset relating to the incoming speech signal determines the set of parameters to use for score normalization.

Tnorm

Still based on the estimate of mean and variance parameters to normalize impostor score distribution, test-normalization (Tnorm), proposed in [44], differs from Znorm by the use of impostor models instead of test speech signals. During testing, the incoming speech signal is classically compared with claimed speaker model as well as with a set of impostor models to estimate impostor score distribution and normalization parameters consecutively. If Znorm is considered as a speaker-dependent normalization technique, Tnorm is a test-dependent one. As the same test utterance is used during both testing and normalization parameter estimate, Tnorm avoids a possible issue of Znorm based on a possible mismatch between test and normalization utterances. Conversely, Tnorm has to be performed online during testing.

HTnorm

Based on the same observation as Hnorm, a variant of Tnorm has been proposed, named HTnorm, to deal with handset-type information. Here, handset-dependent normalization parameters are estimated by testing each incoming speech signal against handset-dependent impostor models. During testing, the type of handset relating to the claimed speaker model determines the set of parameters to use for score normalization.

Cnorm

Cnorm was introduced by Reynolds during NIST 2002 speaker verification evaluation campaigns in order to deal with cellular data. Indeed, the new corpus (Switchboard cellular phase 2) is composed of recordings obtained using different cellular phones corresponding to several unidentified handsets. To cope with this issue, Reynolds proposed a blind clustering of the normalization data followed by an Hnorm-like process using each cluster as a different handset.

This class of normalization methods offers some advantages particularly in the framework of NIST evaluations (text independent speaker verification using long segments of speech—30 seconds in average for tests and 2 minutes for enrollment). First, both the method and the impostor distribution model are simple, only based on mean and standard deviation computation for a given speaker (even if Tnorm complicates the principle by the need of online processing). Secondly, the approach is well adapted to a text-independent task, with a large amount of data for enrollment. These two points allow to find easily pseudo-impostor data. It seems more difficult to find these data in the case of a user-password-based system, where the speaker chooses his password and repeats it three or four times during the enrollment phase only. Lastly, modeling only the impostor distribution is a good way to set a threshold according to the global false acceptance error and reflects the NIST scoring strategy.

For a commercial system, the level of false rejection is critical and the quality of the system is driven by the quality reached for the “worse” speakers (and not for the average).

Dnorm

Dnorm was proposed by Ben et al. in 2002 [45]. Dnorm deals with the problem of pseudo-impostor data availability by generating the data using the world model. A Monte Carlo-based method is applied to obtain a set of client and impostor data, using, respectively, client and world models. The normalized score is given by

$$\tilde{L}_\lambda(X) = \frac{L_\lambda(X)}{\text{KL2}(\lambda, \bar{\lambda})}, \quad (18)$$

where $\text{KL2}(\lambda, \bar{\lambda})$ is the estimate of the symmetrized Kullback-Leibler distance between the client and world models. The estimation of the distance is done using Monte-Carlo generated data. As for the previous normalizations, Dnorm is applied on likelihood ratio, computed using a world model.

Dnorm presents the advantage not to need any normalization data in addition to the world model. As Dnorm is a recent proposition, future developments will show if the method could be applied in different applications like password-based systems.

WMAP

WMAP is designed for multirecognizer systems. The technique focuses on the meaning of the score and not only on normalization. WMAP, proposed by Fredouille et al. in 1999 [46], is based on the Bayesian decision framework. The originality is to consider the two classical speaker recognition hypotheses in the score space and not in the acoustic space. The final score is the a posteriori probability to obtain the score given the target hypothesis:

$$\begin{aligned} \text{WMAP}(L_\lambda(X)) \\ = \frac{P_{\text{Target}} \cdot p(L_\lambda(X)|\text{Target})}{P_{\text{Target}} \cdot p(L_\lambda(X)|\text{Target}) + P_{\text{Imp}} \cdot p(L_\lambda(X)|\text{Imp})}, \end{aligned} \quad (19)$$

where P_{Target} (resp., P_{Imp}) is the a priori probability of a target test (resp., an impostor test) and $p(L_\lambda(X)|\text{Target})$ (resp., $p(L_\lambda(X)|\text{Imp})$) is the probability of score $L_\lambda(X)$ given the hypothesis of a target test (resp., an impostor test).

The main advantage of the WMAP⁶ normalization is to produce meaningful normalized score in the probability space. The scores take the quality of the recognizer directly into account, helping the system design in the case of multiple recognizer decision fusion.

The implementation proposed by Fredouille in 1999 used an empirically approach and nonparametric models for estimating the target and impostor score probabilities.

⁶The method is called WMAP as it is a maximum a posteriori approach applied on likelihood ratio where the denominator is computed using a world model.

4.4. Discussion

Through the various experiments achieved on the use of normalization in speaker verification, different points may be highlighted. First of all, the use of prior information like the handset type or gender information during normalization parameter computation is relevant to improve performance (see [43] for experiments on Hnorm and [44] for experiment on HTnorm).

Secondly, HTnorm seems better than the other kind of normalization as shown during the 2001 and 2002 NIST evaluation campaigns. Unfortunately, HTnorm is also the most expensive in computational time and requires estimating normalization parameters during testing. The solution proposed in [45], Dnorm normalization, may be a promising alternative since the computational time is significantly reduced and no impostor population is required to estimate normalization parameters. Currently, Dnorm performs as well as Znorm technique [45]. Further work is expected in order to integrate prior information like handset type to Dnorm and to make it comparable with Hnorm and HTnorm. WMAP technique exhibited interesting performance (same level as Znorm but without any knowledge about the real target speaker—normalization parameters are learned a priori using a separate set of speakers/tests). However, the technique seemed difficult to apply in a target speaker-dependent mode, since few speaker data are not sufficient to learn the normalization models. A solution could be to generate data, as done in the Dnorm approach, to estimate the score models Target and Imp (impostor) directly from the models.

Finally, as shown during the 2001 and 2002 NIST evaluation campaigns, the combination of different kinds of normalization (e.g., HTnorm & Hnorm, Tnorm & Dnorm) may lead to improved speaker verification performance. It is interesting to note that each winning normalization combination relies on the association between a “learning condition” normalization (Znorm, Hnorm, and Dnorm) and a “test-based” normalization (HTnorm and Tnorm).

However, this behavior of current speaker verification systems which require score normalization to perform better may lead to question the relevancy of techniques used to obtain these scores. The state-of-the-art text-independent speaker recognition techniques associate one or several parameterization level normalizations (CMS, feature variance normalization, feature warping, etc.) with a world model normalization and one or several score normalizations. Moreover, the speaker models are mainly computed by adapting a world/background model to the client enrollment data which could be considered as a “model” normalization.

Observing that at least four different levels of normalization are used, the question remains: is the front-end processing, the statistical techniques (like GMM) the best way of modeling speaker characteristics and speech signal variability, including mismatch between training and testing data? After many years of research, speaker verification still remains an open domain.

5. EVALUATION

5.1. Types of errors

Two types of errors can occur in a speaker verification system, namely, false rejection and false acceptance. A false rejection (or nondetection) error happens when a valid identity claim is rejected. A false acceptance (or false alarm) error consists in accepting an identity claim from an impostor. Both types of error depend on the threshold θ used in the decision making process. With a low threshold, the system tends to accept every identity claim thus making few false rejections and lots of false acceptances. On the contrary, if the threshold is set to some high value, the system will reject every claim and make very few false acceptances but a lot of false rejections. The couple (false alarm error rate, false rejection error rate) is defined as the *operating point* of the system. Defining the operating point of a system, or, equivalently, setting the decision threshold, is a trade-off between the two types of errors.

In practice, the false alarm and nondetection error rates, denoted by P_{fa} and P_{fr} , respectively, are measured experimentally on a test corpus by counting the number of errors of each type. This means that large test sets are required to be able to measure accurately the error rates. For clear methodological reasons, it is crucial that none of the test speakers, whether true speakers or impostors, be in the training and development sets. This excludes, in particular, using the same speakers for the background model and for the tests. However, it may be possible to use speakers referenced in the test database as impostors. This should be avoided whenever discriminative training techniques are used or if across speaker normalization is done since, in this case, using referenced speakers as impostors would introduce a bias in the results.

5.2. DET curves and evaluation functions

As mentioned previously, the two error rates are functions of the decision threshold. It is therefore possible to represent the performance of a system by plotting P_{fa} as a function of P_{fr} . This curve, known as the system operating characteristic, is monotonous and decreasing. Furthermore, it has become a standard to plot the error curve on a normal deviate scale [47] in which case the curve is known as the detection error trade-offs (DETs) curve. With the normal deviate scale, a speaker recognition system whose true speaker and impostor scores are Gaussians with the same variance will result in a linear curve with a slope equal to -1 . The better the system is, the closer to the origin the curve will be. In practice, the score distributions are not exactly Gaussians but are quite close to it. The DET curve representation is therefore more easily readable and allows for a comparison of the system's performances on a large range of operating conditions. Figure 6 shows a typical example of a DET curves.

Plotting the error rates as a function of the threshold is a good way to compare the potential of different methods in laboratory applications. However, this is not suited for the evaluation of operating systems for which the threshold has been set to operate at a given point. In such a case, systems are evaluated according to a cost function which takes into

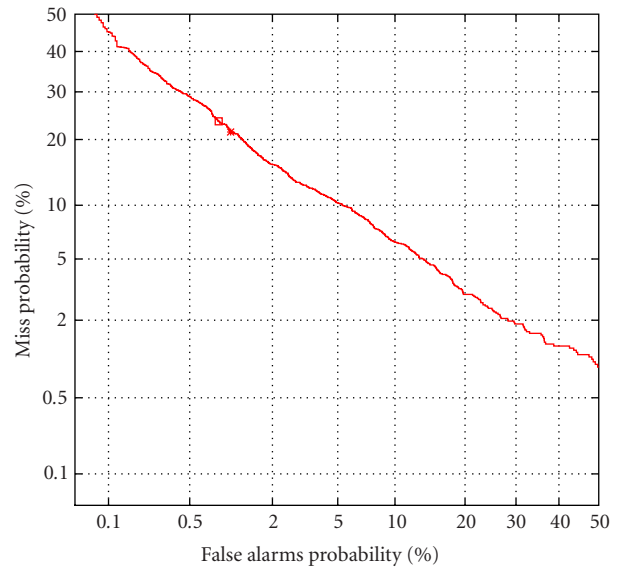


FIGURE 6: Example of a DET curve.

account the two error rates weighted by their respective costs, that is $C = C_{fa}P_{fa} + C_{fr}P_{fr}$. In this equation, C_{fa} and C_{fr} are the costs given to false acceptances and false rejections, respectively. The cost function is minimal if the threshold is correctly set to the desired operating point. Moreover, it is possible to directly compare the costs of two operating systems. If normalized by the sum of the error costs, the cost C can be interpreted as the mean of the error rates, weighted by the cost of each error.

Other measures are sometimes used to summarize the performance of a system in a single figure. A popular one is the equal error rate (EER) which corresponds to the operating point where $P_{fa} = P_{fr}$. Graphically, it corresponds to the intersection of the DET curve with the first bisector curve. The EER performance measure rarely corresponds to a realistic operating point. However, it is a quite popular measure of the ability of a system to separate impostors from true speakers. Another popular measure is the half total error rate (HTER) which is the average of the two error rates P_{fa} and P_{fr} . It can also be seen as the normalized cost function assuming equal costs for both errors.

Finally, we make the distinction between a cost obtained with a system whose operating point has been set up on development data and a cost obtained with a posterior minimization of the cost function. The latter is always to the advantage of the system but does not correspond to a realistic evaluation since it makes use of the test data. However, the difference between those two costs can be used to evaluate the quality of the decision making module (in particular, it evaluates how well the decision threshold has been set).

5.3. Factors affecting the performance and evaluation paradigm design

There are several factors affecting the performance of a speaker verification system. First, several factors have an

impact on the quality of the speech material recorded. Among others, these factors are the environmental conditions at the time of the recording (background noise or not), the type of microphone used, and the transmission channel bandwidth and compression if any (high bandwidth speech, landline and cell phone speech, etc.). Second are factors concerning the speakers themselves and the amount of training data available. These factors are the number of training sessions and the time interval between those sessions (several training sessions over a long period of time help coping with the long-term variability of speech), the physical and emotional state of the speaker (under stress or ill), the speaker cooperativeness (does the speaker want to be recognized or does the impostor really want to cheat, is the speaker familiar with the system, and so forth). Finally, the system performance measure highly depends on the test set complexity: cross gender trials or not, test utterance duration, linguistic coverage of those utterances, and so forth. Ideally, all those factors should be taken into account when designing evaluation paradigms or when comparing the performance of two systems on different databases. The excellent performance obtained in artificial good conditions (quiet environment, high-quality microphone, consecutive recordings of the training and test material, and speaker willing to be identified) rapidly degrades in real-life applications.

Another factor affecting the performance worth noting is the test speakers themselves. Indeed, it has been observed several times that the distribution of errors varies greatly between speakers [48]. A small number of speakers (goats) are responsible for most of the nondetection errors, while another small group of speakers (lambs) are responsible for the false acceptance errors. The performance computed by leaving out these two small groups are clearly much better. Evaluating the performance of a system after removing a small percentage of the speakers whose individual error rates are the higher may be interesting in commercial applications where it is better to have a few unhappy customers (for which an alternate solution to speaker verification can be envisaged) than many ones.

5.4. Typical performance

It is quite impossible to give a complete overview of the speaker verification systems because of the great diversity of applications and experimental conditions. However, we conclude this section by giving the performance of some systems trained and tested with an amount of data reasonable in the context of an application (one or two training sessions and test utterances between 10 and 30 seconds).

For good recording conditions and for text-dependent applications, the EER can be as low 0.5% (YOHO database), while text-dependent applications usually have EERs above 2%. In the case of telephone speech, the degradation of the speech quality directly impacts the error rates which then range from 2% EER for speaker verification on 10 digit strings (SESP database) to about 10% on conversational speech (Switchboard).

6. EXTENSIONS OF SPEAKER VERIFICATION

Speaker verification supposes that training and test are composed of monospeaker records. However, it is necessary for some applications to detect the presence of a given speaker within multispeaker audio streams. In this case, it may also be relevant to determine who is speaking when. To handle this kind of tasks, several extensions of speaker verification to multispeaker case have been defined. The three most common ones are briefly described below.

- (i) The n -speaker detection is similar to speaker verification [49]. It consists in determining whether a target speaker speaks in a conversation involving two speakers or more. The difference from speaker verification is that the test recording contains the whole conversation with utterances from various speakers [50, 51].
- (ii) Speaker tracking [49] consists in determining if and when a target speaker speaks in a multispeaker record. The additional work as compared to the n -speaker detection is to specify the target speaker speech segments (begin and end times of each speaker utterance) [51, 52].
- (iii) Segmentation is close to speaker tracking except that no information is provided on speakers. Neither speaker training data nor speaker ID is available. The number of speakers is also unknown. Only test data is available. The aim of the segmentation task is to determine the number of speakers and when they speak [53, 54, 55, 56, 57, 58, 59]. This problem corresponds to a blind classification of the data. The result of the segmentation is a partition in which every class is composed of segments of one speaker.

In the n -speaker detection and speaker tracking tasks as described above, the multispeaker aspect concerned only the test records. Training records were supposed to be monospeaker. An extension of those tasks consists in having multispeaker records for training too, with the target speaker speaking in all these records. The training phase then gets more complex, requiring speaker segmentation of the training records to extract information relevant to the target speaker.

Most of those tasks, including speaker verification, were proposed in the NIST SRE campaigns to evaluate and compare performance of speaker recognition methods for mono- and multispeaker records (test and/or training). While the set of proposed tasks was initially limited to speaker verification task in monospeaker records, it has been enlarged over the years to cover common problems found in real-world applications.

7. APPLICATIONS OF SPEAKER VERIFICATION

There are many applications to speaker verification. The applications cover almost all the areas where it is desirable to secure actions, transactions, or any type of interactions by identifying or authenticating the person making the transaction. Currently, most applications are in the banking

and telecommunication areas. Since the speaker recognition technology is currently not absolutely reliable, such technology is often used in applications where it is interesting to diminish frauds but for which a certain level of fraud is acceptable. The main advantages of voice-based authentication are its low implementation cost and its acceptability by the end users, especially when associated with other vocal technologies.

Regardless of forensic applications, there are four areas where speaker recognition can be used: access control to facilities, secured transactions, over a network (in particular, over the telephone), structuring audio information, and games. We briefly review those various families of applications.

7.1. On-site applications

On-site applications regroup all the applications where the user needs to be in front of the system to be authenticated. Typical examples are access control to some facilities (car, home, warehouse), to some objects (locksmith), or to a computer terminal. Currently, ID verification in such context is done by mean of a key, a badge or a password, or personal identification number (PIN).

For such applications, the environmental conditions in which the system is used can be easily controlled and the sound recording system can be calibrated. The authentication can be done either locally or remotely but, in the last case, the transmission conditions can be controlled. The voice characteristics are supplied by the user (e.g., stored on a chip card). This type of application can be quite dissuasive since it is always possible to trigger another authentication mean in case of doubt. Note that many other techniques can be used to perform access control, some of them being more reliable than speaker recognition but often more expensive to implement. There are currently very few access control applications developed, none on a large scale, but it is quite probable that voice authentication will increase in the future and will find its way among the other verification techniques.

7.2. Remote applications

Remote applications regroup all the applications where the access to the system is made through a remote terminal, typically a telephone or a computer. The aim is to secure the access to reserved services (telecom network, databases, web sites, etc.) or to authenticate the user making a particular transaction (e-trade, banking transaction, etc.). In this context, authentication currently relies on the use of a PIN, sometimes accompanied by the identification of the remote terminal (e.g., caller's phone number).

For such applications, the signal quality is extremely variable due to the different types of terminals and transmission channels, and can sometimes be very poor. The vocal characteristics are usually stored on a server. This type of applications is not very dissuasive since it is nearly impossible to trace the impostor. However, in case of doubt, a human interaction is always possible. Nevertheless, speaker verification remains the most natural user verification modality in this case and the easiest one to implement, along with PIN

codes, since it does not require any additional sensors. Some commercial applications in the banking and telecommunication areas are now relying on speaker recognition technology to increase the level of security in a way transparent to the user. The application profile is usually designed to reduce the number of frauds. Moreover, speaker recognition over the phone complements nicely voice-driven applications from the technological and ergonomic point of views.

7.3. Information structuring

Organizing the information in audio documents is a third type of applications where speaker recognition technology is involved. Typical examples of the applications are the automatic annotation of audio archives, speaker indexing of sound tracks, and speaker change detection for automatic subtitling. The need for such applications comes from the movie industry and from the media related industry. However, in a near future, the information structuring applications should expand to other areas, such as automatic meeting recording abstracting.

The specificities of those types of applications are worth mentioning and, in particular, the huge amount of training data for some speakers and the fact that the processing time is not an issue, thus making possible the use of multipass systems. Moreover, the speaker variability within a document is reduced. However, since speaker changes are not known, the verification task goes along with a segmentation task eventually complicated by the fact that the number of speakers is not known and several persons may speak simultaneously. This application area is rapidly growing, and in the future, browsing an audio document for a given program, a given topic, or a given speaker will probably be as natural as browsing textual documents is today. Along with speech/music separation, automatic speech transcription, and keyword and key sound spotting, speaker recognition is a key technology for audio indexing.

7.4. Games

Finally, another application area, rarely explored so far, is games: child toys, video games, and so forth. Indeed, games evolve toward a better interactivity and the use of player profiles to make the game more personal. With the evolution of computing power, the use of the vocal modality in games is probably only a matter of time. Among the vocal technologies available, speaker recognition certainly has a part to play, for example, to recognize the owner of a toy, to identify the various speakers, or even to detect the characteristics or the variations of a voice (e.g., imitation contest). One interesting point with such applications is that the level of performance can be a secondary issue since an error has no real impact. However, the use of speaker recognition technology in games is still a prospective area.

8. ISSUES SPECIFIC TO THE FORENSIC AREA

8.1. Introduction

The term "forensic acoustics" has been widely used regarding police, judicial, and legal use of acoustics samples. This

wide area includes many different tasks, some of them being recording authentication, voice transcription, specific sound characterization, speaker profiling, or signal enhancement. Among all these tasks, forensic speaker recognition [60, 61, 62, 63, 64] stands out as far as it constitutes one of the more complex problems in this domain: the fact of determining whether a given speech utterance has been produced by a particular person. In this section, we will focus on this item, dealing with forensic conditions and speaker variability, forensic recognition in the past (speaker recognition by listening (SRL), and “voiceprint analysis”), and semi- and fully-automatic forensic recognition systems, discussing also the role of the expert in the whole process.

8.2. Forensic conditions and speaker variability

In forensic speaker recognition, the disputed utterance, which constitutes the evidence, is produced in crime perpetration under realistic conditions. In most of the cases, this speech utterance is acquired by obtaining access to a telephone line, mainly in two different modalities: (i) an anonymous call or, when known or expected, (ii) a wiretapping process by police agents.

“Realistic conditions” is used here as an opposite term to “laboratory conditions” in the sense that no control, assumption, or forecast can be made with respect to acquisition conditions. Furthermore, the perpetrator is not a collaborative partner, but rather someone trying to impede that any finding derived from these recordings could help to convict him.

Consequently, these “realistic conditions” impose on speech signals a high degree of variability. All these sources of variability can be classified [65] as follows:

- (i) peculiar intraspeaker variability: type of speech, gender, time separation, aging, dialect, sociolect, jargon, emotional state, use of narcotics, and so forth;
- (ii) forced intraspeaker variability: Lombard effect, external-influenced stress, and cocktail-party effect;
- (iii) channel-dependent external variability: type of handset and/or microphone, landline/mobile phone, communication channel, bandwidth, dynamic range, electrical and acoustical noise, reverberation, distortion, and so forth.

Forensic conditions will be reached when these variability factors that constitute the so-called “realistic conditions” emerge without any kind of principle, rule, or norm. So they might be present constantly on a call, or else arise and/or disappear suddenly, so affecting in a completely unforeseeable manner the whole process.

The problem will worsen if we consider the effect of these variability factors in the comparative analysis between the disputed utterances and the undisputed speech controls. Factors like time separation, type of speech, emotional state, speech duration, transmission channel, or recording equipment employed acquire—under these circumstances—a pre-eminent role.

8.3. Forensic recognition in the past decades

Speaker recognition by listening

Regarding SRL [63, 66], the first distinctive issue to consider makes reference to the condition of familiar or unfamiliar voices. Human beings show high recognition abilities with respect to well-known familiar voices, in which a long-term training process has been unconsciously accomplished. In this case, even linguistic variability (at prosodic, lexical, grammatical, or idiolectal levels) can be comprised within these abilities. The problem here arises when approaching the forensic recognition area in which experts always deal with unfamiliar voices. Since this long-term training cannot be easily reached even if enough speech material and time are available, expert recognition abilities in the forensic field will be affected by this lack.

Nevertheless, several conventional procedures have been traditionally established in order to perform forensic SRL-based procedures, depending upon the condition (expert/nonexpert) of the listener, namely,

- (1) by nonexperts: regarding nonexperts, which in the case of forensic cases include victims and witnesses, SRL refer to voice lineups. Many problems arise with these procedures, for both speakers and listeners, like size, auditory homogeneity, age, and sex; quantity of speech heard; and time delay between disputed and lineup utterances. Consequently, SRL by nonexperts is given just an indicative value, and related factors, like concordance with eyewitness, become key issues;
- (2) by experts: SRL by experts is a combination of two different approaches, namely,
 - (i) aural-perceptual approach which constitutes a detailed auditory analysis. This approach is organized in levels of speaker characterization, and within each level, several parameters are analyzed:
 - (a) voice characterization: pitch, timbre, fullness, and so forth;
 - (b) speech characterization: articulation, diction, speech rate, intonation, defects, and so forth;
 - (c) language characterization: dynamics, prosody, style, sociolect, idiolect, and so forth;
 - (ii) phonetic-acoustic approach which establishes a more precise and systematic computer-assisted analysis of auditory factors:
 - (a) formants: position, bandwidth, and trajectories;
 - (b) spectral energy, pitch, and pitch contour;
 - (c) time domain: duration of segments, rhythm, and jitter (interperiod short-term variability).

“Voiceprint” analysis and its controversy

Spectrographic analysis was firstly applied to speaker recognition by Kersta, in 1962 [67], giving rise to the term “voiceprint.” Although he gave no details about his research tests and no documentation for his claim (“My claim to voice pattern uniqueness then rests on the improbability that two speakers would have vocal cavity dimensions and articulator use-patterns identical enough to confound voiceprint

identification methods”), he ensured that the decision about the uniqueness of the “voiceprint” of a given individual could be compared, in terms of confidence, to fingerprint analysis.

Nevertheless, in 1970, Bolt et al. [68] denied that voiceprint analysis in forensic cases could be assimilated to fingerprint analysis, adducing that the physiological nature of fingerprints is clearly differentiated from the behavioral nature of speech (in the sense that speech is just a product of an underlying anatomical source, namely, the vocal tract); so speech analysis, with its inherent variability, cannot be reduced to a static pattern matching problem. These dissimilarities introduce a misleading comparison between fingerprint and speech, so the term voiceprint should be avoided. Based in this, Bolt et al. [69] declared that voiceprint comparison was closer to aural discrimination of unfamiliar voices than to fingerprint discrimination.

In 1972, Tosi et al. [70] tried to demonstrate the reliability of voiceprint technique by means of a large-scale study in which they claimed that the scientific community had accepted the method by concluding that “if trained voiceprint examiners use listening and spectrogram they would achieve lower error rates in real forensic conditions than the experimental subjects did on laboratory conditions.”

Later on, in 1973, Bolt et al. [69] invalidated the preceding claim, as the method showed lack of scientific basis, specifically in practical conditions, and, in any case, real forensic conditions would decrease results with respect to those obtained in the study.

At the request of the FBI, and in order to solve this controversy, the National Academy of Sciences (NAS) authorized in 1976 the realization of a study. The conclusion of the committee was clear—the technical uncertainties were significant and forensic applications should be allowed with the utmost caution. Although forensic practice based on voiceprint analysis has been carried out since then [71]; from a scientific point of view, the validity and usability of the method in the forensic speaker recognition has been clearly set under suspect, as the technique is, as stated in [72], “subjective and not conclusive Consistent error rates cannot be obtained across different spectrographic studies.” And, due to lack of quality, about 65% of the cases in a survey of 2,000 [71] remain inadequate to conduct voice comparisons.

8.4. Automatic speaker recognition in forensics

Semiautomatic systems

Semiautomatic systems refer to systems in which a supervised selection of acoustic phonetic events, on the complete speech utterance, has to be accomplished prior to the computer-based analysis of the selected segment.

Several systems can be found in the literature [66], the most outstanding are the following: (i) SASIS [73], semiautomatic speaker identification system, developed by Rockwell International in the USA; (ii) AUROS [74], automatic recognition of speaker by computer, developed jointly by Philips GmbH and BundesKriminalamt (BKA) in Germany; (iii) SAUSI [75], semiautomatic speaker identification system, developed by the University of Florida; (iv) CAVIS [76],

computer assisted voice identification system, developed by Los Angeles County Sheriff’s Department, from 1985; or (v) IDEM [77], developed by Fondazione Ugo Bordoni in Rome, Italy.

Most of these systems require specific use by expert phoneticians (in order to select and segment the required acoustic phonetic events) and, therefore, suffer a lack of generalization in their operability; moreover, many of them have been involved in projects already abandoned by scarceness of results in forensics.

Automatic speaker recognition technology

As it is stated in [72], “automatic speaker recognition technology appears to have reached a sufficient level of maturity for realistic application in the field of forensic science.” State-of-the-art speaker recognition systems, widely described in this contribution, provide a fully automated approach, handling huge quantities of speech information at a low-level acoustic signal processing [78, 79, 80]. Modern speaker recognition systems include features as mel frequency cepstral coefficients (MFCC) parameterization in the cepstral domain, cepstral mean normalization (CMN) or RASTA channel compensation, GMM modeling, MAP adaptation, UBM normalization, or score distribution normalization.

Regarding speaker verification (the authentication problem), the system is producing binary decisions as outputs (accepted versus rejected), and the global performance of the system can be evaluated in terms of false acceptance rates (FARs) versus miss or false rejection rates (FRRs), shown in terms of DET plots. This methodology perfectly suits the requirements of commercial applications of speaker recognition technology, and has led to multiple implementations of it.

Forensic methodology

Nevertheless, regarding forensic applicability of speaker recognition technology and, specially, when compared with commercial applications, some crucial questions arise concerning the role of the expert.

- (i) Provided that the state-of-the-art recognition systems under forensic conditions produce nonzero errors, what is the real usability of them in the judicial process?
- (ii) Is acceptance/rejection (making a decision) the goal of forensic expertise? If so, what is the role of judge/jury in a voice comparison case?
- (iii) How can the expert take into account the prior probabilities (circumstances of the case) in his/her report?
- (iv) How can we quantify the human cost related with FAR (innocent convicted) and with FRR (guilty freed)?

These and other related questions have led to diverse interpretation of the forensic evidence [81, 82, 83, 84]. In the field of forensic speaker recognition, some alternatives to the direct commercial interpretation of scores have been recently proposed.

- (i) Confidence measure of binary decisions: this implies that for every verification decision, a measure of confidence of that decision is addressed. A practical implementation of this approach is the forensic automatic speaker recognition (FASR) system [72], developed at the FBI, based on standard speaker verification processing, and producing as an output, together with the normalized log LR score of the test utterance with respect to a given model, a confidence measurement associated with each recognition decision (accepted/rejected). This confidence measure is based on an estimate of the posterior probability for a given set of conditional testing conditions, and normalizes the score to a range from 0 to 100.
- (ii) Bayesian approach through LR of opposite hypothesis: Bayesian approach posterior odds (a posteriori probability ratio)—assessments pertaining only to the court—are computed from prior odds (a priori probability ratio)—circumstances related with evidence—and LR (ratio between likelihood of evidence compared with H0 and likelihood of evidence compared with H1)—computed by expert [62]. In this approach, H0 stands for positive hypothesis (the suspected speaker is the source of the questioned recording), while H1 stands for the opposite hypothesis (the suspected speaker is not the source of the questioned recording). The application of this generic forensic approach to the specific field of forensic speaker recognition can be found in [85, 86] in terms of Tippett plots [87] (derived from standard forensic interpretation of DNA analysis); and a practical implementation as a complete system of the LR approach, denoted as *IdentiVox* [64], (developed in Spain by Universidad Politécnica de Madrid and Dirección General de la Guardia Civil) has shown to have encouraging results in real forensic approaches.

8.5. Conclusion

Forensic speaker recognition is a multidisciplinary field in which diverse methodologies coexist, and subjective heterogeneous approaches are usually found between forensic practitioners; although technical invalidity of some of these methods has been clearly stated, they are still used by several gurus in unscientific traditional practices. In this context, the emergence of automatic speaker recognition systems, producing robust objective scoring of disputed utterances, constitutes the milestone of forensic speaker recognition. This does not imply that all problems in the field are positively solved, as issues like availability of real forensic speech databases, forensic-specific evaluation methodology, or role of the expert are still open; but definitively, they have made possible a common-framework unified technical approach to the problem.

9. CONCLUSION AND FUTURE RESEARCH TRENDS

In this paper, we have proposed a tutorial on text-independent speaker verification. After describing the training

and test phases of a general speaker verification system, we detailed the cepstral analysis, which is the most commonly used approach for speech parameterization. Then, we explained how to build a speaker model based on a GMM approach. A few speaker modeling alternatives have been mentioned, including neural network and SVMs. The score normalization step has then been described in details. This is a very important step to deal with real-world data. The evaluation of a speaker verification system has then been exposed, including how to plot a DET curve. Several extensions of speaker verification have then been enumerated, including speaker tracking and segmentation by speakers. A few applications have been listed, including on-site applications, remote applications, applications relative to structuring audio documents, and games. Issues specific to the forensic area have then been explored and discussed.

While it is clear that speaker recognition technology has made tremendous strides forward since the initial work in the field over 30 years ago, future directions in speaker recognition technology are not totally clear, but some general observations can be made. From numerous published experiments and studies, the largest impediment to widespread deployment of speaker recognition technology and a fundamental research challenge is the lack of robustness to channel variability and mismatched conditions, especially microphone mismatches. Since most systems rely primarily on acoustic features, such as spectra, they are too dependent on channel information and it is unlikely that new features derived from the spectrum will provide large gains since the spectrum is obviously highly affected by channel/noise conditions. Perhaps a better understanding of specific channel effects on the speech signal will lead to a decoupling of the speaker and channel thus allowing for better features and compensation techniques. In addition, there are several other levels of information beyond raw acoustics in the speech signal that convey speaker information. Human listeners have a relatively keen ability to recognize familiar voices which points to exploiting long-term speaking habits in automatic systems. While this seems a rather daunting task, the incredible and sustained increase in computer power and the emergence of better speech processing techniques to extract words, pitch, and prosody measures make these high-level information sources ripe for exploitation. The real breakthrough is likely to be in using features from the speech signal to learn about higher-level information not currently found in and complementary to the acoustic information. Exploitation of such high-level information may require some form of event-based scoring techniques, since higher-levels of information, such as indicative word usage, will not likely occur regularly as acoustic information does. Further, fusion of systems will also be required to build on a solid baseline approach and provide the best attributes of different systems. Successful fusion will require ways to adjudicate between conflicting signals and to combine systems producing continuous scores with systems producing event-based scores.

Below are some of the emerging trends in speaker recognition research and development.

Exploitation of higher levels of information

In addition to the low-level spectrum features used by current systems, there are many other sources of speaker information in the speech signal that can be used. These include idiolect (word usage), prosodic measures, and other long-term signal measures. This work will be aided by the increasing use of reliable speech recognition systems for speaker recognition R&D. High-level features not only offer the potential to improve accuracy, they may also help improve robustness since they should be less affected by channel effects. Recent work at the JHU SuperSID workshop has shown that such levels of information can indeed be exploited and used profitably in automatic speaker recognition systems [24].

Focus on real-world robustness

Speaker recognition continues to be data driven, setting the lead among other biometrics in conducting benchmark evaluations and research on realistic data. The continued ease of collecting and making available speech from real applications means that researchers can focus on more real-world robustness issues that appear. Obtaining speech from a wide variety of handsets, channels, and acoustic environments will allow examination of problem cases and development and application of new or improved compensation techniques. Making such data widely available and used in evaluations of systems, like the NIST evaluations, will be a major driver in propelling the technology forward.

Emphasis on unconstrained tasks

With text-dependent systems making commercial headway, R&D effort will shift to more difficult issues in unconstrained situations. This includes variable channels and noise conditions, text-independent speech, and the tasks of speaker segmentation and indexing of multispeaker speech. Increasingly, speaker segmentation and clustering techniques are being used to aid in adapting speech recognizers and for supplying metadata for audio indexing and searching. This data is very often unconstrained and may come from various sources (e.g., broadcast news audio with correspondents in the field).

REFERENCES

- [1] R. N. Bracewell, *The Fourier Transform and Its Applications*, McGraw-Hill, New York, NY, USA, 1965.
- [2] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.
- [3] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking," in *Proc. of the Symposium on Time Series Analysis*, M. Rosenblatt, Ed., pp. 209–243, John Wiley & Sons, New York, NY, USA, 1963.
- [4] A. V. Oppenheim and R. W. Schaffer, "Homomorphic analysis of speech," *IEEE Transactions on Audio and Electroacoustics*, vol. 16, no. 2, pp. 221–226, 1968.
- [5] G. Fant, *Acoustic Theory of Speech Production*, Mouton, The Hague, The Netherlands, 1970.
- [6] D. Petrovska-Delacrétaz, J. Cernocky, J. Hennebert, and G. Chollet, "Segmental approaches for automatic speaker verification," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 198–212, 2000.
- [7] S. Furui, "Comparison of speaker recognition methods using static features and dynamic features," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 29, no. 3, pp. 342–350, 1981.
- [8] R. B. Dunn, D. A. Reynolds, and T. F. Quatieri, "Approaches to speaker detection and tracking in conversational speech," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 93–112, 2000.
- [9] A. Higgins, L. Bahler, and J. Porter, "Speaker verification using randomized phrase prompting," *Digital Signal Processing*, vol. 1, no. 2, pp. 89–106, 1991.
- [10] A. E. Rosenberg, J. DeLong, C.-H. Lee, B.-H. Juang, and F. K. Soong, "The use of cohort normalized scores for speaker verification," in *Proc. International Conf. on Spoken Language Processing (ICSLP '92)*, vol. 1, pp. 599–602, Banff, Canada, October 1992.
- [11] D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," *Speech Communication*, vol. 17, no. 1–2, pp. 91–108, 1995.
- [12] T. Matsui and S. Furui, "Similarity normalization methods for speaker verification based on a posteriori probability," in *Proc. 1st ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*, pp. 59–62, Martigny, Switzerland, April 1994.
- [13] M. Carey, E. Parris, and J. Bridle, "A speaker verification system using alpha-nets," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '91)*, vol. 1, pp. 397–400, Toronto, Canada, May 1991.
- [14] D. A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification," in *Proc. 5th European Conference on Speech Communication and Technology (Eurospeech '97)*, vol. 2, pp. 963–966, Rhodes, Greece, September 1997.
- [15] T. Matsui and S. Furui, "Likelihood normalization for speaker verification using a phoneme- and speaker-independent model," *Speech Communication*, vol. 17, no. 1–2, pp. 109–116, 1995.
- [16] A. E. Rosenberg and S. Parthasarathy, "Speaker background models for connected digit password speaker verification," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '96)*, vol. 1, pp. 81–84, Atlanta, Ga, USA, May 1996.
- [17] L. P. Heck and M. Weintraub, "Handset-dependent background models for robust text-independent speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '97)*, vol. 2, pp. 1071–1074, Munich, Germany, April 1997.
- [18] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.
- [19] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, NY, USA, 1973.
- [20] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech, and Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.
- [21] D. A. Reynolds, *A Gaussian mixture modeling approach to text-independent speaker identification*, Ph.D. thesis, Georgia Institute of Technology, Atlanta, Ga, USA, September 1992.
- [22] M. Newman, L. Gillick, Y. Ito, D. McAllister, and B. P. P. "Speaker verification through large vocabulary continuous speech recognition," in *Proc. International Conf. on Spoken Language Processing (ICSLP '96)*, vol. 4, pp. 2419–2422, Philadelphia, Pa, USA, October 1996.
- [23] M. Schmidt and H. Gish, "Speaker identification via support vector classifiers," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '96)*, vol. 1, pp. 105–108, Atlanta, Ga, USA, May 1996.

- [24] *SuperSID Project at the JHU Summer Workshop, July-August 2002*, <http://www.clsp.jhu.edu/ws2002/groups/supersid>.
- [25] J.-L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. Speech, and Audio Processing*, vol. 2, no. 2, pp. 291–298, 1994.
- [26] J. Hertz, A. Krogh, and R. J. Palmer, *Introduction to the Theory of Neural Computation*, Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, Reading, Mass, USA, 1991.
- [27] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan, New York, NY, USA, 1994.
- [28] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- [29] H. Bourslard and C. J. Wellekens, "Links between Markov models and multilayer perceptrons," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 12, pp. 1167–1178, 1990.
- [30] M. D. Richard and R. P. Lippmann, "Neural network classifiers estimate Bayesian a posteriori probabilities," *Neural Computation*, vol. 3, no. 4, pp. 461–483, 1991.
- [31] J. Oglesby and J. S. Mason, "Optimization of neural models for speaker identification," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '90)*, vol. 1, pp. 261–264, Albuquerque, NM, USA, April 1990.
- [32] Y. Bennani and P. Gallinari, "Connectionist approaches for automatic speaker recognition," in *Proc. 1st ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*, pp. 95–102, Martigny, Switzerland, April 1994.
- [33] K. R. Farrell, R. Mammone, and K. Assaleh, "Speaker recognition using neural networks and conventional classifiers," *IEEE Trans. Speech, and Audio Processing*, vol. 2, no. 1, pp. 194–205, 1994.
- [34] J. M. Naik and D. Lubenskt, "A hybrid HMM-MLP speaker verification algorithm for telephone speech," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '94)*, vol. 1, pp. 153–156, Adelaide, Australia, April 1994.
- [35] D. J. Sebald and J. A. Bucklew, "Support vector machines and the multiple hypothesis test problem," *IEEE Trans. Signal Processing*, vol. 49, no. 11, pp. 2865–2872, 2001.
- [36] Y. Gu and T. Thomas, "A text-independent speaker verification system using support vector machines classifier," in *Proc. European Conference on Speech Communication and Technology (Eurospeech '01)*, pp. 1765–1769, Aalborg, Denmark, September 2001.
- [37] J. Kharroubi, D. Petrovska-Delacrétaz, and G. Chollet, "Combining GMM's with support vector machines for text-independent speaker verification," in *Proc. European Conference on Speech Communication and Technology (Eurospeech '01)*, pp. 1757–1760, Aalborg, Denmark, September 2001.
- [38] S. Fine, J. Navratil, and R. A. Gopinath, "Enhancing GMM scores using SVM "hints"," in *Proc. 7th European Conference on Speech Communication and Technology (Eurospeech '01)*, Aalborg, Denmark, September 2001.
- [39] X. Dong, W. Zhaohui, and Y. Yingchun, "Exploiting support vector machines in hidden Markov models for speaker verification," in *Proc. 7th International Conf. on Spoken Language Processing (ICSLP '02)*, pp. 1329–1332, Denver, Colo, USA, September 2002.
- [40] K. P. Li and J. E. Porter, "Normalizations and selection of speech segments for speaker recognition scoring," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '88)*, vol. 1, pp. 595–598, New York, NY, USA, April 1988.
- [41] T. Matsui and S. Furui, "Concatenated phoneme models for text-variable speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '93)*, vol. 1, pp. 391–394, Minneapolis, Minn, USA, April 1993.
- [42] G. Gravier and G. Chollet, "Comparison of normalization techniques for speaker recognition," in *Proc. Workshop on Speaker Recognition and its Commercial and Forensic Applications (RLA2C '98)*, pp. 97–100, Avignon, France, April 1998.
- [43] D. A. Reynolds, "The effect of handset variability on speaker recognition performance: experiments on the switchboard corpus," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '96)*, vol. 1, pp. 113–116, Atlanta, Ga, USA, May 1996.
- [44] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification system," *Digital Signal Processing*, vol. 10, no. 1, 2000.
- [45] M. Ben, R. Blouet, and F. Bimbot, "A Monte-Carlo method for score normalization in automatic speaker verification using Kullback-Leibler distances," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '02)*, vol. 1, pp. 689–692, Orlando, Fla, USA, May 2002.
- [46] C. Fredouille, J.-F. Bonastre, and T. Merlin, "Similarity normalization method based on world model and a posteriori probability for speaker verification," in *Proc. European Conference on Speech Communication and Technology (Eurospeech '99)*, pp. 983–986, Budapest, Hungary, September 1999.
- [47] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Proc. European Conference on Speech Communication and Technology (Eurospeech '97)*, vol. 4, pp. 1895–1898, Rhodes, Greece, September 1997.
- [48] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves, an analysis of individual differences in speaker recognition performances in the nist 1998 speaker recognition evaluation," in *Proc. International Conf. on Spoken Language Processing (ICSLP '98)*, Sydney, Australia, December 1998.
- [49] M. Przybocki and A. Martin, "The 1999 NIST speaker recognition evaluation, using summed two-channel telephone data for speaker detection and speaker tracking," in *Proc. European Conference on Speech Communication and Technology (Eurospeech '99)*, vol. 5, pp. 2215–2218, Budapest, Hungary, September 1999.
- [50] J. Koolwaaij and L. Boves, "Local normalization and delayed decision making in speaker detection and tracking," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 113–132, 2000.
- [51] K. Sönmez, L. Heck, and M. Weintraub, "Speaker tracking and detection with multiple speakers," in *Proc. 6th European Conference on Speech Communication and Technology (Eurospeech '99)*, vol. 5, pp. 2219–2222, Budapest, Hungary, September 1999.
- [52] A. E. Rosenberg, I. Magrin-Chagnolleau, S. Parthasarathy, and Q. Huang, "Speaker detection in broadcast speech databases," in *Proc. International Conf. on Spoken Language Processing (ICSLP '98)*, Sydney, Australia, December 1998.
- [53] A. Adami, S. Kajarekar, and H. Hermansky, "A new speaker change detection method for two-speaker segmentation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '02)*, vol. 4, pp. 3908–3911, Orlando, Fla, USA, May 2002.
- [54] P. Delacourt and C. J. Wellekens, "DISTBIC: A speaker based segmentation for audio data indexing," *Speech Communication*, vol. 32, no. 1–2, pp. 111–126, 2000.
- [55] T. Kemp, M. Schmidt, M. Westphal, and A. Waibel, "Strategies for automatic segmentation of audio data," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '00)*, vol. 3, pp. 1423–1426, Istanbul, Turkey, June 2000.

- [56] S. Meignier, J.-F. Bonastre, and S. Igonet, "E-HMM approach for learning and adapting sound models for speaker indexing," in *Proc. 2001: A Speaker Odyssey—The Speaker Recognition Workshop*, pp. 175–180, Crete, Greece, June 2001.
- [57] D. Moraru, S. Meignier, L. Besacier, J.-F. Bonastre, and I. Magrin-Chagnolleau, "The ELISA consortium approaches in speaker segmentation during the NIST 2002 speaker recognition evaluation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '03)*, vol. 2, pp. 89–92, Hong Kong, China, April 2003.
- [58] D. A. Reynolds, R. B. Dunn, and J. J. McLaughlin, "The Lincoln speaker recognition system: NIST EVAL2000," in *Proc. International Conf. on Spoken Language Processing (ICSLP '00)*, vol. 2, pp. 470–473, Beijing, China, October 2000.
- [59] L. Wilcox, D. Kimber, and F. Chen, "Audio indexing using speaker identification," in *Proc. SPIE Conference on Automatic Systems for the Inspection and Identification of Humans*, pp. 149–157, San Diego, Calif, USA, July 1994.
- [60] H. J. Kunzel, "Current approaches to forensic speaker recognition," in *Proc. 1st ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*, pp. 135–141, Martigny, Switzerland, April 1994.
- [61] A. P. A. Broeders, "Forensic speech and audio analysis: the state of the art in 2000 AD," in *Actas del I Congreso Nacional de la Sociedad Española de Acústica Forense*, J. Ortega-García, Ed., pp. 13–24, Madrid, Spain, 2000.
- [62] C. Champod and D. Meuwly, "The inference of identity in forensic speaker recognition," *Speech Communication*, vol. 31, no. 2–3, pp. 193–203, 2000.
- [63] D. Meuwly, "Voice analysis," in *Encyclopaedia of Forensic Sciences*, J. A. Siegel, P. J. Saukko, and G. C. Knupfer, Eds., vol. 3, pp. 1413–1421, Academic Press, NY, USA, 2000.
- [64] J. Gonzalez-Rodriguez, J. Ortega-García, and J.-L. Sanchez-Bote, "Forensic identification reporting using automatic biometric systems," in *Biometrics Solutions for Authentication in an E-World*, D. Zhang, Ed., pp. 169–185, Kluwer Academic Publishers, Boston, Mass, USA, July 2002.
- [65] J. Ortega-García, J. Gonzalez-Rodriguez, and S. Cruz-Llanas, "Speech variability in automatic speaker recognition systems for commercial and forensic purposes," *IEEE Trans. on Aerospace and Electronics Systems*, vol. 15, no. 11, pp. 27–32, 2000.
- [66] D. Meuwly, *Speaker recognition in forensic sciences—the contribution of an automatic approach*, Ph.D. thesis, Institut de Police Scientifique et de Criminologie, Université de Lausanne, Lausanne, Switzerland, 2001.
- [67] L. G. Kersta, "Voiceprint identification," *Nature*, vol. 196, no. 4861, pp. 1253–1257, 1962.
- [68] R. H. Bolt, F. S. Cooper, E. E. David Jr., P. B. Denes, J. M. Pickett, and K. N. Stevens, "Speaker identification by speech spectrograms: A scientists' view of its reliability for legal purposes," *J. Acoust. Soc. Amer.*, vol. 47, pp. 597–612, 1970.
- [69] R. H. Bolt, F. S. Cooper, E. E. David Jr., P. B. Denes, J. M. Pickett, and K. N. Stevens, "Speaker identification by speech spectrograms: Some further observations," *J. Acoust. Soc. Amer.*, vol. 54, pp. 531–534, 1973.
- [70] O. Tosi, H. Oyer, W. Lashbrook, C. Pedrey, and W. Nash, "Experiment on voice identification," *J. Acoust. Soc. Amer.*, vol. 51, no. 6, pp. 2030–2043, 1972.
- [71] B. E. Koenig, "Spectrographic voice identification: A forensic survey," *J. Acoust. Soc. Amer.*, vol. 79, no. 6, pp. 2088–2090, 1986.
- [72] H. Nakasone and S. D. Beck, "Forensic automatic speaker recognition," in *2001: A Speaker Odyssey—The Speaker Recognition Workshop*, pp. 139–142, Crete, Greece, June 2001.
- [73] J. E. Paul et al., "Semi-Automatic Speaker Identification System (SASIS) — Analytical Studies," Final Report C74-11841501, Rockwell International, 1975.
- [74] E. Bunge, "Speaker recognition by computer," *Philips Technical Review*, vol. 37, no. 8, pp. 207–219, 1977.
- [75] H. Hollien, "SAUSI," in *Forensic Voice Identification*, pp. 155–191, Academic Press, NY, USA, 2002.
- [76] H. Nakasone and C. Melvin, "C.A.V.I.S.: (Computer Assisted Voice Identification System)," Final Report 85-IJ-CX-0024, National Institute of Justice, 1989.
- [77] M. Falcone and N. de Sario, "A PC speaker identification system for forensic use: IDEM," in *Proc. 1st ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*, pp. 169–172, Martigny, Switzerland, April 1994.
- [78] S. Furui, "Recent advances in speaker recognition," in *Audio- and Video-Based Biometric Person Authentication*, J. Bigun, G. Chollet, and G. Borgefors, Eds., vol. 1206 of *Lecture Notes in Computer Science*, pp. 237–252, Springer-Verlag, Berlin, 1997.
- [79] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1, pp. 19–41, 2000.
- [80] A. F. Martin and M. A. Przybocki, "The NIST speaker recognition evaluations: 1996–2001," in *2001: A Speaker Odyssey—The Speaker Recognition Workshop*, pp. 39–43, Crete, Greece, June 2001.
- [81] B. Robertson and G. A. Vignaux, *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*, John Wiley & Sons, Chichester, UK, 1995.
- [82] K. R. Foster and P. W. Huber, *Judging Science: Scientific Knowledge and the Federal Courts*, MIT Press, Cambridge, Mass, USA, 1997.
- [83] I. W. Evett, "Towards a uniform framework for reporting opinions in forensic science casework," *Science & Justice*, vol. 38, no. 3, pp. 198–202, 1998.
- [84] C. G. C. Aitken, "Statistical interpretation of evidence/Bayesian analysis," in *Encyclopedia of Forensic Sciences*, J. A. Siegel, P. J. Saukko, and G. C. Knupfer, Eds., vol. 2, pp. 717–724, Academic Press, NY, USA, 2000.
- [85] D. Meuwly and A. Drygajlo, "Forensic speaker recognition based on a Bayesian framework and Gaussian Mixture Modelling (GMM)," in *2001: A Speaker Odyssey—The Speaker Recognition Workshop*, pp. 145–150, Crete, Greece, June 2001.
- [86] J. Gonzalez-Rodriguez, J. Ortega-García, and J.-J. Lucena-Molina, "On the application of the Bayesian approach to real forensic conditions with GMM-based systems," in *2001: A Speaker Odyssey—The Speaker Recognition Workshop*, pp. 135–138, Crete, Greece, June 2001.
- [87] C. F. Tippet, V. J. Emerson, M. J. Fereday, et al., "The evidential value of the comparison of paint flakes from sources other than vehicles," *Journal of the Forensic Science Society*, vol. 8, pp. 61–65, 1968.

Frédéric Bimbot after graduating as a Telecommunication Engineer in 1985 (ENST, Paris, France), he received his Ph.D. degree in signal processing (speech synthesis using temporal decomposition) in 1988. He also obtained his B.A. degree in linguistics (Sorbonne Nouvelle University, Paris III) in 1987. In 1990, he joined CNRS (French National Center for Scientific Research) as a Permanent Researcher, worked with ENST for 7 years, and then moved to IRISA (CNRS & INRIA) in Rennes. He also repeatedly visited AT&T Bell Laboratories



between 1990 and 1999. He has been involved in several European projects: SPRINT (speech recognition using neural networks), SAM-A (assessment methodology), and DiVAN (audio indexing). He has also been the Work-Package Manager of research activities on speaker verification in the projects CAVE, PICASSO, and BANCA. From 1996 to 2000, he has been the Chairman of the Groupe Francophone de la Communication Parlée (now AFCP), and from 1998 to 2003, a member of the ISCA board (International Speech Communication Association, formerly known as ESCA). His research focuses on audio signal analysis, speech modeling, speaker characterization and verification, speech system assessment methodology, and audio source separation. He is heading the METISS research group at IRISA, dedicated to selected topics in speech and audio processing.

Jean-François Bonastre has been an Associate Professor at the LIA, the University of Avignon computer laboratory since 1994. He studied computer science in the University of Marseille and obtained a DEA (Master) in artificial intelligence in 1990. He obtained his Ph.D. degree in 1994, from the University of Avignon, and his HDR (Ph.D. supervision diploma) in 2000, both in computer science, both on speech science, more precisely, on speaker recognition. J.-F. Bonastre is the current President of the AFCP, the French Speaking Speech Communication Association (a Regional Branch of ISCA). He was the Chairman of the RLA2C workshop (1998) and a member of the Program Committee of Speaker Odyssey Workshops (2001 and 2004). J.-F. Bonastre has been an Invited Professor at Panasonic Speech Technology Lab. (PSTL), Calif, USA, in 2002.



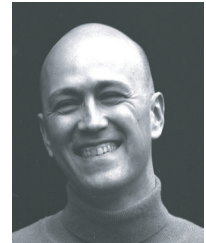
Corinne Fredouille obtained her Ph.D. degree in 2000 in the field of automatic speaker recognition. She has joined the computer science laboratory LIA, University of Avignon, and more precisely, the speech processing team, as an Assistant Professor in 2003. Currently, she is an active member of the European ELISA Consortium, of AFCP, the French Speaking Speech Communication Association, and, of ISCA/SIG SPLC (Speaker and Language Characterization Special Interest Group).



Guillaume Gravier graduated in Applied Mathematics from Institut National des Sciences Appliquées (INSA Rouen) in 1995 and received his Ph.D. degree in signal and image processing from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, in January 2000. Since 2002, he is a Research Fellow at Centre National pour la Recherche Scientifique (CNRS), working at the Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), INRIA, Rennes. His research interests are in the fields of speech recognition, speaker recognition, audio indexing, and multimedia information fusion. Guillaume Gravier also worked on speech synthesis at ELAN Informatique in Toulouse, France, from 1996 to 1997 and on audiovisual speech recognition at IBM Research, NY, USA, from 2001 to 2002.



Ivan Magrin-Chagnolleau received the Engineer Diploma in electrical engineering from the ENSEA, Cergy-Pontoise, France, in June 1992, the M.S. degree in electrical engineering from Paris XI University, Orsay, France, in September 1993, the M.A. degree in phonetics from Paris III University, Paris, France, in June 1996, and the Ph.D. degree in electrical engineering from the ENST, Paris, France, in January 1997. In February 1997, he joined the Speech and Image Processing Services laboratory of AT&T Labs Research, Florham Park, NJ, USA. In October 1998, he visited the Digital Signal Processing Group of the Electrical and Computer Engineering Department at Rice University, Houston, Tex, USA. In October 1999, he went to IRISA (a research institute in computer science and electrical engineering), Rennes, France. From October 2000 to August 2001, he was an Assistant Professor at LIA (the computer science laboratory of the University of Avignon), Avignon, France. In October 2001, he became a Permanent Researcher with CNRS (the French National Center for Scientific Research) and is currently working at the Laboratoire Dynamique Du Langage, one of the CNRS associated laboratories in Lyon, France. He has over 30 publications in the area of audio indexing, speaker characterization, language identification, pattern recognition, signal representations and decompositions, language and cognition, and data analysis, and one US patent in audio indexing. He is an IEEE Senior Member, a Member of the IEEE Signal Processing Society, the IEEE Computer Society, and the International Speech Communication Association (ISCA). <http://www.ddl.ish-lyon.cnrs.fr/membres/imc/index.html>.



Teva Merlin is currently a Ph.D. candidate at the computer science laboratory LIA at the University of Avignon.



Javier Ortega-García received the M.S. degree in electrical engineering (Ingeniero de Telecomunicación), in 1989; and the Ph.D. degree “cum laude” also in electrical engineering (Doctor Ingeniero de Telecomunicación), in 1996, both from Universidad Politécnica de Madrid, Spain. From 1999, he was an Associate Professor at the Audio-Visual and Communications Engineering Department, Universidad Politécnica de Madrid. From 1992 to 1999, he was an Assistant Professor also at Universidad Politécnica de Madrid. His research interests focus on biometrics signal processing: speaker recognition, face recognition, fingerprint recognition, online signature verification, data fusion, and multimodality in biometrics. His interests also span to forensic engineering, including forensic biometrics, acoustic signal processing, signal enhancement, and microphone arrays. He has published diverse international contributions, including book chapters, refereed journal, and conference papers. Dr. Ortega-García has chaired several sessions in international conferences. He has participated in some scientific and technical committees, as in EuroSpeech’95 (where he was also a Technical Secretary), EuroSpeech’01, EuroSpeech’03, and Odyssey’01—The Speaker Recognition Workshop. He has been appointed as General Chair at Odyssey’04—The Speaker Recognition Workshop to be held in Toledo, Spain, in June 2004.

Dijana Petrovska-Delacrétaz obtained her M.S. degree in Physics in 1981 from the Swiss Federal Institute of Technology (EPFL) in Lausanne. From 1982 to 1986, she worked as a Research Assistant at the Polymer Laboratory, EPFL. During a break to raise her son, she prepared her Ph.D. work, entitled “Study of the mechanical properties of healed polymers with different structures,” that she defended in 1990. In 1995, she received a grant for women reinsertion of the Swiss National Science Foundation. That is how she started a new research activity in speech processing at the EPFL-CIRC, where she worked as a Postdoctoral Researcher until 1999. After one year spend as a Consultant in AT&T Speech Research Laboratories and another year in Ecole Nationale Supérieure des Telecommunications (ENST), Paris, she worked as a Senior Assistant at the Informatics Department, Fribourg University (DIUF), Switzerland. Her main research activities are based on applications of data-driven speech segmentation for segmental speaker verification, language identification, and very low-bit speech coding. She has published 20 papers in journals and conferences, and held 3 patents.



Douglas A. Reynolds received the B.E.E. degree and Ph.D. degree in electrical engineering, both from Georgia Institute of Technology. He joined the Information Systems Technology Group at the Massachusetts Institute of Technology Lincoln Laboratory in 1992. Currently, he is a Senior Member of the technical staff and his research interests include robust speaker identification and verification, language recognition, speech recognition, and speech-content-based information retrieval. Douglas has over 40 publications in the area of speech processing and two patents related to secure voice authentication. Douglas is a Senior Member of IEEE Signal Processing Society and has served on the Speech Technical Committee.



3.2 Indexation sonore et multimédia

An Overview of the AT&T Spoken Document Retrieval System (1998)

John Choi, Don Hindle, Julia Hirschberg, Ivan Magrin-Chagnolleau, Christine Nakatani, Fernando Pereira, Amit Singhal, and Steve Whittaker.

An Overview of the AT&T Spoken Document Retrieval System.
Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, United States, December 1998.

7 pages

AN OVERVIEW OF THE AT&T SPOKEN DOCUMENT RETRIEVAL

*John Choi, Don Hindle, Julia Hirschberg, Ivan Magrin-Chagnolleau,
Christine Nakatani, Fernando Pereira, Amit Singhal, Steve Whittaker*

AT&T Labs-Research
180 Park Avenue
Florham Park, NJ 07932-0971

ABSTRACT

We present an overview of a spoken document retrieval system developed at AT&T Labs-Research for the HUB4 Broadcast News corpus. This overview includes a description of the intonational phrase boundary detection, classification, speech recognition, information retrieval and user interface components of the system, along with updated system assessments based on the 49-query task defined for the TREC-6 SDR track. Results from a comparative ranking study, based on queries taken from AP Newswire headlines from the same time period that the Broadcast News corpus was recorded, are presented. For the AP task, retrieval accuracy is assessed by comparing the documents retrieved from ASR generated transcriptions with those from human generated transcriptions.

1. INTRODUCTION

This paper presents an overview of a spoken document retrieval and browsing system developed at AT&T Labs Research. The system was designed for the TREC-6 SDR track [23], which involved a retrieval task consisting of 49 known-item queries submitted over approximately 47 hours of speech from the HUB4 Broadcast News corpus [5]. Automatic speech recognition is used to generate textual transcripts of the HUB4 corpus. Text-based information retrieval techniques are then applied. In addition to supporting research on information retrieval strategies for machine generated speech transcripts, the system is also a testbed for user-interface experiments on intelligent presentation of speech documents to users.

Previous work on spoken document retrieval includes a Video Mail Retrieval system [7,8], radio news broadcast retrieval using subword units [16], a retrieval system for a digital video library [25], a system for Swiss radio news [24], and the systems developed for the TREC-6 SDR track [23], *inter alia*.

2. SYSTEM OVERVIEW

An overview of the system architecture is provided in Figure 1. Speech documents, whose boundaries were prespecified for the

HUB4 retrieval task, are initially processed by an intonational phrase boundary detector. The phrase defined by the boundary detector is then submitted for classification. The classifier provides a hypothesis about the channel conditions in the given phrase. Based on this hypothesis, one of several acoustic models is selected for use in the recognizer. Having generated the transcripts for the speech corpus, an information-retrieval engine indexes the transcripts for retrieval. Boundary detection, classification, recognition and indexing are all conducted off-line. The user-interface currently supports real-time query-submission and retrieval, making calls to the information-retrieval engine which returns a ranked list of hypothesized relevant documents based on the machine generated transcripts. Each of the system components is described in more detail below.

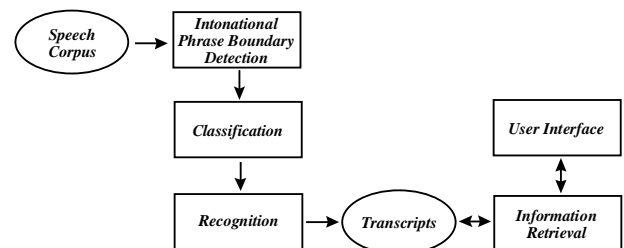


Figure 1: Overview of the spoken document system architecture

2.1. Intonational Phrase Boundary Detection

The speech documents are initially passed through an intonational phrase boundary detector [6]. The detector generates a binary classification for every 10 msec of the speech stream as either (i) occurring within an intonational phrase or (ii) occurring in the break between two intonational phrases. This decision is based on models generated by classification and regression tree techniques [1], which were trained on acoustic vectors consisting of the following parameters: fundamental frequency, RMS energy and autocorrelation-peak. The training corpus consisted of the Boston Directions Corpus [15] which was prosodically transcribed using ToBI conventions [18].

Contiguous frames classified as occurring in the break between intonational phrases are used to define phrase boundaries, with the constraint that an intonational phrase be at least 1 second in duration and no greater than 20 seconds in duration. Selection is biased toward longer phrases, so that if, e.g., a phrase boundary is detected at 5 seconds and another at 19 seconds, the boundary at 5 seconds is ignored in favor of a single larger phrase 19 seconds in length. This boundary detection scheme segmented the speech corpus into 10,480 phrases averaging 14.78 seconds in duration.

The boundary detection serves two purposes. First, the intonational phrases serve as domains over which we can apply one of several acoustic models for recognition. This is discussed in more detail in section 2.2. Second, the frame classifications are also used to identify suitable browsing-units for the user interface of the spoken document retrieval system. This is based on the assumption that presentation of prosodically well-formed segments of speech to the user would be preferable to fixed-size segments which might begin and terminate in the middle of an utterance.

2.2. Classifier

To select the best acoustic model to apply to a given recognition unit, a classifier [13] is used which assigns log-likelihood scores associated with each of four categories, reflecting a four-way partitioning of the training corpus into "high-fidelity", "medium-fidelity", "low-fidelity" and "noise" categories. The high-fidelity partition consists of wideband (0-8kHz) speech recorded in the studio with no background noise. The medium partition also consists of wideband (0-8kHz) speech, but recorded in other recording environments, including field conditions, with no background noise. The low-fidelity partition consists of narrowband speech (0-4kHz) recorded from telephone interviews, again with no background noise. All speech recorded in the presence of background noise is consigned to the noise partition. Partitioning was based on labels provided with the HUB4 training corpus. The purpose of this partitioning was to minimize observed variability in channel conditions when training different acoustic models.

| | D_{HIGH} | D_{MED} | D_{LOW} | D_{NOISE} | CORRECT |
|-------------|------------|-----------|-----------|-------------|---------|
| M_{HIGH} | 123 | 38 | 0 | 15 | 69.88% |
| M_{MED} | 3 | 4 | 0 | 9 | 25.00% |
| M_{LOW} | 10 | 12 | 85 | 0 | 79.44% |
| M_{NOISE} | 44 | 31 | 1 | 167 | 68.72% |

Table 1: Confusion-matrix for the classification models; correct classifications are represented on the diagonal. Each row provides counts of positive classifications by model M_C for each data partition D_C . Accuracy for each model M_C is presented as percent correct.

Classification is based on full covariance Gaussian mixture models [22], initialized using vector quantization [10], and trained using the Expectation-Maximization algorithm [3]. The

training input consists of 31-dimensional vectors of filter-bank coefficients in dB units, derived from Hamming windowed frames of 20 msec with a frame advance rate of 10 msec. The filter-bank coefficients are computed by taking the base 10 logarithm from short term power spectra in the 0-8kHz band from a mel-scaled bank of filters.

Log-likelihood scores are computed for each of the four classification models. Recognition then proceeds with the acoustic model associated with the classification that provides the highest log-likelihood score. When tested on the three hour HUB4 development test partition, overall accuracy for the classifier was assessed at 69.91%, normalizing for the number of observations in each category. Within category accuracy scores, along with a confusion matrix profiling errors, are reported in Table 1.

2.3. Recognizer

The recognizer incorporates a standard time-synchronous beam search algorithm with continuous density, three-state, left-to-right, context-dependent hidden Markov phone models. The models connecting phone HMMs to word sequences are implemented in the general framework of weighted finite-state transducers [14,17]. The probabilities defining the transduction from context-dependent phone sequences to word sequences are estimated on word level grapheme-to-phone mappings generated by a set of heuristics. These mappings are typically one-to-one, although multiple pronunciations are defined for less than 5% of the 29K words considered in the training phase.

Recognition hypotheses are output in the form of word lattices which are derived from model lattices by transducer composition [14]. Given a phone model p for transition a from state s to state s' matching the input from time t to time t' , the corresponding model lattice arc goes from node (s,t) to node (s',t') . In the implementation, for any such t' , only the best t is recorded for reasons of run-time efficiency and lattice size. The resulting lattice reduction does not seem to have negative consequences for recognition accuracy.

Acoustic observations serving as input to the HMMs consist of 39-dimensional vectors taken from 20 msec analysis frames. The frame advance rate is 10 msec. Each acoustic vector contains the first 13 normalized mel-frequency cepstral coefficients, along with their first and second time derivatives.

Two sets of acoustic models have been trained. The first set consists of four acoustic models – AI_{HIGH} , AI_{MED} , AI_{LOW} and AI_{NOISE} – based on the four-way partitioning of the training corpus described in section 2.2. Each of these models was bootstrapped from a single model trained on the channel 1 WSJ training corpus. The second set of models consists of two acoustic models – $A2_{WB}$ and $A2_{NB}$ – trained on wideband speech (the union of the "high-fidelity", "medium-fidelity" and "noise" partitions), and narrowband speech (the "low-fidelity" partition), respectively. The $A2_{WB}$ model was bootstrapped

from $A1_{HIGH}$. In addition, the training corpus for $A2_{WB}$ was expanded to include the “spontaneous speech” partition of the channel 2 WSJ training corpus. This additional training data was weighted by a factor of 0.1 in calculating means and variances; the HUB4 training data was weighted by a factor of 0.9. The WSJ data was not included in defining the context-dependency models. The $A2_{NB}$ model was bootstrapped from a narrowband model trained on the SWITCHBOARD corpus [4].

Training iterations in both sets consisted of eigenvector rotations to decorrelate the training data [11], k-means clustering, normalization of means and variances based on maximum-likelihood, and Viterbi alignment to re-segment the data. The output probability distributions in the HMMs consist of a weighted mixture of Gaussians with diagonal covariance. Each mixture in the $A1$ set of models contains at most 8 components. In the $A2$ set of models, each mixture contains at most 12 components.

2.4. Language Models

Language modeling was based on a 116 million word corpus consisting of text designated for SDR language modeling and the transcriptions from the training partition. The 20K most frequent words in the training corpus comprise our lexicon. The least frequent of the 20K words appear 121 times in the training corpus. In addition, two pseudo words – *pause* and *unknown* – were added to the lexicon. *pause* is freely insertable at any point in recognition, and *unknown* is used for words outside the vocabulary.

Two Markov language models were trained. A standard Katz [9] backoff bigram model was constructed from the 6.1 million bigrams observed in the training corpus. When tested on the retrieval test corpus, this model exhibits an out-of-vocabulary rate of 2.2% and perplexity of 200.

A backoff trigram model was also constructed based on the 9.4 million trigrams observed in the training corpus. This model showed an out-of-vocabulary rate of 2.2% and a perplexity of 144 on the three hour development test corpus. From this model, a more compact trigram language model was constructed following the procedures described in [20]. In particular, trigrams and bigrams were discarded from the model in cases where the difference between the model prediction and the backed-off prediction is less than a threshold T :

$$f * (P_O - P_B) < T$$

where f is the observed n -gram frequency, P_O is the n -gram prediction and P_B is the backed-off $(n-1)$ -gram prediction. In our case, a threshold of 20 was used. This technique reduces the number of trigrams in the model by 85% and the number of bigrams in the model by 83%, with a concomitant increase in perplexity of 18% to 169. Compared to the bigram model, this compacted trigram model has 16% lower perplexity and is 60% smaller.

2.5. Information Retrieval

We use the SMART retrieval system which is a text processing system based on the vector space model [19,2]. SMART automatically generates weighted vectors for any given text using the following indexing scheme:

- Tokenization: The text is first tokenized into individual words and other tokens.
- Stop word removal: Common function words (like *the, of, an, ...*), also called stop words, are removed from this list of tokens. The SMART system uses a predefined list of 571 stop words.
- Stemming: Various morphological variants of a word are normalized to the same *stem* [12]. Usually simple rules for suffix stripping are used in this process.
- Weighting: The term (word) vector, thus created for a text, is weighted using *tf*, *idf*, and length normalization considerations.

We use the *Lnu* term weighting scheme to assign weights to the terms of a document [21]:

$$\frac{(1 + \ln(tf)) / (1 + \ln(\text{average } tf))}{0.8 * \text{pivot} + 0.2 * (\# \text{ of unique terms})}$$

where tf is the number of times a term occurs in the text, and average tf is the average of the tfs of all the terms in a document. The average number of unique terms in a document (computed *across the entire collection*) is used as the *pivot*. The user queries are also indexed using the above steps, and are weighted using *ltn* weights [21]:

$$(1 + \ln(tf)) * idf$$

where tf is once again the frequency of a word in the query and *idf* is $\ln(N/df)$ (N is the total number of documents in the collection, and df is the number of documents that contain the word).

If Q is the query vector and D_i is the vector representation for document- i , a numerical (inner-product) similarity between the query and the document is computed as

$$Sim(Q, D_i) = \sum_{\text{common terms } t_j} q_j * d_{ij}$$

where t_j is a term present in both the query and the document, q_j is the weight of term t_j in the query, and d_{ij} is its weight in document- i . The documents in the information base are ranked by their decreasing similarity to the query and are presented to the user in this order.

2.6. User Interface

The user interface design is based on extensive user testing of a simple browser on a voicemail database and consists of three main components: an Programs Overview window, a Speech

Feedback window and a Player window. These are shown in Figure 1 for the query “whitewater clinton scandal”. Queries are submitted textually at a shell prompt.



Figure 2: The graphical user interface for spoken document retrieval and browsing.

The Programs Overview window presents the speech search results as a relevance ranked list of stories, named by the news program in which they occur. The top 10 most relevant stories are displayed. For each story, the program title (e.g., “NPR All Things Considered”), the date of the program and all instances of keywords in the story that matched the query are displayed. Clicking on one of the program-story buttons loads the corresponding speech into the Speech Feedback window, along with a time-aligned cursor which shows the location of the story that contains the query terms. The Player window provides controls for navigation and play within the program displayed in the Speech Feedback window and includes the following: a play button which plays from the point selected in the Speech Feedback window; a stop-play button; a move-to-beginning button; skip-forward buttons which skip forward an intonational phrase or paratone; skip-back buttons which skip backwards an intonational phrase or paratone. A paratone is a unit larger than the intonational phrase whose pause boundaries are typically longer in duration than those that delimit intonational phrases. Like the intonational phrases, the paratones are automatically detected labeled using the boundary-detection model described in section 2.1.

3. SYSTEM EVALUATION

We evaluate the retrieval effectiveness of our system using two different tasks: the TREC-6 SDR task and an AP Newswire headlines task. Both tasks, described in sections 3.1 and 3.2,

involve two different sets of transcriptions. The first set (*Trans1*) was generated by a recognizer using the $A1$ model set and the bigram language model. Subsequently, another set of transcriptions (*Trans2*) was generated by using the $A2_{WB}$ model, the $A1_{LOW}$ model and the compact trigram language model.¹ The word error rates associated with the two recognizers on the three hour development test partition and the 47 hour test partition are presented in Table 2.

| | Word Error Rate | |
|-------------------|-----------------------------------|-----------------------------------|
| | Recognizer 1 (<i>Trans1</i>) | Recognizer 2 (<i>Trans2</i>) |
| Devtest (3 hours) | 50.50% | 36.64% |
| Test (47 hours) | 42.70% | 30.09% |

Table 2: Word error rates associated with the two different recognizers on both the three hour development-test partition and the 47 hour test partition.

3.1. TREC-6 SDR Task

The first benchmark we use is the TREC-6 SDR task. For this task, each broadcast-show has been manually divided into several documents. There are 49 user queries with each query having a unique “answer document” in the collection of 1452 documents (as indexed by SMART). Two queries, numbered SDR43 and SDR48, have two answer documents in the collection. The aim of the retrieval system is to rank the answer document for a query as high up in rank as possible (in response to the query). This task has also been referred to as “known item searching”. For the two queries with more than one answer document, the rank of the answer document with a better rank (lower absolute rank) is used in the evaluations.

It is generally accepted that this is a relatively easy task compared to other benchmark retrieval tasks. For most of the queries, a typical IR system running on the human transcription of the speech consistently retrieves the answer document within the top few ranks. This pattern holds true even when retrieval is done using machine generated transcription, albeit less accurately. This ease of retrieval makes it difficult to conduct rigorous comparisons of multiple retrieval approaches. Given the ease of retrieval from this collection, using any single evaluation measure does not exemplify meaningful differences between two retrievals. For this reason, we use *five* different evaluation measures in this study to compare the retrieval effectiveness of retrieval from various transcriptions.

Evaluation Measures. The evaluation measure that we use are:

- **E1:** Number of queries for which the answer document is ranked 1. This indicates “perfect” retrieval. High values for this evaluation measure are desired as they

¹ The training of the $A2_{NB}$ model was incomplete at the time the second set of transcripts was generated.

indicate that a retrieval system is doing perfect retrieval for many queries.

- **E2:** *Number of queries for which the answer document is ranked within the top 5.* Perfect retrieval might be too strict an evaluation measure. From a user's perspective, if an answer document is retrieved at rank 2, then the system is still doing a good job. This evaluation measure credits a system if the answer is found within the top five speech documents retrieved.
- **E3:** *Mean answer rank.* This measure uses the answer ranks for all the queries. A system that has a low mean answer rank is better.
- **E4:** *Mean answer rank, after removing outliers.* Since there are only 49 queries, if one of the answers is retrieved at a poor rank, the *mean answer rank* measure for the entire system suffers noticeably. For example, if one of the answers is ranked 400 for a system, the *mean answer rank* of the system falls by almost 8. For this reason, we allow each system to ignore one to two of its worst queries for which the answer is retrieved at a very poor rank when computing the *mean answer rank*.
- **E5:** *Mean reciprocal rank.* This is the known item search variant of the non-interpolated average precision score. Its value is computed as:

$$\frac{\sum_{i=1}^{49} (1 / \text{answer rank for query}_i)}{49}$$

49

Higher values of this measure are better. We must point out that this measure is heavily governed by an answer's rank. For example, if the answer is at rank 1, this measure assigns a credit of 1.0 to the system; whereas if the answer is ranked 2, the credit assigned is just half. It is unclear if, from a user's perspective, ranking an answer at rank 1 is 100% better than ranking the answer at rank 2. On the other hand, this measure makes very little distinction between an answer ranked at low ranks (1/100 is not much different from 1/300 in absolute terms).

If one retrieval run consistently outperforms another on most of the above measures, then, despite the shortcomings of this task, we have strong reason to believe that the better retrieval run is indeed superior.

Results. We performed retrieval using the 49 queries on the first transcription set *Trans1*, the second transcription set *Trans2* and the human transcriptions *Human*. The results are shown in Table 3. The -1 in the **E4** row indicates that each run was allowed to skip its one worst query. Similarly the -2 in the very next row indicates that the two worst queries were skipped for each run. The numbers in the parentheses represent improvement in retrieval (in the case of positive numbers) as compared to the baseline performance associated with *Trans1*.

Table 3 shows that retrieval on *Trans2* outperforms retrieval on

Trans1 for each and every evaluation measure. Using *Trans2*, e.g., 32 out of the 49 queries get their answer document retrieved at rank 1, as opposed to just 29 for *Trans1*. Similar improvements are obtained for other measures. These results are reassuring as they indicate that when a speech recognizer improves, a speech retrieval system should improve as well. As expected, retrieval from recognizer generated transcripts is still poorer than retrieval from human generated transcripts.

| | Trans1 | Trans2 | Human |
|-----------|---------------|---------------|--------------|
| E1 | 29 | 32 | 35 |
| | -- | (+3) | (+6) |
| E2 | 37 | 39 | 45 |
| | -- | (+2) | (+8) |
| E3 | 25.80 | 10.39 | 7.39 |
| | -- | (+15.41) | (+18.41) |
| E4 | 9.83 | 6.10 | 2.63 |
| (-1) | -- | (+3.73) | (+7.20) |
| E4 | 6.06 | 5.21 | 2.00 |
| (-2) | -- | (+0.81) | (+4.06) |
| E5 | 0.6703 | 0.7149 | 0.8020 |
| | -- | (+6.65%) | (+19.65%) |

Table 3: Retrieval results for TREC-6 SDR task.

3.2. AP Newswire Queries

To exercise the system and expand our evaluation protocol, we devised another task to compare the retrieval from the three transcription sets. We selected 94 AP Newswire headlines, published between May 10, 1996 and June 20, 1996, the same period from which the broadcast news shows were recorded. We use these headlines, which are typically short, as potential user queries for retrieval. However, since we do not have relevance assessments for these headlines, we cannot use the standard IR evaluation measures (such as average precision) to evaluate our retrieval. Since most speech retrieval work aspires to achieve results comparable to the results of doing retrieval from the human transcription of speech, in this task, we use the ranking for the human transcriptions as a "gold standard" and evaluate how close retrieval from speech comes to this gold standard.

Evaluation Measures and Results. As a first test, we compare the average number of common documents retrieved per query (within the top K ranks) from the human transcriptions and the machine transcriptions. A higher average number of common documents would signify a higher correlation between retrievals from the *Human* set of transcriptions and those from the *Trans1* and *Trans2* sets. This, in turn, would suggest that the corresponding speech retrieval run is more effective.

We test this measure for the K values 1, 5, 10, 20, 30, 50, and 100. The results are shown in Figure 2. We observe that retrieval from *Trans2* is more closely correlated to the human transcription than retrieval from *Trans1* for all K ranks on this

evaluation measure. On average, *Trans2* retrieves about 75% more in common to *Human* whereas *Trans1* only retrieves about 70%.

We use the following additional test to compare rankings from the human and the machine transcriptions. We assume that the top K documents as retrieved from the human transcriptions are relevant for the corresponding query. Given this assumption, we can compute the average precision for retrieval from the machine transcriptions. Note that by this assumption, retrieval from human transcriptions always has an average precision of 100%.

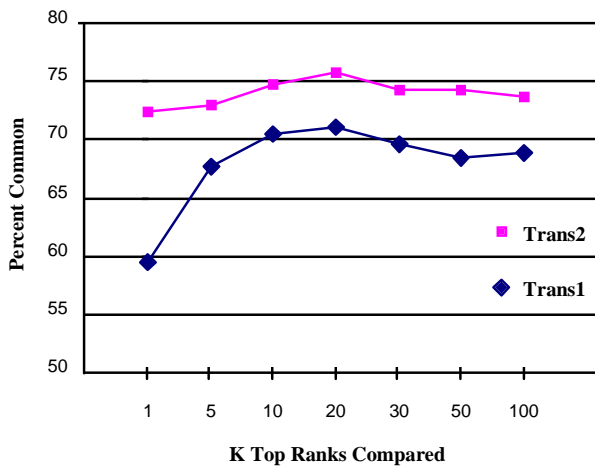


Figure 3: Percent of common documents in top K ranks.

Once again, we test this measure for the K values 1, 5, 10, 20, 30, 50, and 100. The results, shown in Figure 3, again confirm that retrieval from *Trans2* is more effective for all K ranks, as measured by average precision.

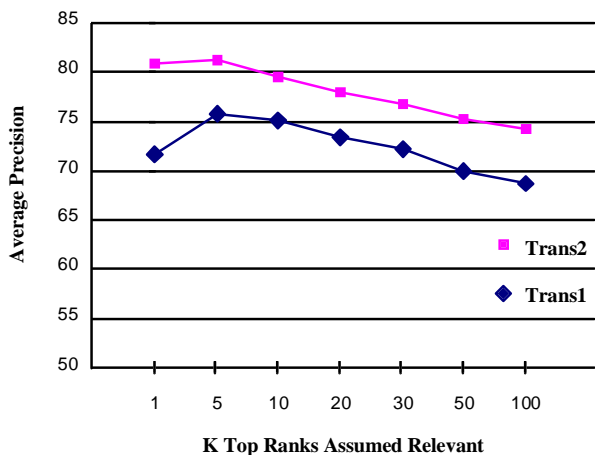


Figure 4: Average precision for retrieval from machine generated transcriptions if the top K ranks of human generated transcriptions are assumed relevant.

4. CONCLUSIONS

We have provided an overview of a spoken document retrieval system designed for the HUB4 corpus and presented experimental results which provide some metric of retrieval efficiency. Two general conclusions can be drawn from these results. First, spoken document retrieval is a tractable problem. As shown in Figures 3 and 4, retrieval from automatic transcriptions is about 70-80% as effective as retrieval from human transcriptions. This number is in agreement with other studies that have compared the retrieval effectiveness on both automatically generated transcriptions and human generated transcriptions [8,25].

The second conclusion we draw is simply that better recognition contributes to better retrieval. While this seems intuitively transparent, we are pleased that our empirical findings support this hypothesis. The relatively small size of the HUB4 corpus for information retrieval purposes initially forced us to interpret our early results with some skepticism. However, with the introduction of the various other evaluation measures outlined in this paper, and the consistency with which the trends were evident along all these measures, we find that speech retrieval systems do stand to gain from better recognition.

We are encouraged by these results and continue to explore various approaches to increase retrieval efficiency and to resolve the ‘out-of-vocabulary’ problem, which was not addressed in this paper. Preliminary results from work with word lattices which provide multiple recognition hypotheses suggest that the use of lattices increases word recall, albeit with noticeable cost in word precision, and that this increase results in better retrieval. We have also conducted experiments exploring the use of subword units and hope to expand this line of research in the coming months. The system is also being utilized to explore different user interfaces to maximize the utility of spoken document retrieval for browsing broadcast news recordings as well as other application domains.

ACKNOWLEDGEMENTS

We wish to thank Andrej Ljolje, Mehryar Mohri and Michael Riley for their help in the development of the recognizer.

REFERENCES

1. Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., *Classification and Regression Trees*, Chapman and Hall, 1984.
2. Buckley, C., “Implementation of the SMART information retrieval system”, *Technical Report TR85-686*, Department of Computer Science, Cornell University, Ithaca, NY, 1985.
3. Dempster, A. P., Laird, N. M., and Rubin, D. B.,

- “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society*, B39(1), 1977, pp. 1-38.
4. Godfrey, J. J., Holliman, E. C., and McDaniel, J., "SWITCHBOARD: Telephone speech corpus for research and development", *Proceedings of the ICASSP 92*, 1992, pp. 517-520.
 5. Graff, D., Wu, Z., MacIntyre, R., and Liberman, M., "The 1996 broadcast news speech and language-model corpus", *Proceedings of the 1997 DARPA Speech Recognition Workshop*, 1997.
 6. Hirschberg, J., and Nakatani, C., "Using machine learning to identify intonational segments", *Proceedings of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, 1998, (forthcoming).
 7. Jones, G. J. F., Foote, J. T., Sparck-Jones, K., and Young, S. J., "Video mail retrieval: The effect of word spotting accuracy on precision", *Proceedings of ICASSP 95*, Vol. 1, 1995, pp. 309-312.
 8. Jones, G. J. F., Foote, J. T., Sparck-Jones, K., and Young, S. J., "Retrieving spoken documents by combining multiple index sources", *Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1996, pp. 30-38.
 9. Katz, S. M., "Estimation of probabilities from sparse data from the language model component of a speech recognizer", *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1987, pp. 400-401.
 10. Linde, Y., Buzo, A., and Gray, R. M., "An algorithm for vector quantization design", *IEEE Transactions on Communications*, 28(1), 1980, pp. 84-95.
 11. Ljolje, A., "The importance of cepstral parameter correlations in speech recognition", *Computer Speech and Language*, 1994.
 12. Lovins, J. B., "Development of a stemming algorithm", *Mechanical Translation and Computational Linguistics*, 1-2(11), 1968, pp. 11-31.
 13. Magrin-Chagnolleau, I., Parthasarathy, S., and Rosenberg, A., "Automatic labeling of broadcast news into different sound classes using gaussian mixture models", manuscript in preparation.
 14. Mohri, M., Riley, M., Hindle, D., Ljolje, A., and Pereira, F. C. N., "Full expansion of context-dependent networks in large vocabulary speech recognition", *Proceedings of ICASSP 98*, 1998, forthcoming.
 15. Nakatani, C., Grosz, B. and Hirschberg, J., "Discourse structure in spoken language: studies on speech corpora", *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, 1995.
 16. Ng, K., and Zue, V., "Subword unit representations for spoken document retrieval", *Proceedings of Eurospeech 97*, 1997, pp. 1607-1610.
 17. Pereira, F., and Riley, M., "Speech recognition by composition of weighted finite automata", in Roche, E., and Schabes, Y., (eds), *Finite-State Language Processing*, MIT Press, Cambridge, 1997, pp. 431-453.
 18. Pitrelli, J., Beckman, M., and Hirschberg, J., "Evaluation of prosodic transcription labeling reliability in the ToBI framework", *Proceedings of the 3rd International Conference on Spoken Language Processing*, Vol. 2, 1994, pp. 123-126.
 19. Salton, G., (ed.), *The SMART Retrieval System - Experiments in Automatic Document Retrieval*. Prentice Hall, Inc., Englewood Cliffs, NJ, 1971.
 20. Seymore, K., and Rosenfeld, R., "Scalable backoff language models", *Proceedings of the Fourth International Conference on Spoken Language Processing*, 1996.
 21. Singhal, A., Buckley, C., and Mitra, M., "Pivoted document length normalization", *Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1996, pp. 21-29.
 22. Titterton, D. M., Smith, A. F. M., and Makov, U. E., *Statistical Analysis of Finite Mixture Distributions*, John Wiley and Sons, 1985.
 23. Vorhees, E. M., and Harman, D. K., "Overview of the sixth Text REtrieval Conference (TREC-6)", in Vorhees, E. M., and Karman, D. K. (eds.), *Proceedings of the Sixth Text Retrieval Conference (TREC-6)*, 1998, forthcoming.
 24. Wechsler, M., and Schauble, P., "Indexing methods for a speech retrieval system", in van Rijsbergen, C. J. (ed.), *Proceedings of the MIRO Workshop*, 1995.
 25. Witbrock, M. J., and Hauptmann, A. G., "Using words and phonetic strings for efficient information retrieval from imperfectly transcribed spoken documents", *Proceedings of the 2nd ACM International Conference on Digital Libraries*, 1997, pp. 30-35.

SCAN - Speech Content Based Audio Navigator : A System Overview (1998)

John Choi, Don Hindle, Julia Hirschberg, Ivan Magrin-Chagnolleau, Christine Nakatani, Fernando Pereira, Amit Singhal, and Steve Whittaker.

SCAN - Speech Content Based Audio Navigator : A System Overview. *Proceedings of ICSLP 98*, Sydney, Australia, December 1998.

4 pages

SCAN - Speech Content Based Audio Navigator: A Systems Overview

John Choi, Don Hindle, Julia Hirschberg, Ivan Magrin-Chagnolleau,
Christine Nakatani, Fernando Pereira, Amit Singhal, Steve Whittaker

AT&T Labs - Research

ABSTRACT

SCAN (Speech Content based Audio Navigator) is a spoken document retrieval system integrating speaker-independent, large-vocabulary speech recognition with information-retrieval to support query-based retrieval of information from speech archives. Initial development focused on the application of SCAN to the broadcast news domain. This paper provides an overview of this system, including a description of its graphical user interface which incorporates machine-generated speech transcripts to provide local contextual navigation and random access for browsing large speech databases.

1. INTRODUCTION

We present an overview of SCAN (Speech Content based Audio Navigator), a system developed at AT&T Labs-Research which supports the retrieval, browsing and navigation of speech archives. The system consists of three components: (i) a speaker-independent large-vocabulary speech recognition engine which segments the speech archive and generates transcripts, (ii) an information-retrieval engine which indexes the transcriptions and formulates hypotheses regarding document relevance to user-submitted queries and (iii) a graphical-user-interface which supports search and local contextual navigation based on the machine-generated transcripts and graphical representations of query-keyword distribution in the retrieved speech transcripts. An overview of the system architecture is provided in Figure 1.

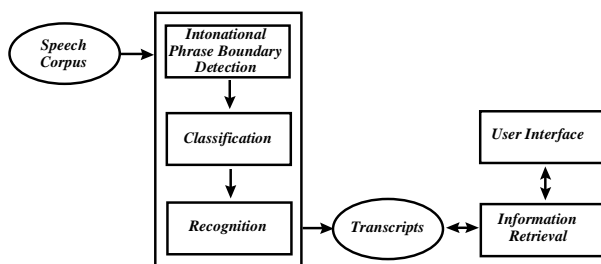


Figure 1: Overview of the spoken document system architecture

Previous work on information retrieval from speech databases include a system for Swiss radio news [14], a system for a digital video library [15], radio news broadcast retrieval using subword units [9], a Video Mail Retrieval system [4,5] and a number of systems developed for the Text Retrieval Conference Spoken Document Retrieval track [13], *inter alia*.

2. SPEECH RECOGNITION

The speech recognition component of SCAN includes an intonational phrase boundary detection module and a classification module. These subcomponents preprocess the speech data before passing the speech to the recognizer itself.

2.1. Intonational Phrase Boundary Detection

The intonational phrase boundary detection module analyzes and classifies the incoming speech stream every 20 msec as either occurring within an intonational phrase or within a break between intonational phrases [3]. This classification is driven by a regression tree model [1] which take as observation vectors values for fundamental frequency, RMS energy and autocorrelation-peaks.

The resulting intonational phrases serve as recognition units by breaking up the incoming speech stream into smaller segments for decoding. In addition, the phrases serve as prosodically well-formed browsing and play-back units in the user-interface. This is arguably preferable to using fixed-size segments, which might begin and terminate within words and otherwise form prosodically ill-formed segments for play-back.

2.2. Channel Condition Classifier

The intonational phrases are subsequently passed to a classifier designed to detect one of four different channel conditions: (i) wideband (0-8kHz) speech recorded in a studio environment with no background noise; (ii) wideband (0-8kHz) speech recorded in non-studio environments, including field conditions, with no background noise; (iii) narrowband (0-4kHz) speech recorded from telephone interviews and (iv) speech with background noise and/or music. The training data come from the 1996 broadcast news corpus distributed by the Linguistic Data Consortium and that comprised the HUB4 continuous speech recognition task for that year. The decision to apply one of potentially several acoustic models is based on this classification.

The classifier itself is based on full covariance Gaussian mixture models, initialized using vector quantization and trained using the Expectation-Maximization algorithm [7]. Input consists of 31-dimensional vectors of filter-bank coefficients in dB units, derived from Hamming windowed frames of 20 msec with a frame advance rate of 10 msec. The filter-bank coefficients are computed by taking the base 10 logarithm from short term power spectra in the 0-8kHz band from a mel-scaled bank of

filters.

2.3. Recognizer

The core speech recognizer is based on a standard time-synchronous beam search algorithm and continuous-density, left-to-right, three-state, context-dependent HMM phone models. The transduction from phone HMMs to word sequences is implemented in the general framework of weighted finite-state transducers [14,17]. The decoder supports multiple recognition hypotheses in the form of word lattices, derived from model lattices by transducer composition.

Acoustic observations serving as input to the HMMs consist of 39-dimensional vectors taken from 20 msec analysis frames with 10 msec overlap. Each acoustic vector contains the first 13 normalized mel-frequency cepstral coefficients, along with their first and second time derivatives.

We have experimented with several acoustic models trained on broadcast news data, partitioning the training data in terms of differing channel conditions. Training iterations for all the acoustic models consist of eigenvector rotations to decorrelate the training data, k-means clustering, normalization of means and variances based on maximum-likelihood, and Viterbi alignment to resegment the data. The output probability distributions in the HMMs consist of a weighted mixture of Gaussians with diagonal covariance. The number of components per mixture varies from model to model, but ranges between 4 and 12.

We have also experimented with several language models, varying in vocabulary size from 20k to 237k. Standard Katz backoff trigram models [6] are constructed from appropriate training corpora. Trigrams and bigrams are then discarded from the model in cases where the difference between the model prediction and backed-off prediction is less than a threshold T :

$$f * (P_o * P_b) < T$$

where f is the observed n -gram frequency, P_o is the n -gram prediction and P_b is the backed-off $(n-1)$ -gram prediction.

3. INFORMATION RETRIEVAL

The information-retrieval engine used in the system is based on a vector space model which generates weighted term (word) vectors for a given transcript and is known as SMART in the information retrieval community [2,11]. SMART initially preprocesses the transcripts by (i) tokenizing the text into individual words, (ii) removing common functions words and (iii) running morphological stemming. The term vectors are then weighted using the lnu term weighting scheme [12]:

$$\frac{(1 + \ln(tf)) / (1 + \ln(average\ tf))}{0.8 * pivot + 0.2 * (\# \text{ of unique terms})}$$

where term frequency tf is the number of times a term occurs in the text, $average\ tf$ is the average of the tf s of all the terms in a document, ln is a length normalization factor, and $pivot$ is the

average number of unique terms in a document, computed across the entire collection.

The user-queries are also preprocessed in the manner described above and indexed using ltn weights [12]:

$$(1 + \ln(tf)) * idf$$

where inter document frequency idf is defined as the ratio of the total number of documents in the collection N and the number of documents df that contain the word, scaled again by a length normalization factor ln :

$$\ln(N / df)$$

An inner-product similarity measure is used to assess the relevance of a document vector D_i and a query vector Q

$$Sim(Q, D_i) = \sum_{\text{common terms } t_j} q_j * d_{ij}$$

where t_j is a term present in both the query and the document, q_j is the weight of term t_j in the query, and d_{ij} is its weight in document i . The transcripts are ranked by their decreasing similarity to the query and presented to the user in this order.

4. USER INTERFACE

The graphical-user-interface is built on the principle of "what you see is almost what you hear" (WYSIAYH) and uses the ASR transcripts extensively. The term "almost" is included because the machine-generated transcripts are errorfull.

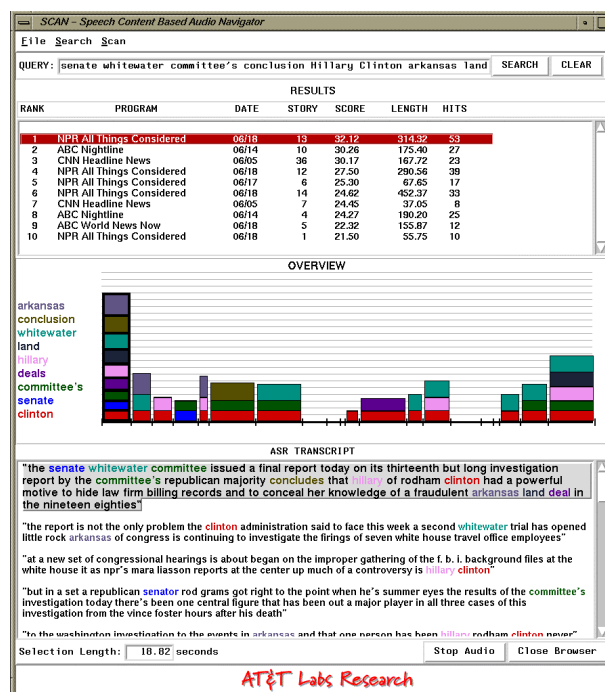


Figure 2: The SCAN graphical user interface.

The interface, presented in Figure 2, consists of three

components: (i) Search, (ii) Overview and (iii) Transcripts. The Search component (top of Figure 2) accommodates user submitted queries. The information returned by the information-retrieval engine is displayed as a ranked list of (potentially) relevant documents. Each of the items displayed in the ranked list provide links to the corresponding documents.

Associated with each document in the ranked list is an Overview display and a Transcript display. The Overview display (middle of Figure 2) provides the user with information about the distribution of query-keywords within the transcripts associated with the selected document. A graphical rendering of this information allows the user to quickly assess regions within the transcripts that are “hotspots” insofar as they exhibit high query-keyword density or occurrences of a certain word or word combinations. This information is rendered as a histogram in Figure 2. The x-axis represents the length of the selected audio document, with the bin-boundaries corresponding to the intonational phrase boundaries detected at recognition time. Within each intonational phrase, we denote the occurrence of a query-keyword in the phrase by adding a color-coded block. The height of the block is determined by the term weight computed by the information-retrieval engine for the specific query-keyword. The query-keywords themselves are displayed to the left of the histogram, again color-coded for easy identification with the blocks in the histogram. Each bin in the histogram is linked to the audio associated with the intonational phrase; simply clicking on a bar in the histogram, or on the baseline in cases of an empty bin, will trigger audio play-back of the corresponding intonational phrase.

The final component of our interface consists of a display of the ASR transcripts (bottom of Figure 2) associated with the document that the user has selected. Within the transcripts, words matching the keywords from the query are color-highlighted to facilitate quick scanning. The transcripts are also formatted so that the text associated with separate intonational phrases are delimited in paragraph type breaks. Again, each paragraph is linked to the audio associated with the intonational phrase; simply clicking on a paragraph will trigger audio play-back of the corresponding intonational phrase.

Both the Overview and Transcript components are designed to provide the user with information about the internal structure and content of the selected transcript. The aim of this design is to afford the user quick access to the underlying speech at any point in the speech document. By visually rendering the underlying speech in terms of keyword distribution and machine-generated transcripts, the interface allows the user to fully exploit the human ability to rapidly scan and browse complex visual data to identify the relevant portions of the lengthy audio material that they wish to listen to. Each transcript can be viewed as a visual analogue to the speech story, and it can be used as a partial index to access relevant speech information. The visual analogue also allows us to exploit existing textual layout conventions, to provide some structure to the underlying speech. Furthermore, the Transcript component allows the played speech to be interpreted in context: i.e., users can scan the transcript to see what was said before and after the soundbite they are currently playing. They no longer hear an isolated soundbite with no idea of where the played speech occurs in relation to the whole document.

Finally, providing the transcript enables users to exercise “checks” to determine the relevance of a given document: by quickly scanning it they can determine whether problems in either speech recognition or information retrieval led to the selection of an irrelevant document.

5. SYSTEM ASSESSMENT

To provide an assessment of the effectiveness of information-retrieval, we present results from two tasks. The first is the TREC-6 SDR task [13] which involves 49 known-item user-queries where each query has associated with it a unique document in a collection of 1452 documents. Performance on the task can be measured by simply counting the number of queries for which the target document is ranked within some K ranks. Figure 3 presents these counts for K=1, i.e., when the correct document was ranked #1 by the information-retrieval engine, and for K=5, i.e., when the correct document was ranked within the top 5 ranked documents. Results are presented for retrieval based on both machine-generated transcripts and human-generated transcripts.

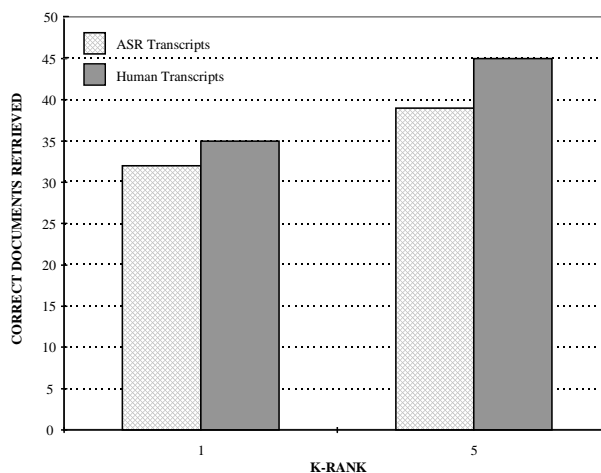


Figure 3: Information retrieval accuracy on a known-item task based on ASR transcripts and human transcripts.

While the TREC-6 SDR task affords us a benchmark with which to compare our system performance against other systems on the same task, it is widely held that the task is too restricted in its scope, due primarily to the small size of the corpus over which retrieval is being conducted, to be a robust measure of spoken document retrieval performance. To compensate for this, we exercised the system further, devising another task in which we selected 94 AP Newswire headlines from the same period that the broadcast news archives were collected. These headlines were then used as user queries. In lieu of relevance assessments for these headlines, which were not available, we simply compared the retrieval results from the machine-generated transcripts with those from the human-generated transcripts. In doing so, we assume that retrieval from human-generated transcripts serves as an upper-bound on our retrieval performance; i.e., average precision is 100%. Having made this assumption, we can calculate average precision based on

retrieval from the machine-generated transcripts and compare this to the upper-bound.

The results from the AP headlines task is illustrated in Figure 4, where average precision based on the machine-generated transcripts is plotted as a function of K ranks, where $K=\{1, 5, 10, 20, 30, 50, 100\}$. The top curve represents average precision when the word-error-rate is 30.0%. For comparison, another curve is plotted which represents average precision when the word-error-rate is 42.7%. This test shows that with a word-error-rate of 30.0%, we perform approximately 81% as well as we would if the transcripts were perfect. This test also shows that better recognition results in higher precision retrieval.

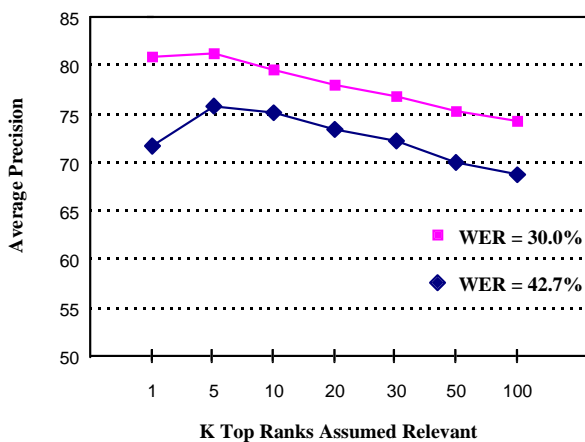


Figure 4: Average precision based on ASR transcripts at different word-error-rates for the AP headlines task.

6. CONCLUSION

We have described a system for querying and retrieving information from speech databases which integrates speech recognition and information retrieval technologies. Preliminary assessments conducted on tasks in the broadcast news domain are encouraging and we continue to explore techniques for improving information retrieval precision through the use of word lattices. We have also described an interface which allows users to identify relevant regions of the underlying speech without having to listen to the entire speech document and are conducting usability tests on this interface.

7. REFERENCES

1. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, Chapman and Hall, 1984.
2. Buckley, C., "Implementation of the SMART information retrieval system", *Technical Report TR85-686*, Department of Computer Science, Cornell University, Ithaca, NY, 1985.
3. Hirschberg, J., and Nakatani, C., "Using machine learning to identify intonational segments", *Proceedings of the AAAI Spring Symposium on*

Applying Machine Learning to Discourse Processing, 1988, forthcoming.

4. Jones, G.J.F., Foote, J.T., Sparck-Jones, K., and Young, S.J., "Video mail retrieval: The effect of word spotting accuracy on precision", *Proceedings of ICASSP 95, Vol. 1*, 1995, pp. 309-312.
5. Jones, G.J.F., Foote, J.T., Sparck-Jones, K., and Young, S.J., "Retrieving spoken documents by combining multiple index sources", *Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1996, pp. 30-38.
6. Katz, S.M., "Estimation of probabilities from sparse data from the language model component of a speech recognizer", *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1987, pp. 400-401.
7. Magrin-Chagnolleau, I., Parthasarathy, S., and Rosenberg, A., "Automatic labeling of broadcast news into different sound classes using gaussian mixtures models", MS in preparation.
8. Mohri, M., Riley, M., Hindle, D., Ljolje, A., and Pereira, F.C.N., "Full expansion of context-dependent networks in large vocabulary speech recognition", *Proceedings of ICASSP 98*, 1998.
9. Ng, K., and Zue, V., "Subword unit representations for spoken document retrieval", *Proceedings of Eurospeech 97*, 1997, pp. 1607-1610.
10. Pereira, F., and Riley, M., "Speech recognition by composition of weighted finite automata", In Roche, E., and Schabes, Y., (eds), *Finite-State Language Processing*, MIT Press, Cambridge, 1997, pp. 431-453.
11. Salton, G., (ed.), *The SMART Retrieval System - Experiments in Automatic Document Retrieval*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1971.
12. Singhal, A., Buckley, C., and Mitra, M., "Pivoted document length normalization", *Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1996, pp. 21-29.
13. Vorhees, E.M., and Harman, D.K., "Overview of the sixth Text Retrieval Conference (TREC-6)", in Vorhees, E.M., and Karman, D.K. (eds.), *Proceedings of the Sixth Text Retrieval Conference (TREC-6)*, 1998, forthcoming.
14. Wechsler, M., and Schauble, P., "Indexing methods for a speech retrieval system", in van Rijsbergen, C.J. (ed.), *Proceedings of the MIRO Workshop*, 1995.
15. Witbrock, M.J., and Hauptmann, A.G., "Using words and phonetic strings for efficient information retrieval from imperfectly transcribed spoken documents", *Proceedings of the 2nd ACM International Conference on Digital Libraries*, 1997, pp. 30-35.

**Detection of Target Speakers in Audio Databases
(1999)**

Ivan Magrin-Chagnolleau, Aaron E. Rosenberg, and S. Parthasarathy.
Detection of Target Speakers in Audio Databases. *Proceedings of
ICASSP 99*, Vol. 2, pp. 821-824, Phoenix, United States, March 1999.

4 pages

DETECTION OF TARGET SPEAKERS IN AUDIO DATABASES

Ivan Magrin-Chagnolleau, Aaron E. Rosenberg and S. Parthasarathy

AT&T Labs Research - Florham Park, New Jersey - USA

ivan@ieee.org - aer@research.att.com - sps@research.att.com

ABSTRACT

The problem of speaker detection in audio databases is addressed in this paper. Gaussian mixture modeling is used to build target speaker and background models. A detection algorithm based on a likelihood ratio calculation is applied to estimate target speaker segments. Evaluation procedures are defined in detail for this task. Results are given for different subsets of the HUB4 broadcast news database. For one target speaker, with the data restricted to high quality speech segments, the segment miss rate is approximately 7%. For unrestricted data, the segment miss rate is approximately 27%. In both cases the segment false alarm rate is 4 or 5 per hour. For two target speakers with unrestricted data, the segment miss rate is approximately 63% with about 27 segment false alarms per hour. The decrease in performance for two target speakers is largely associated with short speech segments in the two target speaker test data which are undetectable in the current configuration of the detection algorithm.

1. INTRODUCTION

The problem addressed in this article is the detection of one or more target speakers in broadcast news programs. Detection, in this context, means the estimation of a beginning and an ending time for each segment in which a target speaker is speaking. This problem is an emerging one which has been reported on recently [6, 7, 8, 5]. It is a potentially important problem since, as more and more audio and multimedia data are recorded and archived, the need grows for useful cues to segment, classify, and organize this data.

In this paper, two sets of experiments are described. In the first, the detection of a single target speaker is addressed; the second generalizes the problem to two target speakers. The evaluation procedures are slightly different in each case, and are described in detail in the article.

2. DATABASE

The HUB4 database [1] is composed of 174 broadcasts from 11 news and commentary programs. For each program, from 6 to 37 broadcasts corresponding to different dates are available. The duration of the longest broadcast is 2 hours, the duration of the shortest is 26 minutes. Speech portions of each broadcast have been transcribed, segmented, and labeled in terms of speaker; mode (spontaneous or planned); fidelity, classifying the quality of the recording environment and transmission channel (High, Medium,

or Low); and background, describing the nature of secondary audio material mixed with the primary speech signal (Speech, Music, or Other). Commercials and sports results have not been transcribed.

For our experiments, the data are classified into 4 categories as follows. The category *high-fidelity* refers to speech labeled High fidelity with no background; the category *clean* refers to speech with all 3 quality categories but no background; the category *all-speech* refers to data of all quality categories with and without background; and the category *alldata* refers to data including the previous *all-speech* category plus all the untranscribed portions. When experiments refer to test data belonging to the *high-fidelity* category only, it means that, for each program, all the data belonging to the *high-fidelity* category are artificially concatenated together, and similarly for the *clean* and *all-speech* categories. The category *alldata* corresponds to the data as originally provided.

The one-target-speaker detection experiments are conducted on the subset aABC_NLI of the HUB4 database, which corresponds to the program ABC Nightline. The target speaker is Ted Koppel. 6 broadcasts are set aside for training the target model and two background models. 12 broadcasts are used as test broadcasts. The duration of the longest test broadcast is 35 minutes, the duration of the shortest is 26 minutes. The remaining 5 programs are not used since they contain no data of the target speaker.

The two-target-speaker detection experiments are conducted on the subset bABC_WNN of the HUB4 database, which corresponds to the program ABC World News Now. The target speakers are Mark Mullen (T1) and Thalia Assures (T2). 8 broadcasts are used to train the two target models and two background models. 16 broadcasts are used as test broadcasts. The duration of the longest test broadcast is 31 minutes; the duration of the shortest is 29 minutes.

3. MODELING

To detect a target speaker in a program, at least two models need to be built: a target speaker model and a background model which is intended to represent speech from speakers other than the target speaker or other types of sounds. These models are built using feature vectors extracted from labeled segments. In the case of several target speakers, at least one model for each target speaker is needed. Finally, more generally, there could be several background models, representing different types of non-target speaker sounds, and several models for each target speaker, representing different speech qualities. In the experiments presented in this article, only one model is built for each target speaker, and one or two background models are used.

The feature vectors consist of cepstral and Δ -cepstral coefficients extracted in the following way: a Winograd Fourier Trans-

form (WFT) is computed on Hamming windowed signal frames of 31.5 ms (504 samples) at a frame rate of 10 ms (160 samples). For each frame, spectral vectors of 31 Mel-Scale Triangular Filter Bank coefficients are then calculated from the Fourier Transform power spectrum, and expressed in logarithmic scale. An algorithm based on a bimodal Gaussian modeling of the energy (sum of the 31 coefficients) is then used to remove low energy vectors. This algorithm removes between 10 % and 30 % of the frames. Cepstral coefficients c_1 to c_{20} are then calculated using cosine transforms. The coefficient c_0 , which corresponds to the energy, is omitted. Finally, Δ -coefficients are calculated over 5 frames, providing 20 additional coefficients. The feature vectors are thus 40-dimensional.

Gaussian mixture models (GMM) [3] are used for modeling target speakers and backgrounds. These models have been used successfully in text-independent speaker recognition [4]. The number of mixture components is set equal to 64, and diagonal covariance matrices are used. The parameters of the GMM are estimated using the Expectation Maximization (EM) algorithm [3], initialized by a binary splitting Vector Quantization (VQ) algorithm [2].

Only *high-fidelity* speech is used to build each target speaker model. For each target speaker, three 90 second segments of speech, extracted from 3 broadcasts uttered by the target speaker, are concatenated together. Feature vectors are then extracted, and a GMM model is trained for each target speaker, as described previously.

For the first background model, *high-fidelity* speech only is used. Samples from 4 female and 4 male speakers, 60 seconds each, are concatenated together. The speakers for this background model are chosen from different broadcasts. Feature vectors are then extracted, and a GMM model is trained, as described previously. This background model is referred to as λ_{B1} .

For some experiments, a second background model is used. Only non-speech segments are used for this second model, that is music only (10 %), noise only (10 %), and commercials (80 %). Three 90 second segments, chosen from 3 broadcasts, are concatenated together. Feature vectors are then extracted, and a GMM model is trained, as described previously. This background model is referred to as λ_{B2} .

None of the broadcasts and none of the speakers used to build the target or the background models are used in the test data.

4. DETECTION ALGORITHM

The simplest experimental problem is the detection of one target speaker, using only one target model (λ_T) and one background model (λ_B).

Given a test broadcast, the task is to detect the segments in which the target speaker is speaking, that is to estimate the beginning and the end of these segments. First of all, feature vectors are extracted from the test broadcast, using the signal analysis described previously. Let $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ denote the sequence of feature vectors. For each vector \mathbf{x}_t , we calculate the likelihood of \mathbf{x}_t given the model λ_T , denoted by $\mathcal{L}(\mathbf{x}_t|\lambda_T)$, the likelihood of \mathbf{x}_t given the model λ_B , denoted by $\mathcal{L}(\mathbf{x}_t|\lambda_B)$, and finally the likelihood ratio of \mathbf{x}_t given λ_T and λ_B , denoted by $\mathcal{R}(\mathbf{x}_t|\lambda_T; \lambda_B)$, whose logarithm is expressed by:

$$\log \mathcal{R}(\mathbf{x}_t|\lambda_T; \lambda_B) = \log \mathcal{L}(\mathbf{x}_t|\lambda_T) - \log \mathcal{L}(\mathbf{x}_t|\lambda_B)$$

Before applying the segmentation algorithm, smoothing is needed to attenuate variations of the logarithm of the likelihood ratio. The smoothing is an arithmetic mean of a specified number of consecutive values of the logarithm of the likelihood ratio. Two

parameters define the smoothing: the number of values used for the average calculation, denoted by τ ; and the delay between two calculations, that is the number of feature vectors between two average calculations, denoted by d . The smoothed value of the logarithm of the likelihood ratio is:

$$\begin{aligned} & \overline{\log \mathcal{R}}(\mathbf{x}_{t_0-t'}, \dots, \mathbf{x}_{t_0+t'}|\lambda_T; \lambda_B) \\ &= \frac{1}{\tau} \sum_{t=t_0-t'}^{t_0+t'} \log \mathcal{R}(\mathbf{x}_t|\lambda_T; \lambda_B) \\ &= \overline{\log \mathcal{L}}(\mathbf{x}_{t_0-t'}, \dots, \mathbf{x}_{t_0+t'}|\lambda_T) - \overline{\log \mathcal{L}}(\mathbf{x}_{t_0-t'}, \dots, \mathbf{x}_{t_0+t'}|\lambda_B) \end{aligned}$$

with $\tau = 2t' + 1$. In our experiments, τ is set to 100 vectors (1 second) and d to 20 vectors (0.2 second).

A segmentation algorithm is then applied on the average values previously calculated. The algorithm is described in detail in the Block 1.

From first block to last block, do:

- If $v > \theta_1^B$ and $detectFlag = 1$ and $endFlag = 1$
endFlag = 0
- If $v > \theta_1^B$ and $beginFlag = 0$
Set the beginning time for a possible segment
 $beginFlag = 1$ (a beginning time has been set)
- If $v > \theta_2^B$ and $detectFlag = 0$
 $detectFlag = 1$ (a target speaker segment has been detected)
- If $v < \theta_1^E$ and $beginFlag = 1$ and $detectFlag = 0$
 $beginFlag = 0$
- If $v < \theta_1^E$ and $detectFlag = 1$ and $endFlag = 0$
Set the ending time for the detected segment
 $endFlag = 1$ (an ending time has been set)
- If $v < \theta_2^E$ and $detectFlag = 1$
Record the detected segment
 $beginFlag = 0$ (reset of $beginFlag$)
 $detectFlag = 0$ (reset of $detectFlag$)
 $endFlag = 0$ (reset of $endFlag$)

Block 1: *Segmentation algorithm.*

v is the average value of the logarithm of the likelihood ratio for a block, θ_1^B and θ_2^B are two thresholds used for the detection of the beginning of a segment, and θ_1^E and θ_2^E are two thresholds used for the detection of the end of a segment. In our experiments, the values of the thresholds are the following: $\theta_1^B = 0$, $\theta_2^B = 3$, $\theta_1^E = 0$, and $\theta_2^E = -0.5$.

Finally, the minimum duration for an estimated segment is set to 2.5 seconds (every segment whose duration is smaller is omitted), and the minimum interval between two consecutive segments to 1 second (two consecutive segments are merged if the interval between them is smaller).

The segmentation algorithm provides the estimated beginning and end times for the target speaker segments.

In the general case, several background models are available ($\lambda_{B1}, \lambda_{B2}, \dots$), as well as several target speakers and several models for each target speaker ($\lambda_{T1}, \lambda_{T1'}, \dots, \lambda_{T2}, \lambda_{T2'}, \dots$). The average value of the logarithm of the log-likelihood of a block is

| Quality | high | | clean | | allspeech | | alldata | |
|----------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Total duration | 141 min | | 194 min | | 242 min | | 359 min | |
| # target segments | 137 | | 257 | | 318 | | 354 | |
| Duration | 61 min | | 69 min | | 78 min | | 78 min | |
| Background | B1 | B1, B2 | B1 | B1, B2 | B1 | B1, B2 | B1 | B1, B2 |
| # estimated segments | 129 | 129 | 195 | 195 | 238 | 237 | 256 | 254 |
| FMIR | 3.90 % | 4.01 % | 4.92 % | 5.07 % | 6.73 % | 7.23 % | 8.89 % | 9.57 % |
| FFAR | 10.75 % | 10.52 % | 9.19 % | 8.80 % | 7.22 % | 7.00 % | 5.66 % | 5.44 % |
| SMIR | 7.30 % | 7.30 % | 19.07 % | 19.46 % | 25.16 % | 26.42 % | 27.40 % | 29.66 % |
| SFAR | 5.53 / hour | 5.53 / hour | 8.35 / hour | 7.42 / hour | 5.95 / hour | 5.21 / hour | 4.35 / hour | 4.18 / hour |

Table 1: Results of the one-target-speaker detection experiments.

then given by:

$$\begin{aligned} & \overline{\log \mathcal{R}}(\mathbf{x}_{t_0-t'}, \dots, \mathbf{x}_{t_0+t'} | \lambda_{T_1}, \lambda_{T_1'}, \dots, \lambda_{T_2}, \lambda_{T_2'}, \dots; \lambda_{B_1}, \lambda_{B_2}, \dots) \\ &= \max_{T \in \{T_1, T_1', \dots, T_2, T_2', \dots\}} \overline{\log \mathcal{L}}(\mathbf{x}_{t_0-t'}, \dots, \mathbf{x}_{t_0+t'} | \lambda_T) \\ & \quad - \max_{B \in \{B_1, B_2, \dots\}} \overline{\log \mathcal{L}}(\mathbf{x}_{t_0-t'}, \dots, \mathbf{x}_{t_0+t'} | \lambda_B) \end{aligned}$$

In the case of several target speakers, each target segment is labeled with the identity of the detected target speaker.

5. EVALUATION

The performance of the detection algorithm has been evaluated in two ways, one giving the performance at the frame level, the other at the segment level. Each frame inside an estimated target speaker segment is referred to as an *estimated target frame*, each frame outside as an *estimated non-target frame*. The estimated target frames and the estimated target segments are then compared with those provided by the database. Each frame inside a labeled target speaker segment provided with the database is referred to as a *labeled target frame*, each frame outside as a *labeled non-target frame*.

The **Frame-level Miss Rate (FMIR)** is the number of labeled target frames which have not been estimated as target frames (that is which have been estimated as non-target frames or as target frames from another speaker in the case of several target speakers), divided by the total number of labeled target frames.

The **Frame-level False Alarm Rate (FFAR)** is the number of estimated target frames which are in fact labeled non-target frames, divided by the total number of labeled non-target frames.

In the case of the detection of several target speakers, the proportion of missed frames due to a confusion with another speaker is also indicated in parentheses by the **Frame-level Confusion Rate (FCOR)**, which is the number of labeled target frames which have been estimated as target frames of another speaker, divided by the total number of labeled target frames. FCOR is a component of FMIR.

A missed segment is a labeled target segment for which the proportion of frames estimated as target frames (of the correct target speaker in the case of several target speakers) is less than the threshold of Frames Correctly Detected (FCD). For our experiments, FCD has been set to 75 %. The **Segment-level Miss Rate (SMIR)** is the number of missed segments divided by the total number of target segments.

In the case of one target speaker, a false alarm segment is an estimated target segment for which the proportion of labeled non-target frames is greater than the proportion of labeled target

frames. The **Segment-level False Alarm Rate (SFAR)** is defined as the total number of false alarm segments divided by the total duration of the broadcast in hours.

The following definitions are needed to define the SFAR and the Segment-level Confusion Rate (SCOR) in the case of the detection of several target speakers. A *false-alarm* frame is an estimated target frame which is actually a non-target frame. A *confusion* frame is an estimated target frame which is actually a target frame of another target speaker. A *hit* frame is an estimated target frame which is actually a target frame of the correct speaker.

A false alarm segment is an estimated target segment for which the proportion of *false alarm* frames is greater than the proportion of *confusion* frames and also greater than the proportion of *hit* frames. The **SFAR** is then the total number of false alarm segments divided by the total duration of the broadcast in hours.

A confusion segment is an estimated target segment for which the proportion of *confusion* frames is greater than the proportion of *false alarm* frames and also greater than the proportion of *hit* frames. The **Segment-level Confusion Rate (SCOR)** is then the total number of confusion segments divided by the total duration of the broadcast in hours.

6. EXPERIMENTAL RESULTS

Table 1 reports results for the one-target-speaker detection experiments. Results are reported for the 4 categories of sound: *high-fidelity*, *clean*, *allspeech*, and *alldata*. For each category, the total duration of test material, all 12 programs together, is indicated, as well as the number of target segments and their total duration. Results are given when B1 only is used as the background model, and when B1 and B2 are used together. In both cases, the number of estimated target segments is indicated, and the 4 error rates are given.

There is not a lot of difference between the results obtained when B1 is used alone, or when B1 and B2 are used together. With B1 only, the FMIR is generally slightly lower, but the FFAR is slightly higher. With data from the *alldata* category, the sum of the two rates is better when B1 is used alone (14.55 % versus 15.01 %). At the segment level, the same trend can be observed.

The FMIR degrades through each category from the *high-fidelity* category to the *alldata* category, because of the increase of the mismatch between the test data and the data used for the training of the models. This problem might be solved by training the target speaker model with data from different categories pooled together, or training several models for the target speaker using data from each category.

Conversely, the FFAR improves through each category from the *high-fidelity* category to the *alldata* category. This may be be-

| Target speakers | T1 | | | T2 | | | T1, T2 | | |
|-------------------|---------|---------|--------------|---------|---------|--------------|-----------|-----------|--------------|
| Segt. duration | > 4s | > 2s | All | > 4s | > 2s | All | > 4s | > 2s | All |
| # target segments | 119 | 200 | 315 | 125 | 191 | 310 | 244 | 391 | 625 |
| Duration | 29 min | 33 min | 35 min | 41 min | 44 min | 46 min | 71 min | 78 min | 82 min |
| # est. segments | – | – | 216 | – | – | 183 | – | – | 349 |
| FMIR | 14.06 % | 19.34 % | 22.70 % | 17.40 % | 21.42 % | 23.90 % | 36.52 % | 38.79 % | 40.27 % |
| (FCOR) | – | – | – | – | – | – | (24.47 %) | (24.55 %) | (24.69 %) |
| FFAR | – | – | 9.75 % | – | – | 7.21 % | – | – | 18.38 % |
| SMIR | 23.53 % | 42.50 % | 58.10 % | 28.00 % | 45.03 % | 64.19 % | 42.62 % | 52.69 % | 62.88 % |
| SCOR | – | – | – | – | – | – | – | – | 5.27 / hour |
| SFAR | – | – | 17.41 / hour | – | – | 12.59 / hour | – | – | 27.51 / hour |

Table 2: Results of the two-target-speaker detection experiments for the alldata category.

cause there is proportionally less and less *high-fidelity* data through each category, *high-fidelity* data being more likely to be confused with the *high-fidelity* target model than data from the other categories.

Similar observations can be made at the segment level.

Table 2 reports results for the two-target-speaker detection experiments. The results are reported only for data from the *alldata* category (total duration = 410 minutes), and when B1 only is used as background model. (The experiments using B1 and B2 do not give better results.) The results for each target speaker alone are given first, and then the results for the two target speakers together. The number of estimated target segments is indicated, and the 6 error rates are given. Because of the nature of the program, the two target speakers are often engaged in a dialog. This produces a high proportion of short target segments and also a small proportion of overlap segments. (4.5 % of all the labeled target frames overlap the other target speaker.) To study the effect of segment duration on performance, FMIR and SMIR results are shown for different values of minimum target segment durations.

Performance approaching the one-target-speaker results with Ted Koppel are obtained only for target segment durations greater than four seconds. In fact, most of the Ted Koppel segments (72 %) are longer than 4 seconds. For Mark Mullen, only 38 % of the segments are longer than 4 seconds, and 40 % for Thalia Assures.

Because the duration of most target segments is smaller than four seconds, the choice of the smoothing block length may be important. If this length is decreased, the FMIR and the SMIR may be improved, at the expense of degrading the FFAR and the SFAR. One solution to this problem could be a two-pass algorithm, each pass with different smoothing parameters. Different values for the minimum duration of an estimated segment have also been tested (from 2.5 seconds to 1 second), but the results do not change significantly. In fact, most of the very short segments are completely missed, whatever the minimum duration of an estimated segment is, due to the block smoothing. Some of them are combined into longer estimated segments comprising two or more short segments.

In some cases, short segments of one target speaker (particularly those which are shorter than the block smoothing length) are merged with longer adjacent segments of the other target speaker, again because of the block smoothing. Because of that, the contribution of the FCOR to the FMIR is high.

7. CONCLUSION

A new approach for detecting target speakers in audio databases has been proposed. This approach, based on Gaussian mixture modeling and a likelihood ratio calculation, gives good results if the duration of the target speaker segments is long enough. However, in the case of short segments, performance degrades, particularly at the segment level. More experiments are needed to provide better understanding of this problem, and the detection algorithm needs to be refined for the detection of short segments.

Further studies in speaker detection will also include the use of more than one model for each target speaker, and the use of other background models. The performance will also be studied as a function of the smoothing parameters and the segmentation algorithm parameters.

8. REFERENCES

- [1] D. Graff, Z. Wu, R. MacIntyre, and M. Liberman. The 1996 broadcast news speech and language-model corpus. In *Proceedings of the 1997 DARPA Speech Recognition Workshop*, February 1997. Chantilly, Virginia.
- [2] Yoseph Linde, Andres Buzo, and Robert M. Gray. An algorithm for vector quantization design. *IEEE Transactions on Communications*, 28(1):84–95, January 1980.
- [3] Geoffrey J. McLachlan and Kaye E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, 1988.
- [4] Douglas A. Reynolds. Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication*, 17(1–2):91–108, August 1995.
- [5] Aaron E. Rosenberg, Ivan Magrin-Chagnolleau, S. Parthasarathy, and Qian Huang. Speaker detection in broadcast speech databases. In *Proceedings of ICSLP 98*, December 1998.
- [6] Man-Hung Siu, George Yu, and Herbert Gish. An unsupervised, sequential learning algorithm for the segmentation of speech waveforms with multiple speakers. In *Proceedings of ICASSP 92*, volume 2, pages 189–192, March 1992. San Francisco, United-States.
- [7] M. Sugiyama, J. Murakami, and H. Watanabe. Speech segmentation and clustering based on speaker features. In *Proceedings of ICASSP 93*, volume 2, pages 395–398, April 1993. Minneapolis, United-States.
- [8] Lynn Wilcox, Francine Chen, Don Kimber, and Vijay Balasubramanian. Segmentation of speech using speaker identification. In *Proceedings of ICASSP 94*, volume 1, pages 161–164, April 1994. Adelaide, Australia.

**Indexing Telephone Conversations by Speakers Using
Time-Frequency Principal Component Analysis
(2000)**

Ivan Magrin-Chagnolleau and Frédéric Bimbot. **Indexing Telephone Conversations by Speakers Using Time-Frequency Principal Component Analysis**. *Proceedings of ICME 2000*, New York, United States, July-August 2000.

4 pages

INDEXING TELEPHONE CONVERSATIONS BY SPEAKERS USING TIME-FREQUENCY PRINCIPAL COMPONENT ANALYSIS

Ivan Magrin-Chagnolleau and Frédéric Bimbot

IRISA (CNRS & INRIA) - Campus universitaire de Beaulieu - 35042 Rennes cedex - France - European Union

ivan@ieee.org - bimbot@irisa.fr

ABSTRACT

In this paper, we present an algorithm for the tracking of target speakers in telephone conversations. Speaker tracking consists in retrieving, in an audio recording, segments which have been uttered by a target speaker. We also compare two speech analysis techniques. The first one is the time-frequency principal component analysis. It is a new speech analysis technique based on the extraction of the principal components of the contextual covariance matrix, which is the covariance matrix of feature vectors expanded by their time context. The other one is the classical cepstral analysis. Experiments are carried out on a subset of the Switchboard database.

1. INTRODUCTION

Speaker tracking is an emerging task in multimedia data processing. It consists in tracking one or several target speakers in an audio recording. This new task has become more and more important with the increase of multimedia data available through Internet. Speaker tracking is one of the numerous new tools used to segment, classify, and organize these data, and therefore to access them faster and more effectively. This emerging problem has been reported on only recently [13, 9, 1, 2, 5, 14].

In this paper, we present a tracking algorithm based on a log-likelihood ratio calculation and a multiple thresholds segmentation algorithm [13, 9]. We also compare two speech analysis techniques in the framework of speaker tracking. The first one is a new speech analysis technique called *time-frequency principal component (TFPC)* analysis [6, 7, 8]. TFPC analysis consists in calculating the principal components of the contextual covariance matrix, which is the covariance matrix of a sequence of vectors (for instance spectral vectors) expanded by their time context. We compare this new analysis to a classical cepstral analysis, using cepstral coefficients [11] augmented by their Δ parameters [3].

The experiments reported in this paper are carried out on a subset of the Switchboard database. This subset is a part of the development data of the NIST 2000 speaker recognition evaluation.

2. TIME-FREQUENCY PRINCIPAL COMPONENTS

TFPC analysis consists in extracting time-frequency patterns which are characteristic of a whole sequence of training vectors, that is,

This work has been done in the framework of the first phase of the AGIR project, funded by the French RNRT Programme.

to summarize the evolution of a sequence of training vectors by a few short sequences extracted from the entire sequence. The original sequence has therefore to be long enough, and representative of the class we want to represent with the time-frequency patterns. This strategy can be applied to any pattern recognition problem, as long as we have enough vectors for each class to calculate the time-frequency patterns. Once the patterns have been extracted, they are used to filter the original vectors of both the training and the test datasets. And any modeling technique can then be applied on the new vectors, as it is done usually on spectral vectors or cepstral vectors, or any other vector representation of the original signal. In this paper, the original vectors are spectral vectors. But the same procedure can be used on any kind of vectors. The only requirement is that the vectors are indexed by time.

Let $\{\mathbf{x}_t\}_{1 \leq t \leq M}$ denote a sequence of vectors, and $\{\mathbf{x}_t^*\}$ the sequence of the corresponding centered vectors.

Let \mathbf{X}_0 denote the covariance matrix of the sequence $\{\mathbf{x}_t\}$:

$$\mathbf{X}_0 = \frac{1}{M} \sum_{t=1}^M (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_t - \bar{\mathbf{x}})^T = \frac{1}{M} \sum_{t=1}^M \mathbf{x}_t^* \cdot \mathbf{x}_t^{*T}$$

and \mathbf{X}_k the lagged covariance matrix at the order k :

$$\mathbf{X}_k = \frac{1}{M} \sum_{t=k+1}^M (\mathbf{x}_t - \bar{\mathbf{x}}) \cdot (\mathbf{x}_{t-k} - \bar{\mathbf{x}})^T = \frac{1}{M} \sum_{t=k+1}^M \mathbf{x}_t^* \cdot \mathbf{x}_{t-k}^{*T}$$

The dimension of the covariance matrix and of the lagged covariance matrices is $p \times p$.

We now define a new matrix, \mathbf{X}_{2q+1} , by:

$$\mathbf{X}_{2q+1} = \begin{bmatrix} \mathbf{X}_0 & \mathbf{X}_1 & \dots & \mathbf{X}_{2q} \\ \mathbf{X}_1^T & \mathbf{X}_0 & \dots & \mathbf{X}_{2q-1} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{X}_{2q}^T & \mathbf{X}_{2q-1}^T & \dots & \mathbf{X}_0 \end{bmatrix}$$

This matrix is block-Toeplitz, and its dimension is $(2q+1)p \times (2q+1)p$.

Let define the sequence of vectors \mathbf{x}_t between time $t-q$ and $t+q$ by:

$$\mathbf{X}_{t-q}^{t+q} = \begin{bmatrix} \mathbf{x}_{t+q} \\ \vdots \\ \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t-q} \end{bmatrix}$$

By convention, $\mathbf{x}_t = 0$ if $t \leq 0$ or $t > T$. The dimension of vector \mathbf{X}_{t-q}^{t+q} is $(2q+1)p$.

The matrix \mathbf{X}_{2q+1} can be interpreted as the covariance matrix of the vectors $\{\mathbf{X}_{t-q}^{t+q}\}_{1 \leq t \leq M}$, and can therefore be called *contextual covariance matrix*.

We now calculate the principal components of this matrix [4]. It is equivalent to the extraction of eigenvalues and eigenvectors of the matrix. The eigenvector associated with the largest eigenvalue is then the direction of projection which conserves the maximum of the variance, the eigenvector associated to the second largest eigenvalue is the direction of projection which conserves the maximum of the variance uncorrelated (that is orthogonal) to the first one, and so on. We have then:

$$\mathbf{X}_{2q+1} = \mathbf{V}_{2q+1} \cdot \mathbf{\Lambda}_{2q+1} \cdot \mathbf{V}_{2q+1}^T$$

with:

$$\begin{aligned} \mathbf{V}_{2q+1} &= (\mathbf{v}_1, \dots, \mathbf{v}_{2q+1}) \\ \mathbf{\Lambda}_{2q+1} &= \text{diag}(\lambda_1, \dots, \lambda_{2q+1}), \quad \lambda_1 \geq \dots \geq \lambda_{2q+1} \end{aligned}$$

The dimension of the matrices \mathbf{V}_{2q+1} and \mathbf{M}_{2q+1} is $(2q+1)p \times (2q+1)p$. The dimension of each vector \mathbf{v}_i , $1 \leq i \leq 2q+1$, is $(2q+1)p$.

Since the principal components are extracted from the contextual covariance matrix instead of the covariance matrix itself, we call them *contextual principal components (CPC)*. When the original vectors represent some information about the frequency content of the signal, as spectral vectors for instance, we can call these components more specifically *time-frequency principal components (TFPC)*.

Once the TFPC have been calculated, we choose some of them to build a filtering matrix. For instance, if we choose to keep the components $\{\mathbf{v}_1, \dots, \mathbf{v}_6, \mathbf{v}_{14}, \dots, \mathbf{v}_{22}\}$, the filtering matrix for the corresponding class will be:

$$\mathbf{H} = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_6^T \\ \mathbf{v}_{14}^T \\ \vdots \\ \mathbf{v}_{22}^T \end{bmatrix}$$

3. SEGMENTATION ALGORITHM

Given a telephone conversation, speaker tracking consists in detecting the segments in which a target speaker is speaking, that is, to estimate the beginning and the end of these segments. First, spectral vectors are extracted from the telephone conversation, using the spectral analysis described in the next section.

Let $\{\mathbf{x}_t\}_{1 \leq t \leq T}$ denote the sequence of spectral vectors.

TFPC Analysis. In the case of TFPC analysis, spectral vectors are then filtered by the TFPC extracted from the training data of the target speaker in order to obtain a new sequence of vectors,

$\{\mathbf{x}_t^T\}$. The spectral vectors are also filtered by the TFPC extracted from the world data, and we obtain another sequence of filtered vectors, $\{\mathbf{x}_t^W\}$.

The world data contain speech from a population of speakers from the same gender as the target speaker, and using the same type of handset¹. This population is supposed to be representative of the world population in term of speaker variability. The model extracted from the world data is used to normalize the likelihood score during the tracking phase.

Let λ_{tfpc}^T and λ_{tfpc}^W denote respectively the model for the target speaker and the world model in the case of TFPC analysis. We then calculate the likelihood of \mathbf{x}_t^T given the model λ_{tfpc}^T , denoted by $\mathcal{L}(\mathbf{x}_t^T | \lambda_{tfpc}^T)$, the likelihood of \mathbf{x}_t^W given the model λ_{tfpc}^W , denoted by $\mathcal{L}(\mathbf{x}_t^W | \lambda_{tfpc}^W)$, and finally the likelihood ratio, denoted by $\mathcal{R}(\mathbf{x}_t^T; \mathbf{x}_t^W | \lambda_{tfpc}^T; \lambda_{tfpc}^W)$, whose logarithm is expressed by:

$$\log \mathcal{R}(\mathbf{x}_t^T; \mathbf{x}_t^W | \lambda_{tfpc}^T; \lambda_{tfpc}^W) = \log \mathcal{L}(\mathbf{x}_t^T | \lambda_{tfpc}^T) - \log \mathcal{L}(\mathbf{x}_t^W | \lambda_{tfpc}^W)$$

Cepstral Analysis. In the case of cepstral analysis, spectral vectors are transformed through cosine functions in order to form cepstral vectors [11]. Then, these cepstral vectors are augmented by their Δ parameters [3]. We finally obtain the sequence $\{\mathbf{c}_t\}$ containing cepstral coefficients and Δ -cepstral coefficients. Let λ_{ceps}^T and λ_{ceps}^W denote respectively the model for the target speaker and the world model in the case of cepstral analysis. We then calculate the likelihood of \mathbf{c}_t given the model λ_{ceps}^T , denoted by $\mathcal{L}(\mathbf{c}_t | \lambda_{ceps}^T)$, the likelihood of \mathbf{c}_t given the model λ_{ceps}^W , denoted by $\mathcal{L}(\mathbf{c}_t | \lambda_{ceps}^W)$, and finally the likelihood ratio, denoted by $\mathcal{R}(\mathbf{c}_t | \lambda_{ceps}^T; \lambda_{ceps}^W)$, whose logarithm is expressed by:

$$\log \mathcal{R}(\mathbf{c}_t | \lambda_{ceps}^T; \lambda_{ceps}^W) = \log \mathcal{L}(\mathbf{c}_t | \lambda_{ceps}^T) - \log \mathcal{L}(\mathbf{c}_t | \lambda_{ceps}^W)$$

Smoothing. Before applying the segmentation algorithm, smoothing is needed to attenuate variations of the log-likelihood ratio. The smoothing is an arithmetic mean of a specified number of consecutive values of the log-likelihood ratio. Two parameters define the smoothing: the number of values used for the average calculation, denoted by τ ; and the delay between two calculations, that is, the number of feature vectors between two average calculations, denoted by d . In our experiments, τ was set to 30 vectors (320 ms) and d to 5 vectors (70 ms).

Segmentation Algorithm. A segmentation algorithm using multiple thresholds [13, 9] is then applied on the average values previously calculated. The algorithm is illustrated on Figure 1.

On this figure, the smoothed log-likelihood ratio is plotted in solid line. When its value becomes higher than the threshold θ_0 , the current time is recorded as the beginning of a possible segment. But the beginning of the segment is marked only if the value of the smoothed log-likelihood ratio becomes also higher than the threshold θ_1 . This procedure allows us to avoid the detection of a segment when the score fluctuates just a little bit around θ_0 , and therefore reduces the number of false alarms. The same principle is applied for the detection of the end of a segment, with the two thresholds θ_2 and θ_3 . Figure 1 shows an example with the values

¹Gender and handset are automatically detected during the tracking phase.

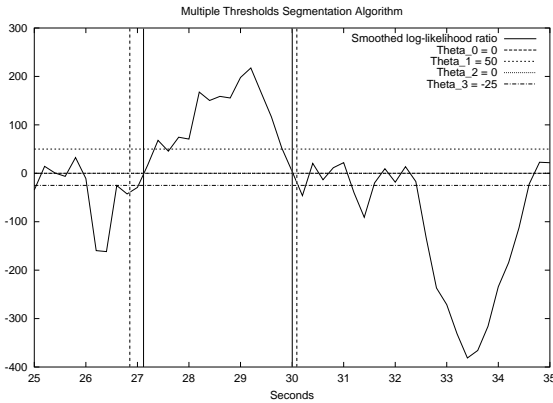


Figure 1: Illustration of the segmentation algorithm using multiple thresholds. The vertical solid lines are the boundaries detected by the segmentation algorithm. The vertical dash lines are the real boundaries of the segment.

$$\theta_0 = \theta_2 = 0, \theta_1 = 50, \text{ and } \theta_3 = -25.$$

Finally, the minimum duration for an estimated segment is set to 2.5 seconds (every segment whose duration is smaller is omitted), and the minimum interval between two consecutive segments to 1 second (two consecutive segments are merged if the interval between them is smaller). The segmentation algorithm provides the estimated beginning and end times for the target speaker segments.

4. EXPERIMENTS

Task. Given a telephone conversation between two persons, and a target speaker, the speaker tracking task consists in tracking the target speaker in the telephone conversation. The target speaker may or may not be one of the two speakers engaged in the conversation.

Database. The database used for these experiments was a subset of the Switchboard 2 - Phase 3 corpus collected by the Linguistic Data Consortium (LDC). The whole corpus consists of 5 minute telephone conversations involving 640 speakers, most of them being college students in the southern United States. The population of speakers used for the experiments was composed of 120 speakers (80 females and 40 males). For each of them, two sentences of approximately 1 minute of speech were available for the training phase. The test data was composed of 66 telephone conversations, which were tested against 4 hypothetical target speakers, 2 of them being the 2 speakers of the conversation.

Data from the Switchboard database was also used to build world models. These models were used to normalize the log-likelihood scores during the tracking phase. These models were gender-dependent and handset-dependent. Therefore, we had one world model for females using electret handset, one world model for females using carbon handset, and so on. Data for these models were obtained from a subset composed of 100 speakers (25 females using electret handset, 25 females using carbon handset, 25 males using electret handset, 25 males using carbon handset).

Spectral Analysis. For each speaker of the training dataset, the two training sentences were concatenated. Then, the speech signal was decomposed in frames of 20 ms at a frame rate of 10 ms. A Hamming window was applied to each frame. The signal was pre-emphasized with a coefficient 0.95. For each frame, a fast Fourier transform was computed and provided 256 square modulus values representing the short term power spectrum in the 0-4 kHz band. This Fourier power spectrum was then used to compute 24 filterbank coefficients, using triangular filters placed on a linear frequency scale in the bandwidth 300-3400 Hz. We finally took the base 10 logarithm of each filter output and multiplied the result by 10, to form a 24-dimensional vector of filterbank coefficients in dB. The same spectral analysis was applied to the data of the world model.

TFPC and Cepstral Analysis. Once spectral vectors have been extracted from a training utterance, we calculated the TFPC corresponding to that sequence using the value $q = 2$ (which corresponds to a context of 5 spectral vectors), and kept only the 24 first components. We then filtered the spectral vectors by these components to obtain a new set of feature vectors. Therefore, for each speaker, we had a set of components for the application of the TFPC filtering, and a sequence of vectors filtered by these components. The TFPC analysis was also applied to the spectral vectors extracted from the world data.

We also extracted 12 cepstral coefficients [11] from the spectral vectors, and computed their Δ parameters [3], obtaining 24-dimensional vectors.

Modeling. Each training sequence was then modeled by a Gaussian mixture model (GMM) [15, 12] using 256 or 512 components and diagonal covariance matrices. The Gaussian mixture models were calculated using an expectation maximization (EM) algorithm, initialized by a vector quantization (VQ) algorithm. The same modeling was used for each world model using 512 components and diagonal covariance matrices.

Tracking Phase. Each test consisted in a telephone conversation between two unknown speakers and the identity of one target speaker that we wanted to track in the telephone conversation. We applied the segmentation algorithm described in Section 3.

5. RESULTS AND DISCUSSION

The performance of the system was measured by a DET curve [10]. This representation is a way to show all of the possible operating points of a system (false alarm rate vs. miss rate) in a scale which makes the result curves rather linear. For this task, the false alarm rate (or false acceptance rate) and the miss rate (or false rejection rate) were defined in the following way:

$$\mathcal{R}_{FA} = \frac{\text{Number of non-target frames labeled as target}}{\text{Number of non-target frames}}$$

$$\mathcal{R}_{MI} = \frac{\text{Number of target frames labeled as non-target}}{\text{Number of target frames}}$$

Results are reported on Figure 2 and show the performance of the system for the two speech parameterizations and for two model sizes, 256 or 512 components.

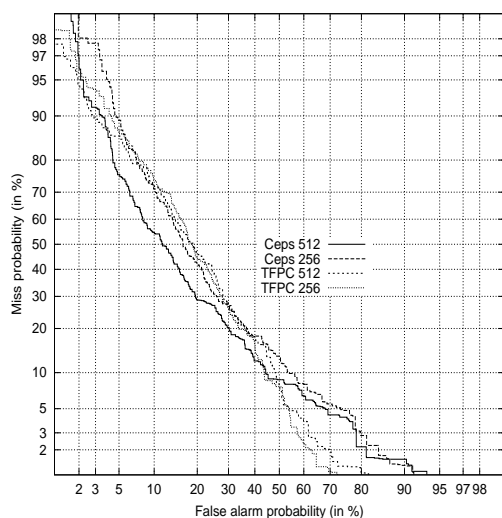


Figure 2: Comparison between a system based on the TFPC analysis and a system based on the cepstral analysis for two model sizes (256 or 512 components).

The first remark is that the four configurations of the system perform very similarly. The cepstral-based systems are better in the middle range of the DET curve, in particular around the EER (Equal Error Rate) which is approximately 26 % for the system using 512 components. But the TFPC-based systems perform slightly better in the extremities of the DET curve, particularly when the false alarm probability is high and the miss probability is low. It is also interesting to notice that, in the case of the TFPC-based systems, the two model sizes give similar performances.

6. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have used a speaker tracking system based on a log-likelihood ratio calculation and a multiple thresholds segmentation algorithm to track target speakers in telephone conversations. Two speech analysis techniques have been tested. The first one, the time-frequency principal component analysis, is a new speech analysis technique based on the extraction of the principal components of the contextual covariance matrix, which is the covariance matrix of feature vectors expanded by their time context. The second one is the classical cepstral analysis.

The results reported show very similar performance between the two speech parameterizations. The cepstral-based system performs slightly better in the middle range of the DET curves, which corresponds to the area where the two types of error rates are close to each other. The TFPC-based system performs better on the extremities of the DET curves, which corresponds to the areas where one error rate is much higher than the other one. If the chosen operating point is in the middle part of the DET curve, one will choose a cepstral-based system. On the contrary, if the chosen operating point is in the extremities of the DET curve, one will rather choose a TFPC-based system.

The performances shown in this paper correspond to a raw calculation of the log-likelihood ratios. It may be interesting to use

a normalization technique like the z-norm or the h-norm in order to improve the performances. These normalization techniques are based on a modeling of the score distributions. It has been shown to work pretty well for speaker verification [1]. The authors intend to test these normalization techniques in the framework of speaker tracking.

Another perspective to this work is the use of a Maximum A Posteriori (MAP) training algorithm, as it is presented in [14], instead of a Maximum Likelihood (ML) training algorithm.

7. REFERENCES

- [1] Robert B. Dunn, Douglas A. Reynolds, and Thomas F. Quatieri. Approaches to speaker detection and tracking in multi-speaker audio. *Digital Signal Processing*, 10(1-3), April 2000.
- [2] Corinne Fredouille, Jean-François Bonastre, and Teva Merlin. AMIRAL: A block-segmental multi-recognizer architecture for automatic speaker recognition. *Digital Signal Processing*, 10(1-3), April 2000.
- [3] Sadaoki Furui. Comparison of speaker recognition methods using static features and dynamic features. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(3):342–350, June 1981.
- [4] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.
- [5] Johan Koolwaaij and Lou Boves. Local normalization and delayed decision making in speaker detection and tracking. *Digital Signal Processing*, 10(1-3), April 2000.
- [6] Ivan Magrin-Chagnolleau and Geoffrey Durou. Time-frequency principal components of speech: application to speaker identification. In *Proceedings of EUROSPEECH 99*, pages 759–762, September 1999. Budapest, Hungary.
- [7] Ivan Magrin-Chagnolleau and Geoffrey Durou. Application of time-frequency principal component analysis to speaker verification. *Digital Signal Processing*, 10(1-3), April 2000.
- [8] Ivan Magrin-Chagnolleau, Geoffrey Durou, and Frédéric Bimbot. Application of time-frequency principal component analysis to text-independent speaker identification. *Submitted to IEEE Transactions on Speech and Audio Processing*.
- [9] Ivan Magrin-Chagnolleau, Aaron E. Rosenberg, and S. Parthasarathy. Detection of target speakers in audio databases. In *Proceedings of ICASSP 99*, pages 821–824, March 1999. Phoenix, Arizona, United States.
- [10] A. Martin et al. The DET curve in assessment of detection task performance. In *Proceedings of EUROSPEECH 97*, volume 4, pages 1895–1898, September 1997. Rhodes, Greece.
- [11] Alan V. Oppenheim and Ronald W. Schaffer. Homomorphic analysis of speech. *IEEE Transactions on Audio and Electroacoustics*, 16(2):221–226, June 1968.
- [12] Douglas A. Reynolds and Richard C. Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83, January 1995.
- [13] Aaron E. Rosenberg, Ivan Magrin-Chagnolleau, S. Parthasarathy, and Qian Huang. Speaker detection in broadcast speech databases. In *Proceedings of ICSLP 98*, December 1998.
- [14] Mouhamadou Seck, Raphaël Blouet, and Frédéric Bimbot. The IRISA / ELISA speaker detection and tracking systems for the NIST'99 evaluation campaign. *Digital Signal Processing*, 10(1-3), April 2000.
- [15] D.M. Titterton, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and sons, 1985.

Audio Indexing : What has been accomplished and the Road Ahead (2002)

Ivan Magrin-Chagnolleau and Nathalie Parlangeau-Vallès. **Audio Indexing : What has been accomplished and the Road Ahead.** *Proceedings of JCIS 2002*, pp. 911-914, Durham, North Carolina, USA, March 2002 (Invited paper).

4 pages

AUDIO INDEXING: WHAT HAS BEEN ACCOMPLISHED AND THE ROAD AHEAD

Ivan Magrin-Chagnolleau⁽¹⁾ and Nathalie Parlangeau-Vallès⁽²⁾

⁽¹⁾Laboratoire Dynamique Du Langage (UMR 5596)

Université Lumière Lyon 2 & CNRS – 14, avenue Berthelot – 69363 Lyon Cedex 07 – France

⁽²⁾LORIA, INRIA Lorraine

Campus Scientifique – BP 239 – 54506 Vandœuvre-les-Nancy – France

ivan@ieec.org - Nathalie.Parlangeau-Valles@loria.fr

ABSTRACT

This paper presents an overview of audio indexing, which has emerged very recently as a research topic with the development of Internet. A lot of data, including audio data, are currently not indexed by web search engines, and audio indexing consists in finding good descriptors of audio documents which can be used as indexes for archiving and search. We discuss speech/music segmentation, language identification, speaker tracking and speaker indexing, and propose some research directions for other audio descriptors which have not been used in the framework of audio indexing, namely key sounds detection, keywords detection, and themes detection. We finally conclude this overview and give a few promising and key perspectives.

1. INTRODUCTION

Internet has become a very important vector of communication during the last few years. There are currently about 400 millions of Internet users, 2 billions of pages directly accessible on 7 millions of public servers¹. There is an urgent need to classify this huge amount of data. Most of the search engines currently access mainly the HTML pages (or equivalent textual data). But there is an important part of the data which is not accessible, because the data are not indexed, have a dynamic content, or belong to a category which is not easily indexable. All these data belong to what is called the *invisible web*, including audio and video documents.

In this paper, we describe some techniques used to structure, and possibly index, audio documents. Although automatic transcription has reached a good performance level recently, it is not the only kind of information which can be extracted from an audio document. A lot of non-verbal information are also very structuring for an audio document, and can lead to the extraction of pertinent descriptors. For instance, as seen on Fig. 1, we can separate speech segments from music segments, detect key sounds (like jingles), identify the language of a segment, track a known speaker or

segment by speakers, detect some keywords, or extract the main themes.

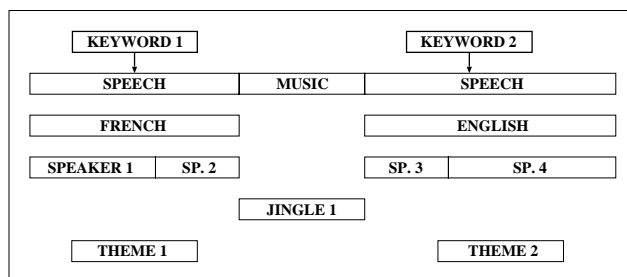


Fig. 1. Some descriptors which can be extracted from an audio document.

2. SPEECH/MUSIC TRACKING AND SEGMENTATION

The first task of interest is the tracking of speech and/or music segments in order to segment an audio document into speech and music portions. Most of broadcast news transcription systems use this kind of separation of speech/music segments in order to confine the application of speech transcription systems on speech segments[1]. This segmentation is often related to a specific structuration of the document (advertisements, jingles, etc.), and it seems important to keep it as a descriptor in an audio indexing system.

Generally, speech/music segmentation is based on the acoustic differences between the two sorts of sounds. [2, 3, 4, 5] have based their discrimination on sets of indices like the mean and the variance of the zero-crossing rate, the spectral "flux" which is defined as the balancing point of the spectral magnitude distribution, etc. All these features usually have interesting discriminating properties as they take very different values for each category of sound.

Various classifiers are commonly used: Gaussian Mixture Models (GMM), k-Nearest-Neighbors (kNN), Neural Networks (NN), etc. In [6, 7], a classical cepstral parame-

¹See <http://www.elec.unice.fr/pages/veille/veille.html>, in French.

terization associated with GMM's gives good results.

What seems important in speech/music separation is the notion of independency between the two tracked cues. In [Fon00], the "differentiated modeling" permits to exploit the structural differences between speech and music. Each class has its own set of parameters (cepstra for speech, filters for music) and models (a 32-component GMM for speech, and a 10-component GMM for music). The decision is totally independent and a module of fusion gives the final segmentation, including mixed segments.

3. LANGUAGE IDENTIFICATION

Language identification consists in identifying the language of an audio document or of a segment in an audio document. This task has emerged recently in the framework of multilingual speech recognition, but has not been used in the framework of audio indexing.

The state-of-the-art approach uses a cepstral-based acoustic analysis and several language-dependent phone recognizers followed by n-gram language models specific of each language to identify [8]. This approach requires that labeled training data be available for several languages, but not necessary for all.

Among the alternative approaches that have been proposed, some researchers work at the level of the acoustic analysis [9, 10], and other propose to explore various cues to identify a language, as rythm [11] for instance.

4. SPEAKER TRACKING

Speaker tracking consists in estimating a beginning and an ending time for each segment in which a target speaker is speaking.

This task is usually tackled by a statistical approach. To detect a target speaker in an audio document, two models are usually built: a target speaker model and a background model which is intended to represent speech from speakers other than the target speaker or other types of sounds. These models are built using feature vectors extracted from labeled segments. In the case of several target speakers, at least one model for each target speaker is needed. Finally, more generally, there could be several background models, representing different types of non-target speaker sounds, and several models for each target speaker, representing different speech qualities. This approach has been adopted for instance in [12, 13].

5. SPEAKER INDEXING

Speaker indexing of an audio database consists in organizing the audio data according to the speakers present in the database. It is composed of three steps (Figure 2). The first step is the segmentation of each audio document by speakers. The segmentation produces a set of speaker-based segmented portions, that is, a set of *speaker utterances*. The

second step consists in tying the various speaker utterances among several previously segmented audio documents. During this stage, one label is attributed to all the speaker utterances matched together. The last stage corresponds to the creation of a speaker-based index.

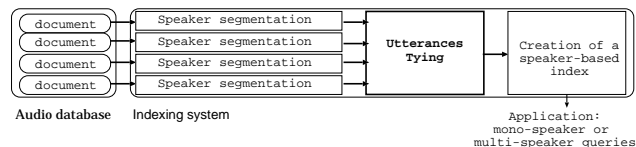


Fig. 2. Block-diagram of a speaker indexing system (after [19]).

The speaker segmentation problem is usually addressed by one of the two following methods. The first one (described in [14][15][16]) relies on a speaker change detection followed by a clustering step. The second one (see for instance [17][18]) does the segmentation and the clustering simultaneously using a hidden Markov model. In both cases, the system has to determine the set of speakers present within a given audio document as well as the boundaries of each intervention. No *a priori* information on speakers is used in these approaches, *i.e.* the speaker utterance models are built during the segmentation process.

Speaker utterances tying [19] is a classification problem similar to speaker clustering [15][16]. Speaker clustering is usually done inside one audio document, whereas, in speaker utterances tying, utterances are matched among several audio documents. However, similar segments inside one audio document are already matched during the preliminary segmentation process. Moreover, the matched utterances are longer than the segments used in speaker clustering. But the great number of channels in the set of audio documents represents an additional difficulty.

The last step of speaker indexing is the creation of a speaker-based index which remains an open and difficult problem for real applications. There are very few papers on this topic. The aim of a speaker-based index is to organize the matched speakers to make the search in a database more efficient.

6. OTHER TYPES OF INFORMATION

Broadcast news systems are based on a complete transcription of speech segments. As far as indexing systems are concerned, moreover for web search applications, keyword detection is preferred. Keyword spotting consists in detecting a more or less important set of keywords from the speech stream. This process gives the exact time position of a keyword.

Word-spotting systems based on hidden Markov models are considered more efficient at modeling arbitrary speech than template based systems [20, 21]. Two main approaches

are found in the literature. The most obvious is to use a large vocabulary continuous speech recognition system (LVCSR) to produce a word string. Then, search algorithms are applied for keyword detection in that string [22]. This approach is considered as giving the best results [23, 24]. Another common approach is based on the use of keyword and filler models. These latest represent the non-keyword intervals of the utterance [23]. Models can be sub-word keyword models like phonetic models or can be whole-word models. [25] gives a very interesting overview of re-scoring methods applied to these two kinds of models: sub-word models are shown to yield a higher hit rate. An original approach based on a HMM-based acoustic decoder combines a multi-keyword spotter and a RNN prosodic model [26]. Prosodic information intends to give better boundaries of words assuming that speakers emphasize keywords more than non-keywords in a sentence. The last relevant approach is based on a modification of the Viterbi search to compute normalized scores which indicate the matching of the keywords at distinct time positions in the utterance. Scores of the keywords are compared with decision thresholds [27]. This is used efficiently as input in a post-processor stage in [28]. Although LVCSR approaches give better results, it requires task-specific knowledge and large training databases. This have a high computational cost. The filler approach is interesting to avoid this last inconvenient and also to maintain domain independence. But detailed filler models increase the performances and then most of the disadvantages of this approach are lost. Systems based on a real acoustic decoder are interesting in terms of domain independence too. This last method seems to have the advantages of needing quite a small amount of training data and does not require filler models.

The topic detection and tracking (TDT) program proposes a definition of a topic as "a seminal event or activity along with all directly events and activities" [29]. As far as speech is concerned, a topic is commonly defined by a set of relevant words [30]. This implies the use of an automatic speech recognizer as a front-end. Recent approaches consider that phoneme-based methods are more accurate. Major advantages of such an approach have been pointed out in [31]. An interesting article presents a method based on phoneme n-grams [32]. Documents or parts of it will be considered to be "on topic" whenever the story is directly connected to the associated event. We can consider three different tasks: story segmentation, topic detection, and topic tracking. The story segmentation is the task of segmenting the stream of speech into topically cohesive stories. The tracking task consists in associating incoming stories with known topics. Topic detection is the task of detecting and tracking topics not previously known by the system. These three tasks can be integrated into an automatic speech indexing system, even if the most useful is usually the track-

ing task. The association between a basic unit (word or sequence of phonemes) and a topic is then based on similarity measures, or occurrence frequencies between what is called the conversational vector and the topic-specific vector. A lot of problems are addressed through the topic tracking task and not all problems are yet resolved.

Finally, if speech cues are well known, music segments can also be exploited to find particular cues like rythm [33], musical sentences [34], pitch [35], notes, instrumental parts, singing voices, melody [36, 37], etc.

7. CONCLUSION AND PERSPECTIVES

In this paper, we have presented the current state of research in audio indexing. We have discussed speech/music segmentation, language identification, speaker tracking and speaker indexing, and proposed some research directions for other audio descriptors which have not been used in the framework of audio indexing, namely key sounds detection, keywords detection, themes detection, and other musical descriptors. A lot of work needs to be done in audio indexing. In particular, after the extraction of audio descriptors, there is another important step consisting in organizing all these descriptors to make the search and the navigation in an audio database more efficient. We intend to tackle some of these issues in the ongoing RAIVES project, which is a research project on audio indexing. We also hope that this paper will stimulate some research in that field, which still needs to be largely explored.

8. ACKNOWLEDGMENT

We would like to thank the CNRS (French national center for scientific research) for its support to the RAIVES project. We also would like to thank Régine André-Obrecht, Frédéric Bimbot, Jean-François Bonastre, Sylvain Meignier, François Pellegrino, and Christine Sénac for their help and the fruitful discussions.

9. REFERENCES

- [1] Jean-Luc Gauvain, Lori Lamel, and Gilles Adda, "Audio partitioning and transcription for broadcast data indexing," in *Proceedings of CBMI 99*, 1999, pp. 67–73, Toulouse, France.
- [2] John Saunders, "Real-time discrimination of broadcast speech/music," in *Proceedings of ICASSP 96*, 1996, pp. 993–996, Atlanta, Georgia, United States.
- [3] Eric Scheirer and Malcolm Slaney, "Construction and evaluation of a robust multifeatures speech/music discriminator," in *Proceedings of ICASSP 97*, 1997, pp. 1331–1334, Munich, Germany.
- [4] Eluned S. Parris, Michael J. Carey, and Harvey Lloyd-Thomas, "Feature fusion for music detection," in *Proceedings of EUROASPEECH 99*, 1999, pp. 2191–2194, Budapest, Hungary.
- [5] Michael J. Carey, Eluned S. Parris, and Harvey Lloyd-Thomas, "A comparison of features for speech, music discrimination," in *Proceedings of ICASSP 99*, 1999, pp. 149–152, Phoenix, Arizona, United States.
- [6] Mouhamadou Seck, Frédéric Bimbot, Didier Zudaj, and Bernard Delyon, "Two-class segmentation for speech/music detection in audio tracks," in *Proceedings of EUROASPEECH 99*, 1999, pp. 2801–2804, Budapest, Hungary.

- [7] Mouhamadou Seck, Ivan Magrin-Chagnolleau, and Frédéric Bimbot, "Experiments on speech tracking in audio documents using gaussian mixture modeling," in *Proceedings of ICASSP 01*, 2001, pp. 601–604, Salt Lake City, Utah, United States.
- [8] Marc A. Zissman, "Comparison of four approaches to automatic language identification for telephone speech," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 1, pp. 31–44, January 1996.
- [9] Michal Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot, "Language recognition using time-frequency principal component analysis and acoustic modeling," in *Proceedings of ICSLP 00*, Dec. 2000, Beijing, China.
- [10] Michel Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot, "Acoustic modeling of spoken languages using time-frequency principal component analysis and hidden markov models: Application to language identification," *Submitted to IEEE Transactions on Speech and Audio Processing*.
- [11] Jérôme Farinas and François Pellegrino, "Automatic rhythm modeling for language identification," in *Proceedings of EUROSPEECH 01*, 2001, pp. 2539–2542, Aalborg, Denmark.
- [12] Aaron E. Rosenberg, Ivan Magrin-Chagnolleau, S. Parthasarathy, and Qian Huang, "Speaker detection in broadcast speech databases," in *Proceedings of ICSLP 98*, 1998, pp. 202–205.
- [13] Ivan Magrin-Chagnolleau, Aaron E. Rosenberg, and S. Parthasarathy, "Detection of target speakers in audio databases," in *Proceedings of ICASSP 99*, 1999, pp. 821–824, Phoenix, Arizona, United States.
- [14] S. Chen, J.F. Gales, P. Gopalakrishnan, R. Gopinath, H. Printz, D. Kanevsky, P. Olsen, and L. Polymenakos, "IBM's LVCSR system for transcription of broadcast news in the 1997 HUB4 English evaluation," in *Proceedings of the DARPA speech recognition workshop*, 1998, <http://www.nist.gov/speech/publications/darpa98/>.
- [15] A. Solomonoff, A. Mielke, M. Schmidt, and H. Gish, "Clustering speakers by their voices," in *Proceedings of ICASSP 98*, 1998, Munich, Germany.
- [16] D.A. Reynolds, E. Singer, B.A. Carlson, J.J. McLaughlin G.C. O'Leary, and M.A. Zissman, "Blind clustering of speech utterances based on speaker and language characteristics," in *Proceedings of ICSLP 98*, 1998, pp. 610–613, Sydney, Australia.
- [17] Sylvain Meignier, Jean-François Bonastre, and Stéphane Igounet, "E-HMM approach for learning and adapting sound models for speaker indexing," in *Proceedings of 2001: A Speaker Odyssey*, June 2001, pp. 175–180, Chania, Crete, Greece.
- [18] Lynn D. Wilcox, Don Kimber, and Francine Chen, "Audio indexing using speaker identification," in *Proceedings of the SPIE Conference on Automatic Systems for the Inspection and Identification of Humans*, 1994, pp. 149–157, San Diego, California, United States.
- [19] Sylvain Meignier, Jean-François Bonastre, and Ivan Magrin-Chagnolleau, "Speaker utterances tying among speaker segmented audio documents using hierarchical classification: Towards speaker indexing of audio databases," in *Submitted to ICASSP 02*, 2002.
- [20] Richard C. Rose and D. B. Paul, "A hidden Markov model based keyword recognition system," in *Proceedings of ICASSP 90*, 1990, pp. 129–132, Albuquerque, New Mexico, United States.
- [21] Lynn D. Wilcox and Marcia A. Bush, "HMM based wordspotting for voice editing and indexing," in *Proceedings of EUROSPEECH 91*, 1991, pp. 25–28, Genova, Italy.
- [22] Mitch Weintraub, "Keyword-spotting using SRI's DECIPHER large-vocabulary speech recognition system," in *Proceedings of EUROSPEECH 93*, 1993, pp. 1265–1268, Berlin, Germany.
- [23] Alexandros Manos and Victor Zue, "A segment-based wordspotter using phonetic filler models," in *Proceedings of ICASSP 97*, 1997, pp. 899–902, Munich, Germany.
- [24] Mazin Rahim, "Recognizing connected digits in a natural spoken dialog," in *Proceedings of ICASSP 99*, 1999, pp. 153–156, Phoenix, Arizona, United-States.
- [25] K. M. Knill and Steve J. Young, "Speaker dependent keyword spotting for accessing stored speech," Tech. Rep., CUED/F-INFENG/TR-193, Cambridge University Engineering Department, 1994.
- [26] Wern-Jun Wang, Chun-Jen Lee, Eng-Fong Huang, and Sin-Horng Chen, "Multi-keyword spotting of telephone speech using orthogonal transform-based sbr and rnn prosodic model," in *Proceedings of EUROSPEECH 01*, 2001, pp. 2773–2776, Aalborg, Denmark.
- [27] Jochen Junkawitsch, Günter Ruske, and Harald Höge, "Efficient methods for detecting keywords in continuous speech," in *Proceedings of EUROSPEECH 97*, 1997, pp. 259–262, Patras, Greece.
- [28] Luciana Ferrer and Claudio Estienne, "Improving performance of a keyword spotting system by using a new confidence measure," in *Proceedings of EUROSPEECH 01*, 2001, pp. 2561–2564, Aalborg, Denmark.
- [29] Yiming Yang, Tom Pierce, and Jaime Carbonell, "A study on retrospective and on-line event detection," in *Proceedings of SIGIR 98, 21st ACM International Conference on Research and Development in Information Retrieval*, 1998, pp. 28–36, Melbourne, Australia.
- [30] Hubert Jin, Rich Schwartz, Sreenivasa Sista, and Frederick Walls, "Topic tracking for radio, TV broadcast, and newswire," in *Proceedings of EUROSPEECH 99*, 1999, pp. 2439–2442, Budapest, Hungary.
- [31] Jeremy H. Wright, Michael J. Carey, and Eluned S. Parris, "Statistical models for topic identification using phoneme substrings," in *Proceedings of ICASSP 96*, 1996, pp. 307–310, Atlanta, Georgia, United States.
- [32] Marius W. Theunissen, Konrad Scheffler, and Johan A. du Preez, "Phoneme-based topic spotting on the switchboard corpus," in *Proceedings of EUROSPEECH 01*, 2001, pp. 283–286, Aalborg, Denmark.
- [33] Erling Wold, Thom Blum, Douglas Keislar, and James Wheaton, "Classification, search, and retrieval of audio," in *Handbook of Multimedia Computing*. CRC Press, 1999.
- [34] Chih-Chin Liu, Jia-Lien Hsu, and Arbee L.P. Chen, "Efficient theme and non-trivial repeating pattern discovering in databases," in *Proceedings of ICDE 99*, 1999, pp. 14–21, Sydney, Australia.
- [35] Yoav Medan, Eyal Yair, and Dan Chazan, "Super resolution pitch determination of speech signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 39, no. 1, pp. 40–48, 1991.
- [36] Tomonari Sonoda, Masataka Goto, and Yoichi Muraoka, "A www-based melody retrieval system," in *Proceedings of ICMC 98*, 1998, pp. 349–352, Ann Arbor, Michigan, United States.
- [37] Rodger J. McNab, Lloyd A. Smith, David Bainbridge, and Ian H. Witten, "The New Zealand digital library: MELody inDEX," *D-Lib Magazine*, May 1997, <http://www.dlib.org/>.

3.3 Analyse des signaux sismiques

Empirical mode decomposition based frequency attributes (1999)

Ivan Magrin-Chagnolleau and Richard G. Baraniuk. **Empirical mode decomposition based frequency attributes.** *Proceedings of the 69th SEG Meeting*, Houston, Texas, United States, November 1999.

4 pages

Empirical mode decomposition based time-frequency attributes

Ivan Magrin-Chagnolleau and Richard G. Baraniuk
Rice University, Houston, Texas

Summary

This paper describes a new technique, called the *empirical mode decomposition (EMD)*, that allows the decomposition of one-dimensional signals into intrinsic oscillatory modes. The components, called *intrinsic mode functions (IMFs)*, allow the calculation of a meaningful multicomponent instantaneous frequency. Applied to a seismic trace, the EMD allows us to study the different intrinsic oscillatory modes and instantaneous frequencies of the trace. Applied to a seismic section, it provides new time-frequency attributes.

Introduction

Seismic signals, as with many real-world signals, are non-stationary, making Fourier analysis unsatisfying since the frequency content changes across the time. Exploration of the frequency content of such signals dictates a local study of the frequency content. In time-frequency analysis, we analyze the frequency content across a small span of time and then move to another time position (Flandrin, 1993; Cohen, 1995). The major drawback of most time-frequency *transforms* is that the rectangular tiling of the time-frequency plane does not match the shape of many signals.

On the other hand, basis *decomposition* techniques such as the Fourier decomposition or the Wavelet decomposition (Daubechies, 1992) have also been used to analyze real-world signals. The main drawback of these approaches is that the basis functions are fixed, and do not necessarily match varying nature of signals.

In this paper, we use a new technique, called the *empirical mode decomposition (EMD)*, first introduced by Huang et al. (1998). This technique adaptively decomposes a signal into oscillating components. The different components match the signal itself very well, and the time-frequency representation extracted from the decomposition, called the *Hilbert amplitude spectrum*, does not have any constraint on the way to tile the time-frequency plane. Because the approach is algorithmic, it does not allow to express the different components in closed form.

The EMD is in fact type of adaptive wavelet decomposition whose subbands are built up as needed to separate the different components of the signal.

Empirical Mode Decomposition

The empirical mode decomposition was first introduced by Huang et al. (1998). The principle of this technique is to decompose a signal $x(t)$ into a sum of functions that: (1) have the same numbers of zero crossings and extrema; and (2) are symmetric with respect to the local mean.

The first condition is similar to the narrow-band requirement for a stationary Gaussian process. The second condition modifies a global requirement to a local one, and is necessary to ensure that the instantaneous frequency will not have unwanted fluctuations as induced by asymmetric waveforms. These functions are called *intrinsic mode functions (IMFs)*, and denoted $imf_i(t)$. The intrinsic mode functions are obtained using the algorithm of Table 1.

1. Initialize: $r_0(t) = x(t)$, $i = 1$
2. Extract the i -th IMF:
 - (a) Initialize: $h_0(t) = r_i(t)$, $j = 1$
 - (b) Extract the local minima and maxima of $h_{j-1}(t)$
 - (c) Interpolate the local maxima and the local minima by a cubic spline to form upper and lower envelopes of $h_{j-1}(t)$
 - (d) Calculate the mean $m_{j-1}(t)$ of the upper and lower envelopes
 - (e) $h_j(t) = h_{j-1}(t) - m_{j-1}(t)$
 - (f) if stopping criterion is satisfied then set $imf_i(t) = h_j(t)$ else go to (b) with $j = j + 1$
3. $r_i(t) = r_{i-1}(t) - imf_i(t)$
4. if $r_i(t)$ still has at least 2 extrema then go to 2 with $i = i + 1$ else the decomposition is finished and $r_i(t)$ is the residue.

Table 1: Algorithm for the EMD.

At the end of the algorithm, we have:

$$x(t) = \sum_{i=1}^n imf_i(t) + r_n(t)$$

where $r_n(t)$ is the residue of the decomposition.

Another way to explain how the empirical mode decomposition works is that it picks out the highest frequency oscillation that remains in the signal. Thus, locally, each IMF contains lower frequency oscillations than the one extracted just before. This property can be very useful to pick up frequency changes, since a change will appear even more clearly at the level of an IMF.

To illustrate the EMD, we plot in Fig. 1 a seismic trace, the first three IMFs, and the residue. We can see easily

Empirical mode decomposition based time-frequency attributes

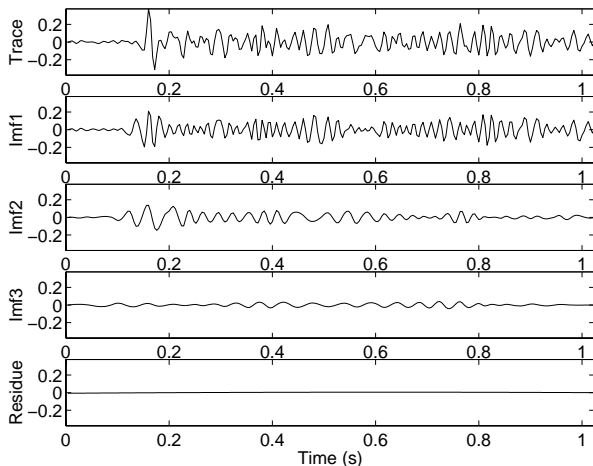


Fig. 1: Illustration of the EMD: seismic trace, first three IMFs, and the residue of the decomposition (the IMFs 4 to 9 are not shown).

that each component has the same number of zero crossings as extrema and is symmetric with respect to the zero line. Hence, we can calculate for each component a meaningful instantaneous frequency. The residue, in this case, has almost a zero value which means that the original signal had a zero mean, as it is the case for most seismic traces. More generally, the residue captures the trend of the original signal.

Complex Traces and Instantaneous Frequency

The notions of complex signals associated with real signals and of instantaneous frequency (Flandrin, 1993; Cohen, 1995) have been first used in seismic analysis by Taner et al. (1979).

Given a real signal $x(t)$, we can build the corresponding analytic signal (or complex trace):

$$X(t) = x(t) + j\mathcal{H}\{x(t)\}$$

where $\mathcal{H}\{x(t)\}$ is the quadrature signal corresponding to $x(t)$ obtained using the Hilbert transform

$$\mathcal{H}\{x(t)\} = \frac{1}{\pi} p.v. \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau$$

with $p.v. \int_{-\infty}^{+\infty}$ being the Cauchy principle value of the integral.

The analytic signal can also be obtained by: (1) taking the Fourier transform of $x(t)$; (2) zeroing the amplitude for negative frequencies and doubling the amplitude for positive frequencies; and (3) taking the inverse Fourier transform.

The complex trace can be expressed as:

$$X(t) = A(t) e^{j\theta(t)}$$

where $A(t)$ is the instantaneous amplitude (or reflection strength) and $\theta(t)$ is the instantaneous phase. We can then define the instantaneous frequency $\omega(t)$ as

$$\omega(t) = \frac{d\theta(t)}{dt}.$$

Instead of calculating the complex trace associated with the seismic trace $x(t)$ itself, we can calculate the complex trace associated with each of the intrinsic mode functions, and thus the instantaneous frequency of each of them, in order to obtain a multicomponent instantaneous frequency.

Fig. 2 represents the instantaneous frequency (normalized between 0 and 0.5) of the original trace and of the first three IMFs.

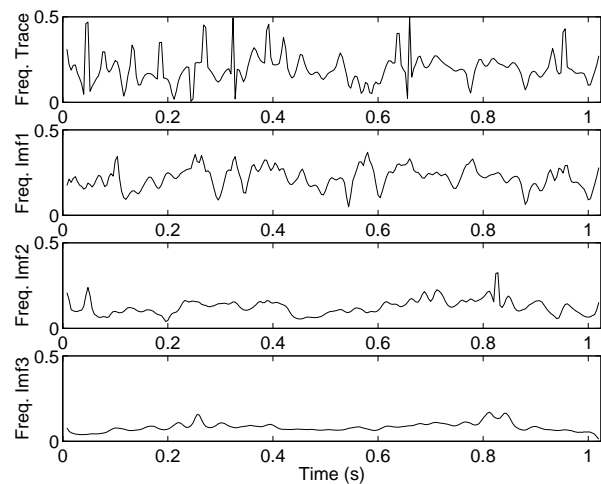


Fig. 2: Normalized instantaneous frequencies of the original trace and of the first three IMFs.

We observe that the instantaneous frequency of the first IMF is very similar to the instantaneous frequency of the original trace, but with much fewer peaks due to noise blow-up, which is a great advantage. The instantaneous frequencies of the other components give additional information about the different oscillatory modes of the original trace.

The fact that the instantaneous frequency of the first IMF possesses so few noise peaks suggests we use it as a time-frequency attribute instead of the instantaneous frequency of the original trace itself.

Hilbert Amplitude Spectrum

Once we have calculated the instantaneous frequency of each IMF, we can represent the Hilbert amplitude spectrum of the initial seismic trace as in Huang et al. (1998), which is obtained in the following way: for each intrinsic mode function $imf_i(t)$, if $\omega_i(t)$ is the corresponding instantaneous function, we represent in the time-frequency plane the triplet $\{t, \omega_i(t), A_i(t)\}$ where $A_i(t)$ is the amplitude of the complex trace associated to $imf_i(t)$.

Empirical mode decomposition based time-frequency attributes

Fig. 3 represents the Hilbert amplitude spectrum of the previous seismic trace, using 512 frequency bins.

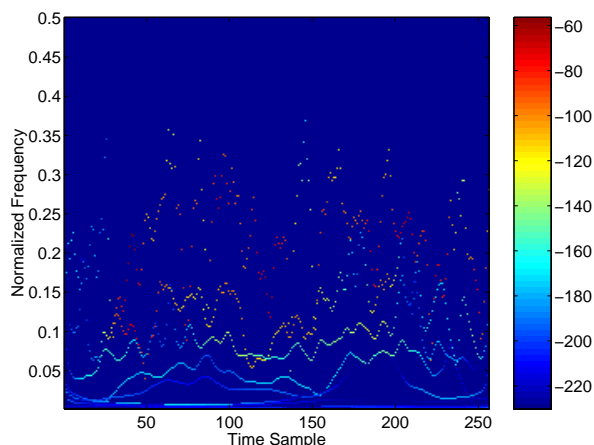


Fig. 3: *Hilbert amplitude spectrum using 512 frequency bins. The amplitude is represented in dB.*

EMD-Based Time-Frequency Attributes

The next step of this analysis is to define new time-frequency attributes based on the EMD. Fig. 4 represents a migrated seismic section from the Alboran sea composed of 100 traces of 256 samples each.

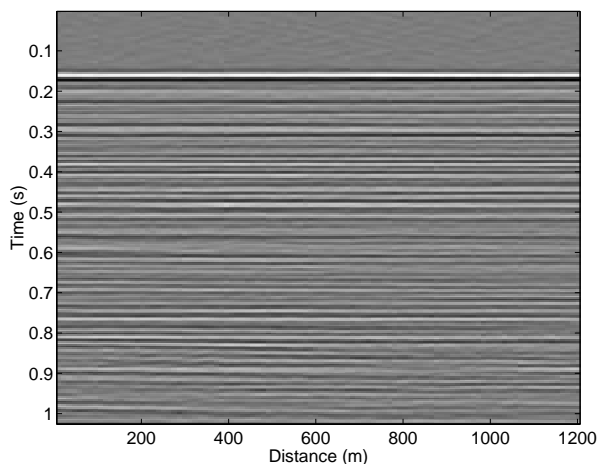


Fig. 4: *Migrated seismic section from the Alboran sea.*

On each trace, we compute the instantaneous frequency as defined in *Taner et al. (1979)*. Fig. 5 represents the instantaneous frequency attribute for the Alboran sea section.

We also compute for each trace the instantaneous frequency of the IMFs obtained by the EMD. The time-frequency attributes represented Fig. 6, Fig. 7, and Fig. 8 were obtained respectively by computing for each trace the instantaneous frequency of the first, second, and third IMFs.

The new time-frequency attributes do not have as many saturated lines as the instantaneous frequency attribute. These saturated lines, corresponding to the peaks which are due to noise blow-up, tend to disappear on the new time-frequency attributes.

The second time-frequency attribute makes apparent a change around 0.42 seconds that may correspond to a change in the stratigraphy of the section.

Conclusion and Future Directions

We have proposed a new methodology to decompose a seismic trace into different oscillatory modes and to extract the instantaneous frequencies of these modes. If we represent all these frequency functions in the time-frequency plane, we have a complete description of the frequency behavior of the trace. We have also extracted new time-frequency attributes from this decomposition, based on an instantaneous frequency calculation of each component of the decomposition.

The next step in this study will be to extract other time-frequency attributes based on different calculations in the time-frequency plane, such as the mean frequency, the local bandwidth, or more generally some higher moments. We can also use other methods for calculating the instantaneous frequencies of the IMFs, for example by using a pseudo-Wigner-Ville transform.

Acknowledgement

The authors would like to thank Guy Drijkoningen for his precious comments during this work. This work has been sponsored by the Rice Consortium for Computational Seismic/Signal Interpretation (www.dsp.rice.edu/ccsi), by the National Science Foundation, grant no. MIP-9457438, and by DARPA/AFOSR, grant no. F49620-97-1-0513.

References

- Cohen, L., 1995, *Time-frequency analysis*: Prentice Hall.
- Daubechies, I., 1992, *Ten lectures on wavelets*: SIAM.
- Flandrin, P., 1993, *Temps-fréquence*: Hermes.
- Huang, N. E., et al., 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis: *Proceedings of the Royal Society Lond. A*, **454**, 903–995.
- Taner, M., Koehler, F., and Sheriff, R., June 1979, Complex seismic trace analysis: *Geophysics*, **44**, no. 6, 1041–1063.

Empirical mode decomposition based time-frequency attributes

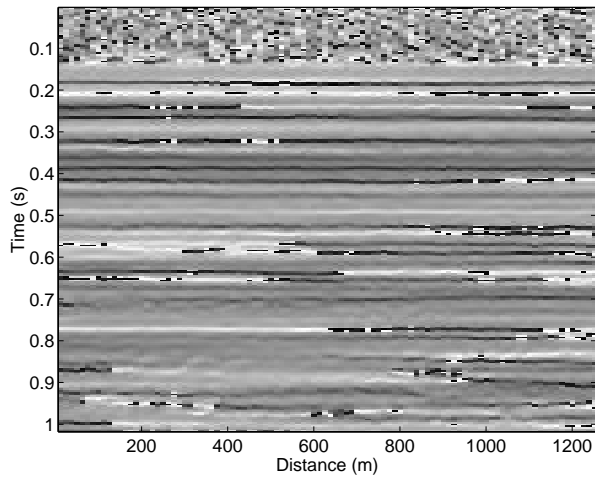


Fig. 5: Frequency attribute based on the calculation of the instantaneous frequency as proposed in Taner et al. (1979).

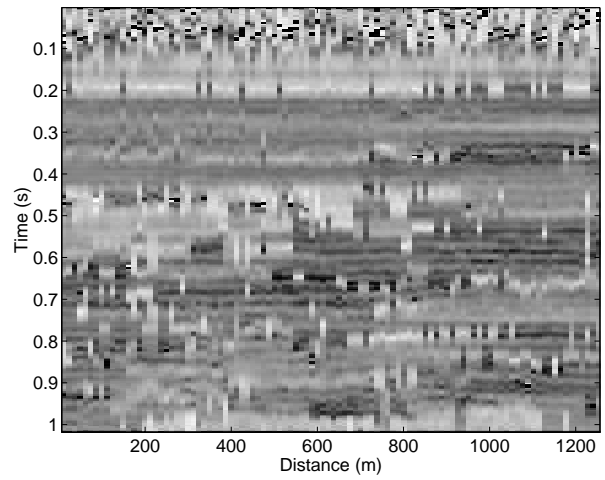


Fig. 7: Time-frequency attribute based on the calculation of the instantaneous frequency of the second IMF.

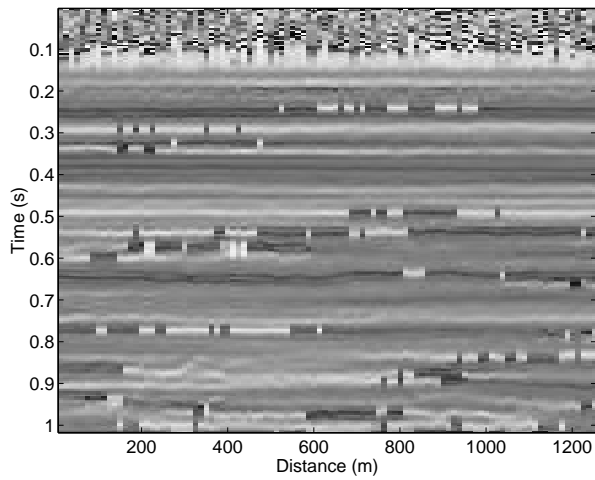


Fig. 6: Time-frequency attribute based on the calculation of the instantaneous frequency of the first IMF.

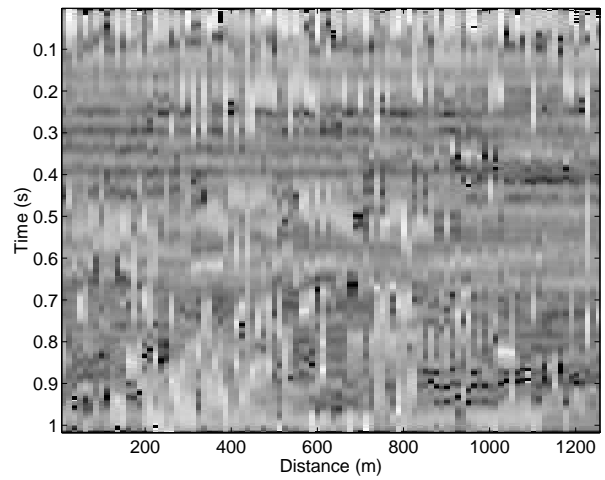


Fig. 8: Time-frequency attribute based on the calculation of the instantaneous frequency of the third IMF.

**Multiscale texture segmentation of dip-cube slices
using Wavelet-domain hidden Markov trees (1999)**

Ivan Magrin-Chagnollean, Hyeokho Choi, Rutger van Spaendonck, Philippe Steeghs and Richard G. Baraniuk. **Multiscale texture segmentation of dip-cube slices using Wavelet-domain hidden Markov trees.** *Proceedings of the 69th SEG Meeting*, Houston, Texas, United States, November 1999.

4 pages

Multiscale texture segmentation of dip-cube slices using wavelet-domain hidden Markov trees

Ivan Magrin-Chagnolleau, Hyeokho Choi, Rutger van Spaendonck, Philippe Steeghs, and Richard G. Baraniuk
Rice University, Houston, Texas

Summary

Wavelet-domain hidden Markov models (HMMs) are powerful tools for modeling the statistical properties of wavelet transforms. By characterizing the joint statistics of wavelet coefficients, HMMs efficiently capture the characteristics of many real-world signals. When applied to images, the model can characterize the joint statistics between pixels, providing a very good classifier for textures. Utilizing the inherent tree structure of wavelet-domain HMMs, classification of textures at various scales is possible, furnishing a natural tool for multiscale texture segmentation. In this paper, we introduce a new multiscale texture segmentation algorithm based on wavelet-domain hidden Markov trees (HMTs). We apply this new technique to the segmentation of dip-cube time slices.

Introduction

The goal of an image segmentation algorithm is to assign a class label to each pixel of the image based on the properties of the pixel and its relation with neighborhood pixels. To capture the properties of each region of the image to be segmented, both the large and small scale behaviors should be utilized to properly segment both large, homogeneous regions and detailed boundary regions. In view of the utilization of properties in different scales, it is natural to approach the segmentation problem using multiscale analysis. There have been efforts to segment using multiscale autoregressive models (Basseville et al., 1992; Fosgate et al., 1997). In Bouman et al. (1994), a multiscale random field model was used to statistically analyze the multiscale structure of the segmentation labels. In this paper, we propose a multiscale texture segmentation algorithm based on the wavelet transform of the image.

For many classes of signals and images, wavelets provide a compact¹ and approximately decorrelated signal representation (Daubechies, 1992). For statistical applications such as detection and estimation, an accurate joint probability model for the wavelet coefficients of the signals plays an important role. In view of the compaction and decorrelation properties of wavelet transforms, modeling the wavelet coefficients instead of direct modeling of the signal may be more efficient and naturally provide a multiscale structure of the involved signal processing.

Recently, a wavelet-domain hidden Markov model (HMM) was proposed as a powerful tool for modeling the probability structure of wavelet transforms (Crouse et al., 1998; Crouse and Baraniuk, 1997). By modeling each wavelet coefficient as a Gaussian mixture and cap-

turing the dependencies between the coefficients as hidden state transitions, HMMs provide a natural setting for exploiting the structure inherent in real-world signals and images for signal detection and classification.

In this paper, we employ the tree structure of wavelet-domain HMMs for multiscale signal classification (Choi and Baraniuk, 1999). By computing the likelihoods of each subblocks of the image in different scales, a “raw” segmentation is obtained. Segmentations in coarser scales are more reliable for large, homogeneous regions, while finer scale segmentations are more appropriate around boundaries between different textures. By combining these raw segmentations in different scales, we can obtain a robust and accurate segmentation result. We accomplish this interscale fusion by setting up a tree-structured model for segmentation results in different scales. By computing the likelihoods of each subblock in fine scale conditioned on the next coarser scale segmentation result, we can make a Bayesian multiscale classification. Unlike the approach in Bouman et al. (1994), we use the concept of “context” to capture the dependencies between segmentation results across different scales, resulting in an algorithm that is more flexible and easier to implement.

Wavelet transform

The discrete wavelet transform (DWT) represents a 1-D signal $z(t)$ in terms of shifted versions of a lowpass scaling function $\phi(t)$ and shifted and dilated versions of a prototype bandpass wavelet function $\psi(t)$ (Daubechies, 1992). For special choices of $\phi(t)$ and $\psi(t)$, the functions $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$, $\phi_{j,k}(t) = 2^{-j}\phi(2^{-j}t - k)$, for j and $k \in \mathbb{Z}$, form an orthonormal basis, and we have the representation:

$$z(t) = \sum_k u_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty} \sum_k w_{j,k} \psi_{j,k}(t)$$

with $u_{j,k} = \int z(t) \phi_{j,k}^*(t) dt$ and $w_{j,k} = \int z(t) \psi_{j,k}^*(t) dt$.

To keep the notation manageable in the sequel, we will adopt an abstract index scheme for the DWT coefficients: $u_{j,k} \rightarrow u_i$, $w_{j,k} \rightarrow w_i$.

For multidimensional signals (e.g. images), the DWT can be easily generalized by alternately applying the 1-D DWT to each coordinate in order to get a 2-D wavelet transform.

Hidden Markov Models

The DWTs of many real-world signals tend to be sparse, with just a few non-zero coefficients containing most of the signal energy. Hence, the probability density function

¹lots of small coefficients and only a few large coefficients

Multiscale texture segmentation of dip-cube slices

(pdf) $f_{W_i}(w_i)$ of each wavelet coefficient is well approximated by a *Gaussian mixture* with two components, one corresponding to the low values of the energy, one to the high values. To each wavelet coefficient W_i , we associate a discrete hidden state S_i that takes on values $m = 1$ or 2 with probability mass function $p_{S_i}(m)$. Conditioned on $S_i = m$, W_i is Gaussian with mean $\mu_{i,m}$ and variance $\sigma_{i,m}^2$. Thus, its overall pdf is given by

$$f_{W_i}(w_i) = \sum_{m=1}^2 p_{S_i}(m) f_{W_i|S_i}(w_i|S_i = m)$$

Wavelet-domain HMMs are multidimensional mixture models in which the hidden states have a Markov dependency structure (Crouse et al., 1998). The idea is to capture the dependencies in the wavelet coefficients through the transition probabilities between the states.

For example, the hidden Markov tree (HMT) model places a tree structure on the hidden states to capture wavelet dependencies across scale. The HMT model is specified via the mixture parameters $(\mu_{i,m}, \sigma_{i,m}^2)$ and transition probabilities $p_{S_i|S_{\rho(i)}}(m|n)$, where $\rho(i)$ denotes the parent of node i . These parameters can be grouped into a model parameter vector Θ . The wavelet-domain HMT is trained to capture the wavelet-domain properties of the signals of interest using the iterative expectation maximization (EM) algorithm.

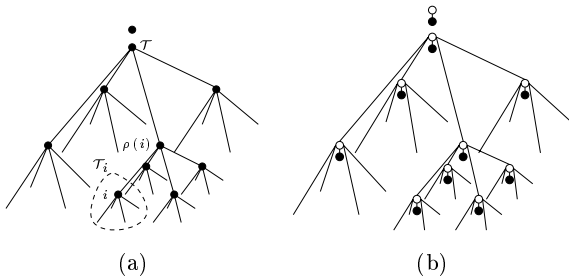


Fig. 1: (a) Quadtree of 2-D wavelet coefficients for each subband, (b) 2-D wavelet-domain HMT model. Black nodes represent wavelet coefficients; white nodes represent hidden state variables.

For 2-D images, quadtrees are used for HMT models (see Fig. 1). In addition, because we have three different subbands in 2-D DWT, we have a distinct quadtree for each subband. The subbands are treated as independent although they are not, but it works well in practice.

Multiscale Classification

For simplicity, we assume that Haar wavelet is used for the underlying wavelet transform. Then, each dyadic subblock of pixels can be associated with a wavelet coefficient, i.e., each node of the quadtree (Fig. 1) corresponds to a dyadic subblock of the entire image. We define \mathcal{T}_i to be the subtree of observed wavelet coefficients with root at node i , so that the subtree \mathcal{T}_i contains coefficient w_i and all of its descendants (Fig. 1). Due to the quadtree structure, each \mathcal{T}_i corresponds to a subblock of the entire image.

Given a HMT model Θ , for each subtree \mathcal{T}_i , we compute the conditional likelihood $\beta_i(m) = f(\mathcal{T}_i|S_i = m, \Theta)$ and the conditional probability $p(S_i = m|\mathbf{w}, \Theta)$, where $\mathbf{w} = \{w_i\}$ is the collection of all wavelet coefficients. Then, the conditional likelihood $f(\mathcal{T}_i|\Theta)$ can be computed as:

$$f(\mathcal{T}_i|\Theta) = \sum_{m=1}^2 \beta_i(m) p(S_i = m|\mathbf{w}, \Theta)$$

Suppose we have HMT models for two different textures, T_1 and T_2 , whose parameters are respectively Θ_{T_1} and Θ_{T_2} . The classification in each scale is accomplished by comparing the likelihood functions $f(\mathcal{T}_i|\Theta_{T_1})$ and $f(\mathcal{T}_i|\Theta_{T_2})$ for every node i . For each i , the subblock of image corresponding to the subtree \mathcal{T}_i can be classified, producing a raw segmentation of the image up to 2×2 pixels image blocks. We call this segmentation a “raw” segmentation because we do not use any relationship between segmentations across different scales.

We expect this “raw” segmentation to be more reliable at coarse scales, because classification of larger image blocks is more reliable due to its richness in statistical information. However, at coarse scale, the boundaries between different textures cannot be captured accurately. At fine scales, the segmentation is less reliable, while the boundary is better captured.

When we compute the likelihoods $f(\mathcal{T}_i|\Theta)$, we ignore the scaling coefficients in the wavelet transform. This implies that we do not rely on the local brightness levels of the image for segmentation. Only the joint statistics between pixels are used as the discriminating feature. This is why only up to 2×2 blocks can be classified. The brightness of each pixel will be used to obtain the pixel level segmentation at the final stage of interscale fusion algorithm explained next.

Interscale Fusion

By combining the segmentation results from different scales, we can obtain better segmentation of the image. Because the “raw” segmentation is more reliable in coarse scales, the segmentation result in coarse scales can be used as extra information for segmentation in finer scales. When a subblock of an image at a given scale is classified as a certain texture, it is very likely that the corresponding subblocks in the next finer scale should be also classified as the same texture. This coarse-to-fine dependencies can be modeled as a 2-D tree (labeling tree), where each node corresponds to a subblock of image. The interscale dependencies are described as connecting branches between nodes. However, as the number of branches increases, analysis of this tree becomes extremely difficult.

Here, we propose a context-based method (Crouse and Baraniuk, 1997) to capture the interscale dependencies, which is easy to manipulate even when the number of incorporated neighborhood nodes is large. Let i be a node of the 2-D labeling tree. Denote the class label for the subblock of image corresponding to i as C_i , which takes values $1, \dots, N$, where N is the number of different textures to be classified. Let v_i denote a context vector that

Multiscale texture segmentation of dip-cube slices

describes the segmentation result in the next coarser scale. That is, v_i represents the segmentation results in the previous coarse scale in the neighborhood of the region centered at i .

The choice of v_i is not restricted at all, as long as it describes the results of the coarse scale segmentation well. For example, in the example of two textures T_1 and T_2 , v_i can be chosen to be the number of neighborhood blocks in the previous coarse scale that were classified as T_1 , with appropriately pre-specified neighborhood region for i .

In some sense, the context v_i indicates whether the part of image corresponding to node i falls in the interior of a homogeneous texture, or whether it is near a boundary between two different textures. Let $J(i)$ denote the scale of the node i , and denote the entire image (or equivalently the wavelet transform) as \mathcal{T} . Let k be the scale under consideration. And, let \mathbf{v}^k denote the collection of all contexts at scale k . The contextual Bayes classification starts by estimating the probabilities $p_{C_i|V_i}(n|v_i)$ to maximize the likelihood of the entire image computed in scale k , given as

$$f(\mathcal{T}|\mathbf{v}^k) = \prod_{i \text{ s.t. } J(i)=k} \sum_{n=1}^N p_{C_i|V_i}(n|v_i) f(\mathcal{T}_i|C_i = n)$$

where $f(\mathcal{T}_i|C_i = n)$'s are given by the EM algorithm applied to HMTs for each candidate texture. The values of $p_{C_i|V_i}(n|v_i)$ are obtained by assuming that they are constant in each scale and averaging over all the nodes in that scale. In practice, we do not specify $p_{C_i|V_i}(n|v_i)$ directly, but rather specify $p_{V_i|C_i}(v_i|n)$ and apply Bayes rule. Thus, the actual probabilities that should be computed are $\epsilon_{i,n} = p_{C_i}(n)$ and $\alpha_{i,v,n} = p_{V_i|C_i}(v_i|n)$. Once $\epsilon_{i,n}$ and $\alpha_{i,v,n}$ are computed, the context-based Bayes classification is performed by computing

$$p_{C_i|V_i,\mathcal{T}}(n|v_i) = \frac{\epsilon_{i,n} \alpha_{i,v,n} f(\mathcal{T}_i|C_i = n)}{\sum_{n=1}^N \epsilon_{i,n} \alpha_{i,v,n} f(\mathcal{T}_i|C_i = n)}$$

Then, we label node i so that $p_{C_i|V_i,\mathcal{T}}(n|v_i)$ is maximized. In practice, the computation of $\epsilon_{i,n}$, $\alpha_{i,v,n}$, and $p_{C_i|V_i,\mathcal{T}}$ is performed simultaneously by an iterative method similar to the EM algorithm. To obtain pixel level segmentation, the probability distribution of a single pixel for each of the textures is necessary. This distribution can be modeled from the histogram of pixel intensities for each texture. For textures, it can be well approximated as a Gaussian mixture model. The problem of pixel-level modeling and segmentation is discussed in more detail in Choi and Baraniuk (1999).

Application to dip-cube slices

Fig. 2 and Fig. 5 represent two time slices of a dip-cube obtained by using a 3-D local Radon power spectra calculation on a 3-D data set from the Ameland survey, located in the Dutch sector of the North Sea (Steeghs et al., 1998). We have selected two 64×64 blocks from each of the two original images to train the HMTs, each block corresponding to two different kinds of texture. The training was performed with intra-scale tying (Crouse et al., 1998)

to avoid overfitting the models. Fig. 3 and Fig. 6 show the raw segmentation results without the interscale processing. Fig. 4 and Fig. 7 show the segmentation results after using the interscale fusion algorithm. As shown on the figures, we obtain excellent results for this two texture segmentation example, particularly on the second image, and we can see more clearly the channel on Fig. 4, and even more clearly the Salt structure on Fig. 7. A better choice of the training images may improve the results for the first image.

Conclusion

We have introduced a new framework for texture segmentation based on wavelet-domain hidden Markov models. By concisely modeling the statistical behavior of textures at multiple scales and by combining segmentations obtained at these scales, the algorithm produces a robust and accurate segmentation of texture images. We have applied this new segmentation technique to dip-cube time slices in order to help the interpretation of the slices. The next step will be to produce a 3-D texture segmentation of dip-cubes, including more than two classes.

Acknowledgement

This work has been sponsored by the Rice Consortium for Computational Seismic/Signal Interpretation (www.dsp.rice.edu/ccsi), by the National Science Foundation, grant no. MIP-9457438, and by DARPA/AFOSR, grant no. F49620-97-1-0513.

References

- Basseville, M., et al., March 1992, Modeling and estimation of multiresolution stochastic processes: IEEE Transactions on Information Theory, **38**, no. 2, 766-784.
- Bouman, C., and Shapiro, M., March 1994, A multiscale random field model for Bayesian image segmentation: IEEE Transactions on Image Processing, **3**, no. 2, 162-177.
- Choi, H., and Baraniuk, R. G., 1999, Multiscale texture segmentation using wavelet-domain hidden Markov models: Submitted to IEEE Transactions on Image Processing.
- Crouse, M., and Baraniuk, R., November 1997, Contextual hidden Markov models for wavelet-domain signal processing: Proceedings of the 31st Asilomar Conference.
- Crouse, M., Nowak, R., and Baraniuk, R., April 1998, Wavelet-based statistical signal processing using hidden Markov models: IEEE Transactions on Signal Processing, **46**, no. 4, 886-902.
- Daubechies, I., 1992, Ten lectures on wavelets: SIAM.
- Fosgate, C., et al., January 1997, Multiscale segmentation and anomaly enhancement of SAR imagery: IEEE Transactions on Image Processing, **6**, no. 1, 7-20.
- Steeghs, P., Fokkema, J. T., and Diephuis, G., 1998, 3-D local Radon power spectra for seismic attribute extraction: Proceedings of the 68th SEG Meeting.

Multiscale texture segmentation of dip-cube slices

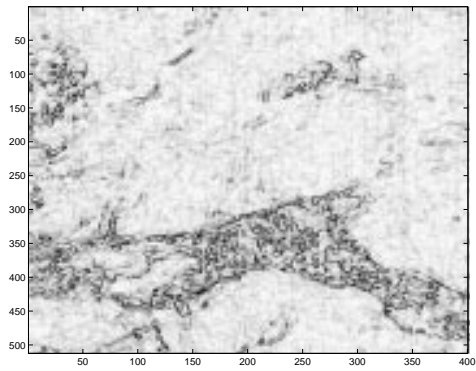


Fig. 2: *First time slice of the dip-cube.*

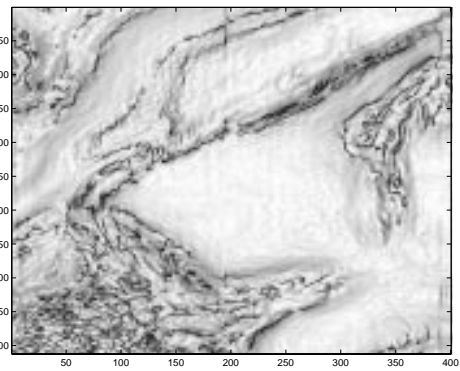


Fig. 5: *Second time slice of the dip-cube.*

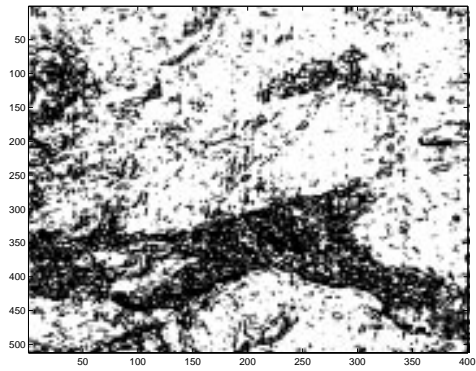


Fig. 3: *Raw segmentation without the interscale processing for the first time slice.*

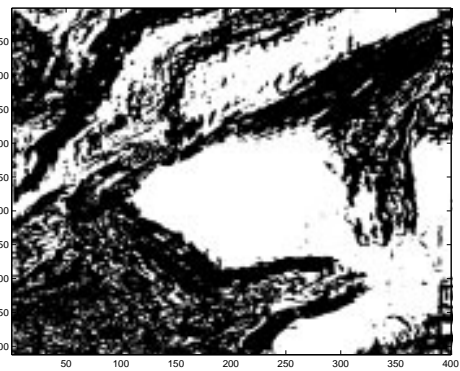


Fig. 6: *Raw segmentation without the interscale processing for the second time slice.*

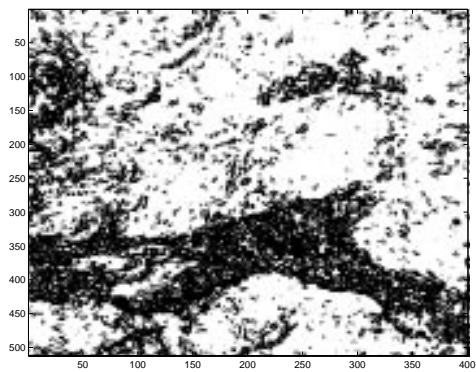


Fig. 4: *Segmentation after using the interscale fusion algorithm for the first time slice.*

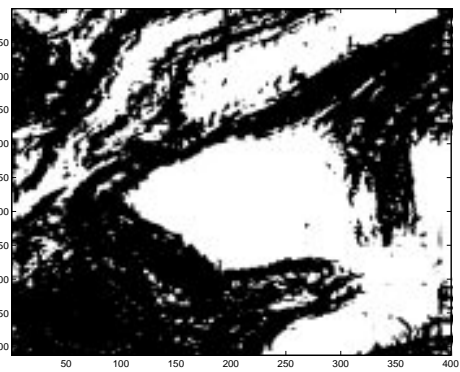


Fig. 7: *Segmentation after using the interscale fusion algorithm for the second time slice.*

3.4 Reconnaissance automatique des langues

Language Recognition using Time-Frequency Principal Component Analysis and Acoustic Modeling (2000)

Michel Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot. **Language Recognition using Time-Frequency Principal Component Analysis and Acoustic Modeling.** *Proceedings of ICSLP 2000*, Beijing, China, October 2000.

4 pages

LANGUAGE RECOGNITION USING TIME-FREQUENCY PRINCIPAL COMPONENT ANALYSIS AND ACOUSTIC MODELING

Michel DUTAT ⁽¹⁾⁽²⁾, Ivan MAGRIN-CHAGNOLLEAU ⁽³⁾, Frédéric BIMBOT ⁽³⁾

⁽¹⁾ LSCP / CNRS URA 8554, 54 Boulevard Raspail, 75270 Paris cedex, France – European Union.

⁽²⁾ ENST / CNRS URA 820, Dépt. Signal et Image, 46 Rue Barrault, 75634 Paris cedex 13, France – European Union.

⁽³⁾ IRISA (CNRS & INRIA), Campus universitaire de Beaulieu, 35042 Rennes cedex, France – European Union.

dutat@lscp.ehess.fr - ivan@ieee.org - bimbot@irisa.fr

ABSTRACT

Time-Frequency Principal Component (TFPC) is a speech parameterization technique based on a principal component analysis applied to acoustic feature parameters augmented by their time context. In this paper, we investigate on the performance of TFPC in the framework of automatic language recognition. In our experiments, identification rate is improved compared to the use of the conventional cepstral coefficients augmented by their Δ coefficients.

1. INTRODUCTION

Several speech parameterizations have been tested in the past for language recognition. However, best results are usually achieved by classical acoustic features, namely cepstral coefficients plus dynamic information using approximations of their first derivative (Δ coefficients).

In this work, we study the benefit of using TFPC (Time-Frequency Principal Components) [4, 5] for incorporating dynamic information in a data-driven manner. The approach can be viewed as a language-dependent selection of acoustical information both in the time and frequency dimensions. This selection is done using time-frequency vector filters, operating on the sequence of acoustic features, the coefficients of which are estimated on a training corpus. The underlying hypothesis is that a filtered utterance is better modeled when the filter is the one corresponding to its language.

After presenting in more details the modeling and recognition schemes, we apply them to a language identification task. In our experiments, we observe an improvement of the identification rate compared to the one obtained with a reference experiment, using a cepstral parameterization and the Δ coefficients.

2. PRESENTATION OF THE METHOD

2.1. Classical Approach

Acoustic modeling of a language consists in analysing, capturing, and modeling the sound structure of this language. In statistical approaches, this is done by using a great number of training utterances covering a large variety of speakers. In a first step, the utterances have to be parameterized into acoustic features corresponding to successive speech frames. Then, the speech frame distribution is modeled by a statistical model, usually a Hidden

Markov Model (HMM).

Figure 1 presents the classical procedure of a training phase. The set of training utterances $E_{app}^{(i)}$ of the language i is transformed through an acoustical analysis into a sequence of parameter vectors $\{\mathbf{x}_t\}^{(i)}$. In training mode, this sequence is used for the estimation of the parameters $\kappa^{(i)}$ of the model of language i . Once the

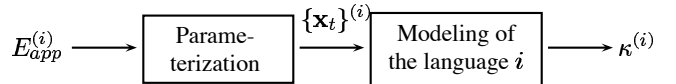


Figure 1: Language recognition : classical training procedure.

parameters of all models have been calculated, n statistical models are obtained, each of them representing one language. The recognition phase then consists in parameterizing the test utterance (as was done in the training phase) and calculating a score with each of the n models. In the case of language identification, a decision algorithm combines the n scores in order to select the most probable language.

2.2. TFPC Approach

The method that we present here was originally proposed in [4, 5]. Its goal is to use a language-dependent parameterization of the utterance, with the goal to make it thus more easy to recognize. Instead of using a common acoustic feature analysis, the sequence of cepstral coefficients of the speech utterance is filtered by a time-frequency filter depending on the language.

The training phase is therefore composed of two steps. The first one consists in a principal component analysis of the training data, where the principal component analysis is applied to sequences of consecutive vectors, so that not only the static information is taken into consideration but also the dynamic information. A language-dependent projection matrix is thus obtained (also called filtering matrix). The second step is the estimation of the statistical model $\lambda^{(i)}$ of the language i using the vector-filtered training acoustic features, with the filter corresponding to the language of the training utterance.

Figure 2 shows the training phase for the language i . The set of training utterances $E_{app}^{(i)}$ is transformed by an acoustic analysis

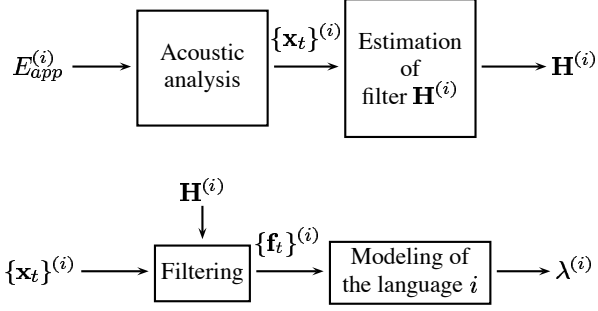


Figure 2: Language recognition : training procedure with TFPC.

(spectral, cepstral, linear prediction, ...) into a sequence $\{\mathbf{x}_t\}^{(i)}$ of p -dimensional vectors. This sequence is then used to compute the filtering matrix $\mathbf{H}^{(i)}$. Once the filtering matrix has been calculated, the sequence $\{\mathbf{x}_t\}^{(i)}$ is filtered by this matrix and it provides a new sequence of r -dimensional vectors $\{\mathbf{f}_t\}^{(i)}$ used to estimate the model $\lambda^{(i)}$ for the language i .

2.3. Construction of the Filter

For a given language and corresponding training data, the TFPC approach extracts the representation which maximizes the inertia of the data in a language-characteristic sub-space. The underlying idea is to enhance static and dynamic language-dependent correlations which may exist between the acoustic features and which may characterize the language. In order to capture the dynamic information, we take the time context of each vector into consideration.

The successive steps for the construction of the filter are the following : let $\{\mathbf{x}_t\}$ be the sequence of p -dimensional acoustic vectors extracted from the training utterances. From this sequence, we extract the expanded vectors \mathbf{X}_{t-q}^{t+q} :

$$\mathbf{X}_{t-q}^{t+q} = \begin{bmatrix} \mathbf{x}_{t+q}^* \\ \vdots \\ \mathbf{x}_t^* \\ \vdots \\ \mathbf{x}_{t-q}^* \end{bmatrix} \quad \text{with } \mathbf{x}_t^* = (\mathbf{x}_t - \bar{\mathbf{x}})$$

Then we compute the covariance matrix of the expanded vectors (contextual covariance matrix). This matrix can be calculated by computing the lagged covariance matrices \mathbf{R}_q :

$$\mathbf{R}_q = \frac{1}{T} \sum_{t=q+1}^T (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_{t-q} - \bar{\mathbf{x}})^T \quad \text{with } \bar{\mathbf{x}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t$$

The number of lagged covariance matrices corresponds to the order of the TFPC analysis. The contextual covariance matrix is then obtained by combining the lagged covariance matrices in a block-Toeplitz matrix \mathbf{R}_{2q+1} .

$$\mathbf{R}_{2q+1} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \cdots & \mathbf{R}_{2q} \\ \mathbf{R}_1^T & \mathbf{R}_0 & \cdots & \mathbf{R}_{2q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{2q}^T & \mathbf{R}_{2q-1}^T & \cdots & \mathbf{R}_0 \end{bmatrix}$$

The principal component analysis of the contextual covariance matrix is obtained by calculating its eigenvalues and eigenvectors, and ordering the eigenvalues (and the corresponding eigenvectors) in decreasing order. The sub-space generated by the eigenvectors corresponding to the highest eigenvalues has the highest inertia. By construction, all the directions of the eigenvectors are orthogonal. The principal components associated with the smallest eigenvalues often correspond to noise. Therefore, they are usually removed. Other strategies can be adopted to select the components [3].

The filtering matrix is finally obtained by transposing the matrix containing the selected eigenvectors. For instance, if the 5 first components and the components 10 to 12 are to be kept, the filtering matrix \mathbf{H} will be :

$$\mathbf{H} = [\mathbf{v}_1 \cdots \mathbf{v}_5 \mathbf{v}_{10} \cdots \mathbf{v}_{12}]^T$$

The filtering of an utterance is ultimately obtained by calculating the convolution between the matrix \mathbf{H} and the expanded vectors extracted from the utterance. For instance, when only the first r principal components are kept, we have :

$$\mathbf{H} = [\mathbf{H}_{-q} \cdots \mathbf{H}_0 \cdots \mathbf{H}_q]$$

And the filtered vectors are obtained as :

$$\mathbf{f}_t = \mathbf{H} \cdot \mathbf{X}_{t-q}^{t+q} = \sum_{k=-q}^{+q} \mathbf{H}_k \cdot \mathbf{x}_{t-k}^*$$

2.4. Recognition Phase

Once the training phase is completed, each language is characterized by a filtering matrix and a statistical model. Identifying the language of a test utterance is carried out by filtering the utterance using the filtering matrix of each language, and then computing for each corresponding model a likelihood measure.

This procedure is illustrated on Figure 3. A first acoustic analysis transforms the test utterance E_{test} into a sequence of vectors $\{\mathbf{x}_t\}$. For each language, this sequence is filtered. Then an estimation of the conditional probability $P(\{\mathbf{f}_t\}^{(i)} | \lambda^{(i)})$ of observing $\{\mathbf{x}_t\}$ knowing its language is $L^{(i)}$ is calculated. The language which gives the maximum likelihood is finally chosen.

With this approach, several parameters can vary:

- the initial acoustic analysis of the utterances (filterbank analysis, LPCC, cepstral coefficients, ...);
- the order of the TFPC analysis, that is, the number of frames taken into account;
- the choice of the principal components to keep for the construction of the filtering matrix.

In the next sections, we present some results obtained by varying some of these parameters. We also present results obtained with a reference parameterization, the cepstral parameterization, in order to compare the TFPC approach to a more conventional one.

3. EXPERIMENTS

3.1. Objectives

Our objective is to evaluate the quality of the new parameterization on a language identification task. We test two language recognition

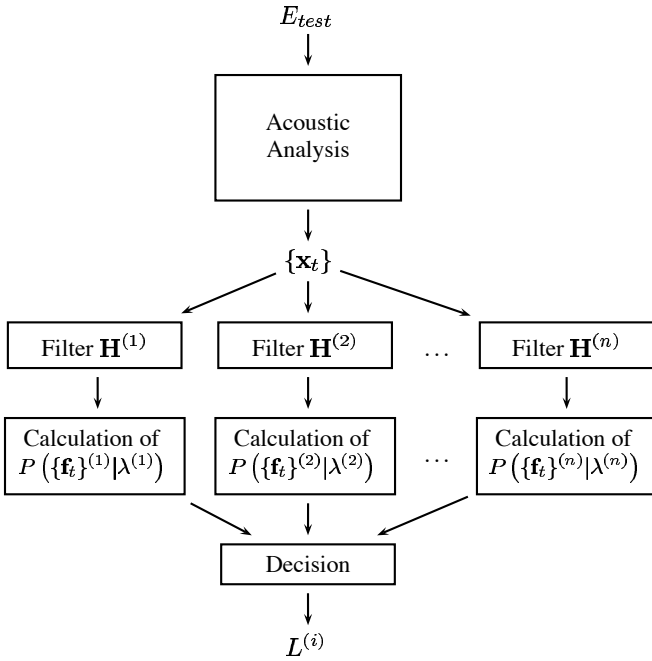


Figure 3: Language recognition : test procedure using the TFPC approach.

approaches which only differ by the parameterization step, one using the classical cepstral parameterization (reference experiment), the other using the TFPC parameterization.

3.2. Corpus

The database used for our experiments is a subset of the OGI MLT-S (Oregon Graduate Institute Multi-language Telephone Speech) database [6]. We selected four languages: English, French, Japanese, and Spanish. For each language, we use 80 utterances of 15 seconds in average pronounced by 20 speakers for the training phase. This corresponds approximately to 20 min. of speech for each language. For the test, we use, for each language, 54 utterances of 15 seconds in average pronounced by 12 (other) speakers. We use therefore a total of 216 test utterances.

3.3. Characteristics of the Reference Experiment

The acoustic model consists in an Ergodic Hidden Markov Model (E-HMM) for each language. The characteristics of this model are as follows :

| | |
|------------------------------|----------|
| Structure | Ergodic |
| State number | 24 |
| Number of Gaussians by state | 2 |
| Covariance matrices | Diagonal |

For the parameterization of the speech signal, we use the cepstral coefficients rather than the spectral coefficients according to a preliminary study [1]. The characteristics of this analysis are the following :

| | |
|-----------------------------|---------------------------------|
| Type of parameterization | MFCC ¹ |
| Frequency scale | Mel scale |
| Number of coefficients | 12 (optionally + 12 Δ s) |
| Size of the analysis window | 30 ms |
| Shift between two windows | 10 ms |
| Type of window | Hamming |
| Cepstral mean subtraction | Yes |

Some experiments with TFPCs take into account the temporal context. In order to make the experiments comparable, we did similarly for the reference experiments by using, in this case, the Δ parameters, calculated as follows [2] :

$$\Delta c_t^i = \frac{\sum_{k=1}^q k(c_{t+k}^i - c_{t-k}^i)}{2 \cdot \sum_{k=1}^q k^2} \quad \text{avec } i = 1, 2, \dots, p$$

where q is half the size of the temporal context.

3.4. Experiments with TFPC

The acoustic model used for these experiments is exactly the same as for the reference experiment (Ergodic HMM). We also use, as initial acoustic analysis, the cepstral parameterization as described above.

Then, we apply the TFPC analysis to the cepstral vectors for various sizes of context and, in each case, we vary the number of principal components which are kept. We present in the next sections the most significant results for the order 0, which corresponds to no context, and for the order $q = 4$, which provides the best results. The order 4 corresponds to a time context of 9 frames (110 ms).

3.4.1. Order 0

The TFPC analysis at order $q = 0$ does not use any time context. The contextual covariance matrix is therefore the usual covariance matrix \mathbf{R}_0 , with dimension 24×24 . The filtering matrix has the same dimension or is smaller, and the filtered vectors are 24-dimensional or smaller. In that case, the reference experiment uses 24-dimensional MFCCs and no Δ parameters. It yields an identification rate of **53.1 %**.

The table below shows the average identification score in percentage for various numbers of TFPC components. The last row shows the relative gain compared to the reference experiment. It can be noticed that the system with the TFPC analysis performs better than the reference experiment, and that 16 is the optimal number of components to keep in this case. The TFPC approach seems therefore to capture language-dependent information in the first principal components, whereas the last components seem to be less meaningful for the task.

| TFPC components kept | 12 | 16 | 20 | 24 |
|------------------------------|------|-------------|------|------|
| Identification rate (in %) | 54.3 | 58.0 | 55.8 | 53.5 |
| Relative gain to ref. (in %) | 2.3 | 9.3 | 5.1 | 0.7 |

3.4.2. Order 4

The order $q = 4$ corresponds to a context of 9 frames (110 ms). The dimension of the contextual covariance matrix \mathbf{R}_9 is now

216×216 . The reference experiment includes 12 MFCC plus 12 Δ MFCC calculated with $q = 4$. The corresponding identification rate is **60.3 %**. The table below shows the results for the TFPC experiment. Here too, a noticeable improvement is observed as compared to the reference experiment, and the optimal number of components in that case is 24.

| | | | | |
|------------------------------|------|-------------|-------|------|
| TFPC components kept | 20 | 24 | 28 | 32 |
| Identification rate (in %) | 58.6 | 63.5 | 53.6 | 60.8 |
| Relative gain to ref. (in %) | -3.0 | 5.2 | -11.2 | 0.7 |

3.5. Results using a Voice Activity Detector

We carried out the same experiments after implementing a voice activity detector as a pre-processing. The detector is based on a bi-Gaussian modeling of the energy, and on a dynamic threshold calculation. A post-processing is added to penalize the detection in silence or noise areas and to reinforce the detection in the speech areas. The results are presented in the next sections.

3.5.1. Order 0

For the order $q = 0$, the identification rate for the reference experiment increases to **59.1 %**. For the TFPC experiments, we obtain:

| | | | |
|------------------------------|------|------|-------------|
| TFPC components kept | 16 | 20 | 24 |
| Identification rate (in %) | 59.1 | 62.6 | 63.6 |
| Relative gain to ref. (in %) | 0.0 | 6.0 | 7.7 |

As could be expected, the results are globally better when we use a voice activity detector. It is also interesting to note that, in that case, all the coefficients seem to contain useful information, although the first 16 coefficients seem to contain already a great part of this information. Finally, we note that the system with the TFPC analysis performs again better than the reference system.

3.5.2. Order 4

With a time context of 9 frames, the reference experiment reaches an identification rate of **69.2 %**. The results of the TFPC experiments are given in the following table:

| | | | | |
|------------------------------|------|------|-------------|------|
| Composantes | 24 | 36 | 44 | 50 |
| Identification rate (in %) | 65.7 | 69.7 | 75.3 | 74.8 |
| Relative gain to ref. (in %) | -5.1 | 0.7 | 8.8 | 8.0 |

This time, we need to take more components (namely 44) to achieve the best performance. If we take only 24 components as previously, the score is significantly worse than the reference (using also 24 coefficients). The system using the TFPC analysis with 44 components yields an identification score of **75.3 %**, which is better than the reference system, but at the expense of a higher dimensionality. However, an additional experiment using 24 MFCC + 24 Δ -MFCC (i.e 48 coefficients in total) gives an identification rate of **68.7 %** only, which shows that the performance improvement observed with 44 TFPCs is not owed to the higher dimensionality, but to the acoustic features themselves.

4. SUMMARY AND PERSPECTIVES

In this paper, we have presented the use of the Time-Frequency Principal Component (TFPC) analysis, for performing a language-dependent vector-filtering on acoustic feature parameters. The Principal Component Analysis is applied to the contextual covariance matrix which is the covariance matrix of a sequence of vectors augmented by their time context.

The filters can be interpreted as projection matrices on language-characteristic sub-spaces. When we apply the TFPC parameterization to language identification, the best configuration of our system gives an identification score of **75.3 %**, outperforming by approximately 9 % a reference system using 12 cepstral and 12 Δ -cepstral coefficients.

We now intend to study the influence of such a parameterization on the statistics of the state *sequences* in the Hidden Markov Models. This is particularly important in the perspective of identifying languages via typical regularities of acoustic events.

5. REFERENCES

- [1] Michel Dutat. Reconnaissance automatique de la langue parlée. Rapport d'avancement de thèse, Ecole Nationale Supérieure des Télécommunications, 1997.
- [2] Sadaoki Furui. Comparison of speaker recognition methods using static features and dynamic features. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(3):342–350, June 1981.
- [3] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.
- [4] Ivan Magrin-Chagnolleau. *Approches statistiques et filtrage vectoriel de trajectoires spectrales pour l'identification du locuteur indépendante du texte*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1997.
- [5] Ivan Magrin-Chagnolleau, Geoffrey Durou, and Frédéric Bimbot. Application of time-frequency principal component analysis to text-independent speaker identification. Submitted to *IEEE Transactions on Speech and Audio Processing*.
- [6] Y. K. Muthusamy, R. A. Cole, and B. T. Oshika. The OGI multi-language telephone speech corpus. Technical report, Center for Spoken Language Understanding Oregon Graduate Institute of Science and Technology., Portland, 1993.

3.5 Perception du langage

Intelligibility of Reverse Speech in French : A Perceptual Study (2002)

Ivan Magrin-Chagnolleau, Melissa Barkat, and Fanny Meunier. **Intelligibility of Reverse Speech in French : A Perceptual Study.** *Proceedings of ICSLP 2002*, Denver, Colorado, USA, September 2002.

4 pages

INTELLIGIBILITY OF REVERSE SPEECH IN FRENCH: A PERCEPTUAL STUDY

Ivan Magrin-Chagnolleau, Melissa Barkat, and Fanny Meunier

Laboratoire Dynamique Du Langage - UMR 5596

Université Lumière Lyon 2 & CNRS - 14, avenue Berthelot - 69363 Lyon Cedex 07 - France

ivan@ieee.org - melissa.barkat@univ-lyon2.fr - fanny.meunier@ish-lyon.cnrs.fr

ABSTRACT

We ran an experiment focusing on cognitive implication of reverse speech segments. Nine durations of reverse speech plus a non-distorted control condition have been considered (varying between 20 ms and 180 ms) in order to test the pattern of intelligibility degradation in French. We observed an overall strong negative correlation between the degree of intelligibility and the size of reverse speech windows. These results appear to be very comparable to those obtained in English by Greenberg and Arai (2001), at least on the slope of intelligibility performance decrease. However, intelligibility loss in French is delayed by 20 ms. Apart from confirming the cognitive ability to restore reverse speech up to a certain point, our study revealed differences that could be interpreted as language specific.

1. INTRODUCTION

The perception of spoken language is a daily task and requires the action of complex cognitive functions. The intelligibility of a speech utterance depends on the quality of the emitted signal but also on the cognitive ability to process this signal. The processing is composed, among other things of: 1) speech / non-speech discrimination to get rid of useless information during a conversation in a noisy environment (reverberation, background noise, multiple-speakers, etc.); 2) the ability to correct some errors due to the distortion of the signal.

Several studies have shown that language remains understandable in spite of certain acoustic degradations: there is a cognitive ability to reconstruct a speech signal which has been corrupted by various factors. For instance, if a phoneme is replaced by noise, subjects continue to perceive the word in its integrity [1]. This reconstruction ability depends on the type of distortion and on its importance.

There are at least two limits concerning the ability to reconstruct a speech signal, a physiological one and a cognitive one. The physiological limit is the maximum quantity

of information which can be processed by time unit. The cognitive limit is at a higher level and is necessary to prevent the system to build purely fictive representations which would lead to a loss of comprehension of the received message. Therefore, it is very important to establish what are the limits of the restoration ability.

Temporal inversion has been qualified as the best way to deteriorate speech [2]. It is very common to use reverse speech as a control condition in experiments on the human ability to process language (see for instance [3]). However, some experiments on English have shown that the human perceptual system was able to handle speech inversion in a certain measure [2, 4]. According to these results, intelligibility depends on the length of inversion windows.

The aim of the experiment presented in this paper is to quantify the ability of the human perceptual system to restore comprehension when stimulated with reverse speech stimuli in French.

2. EXPERIMENT

The principle of the experiment is to let voluntary subject listen to some sentences more or less altered and to ask them to transcribe what they have understood.

2.1. Material

We used a corpus of 50 phonetically balanced sentences [5], extracted from the database BD-SONS [6, 7] (sentences groups PEQ01 to PEQ05), and uttered by 10 different speakers (5 females and 5 males). The average duration of the sentences was 3 seconds.

The preparation of the stimuli has been done using MATLAB. It consists in reversing windows of n ms every n ms along a sentence (see Fig. 1). The values for n were respectively 20, 40, 50, 60, 70, 80, 100, 140, and 180 ms. The control condition was for $n = 0$, which corresponds to the

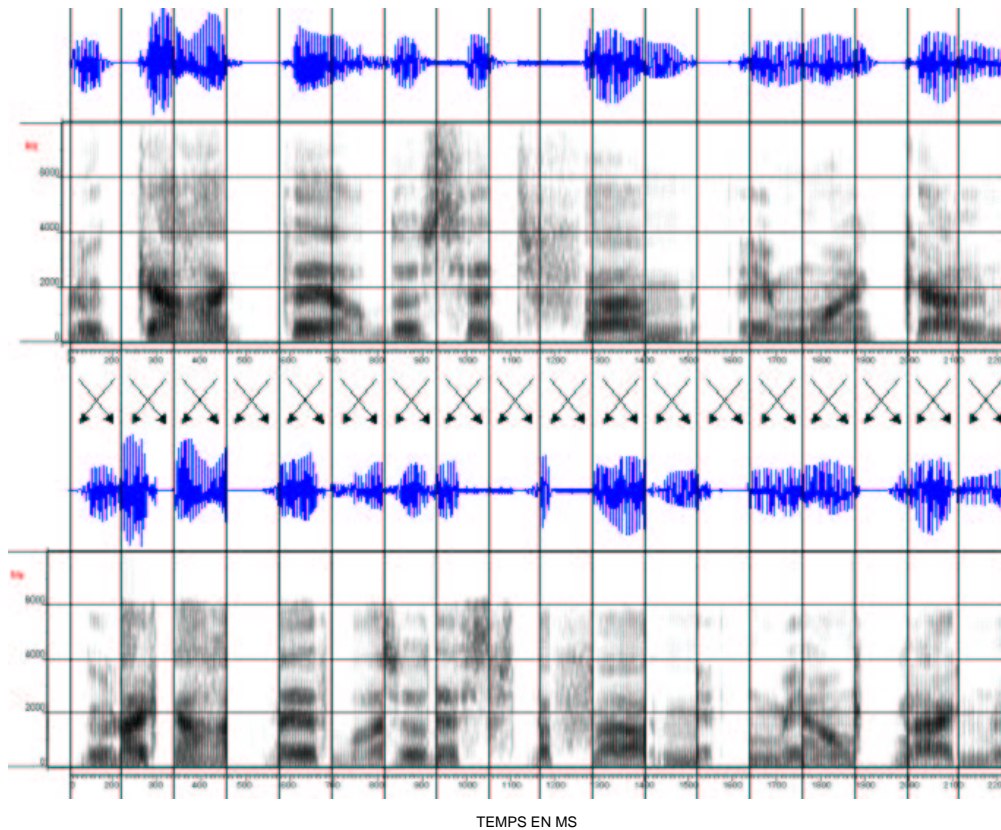


Fig. 1. Principle and example of speech inversion. The top part represents the original signal and its spectrogram. The bottom part represents the inverted signal (windows of 120 ms) and its spectrogram.

original sentence with no alteration.

We also made some tests to determine the predictability of the sentences. The factor can have an influence on the performance of the subjects. Some words may be recognized because they are highly predictable, although the signal has been considerably deteriorated. We gave the list of sentences to 12 persons, but the last word of each sentence was not written and had to be completed. We then used the predictability score of the sentences to balance the construction of the lists for the perceptual test.

We had 10 experimental conditions (including the control condition). We split the 50 sentences in 10 subsets of 5 sentences each. These subsets were built so that the following factors were balanced in each subset: 1) the number of syllables; 2) the predictability degree; 3) the logarithm of the word frequencies; 4) the number of words; 5) the gender of the speaker.

Then we built 5 lists composed of the 50 sentences but in various conditions among the lists. For instance, the sentence “Il se garantira du froid avec un bon capuchon.” was

presented with an inversion condition of 40 ms in the first list, an inversion condition of 140 ms in the second list, an inversion condition of 0 ms in the third list, etc. And the sentence “Le train entre déjà en gare.”, which belonged to another subset, was presented with an inversion condition of 140 ms in the first list, an inversion condition of 40 ms in the second list, an inversion condition of 80 ms in the third list, etc. Each subject, who was listening to only one list, could listen to all the sentences (each sentence appearing only once in each list) and all the conditions (5 sentences by condition). For each list, sentences were distributed quasi-randomly. This distribution guaranteed to have all the experimental factors well distributed inside the list and not to have the same inversion condition two times consecutively.

2.2. Description of the Experiment

The perceptual experiment was conducted at the “Institut des Sciences de l’Homme (ISH)”, Lyon, France. Each subject, equipped with a stereo headphone, was facing a computer. He was presented with sentences in various inversion conditions and had to type what he had understood. The subject could listen the sentence up to 4 times.

The experiment was divided into 2 phases. The first phase was a training phase, consisting in 5 sentences presented in various conditions to let the subject get used to the task and the conditions. Then, after controlling that the subject had understood perfectly the task, the experiment started and the list of 50 sentences was presented. The total duration of the whole experiment, including the repetition of some sentences was 30 minutes in average.

2.3. Subjects

The 28 subjects of the experiment were all French native speakers, between the ages of 19 and 30. None of them had a known auditory trouble. They received an indemnity for their participation in the experiment.

3. RESULTS

Results were obtained by comparing the transcriptions of the subjects with the original sentences. A percentage of word recognition was calculated for each sentence. This percentage was calculated manually and spelling mistakes were accepted.

The curve giving the intelligibility in function of the length of the inversion windows is given Fig. 2. As was expected, the recognition rate corresponding to the control condition (no inversion) is 100 %. The rest of the curve can be divided in three parts: 1) a part where the intelligibility is pretty well conserved [100 %; 90 %] for inversion windows between 0 and 50 ms; 2) a part where a degradation of the intelligibility can be observed [75 %; 25 %] for inversion windows between 60 and 100 ms; and 3) a part where the intelligibility has been lost [10 %; 2 %] for inversion windows of 140 and 180 ms.

We observe a strong negative correlation between the degree of intelligibility (measured by the word recognition rate) and the length of the inversion windows ($r = -0.95$).

4. DISCUSSION

Our results indicate a loss of intelligibility with the increase of the length of the inversion windows.

A similar experiment has been conducted in English by Greenberg and Arai [4]. These authors observe (after an estimation based on the curve published in their paper): 80 % of intelligibility for the condition 40 ms; 60 % for 50 ms; 50 % for 60 ms; 30 % for 70 ms; 25 % for 80 ms. The loss of intelligibility was observed at 100 ms (6 % for 100 ms; 4 % for 140 ms; and 3.5 % for 180 ms). They also report

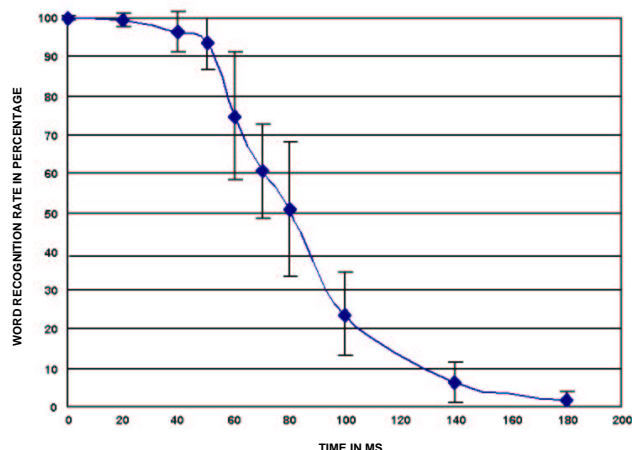


Fig. 2. Intelligibility (in percentage of word recognition) in function of the length of inversion windows (in ms).

a negative correlation between the degree of intelligibility and the length of the inversion windows.

Fig. 3 shows a comparison between English and French. The slopes of the linear regressions are very similar ($a = -0.67$ for French and $a = -0.62$ for English). However, the value of the ordinate at the origin is different ($b = 110.58$ for French and $b = 90.74$ for English). This difference means that the loss of intelligibility occurs about 20 ms later in French.

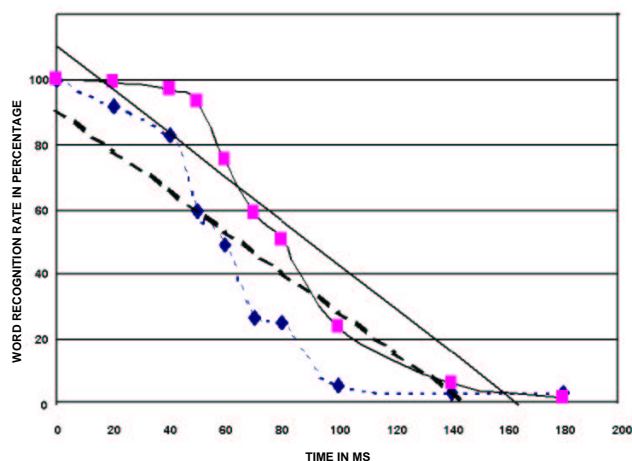


Fig. 3. Intelligibility (in percentage of word recognition) and linear regression for English (after [4]) (dashed line) and for French (solid line).

For instance (see Table 1), in English with windows of 60 ms, the intelligibility is about 50 %, whereas this value is reached in French 2 ms later, that is, with windows of 80 ms.

5. CONCLUSION AND PERSPECTIVES

| Length of windows | English | French |
|-------------------|---------|--------|
| 0 ms | 100 % | 100 % |
| 20 ms | 92 % | 99 % |
| 40 ms | 83 % | 97 % |
| 50 ms | 60 % | 93 % |
| 60 ms | 49 % | 75 % |
| 70 ms | 27 % | 59 % |
| 80 ms | 25 % | 51 % |
| 100 ms | 6 % | 24 % |
| 140 ms | 4 % | 7 % |
| 180 ms | 3.5 % | 2 % |

Table 1. *Intelligibility (measured in percentage of word recognition) for English and French in function of the length of the inversion windows.*

Our experiment confirms the ability of the human cognitive system to restore, in a certain measure, speech segments which have been inverted. But there is a difference as to when the loss of intelligibility occurs. There are two possibilities to explain this difference. The first one is methodological, the second one concerns the intrinsic differences between the two languages.

The methodological difference between the two experiments can be at the level of the intelligibility calculation. Although the two experiments use a word recognition rate to measure this intelligibility, and although we tried to measure this word recognition rate as close as possible to the way it is done in the paper of Greenberg and Arai, it is still possible that the judgment on what is a spelling mistake and what is a word mistake could have been different. In the paper by Greenberg and Arai, it is said that “The number of correct word was scored using an algorithm that automatically compensated for minor errors in spelling”. In our study, we did this correction manually, accepting some spelling mistakes affecting several graphemes (for instance “accour” instead of “accourent”) as long as the pronunciation was exactly the same. This example shows a good phonological decoding even if the orthographic transcription was wrong. It is not clear what were the rules used by Greenberg and Arai, but in the previous example, the word could have been judged as wrong by an automatic algorithm. This difference could explain part (but not all) the observed gap between the two experiments.

The other possibility can come from intrinsic differences between English and French, like phonological differences (the fact that French has a syllabic structure whereas English is a stressed language, etc.).

In this paper, we ran a perceptual experiment on the perception of reverse speech in French. Subjects were presented with sentences more or less altered, the alteration consisting in reversing n ms of speech every n ms, n varying from 0 ms (the control condition) to 180 ms. We found similar results to the experiment ran in English by Greenberg and Arai (2001), that is, there is a strong negative correlation between intelligibility and the size of the inversion windows. But our results show that the loss of intelligibility occurs in French 20 ms later than in English. A part of this difference could be due to a difference in the calculation of the percentage of correctly recognized words. But this difference is most probably due mainly to intrinsic differences between the two languages, French having a syllabic structure whereas English is a stressed language. However, this hypothesis needs to be validated by additional measurements on the data and additional experimentation.

6. ACKNOWLEDGEMENT

The authors would like to thank Jean-Marie Hombert, François Pellegrino, Tristan Cenier, and Leonid Synyukov for their help and their contribution. This research has been funded by CNRS (the French National Center for Scientific Research) APN 2001 No.26635 and by the French minister of research ACI - MSH 2001.

7. REFERENCES

- [1] R. M. Warren, “Perceptual restoration of missing speech sounds,” *Science*, pp. 392–393, 1970.
- [2] K. Saberi and D. R. Perrott, “Cognitive restoration of reverse speech,” *Nature*, 1999.
- [3] J. Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, and Amiel-Tison, “A precursor of language acquisition in young infants,” *Cognition*, vol. 29, pp. 143–178, 1988.
- [4] S. Greenberg and T. Arai, “The relation between speech intelligibility and the complex modulation spectrum,” in *Proceedings of EUROSPEECH 2001*, 2001, pp. 473–476, Aalborg, Denmark.
- [5] P. Combescure, “20 listes de 10 phrases françaises phonétiquement équilibrées,” *Revue d’acoustique*, pp. 34–38, 1981.
- [6] R. Carré, R. Descout, M. Eskenazi, J. Mariani, and M. Rossi, “The French language database: defining, planning, and recording a large database,” in *Proceedings of ICASSP 84*, 1984, pp. 42.10.1–42.10.4.
- [7] “The BDSOONS database,” <http://www.icp-gre-net.fr/Relator/french/bdson.html>.

3.6 Langue des signes

La dénomination d'objets manipulables en LSF au sein d'un discours narratif : peut-on parler de classificateurs verbaux ? (2003)

Karine Grande, Colette Grinevald, Ivan Magrin-Chagnolleau, and Annie Risler. **La dénomination d'objets manipulables en LSF au sein d'un discours narratif : peut-on parler de classificateurs verbaux ?** *La linguistique de la LSF : recherches actuelles*, Lille, France, September 2003.

2 pages

La dénomination d'objets manipulables en LSF au sein d'un discours narratif : peut-on parler de classificateurs verbaux ?

Karine Grande^α, karine_grande@hotmail.com

Colette Grinevald^α, colette.grinevald@univ-lyon2.fr

Ivan Magrin-Chagnolleau^α, ivan@ieee.org

Annie Risler^β, annie.risler@free.fr

[α : UMR 5596 (Un. Lyon 2 & CNRS) ; β : Lab. Lordat (Un. Toulouse-le Mirail) et Lille 3]

En contraste avec les langues orales, la dénomination des objets et leur mise en discours en LSF sont des processus de nature iconique et compositionnelle. En contraste avec l'arbitraire du signe linguistique saussurien, la mise en langue d'objets du monde en langues des signes repose sur des compositions lexicales qui incluent des éléments appelés par certains des "spécificateurs de formes et/ou de taille" (SASS en anglais = Size And Shape Specifiers), qui, comme leur nom l'indique, spécifient certains aspects physiques de l'objet, tels la forme d'une partie de l'objet ou le contour de l'objet entier, tout en précisant la taille de la partie ou du tout. Il reste à étudier systématiquement les dimensions de ces systèmes de "spécificateurs" des langues des signes et leurs rôles dans la dénomination des objets.

PROBLEMATIQUE :

En langue des signes, la reprise anaphorique des éléments du discours se fait à travers des signes qui reprennent certaines caractéristiques du référent. De tels signes ont été appelés « classificateurs verbaux » dans la littérature (EMMOREY, 2002 ; SUPPALLA, 1986), terme emprunté à la littérature sur certaines langues orales du monde dites langues à classificateurs (CRAIG, 1986 ; GRINEVALD, 1999, 2000, 2003). On les appelle également proformes (ENDGBERG-PEDERSEN 1985 ; RISLER 2003). Ces éléments anaphoriques sont probablement à rapprocher de phénomènes dans des langues orales mais peut-on parler de classificateurs dans les langues des signes ? Il s'agit d'étudier le lien entre le signe standard hors discours et sa reprise en discours, soit en tant que signe lexical, soit par un anaphorique. Le lien entre le signe et ses anaphoriques est-il syntaxique, sémantique, formel ? On observe des variations individuelles dans la forme de l'anaphorique pour un même emploi. On observe aussi en discours des variations syntaxiques : selon la fonction syntaxique assumée par l'anaphorique, mais aussi pour une même fonction, dans des variantes contextuelles. Le degré de grammaticalisation et de normalisation de la langue semble en jeu dans ces variations, ainsi que la liberté offerte par l'iconicité.

METHODOLOGIE :

Le protocole expérimental consiste à réaliser un film qui met en jeu les deux processus de dénomination de certains objets hors contexte et leur reprise anaphorique dans une chaîne de discours. Les objets ont été choisis en fonction des formes 1D (long et rigide), 2D (plat et souple) et 3D (rond) des classificateurs des langues orales. Ils sont inclus dans un scénario, insérés dans diverses situations (localisation, déplacement, mise en action, changement d'état, action sur l'objet) puis sont présentés séparément au travers d'images épurées. Le scénario est montré à un 1^{er} sourd qui signe ce qu'il voit à un 2^{ème} sourd, simple observateur. Ce dernier restitue ce qu'il a compris pour contrôle. Les deux prestations sont filmées sur vidéo afin de pouvoir les analyser ensuite, et constituer en même temps un corpus de LSF. Nous avons recueilli les énoncés de 6 sourds (3 paires).

PRINCIPALES CONCLUSIONS :

D'après la typologie de Grinevald (CRAIG 1996 ; GRINEVALD 2003), un classificateur est un morphème (critère1) grammatico-lexical (critère2), dont la fonction catégorisante est motivée par certaines caractéristiques du référent (critère3) et qui apparaît dans certains aspects du discours (critère4). Notre étude sur les verbes de localisation et de déplacements sans agent a permis de mettre en évidence diverses proformes qui répondaient à ces critères. La proforme en main plate semble regrouper les objets plats et rectangulaires (répondant donc au critère3) tels un « livre », une « table », un « cadre-photo », un « cadre » traditionnel et un « miroir ». Sa configuration manuelle donne des indices sur le référent qu'elle localise ou qu'elle déplace (critère4) et comporte des propriétés communes avec son signe lexical (critère2). D'autres proformes étudiées sur ce même corpus présentent les mêmes propriétés.

Si l'on considère comme Supalla (SUPALLA 1986) que la configuration manuelle en langue des signes s'apparente au morphème des langues orales, il nous est alors possible de parler de classificateurs de verbes de localisation et de déplacement en LSF. Nous obtenons un résultat identique avec les verbes de manipulation. Dans ce cas, la fonction catégorisante du classificateur a été plus distinctement repérable.

Il semble donc que certaines proformes de localisation et de déplacement ou de manipulation correspondent bien aux différents critères qui définissent un classificateur dans les langues orales dites à classificateur.

Néanmoins, l'analyse des classificateurs verbaux en LSF nécessiterait des études complémentaires qui permettraient de confirmer les tendances dégagées par les résultats obtenus. Il serait intéressant de faire passer l'expérience à plus grande échelle et d'évaluer voire homogénéiser le niveau de langue de chaque participant : cela éviterait toute possibilité de mime et permettrait d'amoindrir une variabilité individuelle conséquente face à une connaissance hétérogène de la langue.

Bibliographie

- CRAIG, C. 1986. *Noun classes and categorization*. Amsterdam: John Benjamins.
- ENDGBERG-PEDERSEN, E. 1985. *Proformes en morphologie syntaxe et discours, Etudes Européennes en langue des signes*. Edirsa Bruxelles.
- EMMOREY, K. 2002. *Language, cognition, and the brain: Insights from sign language research*. Lawrence Erlbaum and Associates: Mahwah, NJ. Chap 3 & 8
- GRINEVALD, C. 1999. "Language Endangerment in South America: A Programmatic Approach". In *Endangered Languages*. L. Grenoble & L. Whaley eds. Cambridge University Press. A paraître.
- GRINEVALD, C. 2000. *A morpho-syntactic typology of classifiers in Systems of nominal classification*. A paraître.
- GRINEVALD, C. 2003. "Classifier Systems in the Context of a Typology of Nominal Classification". In *Perspectives on Classifier Constructions in Sign Languages*, Karen Emmorey (Ed.), pp. 91-110.
- RISLER, A. 2003. « Point de vue cognitiviste sur les espaces créés en LSF : espace lexical, espace syntaxique ». In *Gestualité et Syntaxe*, LIDIL n°26, pp 45-61.
- SUPALLA, T. 1986. "the classifier system in American Sign Language". In C. Craig (Ed.), *Noun Classification and Categorization*. J. Benjamins.

3.7 La méthode de l'Actors Studio / Le jeu de l'acteur

L'incorporation animale chez l'acteur (2015)

Ivan Magrin-Chagnolleau. **L'incorporation animale chez l'acteur.** *performance, 2*, 2015.

12 pages

L'incorporation animale chez l'acteur

Ivan MAGRIN-CHAGNOLLEAU

Résumé

Dans cet article, je présente en détail l'exercice d'incorporation animale tel qu'il a été proposé par Lee Strasberg et est enseigné aux acteurs qui apprennent la méthode de l'Actors Studio. L'exercice est détaillé de telle sorte qu'une personne lisant cette description puisse pratiquer elle-même l'exercice. Cet exercice est ensuite mis en perspective dans le cadre du travail d'un acteur, d'un metteur en scène, ou d'un coach d'acteur. Je fais pour finir un lien entre la pratique de l'incorporation animale et certaines pratiques chamaniques, notamment celles des indiens d'Amérique du nord.

Mots clés

Incorporation, incorporer, animal, animalité, acteur, jeu, technique de jeu, mise en scène, coaching, chamanisme, intuition, inconscient.

1 Introduction

L'incorporation animale est une technique très utilisée chez les acteurs, et notamment ceux qui ont été formés à la méthode de l'Actors Studio. Il s'agit en effet d'un des exercices phares développés par Lee Strasberg. Mon objectif dans cet article est d'abord de décrire en détails cet exercice afin que toute personne puisse le pratiquer pour elle-même. Je m'appuie pour cela sur ma propre pratique de cet exercice en tant qu'acteur, sur ma formation d'acteur suivie à l'Actors Studio à New York pendant trois ans, et sur ma pratique de cet exercice en tant que metteur en scène et coach d'acteur. Je propose notamment cet exercice régulièrement sous forme d'ateliers, comme par exemple lors du congrès international IDEA qui a eu lieu à Paris en juillet 2013, et au cours duquel j'ai proposé à 25 participants une version de deux heures de cet atelier. La première dimension de cet article sera donc didactique.

Mon second objectif est de mettre en lien cet exercice avec la notion de développement de soi et de certaines compétences personnelles comme par exemple l'intuition. À ce titre, j'essaie de montrer tout au long de l'article, et plus particulièrement dans les dernières parties, comment une pratique régulière de cet exercice peut développer certaines compétences physiques et mentales, en faisant notamment des liens avec des pratiques anciennes comme le chamanisme.

L'article se divise en six parties. La première donne les définitions des trois concepts principaux de l'article : animal, incorporation, incorporation animale. La deuxième partie décrit en détail l'exercice d'incorporation animale. La troisième partie explicite l'utilisation de l'exercice dans le travail de l'acteur. La quatrième partie explore l'exercice du point de vue du metteur en scène. La cinquième partie, du point de vue du travail de coaching. Enfin, la sixième partie tisse des liens entre cet exercice et certaines pratiques chamaniques.

2 Définitions

Je voudrais d'abord tenter de définir quelques termes qui reviendront souvent dans cet article.

2.1 Animal

Le Robert donne deux définitions du mot « animal », l'une incluant l'homme, « être vivant organisé, doué de sensibilité et de motilité, hétérotrophe », l'autre excluant l'homme, « être vivant non végétal, ne possédant pas les caractéristiques de l'espèce humaine (langage articulé, fonction symbolique, etc.) ». Il me paraît plus intéressant d'adopter ici la première définition, celle qui est inclusive. Elle permet, il me semble, une connexion plus profonde entre les animaux et nous, à quelque niveau que ce soit. De plus, dans le cadre d'une pratique d'incorporation animale, il est plus facile de concevoir cette incorporation, et donc de la pratiquer, si on considère que les animaux et nous faisons partie d'un même ensemble.

2.2 Incorporation

Le Robert définit le mot « incorporation » comme « l'action de faire entrer (une matière) dans une autre », et le mot « incorporer » comme « faire qu'une chose fasse corps avec une autre » et « unir intimement (une matière à une autre) ». Le mot « incorporation » regroupe donc plusieurs notions, celle de « faire entrer », celle de « faire corps », et celle d' « unir ».

2.3 Incorporation animale

On pourrait donc dire que l' « incorporation animale » consiste à « faire entrer un animal en soi », « faire corps avec un animal », « unir son énergie avec celle d'un animal ». La première formulation semble être la plus proche de la réalité que j'essaierai d'appréhender tout au long de cet article. Mais je pourrais dire aussi « fusionner avec un animal ».

3 L'incorporation animale à l'Actors Studio

Après avoir tenté de définir les quelques notions principales que nous allons utiliser tout au long de cet article, je vais m'attacher à décrire l'incorporation animale telle qu'elle se pratique à l'Actors Studio. Cet exercice de l'incorporation animale a été développé par Lee Strasberg (Cohen, 2010; Strasberg, 1986, 1987), et est le fruit de tout son questionnement sur le jeu de l'acteur, ainsi que de son expérience d'acteur, de metteur en scène, de professeur et coach d'acteurs.

Elle repose aussi sur sa lecture de plusieurs théoriciens du jeu de l'acteur, dont notamment Stanislavski (Stanislavski, 1937, 1950, 1988, 1996), Meyerhold (Meyerhold, 1980-1992) et Michael Chekhov (Chekhov, 1997a, 1997b). Je l'ai moi-même pratiqué pendant mes trois années à l'Actors Studio, et continue à le pratiquer depuis comme acteur, metteur en scène, et coach d'acteurs. Il se décompose en trois parties, que nous allons détailler : une première partie d'observation, une seconde partie d'incorporation, et une troisième partie de transformation en personnage.

3.1 Observation

Je reviendrai par la suite sur les critères qui poussent à choisir tel animal plutôt qu'un autre. Supposons, pour le moment, qu'un animal a été choisi. La première partie de l'exercice consiste à observer cet animal. Cela peut prendre différentes formes, en fonction de l'animal choisi. Si l'animal est, par exemple, un chien ou un chat, il est assez facile d'en observer directement, surtout si l'on en a un à la maison. Il est possible, également, d'aller observer des animaux dans un zoo. Bien sûr, cela ne remplace pas tout à fait leur habitat naturel, mais pour certains animaux, c'est tout de même utile. Enfin, quand aucune de ces deux solutions n'est possible, il est toujours envisageable de trouver des documentaires animaliers sur l'animal choisi. On en trouve aujourd'hui beaucoup sur Internet. Le plus important, dans cet exercice, est d'observer l'animal dans le plus de situations différentes possibles, et de laisser ces observations s'imprégner. On ne cherchera pas, par la suite, à produire simplement une imitation de l'animal, mais plutôt à l'incorporer. Cette observation a donc pour but de remplir l'inconscient de plein d'éléments à peine perceptibles, mais qui aideront ensuite à l'incorporation. Cela signifie qu'il faut passer beaucoup de temps à observer l'animal choisi, éventuellement à prendre des notes pendant l'observation, ou après l'observation. Mais la prise de notes n'est pas du tout obligatoire. Elle peut aider à fixer l'attention, mais elle ne permettra certainement pas de capturer sur la page tout ce qui est observé et ressenti. Il est conseillé que cette phase d'observation dure jusqu'à ce que l'animal choisi ne semble plus avoir de secret pour vous.

3.2 Incorporation

La phase d'incorporation peut démarrer une fois l'observation terminée, ou en parallèle de cette dernière. Plus vous aurez observé l'animal et le connaîtrez,

plus cette phase d'incorporation sera pertinente et efficace. L'incorporation se déroule en plusieurs étapes. Elle commence toujours par une étape de relaxation. Tout travail d'acteur commence toujours par une étape de relaxation. La relaxation permet de faire le vide, de se mettre à disposition, de se rendre présent au moment. Il existe de nombreuses techniques pour cela, mais je ne les détaillerai pas dans le cadre de cet article (Cohen, 2010). On peut faire une relaxation allongée ou assise. Dans le cadre d'un exercice d'incorporation, il est préférable de faire la relaxation assise, ou tout au moins de la terminer en position assise. Une fois la relaxation effectuée, on procède à l'étape d'incorporation à proprement parler. Concentrez-vous sur une partie de votre corps, et imaginez qu'elle se transforme pour devenir petit à petit la partie du corps correspondante de l'animal choisi. Il est recommandé de choisir au départ des parties de corps plus faciles, comme par exemple un bras entier ou une jambe entière. Au fur et à mesure que votre partie du corps se transforme, vous pouvez commencer à la faire bouger pour prendre conscience de cette transformation au fur et à mesure qu'elle se produit. Puis vous relâchez la partie du corps transformée, et passez à une autre. Vous passez ainsi en revue toutes les parties du corps, les unes après les autres, en commençant par exemple par des membres entiers, et en terminant par des parties du corps très spécifiques, comme par exemple un doigt ou une oreille. Toute cette étape de l'exercice se fait en restant en position assise. Une fois que vous avez passé en revue chaque partie de votre corps, y compris des parties du corps très spécifiques, vous pouvez commencer à les combiner entre elles, par deux, par trois, jusqu'à arriver à votre corps tout entier. Votre corps tout entier est alors transformé, et vous allez pouvoir commencer à bouger de la chaise et à vous déplacer dans l'espace. Il peut être nécessaire de passer par des étapes intermédiaires où vous êtes, par exemple, debout sur vos membres postérieurs avec vos membres antérieurs appuyés sur la chaise. À vous de trouver la bonne configuration en fonction de l'animal que vous avez choisi.

Vous pouvez alors commencer à explorer l'espace et l'environnement typiques de l'animal choisi. Comment votre animal se déplace-t-il ? Comment votre animal court-il ? Comment mange-t-il ? Comment boit-il ? Comment se repose-t-il ? Comment dort-il ? Comment joue-t-il ? Comment se bat-il ? etc. Cherchez à explorer le plus de situations possibles, qui sont typiques du mode de vie de l'animal choisi. Plus vous explorerez de situations différentes, plus vous incorporerez l'animal. Un mot ici pour essayer d'expliquer en quoi consiste l'incorporation de l'animal dans ce contexte. Ce n'est pas une imitation. Il s'agit bien de quelque chose qui se ressent de l'intérieur. L'imagination joue très certainement une part importante, mais il y a bien une expérience réelle de devenir l'animal. C'est toujours de votre corps qu'il s'agit, mais il n'a plus la même énergie, ne fait plus les mêmes mouvements, ne ressent plus les choses de la même façon, ne réagit plus de la même façon. Des expériences en neurosciences nous montreraient sans doute que le cerveau ne fonctionne plus non plus de la même façon. Cette transformation se fera graduellement au fur et à mesure que vous pratiquerez l'exercice. Il se peut que très peu de choses se passent la première fois, ou que vous ressentiez une transformation très minime. Pour arriver à une transformation totale, il faut beaucoup de lâcher

prise, faire confiance à l'instinct animal que nous avons déjà en nous, et ne pas chercher à comprendre à tout prix ce qui se passe, tout du moins pendant la réalisation de l'exercice. Être guidé par un professeur ou un coach expérimenté est bien entendu fortement recommandé. J'ajoute que cet exercice nécessite d'avoir déjà une bonne maîtrise d'un autre exercice pratiqué dans la méthode de l'Actors Studio, celui de la mémoire sensorielle (Cohen, 2010). En effet, c'est en ayant recours à des mémoires sensorielles que vous arriverez à reconstituer l'environnement dans lequel vous ferez évoluer votre animal, puisque vous ne serez pas en réalité dans cet environnement.

Pendant que vous réalisez cet exercice, il est très important de laisser aussi sortir des sons. Une fois encore, il ne s'agit pas d'imitation. Il s'agit plutôt de laisser sortir des sons en rapport avec l'expérience intérieure que vous vivez, et avec le besoin d'exprimer ce que vous ressentez de l'intérieur. Les sons qui sortiront alors ressembleront peut-être aux sons que produit l'animal, mais ce n'est pas certain ni nécessaire. Ce qui est important, c'est d'utiliser les sons qui viennent pour exprimer les ressentis que vous vivez.

Après avoir fait de nombreuses fois l'exercice en vous projetant dans un environnement qui est proche de l'habitat naturel de l'animal choisi, vous pourrez pratiquer l'exercice plusieurs fois en choisissant cette fois-ci de vous trouver dans l'environnement de la pièce de théâtre ou du film ou du personnage que vous travaillez. Cette transition permet de préparer la troisième partie de l'exercice, tout en mélangeant le travail sur l'animal avec un travail de mémoire sensorielle classique (Cohen, 2010).

3.3 Transformation en personnage

Cette troisième partie de l'exercice consiste à glisser progressivement de l'animal vers le personnage. Vous êtes arrivés au point de l'exercice où vous pratiquez l'exercice d'incorporation animale dans l'environnement d'un personnage humain. Pour glisser progressivement vers le personnage, plusieurs choses sont possibles, qui ne se font pas toujours dans le même ordre selon les cas. Vous pouvez, d'une part, humaniser progressivement les sons qui sortent de votre bouche d'animal, pour qu'ils deviennent progressivement langage humain, en passant par des étapes intermédiaires, et en conservant au final des caractéristiques du langage animal que vous aviez mises en place. Cela devrait contribuer tout naturellement à modifier votre voix dans une certaine mesure, ainsi que votre façon de parler, et éventuellement votre accent. Vous pouvez, d'autre part, humaniser progressivement votre corps. Là encore, plusieurs étapes sont possibles. Vous pouvez retransformer différentes parties, ou l'ensemble, mais en gardant des empreintes corporelles ou des attitudes corporelles qui sont résiduelles de l'animal. Cela veut dire que votre posture ne sera plus tout à fait la même, ni votre façon de bouger, et qu'elles intégreront des caractéristiques physiques de l'animal que vous avez travaillé. Au final, la caractérisation physique et vocale qui en découle contiendra des caractéristiques physiques et

vocales de l'animal travaillé. C'est une transformation assez étonnante à observer, de l'intérieur comme de l'extérieur, quand l'exercice a été pratiqué suffisamment pour qu'une vraie transformation s'opère, c'est-à-dire qu'une véritable incorporation animale ait lieu.

4 Jeu d'acteur et incorporation animale

Abordons maintenant la question du travail de l'acteur, et plus particulièrement l'utilisation de l'incorporation animale pour créer un personnage. Autrement dit, comment l'acteur va-t-il utiliser l'exercice décrit précédemment pour créer un personnage ?

4.1 Créer un personnage

Revenons d'abord sur la question du choix de l'animal. Comment choisit-on un animal pour travailler cette incorporation animale ? Il y a plusieurs façons de procéder. La première consiste à faire une liste de caractéristiques physiques et comportementales du personnage à créer, et de rechercher ensuite un animal aux caractéristiques semblables. Une autre possibilité consiste à laisser venir le premier animal qui vient à l'esprit, et à commencer à travailler avec celui-là. Une troisième possibilité consiste à choisir un animal qui va faire travailler certaines caractéristiques que l'acteur souhaite développer. Ce qui compte, c'est surtout de démarrer le travail avec un premier animal. Par la suite, il arrive souvent qu'un autre animal s'impose, puis un autre, etc. Il se peut aussi que le personnage à incarner nécessite le travail de plusieurs animaux. Il est possible d'en travailler plusieurs les uns à la suite des autres, jusqu'à ce qu'éventuellement l'un d'eux s'impose comme le plus représentatif. Au-delà de la pertinence du choix de l'animal, cet exercice présente suffisamment d'intérêt pour être travaillé en tant que tel, même si le choix de l'animal ne semble pas satisfaisant ou pertinent. Pour créer un personnage, l'acteur peut donc recourir à cet exercice de l'incorporation animale, en se laissant guider par son instinct pour le choix de l'animal.

4.2 Incarnation, incorporation et physicalité

Je voudrais revenir sur les notions d'incarnation et de physicalité dont on entend souvent parler au théâtre et au cinéma. On parle d'incarner un personnage, et on pourrait aussi dire incorporer un personnage. On dit aussi travailler la physicalité d'un personnage. Pour moi, ces trois notions parlent de la même chose. La question qui se pose à l'acteur est la question de son corps physique. Comment peut-il transformer son corps physique ? Il s'agit de prendre conscience de ses limites, de ses habitudes, de ses conditionnements, afin de s'en affranchir le plus possible. Sinon, tous les personnages que créera un acteur auront tous les mêmes tics, ce que l'on peut voir chez certains acteurs. En travaillant l'incarnation - l'incorporation - la physicalité, l'acteur travaille sur son corps et le met au service de la création d'un personnage. Pour cela, l'exercice

de l'incorporation animale est à mon avis l'un des plus efficaces. Pour autant, l'acteur travaille toujours avec son propre corps, mais auquel il va faire faire d'autres mouvements, d'autres actions, et avec lequel il va ressentir d'autres états énergétiques.

4.3 Préparation

Une autre question se pose à l'acteur : celle de la préparation. Préparation avant une répétition, ou avant une représentation, ou avant un tournage. Chaque acteur a son rituel préparatoire avant de répéter, d'entrer en scène, ou sur un plateau de tournage. Je ne détaillerai pas ici les différentes préparations possibles ni les composantes de cette préparation. Mais pour ma part, l'exercice de l'incorporation animale en fait presque toujours partie. Et c'est d'ailleurs pour moi la dernière étape de la préparation. C'est l'exercice que je pratique généralement juste avant d'entrer en scène. C'est en effet un exercice qui me permet de trouver très rapidement l'énergie du personnage, c'est-à-dire l'énergie de l'animal sous-jacent à mon personnage. C'est aussi un exercice qui me met immédiatement dans mon corps, et qui me permet d'entrer en scène dans un état de présence intense. C'est donc un excellent exercice pour faire le meilleur usage possible de l'adrénaline qui est générée juste avant une représentation. Sur un plateau de tournage, il est un peu plus difficile d'incorporer cet exercice en préparation car on ne sait jamais vraiment à quel moment on va être appelé. C'est donc un exercice que je vais pratiquer beaucoup pendant les répétitions, mais peu pendant le tournage lui-même, ou bien simplement le matin afin de préparer mon corps à la journée de travail qui l'attend. Je vais néanmoins compter sur l'habituation de mon corps à l'énergie de l'animal pour que je puisse me la remémorer en quelques secondes à peine à chaque fois que je tournerai une scène.

5 Mise en scène et incorporation animale

L'enjeu d'un metteur en scène est différent de celui d'un acteur. Le metteur en scène doit avoir une vision globale de la pièce ou du film, et notamment de tous ses personnages. Il doit donc imaginer une incarnation, et donc une incorporation animale, pour chaque personnage.

5.1 Direction d'acteurs

L'exercice de l'incorporation animale est une composante importante de mon travail de metteur en scène, que ce soit pour le théâtre ou le cinéma. Je l'utilise en effet presque à chaque fois, et pour presque tous les personnages. C'est un exercice excellent pour sortir l'acteur de sa tête, et le connecter à son corps. C'est aussi un exercice très ludique qui permet de tisser des liens rapidement. Cela est dû, à mon avis, au fait que les corps tissent des liens plus rapidement que les têtes. Et les liens qui se tissent à travers cet exercice sont généralement plus profonds. C'est aussi un excellent exercice pour stimuler l'imaginaire de

l'acteur. Même lorsqu'un acteur n'a pas reçu de formation à la méthode de l'Actors Studio, c'est un exercice qui est facilement abordable et qui, même s'il demande du travail, porte toujours des fruits. C'est enfin un exercice qui permet à l'acteur de trouver rapidement une physicalité de personnage sur laquelle travailler.

5.2 Mise en corps

Comme je l'ai dit précédemment, l'un des enjeux majeurs de l'acteur est ce que j'appelle la mise en corps. En tant que metteur en scène, j'essaie d'aider l'acteur à sortir de ses habitudes corporelles. Une partie importante du travail sur le personnage consiste à trouver le corps de ce personnage, sa façon de bouger, de se déplacer, de parler, etc. Mais cela toujours à partir du corps de l'acteur, puisque c'est le seul outil qu'il a à sa disposition, son propre corps. L'exercice de l'incorporation animale permet de trouver tout de suite quelque chose en lien avec le corps du personnage. C'est une piste de travail parmi d'autres, mais elle est souvent une clé importante pour aller à la rencontre de son personnage.

5.3 Mise en espace

L'exercice de l'incorporation animale offre en outre un autre avantage. Une étape de l'exercice consiste en effet à faire explorer l'environnement du personnage à l'acteur incorporé en animal, ce qui permet de lui faire prendre en compte l'espace scénique choisi par le metteur en scène. Du coup, tout en travaillant son personnage et sa physicalité, l'acteur travaille également sur son appréhension de l'espace scénique tel qu'il a été conçu pour ce spectacle ou ce film. L'incorporation animale permet donc de travailler de nombreuses choses en même temps.

6 Coaching et incorporation animale

J'utilise également l'exercice de l'incorporation animale dans mon travail de coach.

6.1 Coaching d'acteurs

Le travail d'un coach d'acteurs est très semblable à celui d'un directeur d'acteurs. Il est donc tout à fait logique que chacun recourt à des exercices semblables. Il y a néanmoins quelques différences. Dans le cas d'une direction d'acteurs, qui intervient le plus souvent dans le cadre d'une mise en scène pour le théâtre ou le cinéma, c'est souvent le metteur en scène qui a choisi le sujet, les personnages, et les acteurs pour les incarner. Il a donc une vision très précise des directions qu'il souhaite voir prendre à ses acteurs pour incarner leurs personnages. Dans le cadre d'un coaching d'acteurs, outre le fait que c'est l'acteur

qui vient avec un personnage à travailler, que le coach n'a donc pas choisi lui-même, il y a également beaucoup moins de temps à disposition pour faire le travail, par rapport à un metteur en scène qui dispose des acteurs plusieurs heures par jour et plusieurs jours par semaine. En outre, le rôle d'un coach d'acteurs dépasse généralement le simple cadre d'un travail de personnages. Il est souvent question aussi de gestion de carrière, de motivation, de blocages artistiques, etc. L'exercice de l'incorporation animale offre néanmoins, dans ce cadre-là, de nombreux avantages. Outre le fait que c'est un exercice très efficace pour explorer la physicalité d'un personnage, comme nous l'avons déjà vu, c'est aussi un exercice très intéressant pour améliorer la connaissance de soi : d'une part, en notant les choix d'animaux qui sont faits par les acteurs, et qui ne sont jamais anodins ; d'autre part, en observant comment l'acteur aborde l'exercice, et où il en est dans son travail de lâcher prise. C'est donc un exercice qui peut s'avérer très révélateur des problématiques personnelles et professionnelles sous-jacentes de l'acteur.

6.2 Instincts

D'autre part, puisque l'incorporation animale consiste à faire appel à un animal pour développer des caractéristiques physiques et corporelles de cet animal, c'est aussi une technique qui permet de développer, ou tout du moins de contacter son instinct animal. C'est une chose qui est naturellement très intéressante dans le cadre d'un coaching d'acteur. Le recours à l'animal permet également de mettre une certaine distance par rapport à soi, ce qui permet souvent de pouvoir lâcher prise plus facilement.

6.3 Pratique sous forme d'ateliers

Il m'arrive régulièrement de proposer des ateliers de pratique d'incorporation animale. Cela a été le cas par exemple lors du congrès IDEA 2013, congrès autour du thème théâtre éducation qui a eu lieu à Paris en juillet 2013. J'ai proposé à ce congrès un atelier d'incorporation animale simplifié et destiné à un groupe d'environ 25 participants. L'atelier ayant une durée de deux heures, et existant en tant que tel, c'est-à-dire notamment sans travail d'observation d'animal préalable, j'ai donc proposé une version modifiée. Les participants ont commencé par un exercice de relaxation. Puis je les ai guidés à travers une incorporation animale, l'animal choisi étant le premier qui leur venait à l'esprit. Les participants ont d'abord tenté de ressentir l'animal de l'intérieur dans différentes parties de leur corps, puis dans leur corps tout entier. Je les ai ensuite fait se déplacer avec ce ressenti animal intérieur et, pour ceux qui le souhaitaient, je les ai finalement fait se rencontrer. Le dernier temps de l'atelier a été dédié à une discussion sur ce qui venait de se passer. Même en si peu de temps de pratique de l'exercice, on peut déjà observer des transformations étonnantes et radicales dans la physicalité des participants, dans leur énergie et dans leur lâcher prise.

7 Incorporation animale et chamanisme

Je souhaiterais terminer cet article en mentionnant le lien qui existe entre incorporation animale et chamanisme. Pour cela, j'aborderai tour à tour la question des animaux totems, et celle de la relation des Indiens d'Amérique du Nord avec les animaux.

7.1 Animaux totems

Pour de nombreuses traditions chamaniques, et c'est notamment le cas chez les Indiens d'Amérique du Nord, chacun d'entre nous a un animal totem. De nombreuses choses ont été écrites au sujet des animaux totems, mais je m'en tiendrai ici à quelque chose de très simple. Un animal totem est un animal qui représente, symbolise notre essence. C'est l'animal dont les caractéristiques (physiques, comportementales, énergétiques) vont nous guider tout au long de notre vie. Certaines personnes diront que l'esprit de cet animal nous habite, et que nous pouvons le solliciter à chaque fois que nous le souhaitons. Cet animal totem est une sorte de guide, et il pointe une direction de cheminement. Est-il nécessaire d'incorporer son animal totem pour le comprendre ? Ou pour puiser dans ses caractéristiques diverses ? Cela peut être intéressant afin d'expérimenter de l'intérieur certaines de ses caractéristiques, et afin de développer une meilleure compréhension de cet animal totem.

7.2 White Eagle

Il existe plusieurs moyens pour appréhender les caractéristiques animales telles qu'elles sont perçues par les Indiens d'Amérique du Nord. Plusieurs jeux de cartes sont disponibles, qui reprennent chacun un ensemble d'animaux, et qui donne des clés pour comprendre leurs énergies, leurs caractéristiques, etc. L'un d'entre eux, le jeu de White Eagle (Wa-Na-Nee-Che, 2007), comporte 28 animaux parmi lesquels l'aigle, le serpent, ou encore le colibri. Sur chaque carte, il y a un travail d'artiste représentant l'animal fait à l'aide de petites perles de couleur, ainsi qu'une photo représentant l'environnement et les attributs correspondant à cet animal. Dans le petit livret qui accompagne le jeu de cartes, on trouve une description du message qui est relié à cet animal, résumé en haut de page par trois caractéristiques principales. Les cartes peuvent être utilisées individuellement pour appréhender chacun de ces animaux, ou sous forme de tirages tels qu'ils sont décrits dans le petit livret. C'est un outil ludique qui permet d'entrer en contact avec une sagesse ancestrale, fruit de plusieurs millénaires de pratique expérimentale. Il peut être intéressant de coupler l'exercice d'incorporation animale avec l'utilisation de ces cartes. Cela permet de développer des liens plus étroits, plus intimes avec les animaux représentés et incorporés, et donc aussi avec l'environnement.

7.3 Medecine Cards

L'autre jeu est celui des Medecine Cards de Jamie Sams & David Carson (Sams & Carson, 1999). Il contient 52 animaux dont par exemple le papillon, la libellule ou la baleine. Le principe est exactement le même : le jeu est accompagné d'un petit livre qui donne des descriptions pour chaque animal, résumé en haut de page par une caractéristique principale. Plusieurs tirages sont proposés en début de livre pour utiliser les cartes. La plupart des animaux du jeu de White Eagle se trouvent aussi dans le jeu des Medecine Cards, et l'on retrouve d'ailleurs assez souvent des similarités entre les deux descriptions. Mais il y a également quelques différences, ces jeux étant issus de la tradition de tribus indiennes différentes. Ces jeux constituent une façon ludique d'aborder les rapports qu'ont les traditions chamaniques aux animaux. Il permet aussi d'inclure facilement cette exploration dans un autre cadre, comme par exemple celui du travail de l'acteur, de la direction ou du coaching d'acteurs. Le côté ludique des cartes et la portée symbolique des descriptions permettent d'utiliser facilement ces jeux en conjonction d'un exercice d'incorporation animal.

8 Conclusion

Je voudrais conclure cet article en proposant quelques pistes pour prolonger cette réflexion sur l'incorporation animale.

8.1 S'incarner

On parle souvent chez les acteurs d'incarner un personnage ou un rôle. Le Robert nous donne comme définition première du mot « incarner » : « revêtir (un être spirituel) d'un corps charnel, d'une forme humaine ou animale ». Lorsqu'un acteur pratique une incorporation animale, on pourrait dire qu'il revêt une forme animale. Mais cela va d'après moi bien plus loin que le simple fait de revêtir. Cela se passe aussi de l'intérieur, et également au niveau de l'énergie, qui n'est donc pas uniquement quelque chose qui recouvre, qui vêt. Je pense que cette transformation qui s'opère n'est pas seulement corporelle, mais aussi spirituelle.

8.2 L'inconscient

Dans ce processus d'incorporation animale, l'inconscient (Freud, 2000, 2004) joue très probablement un rôle. La phase d'observation est une phase d'imprégnation. Une empreinte se forme par osmose, et va imprimer, il me semble, directement l'inconscient. C'est aussi ce qui, à mon avis, permet cette incorporation animale. L'acteur devient alors autre chose que lui-même, tout en étant toujours lui-même. Là est le paradoxe que je n'essaierai pas de lever dans cet article.

8.3 S'inventer et s'envoler

Pratiquer l'exercice de l'incorporation animale est une façon de s'inventer, de se réinventer. À travers l'incorporation animale, il est possible de percevoir son corps autrement, et donc de se percevoir autrement. Il est possible, selon les animaux choisis, d'inventer une autre relation à son corps, donc à soi-même, et aux autres.

Il est possible de s'affranchir de certains conditionnements, ce qui est finalement l'un des buts communs de tous les exercices d'acteurs. Et en s'affranchissant de ses conditionnements, on devient plus léger, et on peut donc s'envoler, surtout si on a choisi comme animal à incorporer un animal qui vole, comme par exemple le colibri. Pour la tribu de White Eagle, le colibri représente l'amour, la joie et la musique. Je vous souhaite de très belles incorporations animales !

9 Références

- Chekhov, M. (1997a). *Être acteur: Pygmalion*.
- Chekhov, M. (1997b). *L'imagination créatrice de l'acteur: Pygmalion*.
- Cohen, L. (2010). *The Lee Strasberg Notes: Routledge*.
- Freud, S. (2000). *Œuvres complètes XIV. Leçons d'introduction à la psychanalyse: PUF*.
- Freud, S. (2004). *Œuvres complètes XIX. Nouvelle suite des leçons d'introduction à la psychanalyse: PUF*.
- Meyerhold, V. (1980-1992). *Écrits sur le théâtre : tomes 1 à 4: L'âge d'homme*.
- Sams, J., & Carson, D. (1999). *Medecine Cards: The Discovery of Power Through the Ways of Animals: Saint Martin's Press*.
- Stanislavski, C. (1937). *An Actor Prepares: Methuen Drama*.
- Stanislavski, C. (1950). *Building A Character: Methuen Drama*.
- Stanislavski, C. (1988). *Creating A Role: Methuen*.
- Stanislavski, C. (1996). *My Life in Art: Routledge*.
- Strasberg, L. (1986). *Le travail à l'Actors Studio: Gallimard, NRF*.
- Strasberg, L. (1987). *A Dream of Passion: Plume*.
- Wa-Na-Nee-Che. (2007). *White Eagle Medecine Wheel Deck: Connections Book*.

L'Héritage de Lee Strasberg (2015)

Ivan Magrin-Chagnolleau. **L'Héritage de Lee Strasberg**. *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.

7 pages

L'héritage de Lee Strasberg

Ivan MAGRIN-CHAGNOLLEAU

Résumé

Dans cet article, je discute l'importance de Lee Strasberg et de sa méthode de jeu pour le théâtre et le cinéma contemporains. Une première partie est consacrée à la genèse de cette méthode (souvent appelée méthode de l'Actors Studio). Puis, dans une deuxième partie, le travail en laboratoire est décrit succinctement. La troisième partie détaille les différentes façons dont cette méthode nous est transmise aujourd'hui.

Mots Clés

Lee Strasberg, La méthode de l'Actors Studio, Théâtre, Cinéma, Technique de jeu, Jeu de l'acteur, Acteur, Comédien, Laboratoire d'acteurs, Archives, Lee Strasberg Institute, Actors Studio, Constantin Stanislavski, Group Theater.

1 Introduction

Lee Strasberg a été une figure prééminente du théâtre américain du vingtième siècle, et par extension, du cinéma américain. Il a été un grand coach d'acteurs, et a développé une méthode de jeu reposant sur des exercices très concrets, dont le but est essentiellement d'aider l'acteur à être dans le moment présent. Quarante ans après, que reste-t-il de ces enseignements ? Quels documents ? Quel héritage ? Je tente, dans cet article, de donner quelques éléments de réponse.

Éléments biographiques sur Lee Strasberg

Lee Strasberg est né Israël Strasberg en 1901 dans l'ancien empire austro-hongrois (aujourd'hui Ukraine). Son père immigré aux États-Unis, à New York, peu de temps après sa naissance. À force de travail et d'économies, il fait venir progressivement toute la famille. C'est en 1909 que Lee rejoint son père à New York. Lee Strasberg passe alors toute sa vie à New York, en tant qu'acteur, metteur en scène, et coach d'acteurs. Il y meurt en 1982.

2 La genèse de la méthode

2.1 Contact avec Stanislavski

En 1923, Lee Strasberg assiste à une représentation du Moscow Art Theater dans une mise en scène de Stanislavski. Il est émerveillé par le jeu des acteurs, et remarque notamment que même les acteurs jouant des petits rôles sont également impliqués émotionnellement. Il fait alors tout pour apprendre cette technique de jeu. C'est avec Maria Ouspenskaya et Richard Boleslavsky, tous deux anciens élèves de Stanislavski, que Lee prend des cours. En parallèle, il lit tout ce qui a été écrit par ou sur Stanislavski.

Il comprend assez vite l'intérêt de cette technique, et la nécessité de la développer et de l'adapter pour les acteurs américains. Cela le conduit à mettre un peu sa carrière d'acteur de côté, au profit de celle de metteur en scène et de coach d'acteurs.

2.2 La recherche de Lee Strasberg

Lee Strasberg a passé sa vie à faire de la recherche sur les théories de jeu, et à essayer de répondre à plusieurs questions qu'un acteur se pose. Il s'est notamment intéressé à la vérité émotionnelle de l'acteur, et à tout ce qui peut l'aider à être dans le moment présent. Il est d'abord parti du système de Stanislavski, tel que ce dernier l'avait décrit, non seulement dans ses livres mais aussi au cours de ses stages et conférences. Puis il a étudié en particulier certains problèmes que Stanislavski lui-même n'avait pu résoudre. Il s'est également appuyé sur plusieurs auteurs dont les recherches découlaient directement du système Stanislavski : Meyerhold, Michael Chekhov, et Vakhtangov.

Parallèlement à cette recherche, il voulait adapter cette approche aux acteurs américains. Il a donc mis en place une série d'exercices dont le but était de donner à l'acteur des moyens concrets de travailler sur sa vérité émotionnelle, sa concentration, sa présence, etc. Cette méthode s'est développée au sein de trois structures, le Group Theater, l'Actors Studio et, vers la fin de sa vie, le Lee Strasberg Institute.

2.3 Le Group Theater

Le Group Theater a été créé en 1931 par un groupe de jeunes comédiennes et comédiens qui avaient découvert le système Stanislavski à travers ses livres, ses stages, ses conférences. Il était composé notamment de Lee Strasberg, d'Elia Kazan, de Cheryl Crawford, d'Harold Clurman, de Stella Adler, de Sanford Meisner, de Robert Lewis, etc. Il fonctionnait comme le studio d'art de Moscou. Un groupe de comédiennes et de comédiens se réunissait régulièrement pour travailler des scènes, faire des exercices, improviser des situations, et créer de nouvelles pièces. Cette façon de travailler a permis de développer plusieurs pièces importantes du répertoire américain. L'auteur Clifford Odets

faisait notamment parti de ce groupe. Harold Clurman a raconté l'aventure du Group Theater dans son livre *The Fervent Years*. Le Group Theater a été interrompu par la seconde guerre mondiale.

2.4 L'Actors Studio

En 1947, juste après la seconde guerre mondiale, Elia Kazan et Cheryl Crawford décident de reconstituer un studio laboratoire à la manière du Group Theater. Ils créent l'Actors Studio. En 1949, ils demandent à Lee Strasberg de les rejoindre. Lee Strasberg devient alors la figure prééminente de l'Actors Studio. Sa forte personnalité et son envie de faire évoluer la technique de jeu de l'acteur l'imposent assez vite comme un coach incontournable pour de nombreux acteurs américains. Certains d'entre eux vont devenir très connus au cinéma grâce à la maîtrise de cette méthode.

3 Le travail en laboratoire d'acteurs

L'Actors Studio devient pour Lee Strasberg un espace d'exploration et d'expérimentation. Il peut y tester de nouveaux exercices, en les proposant aux participants du laboratoire. Il cherche notamment à répondre à certaines questions de Stanislavski restées sans réponse. La séance de travail se structure assez vite et trouve sa forme finale au bout de quelques années d'expérimentation.

Un laboratoire d'acteurs, tel que conçu par Stanislavski, et repris par les membres du Group Theater, puis de l'Actors Studio, est avant tout un espace d'exploration et d'entraînement (training) pour les acteurs. C'est un espace où un acteur peut travailler sur sa technique, sur des personnages et des scènes. C'est aussi un espace où il peut travailler avec ses pairs sans peur du jugement d'un metteur en scène, d'un réalisateur, ou d'un directeur de casting.

Lee Strasberg concevait des séances de travail de 4h. Les deux premières heures étant consacrées à des exercices, et les deux suivantes à un travail de personnages et de scènes. Les exercices étaient supervisés par Lee Strasberg, et avaient pour but d'aider l'acteur à développer sa technique. Il y avait des exercices pour se relaxer, développer sa concentration, entraîner sa mémoire sensorielle, développer son animalité, travailler sur le lâcher-prise, etc.

Par exemple, dans l'exercice du moment privé, l'acteur doit choisir une activité très privée, qu'il ne ferait pas devant quelqu'un, et exercer cette activité devant la classe, après avoir d'abord recréé le cadre de cette activité grâce à sa mémoire sensorielle. L'exercice de l'animal, quant à lui, consiste à devenir l'animal choisi de l'intérieur, puis à le transformer progressivement en personnage.

Les deux dernières heures étaient consacrées à un travail de personnages et de scènes. Deux acteurs proposaient par exemple une scène d'une pièce du

répertoire. Ils pouvaient prendre du temps pour se préparer avant de présenter la scène. Une fois la scène terminée, les autres acteurs pouvaient faire des retours, en tenant compte de ce que les acteurs présentant la scène avaient voulu particulièrement travailler. Lee Strasberg donnait des pistes pour aider les acteurs à avancer dans leur travail.

4 Transmission de la méthode de Lee Strasberg

4.1 Les cassettes vidéo

Dans les dernières années de sa vie, Lee Strasberg enseignait et coachait encore. Il était perçu par ses amis et ses proches comme quelqu'un de très dynamique et en parfaite santé. De nombreuses séances de travail avec ses élèves ont été filmées. La plupart de ces cassettes sont conservées au Lee Strasberg Institute. En revanche, leur accès est restreint et privé. Les archives du Lee Strasberg Institute comportent aussi de très nombreux documents écrits, dont une grande partie a maintenant été confiée à la bibliothèque du Congrès américain à Washington (Library of Congress).

4.2 Les livres

De très nombreux livres ont été écrits sur la méthode de Lee Strasberg, souvent appelée « la méthode ». Mais plusieurs livres sont en lien direct avec Lee Strasberg, sous une forme ou sous une autre. Il y a tout d'abord le livre écrit par Lee Strasberg, *A Dream of Passion*, qui décrit comment Lee Strasberg a développé sa méthode, quelles ont été ses influences, ses recherches, et les exercices qu'il a finalement mis en place pour l'entraînement des acteurs. Il y a aussi le livre intitulé *Strasberg at the Actors Studio*, édité par Robert H. Hethmon à partir des cassettes du Lee Strasberg Institute, et qui décrit ce qu'est un studio-laboratoire d'acteurs, et comment Lee Strasberg travaillait avec ses étudiants, ainsi que les exercices qu'il leur faisait faire. Et plus récemment, *The Lee Strasberg Notes* édité par Lola Cohen, également à partir des cassettes du Lee Strasberg Institute, qui synthétise l'enseignement de Lee Strasberg ainsi que la plupart de ses idées sur le théâtre, la mise en scène, et quelques grands auteurs du répertoire dramatique. Ainsi que *The Method Acting Exercises Handbook* de Lola Cohen, qui reprend en détail tous les exercices de la méthode. Ces livres constituent une source importante d'informations sur ce qu'est la méthode.

4.3 L'enseignement de ses élèves

Mais un livre ne saurait remplacer l'enseignement d'un professeur expérimenté. Et un certain nombre d'anciens élèves de Lee Strasberg enseignent encore aujourd'hui sa méthode au sein de plusieurs institutions.

4.3.1 L'Actors Studio

L'Actors Studio est le lieu privilégié où l'on peut recevoir, aujourd'hui, l'enseignement de la méthode de Lee Strasberg. L'Actors Studio fonctionne toujours comme un studio-laboratoire d'acteurs. Il est donc possible, pour ses membres, de proposer des scènes, qui sont modérées par un membre plus expérimenté du studio. L'emphase est mise sur l'utilisation des outils de la méthode. Seuls les membres du studio peuvent y présenter des scènes. Mais la procédure pour devenir un membre du studio est réservée à des acteurs professionnels, et elle est très compétitive. Il existe heureusement deux autres structures qui, elles, s'adressent à des acteurs en début de carrière.

4.3.2 Le Lee Strasberg Institute

Le Lee Strasberg Institute a été fondé par Lee Strasberg et sa femme Anna Strasberg en 1969. Lee Strasberg ressentait en effet le besoin d'ouvrir son enseignement à des acteurs en début de carrière. Les séances de travail se déroulent de la même façon que les sessions de travail à l'Actors Studio, du temps où Lee Strasberg les animait. Chaque session dure 4h. Les deux premières heures sont consacrées aux exercices de base, et les deux suivantes au travail de scènes et de personnages. Bien que le Lee Strasberg Institute ait été fondé par Lee Strasberg, il n'a pas de lien formel avec l'Actors Studio.

4.3.3 L'école de l'Actors Studio

L'Actors Studio Drama School, en revanche, est une émanation de l'Actors Studio. Là encore, il s'agit de proposer l'enseignement de la méthode à des acteurs en début de carrière. Mais cet enseignement se fait dans le cadre d'un Master of Fine Arts (M.F.A.), et le cycle d'enseignement de trois ans est validé par un diplôme. Les enseignants d'art dramatique sont en majorité des membres de l'Actors Studio. La première année est consacrée essentiellement à l'apprentissage des techniques de base, la deuxième année à leur application à un travail de scènes ou de personnages, et la troisième année à la préparation d'un rôle.

Enfin, on trouve aussi l'enseignement de la méthode de Lee Strasberg dans de nombreuses écoles d'art dramatique aux États-Unis, la plupart dans le cadre d'un cursus universitaire, même si cette méthode n'est généralement qu'une méthode enseignée parmi d'autres.

4.4 Archives vivantes et héritage

“La méthode est un amalgame du travail de Stanislavski, Vakhtangov, Meyerhold, et le Group Theater. Observer et analyser leur travail est essentiel pour se former une compréhension pratique d'une approche théorique de notre travail, l'entraînement de l'acteur, et de son utilisation dans diverses formes de production. Les gens feraient une erreur s'ils

considéraient une seule de ces sources comme pouvant représenter à elle seule toute l'essence de la méthode.”

Cette citation de Lee Strasberg permet de comprendre que la méthode n'est pas figée. D'une part, elle est elle-même le résultat d'un mélange de plusieurs approches passées au filtre de la réflexion et de l'expérimentation de Lee Strasberg. D'autre part, il y a autant de méthodes qu'il y a d'acteurs, chaque acteur incorporant à sa façon les outils qui sont mis à sa disposition au cours de son apprentissage.

La question de l'héritage de Lee Strasberg et des archives qu'il a laissées derrière lui n'est donc pertinente que dans la mesure où elle permet à un acteur de trouver son propre chemin dans l'acquisition de sa propre méthode. Et c'est bien dans ce sens que j'ai proposé cette réflexion. La majorité des archives concernant Lee Strasberg est rassemblée au Lee Strasberg Institute, et ces archives sont gérées par Anna Strasberg, la dernière épouse de Lee Strasberg, et plus récemment par Victoria Krane, la sœur d'Anna Strasberg. Il existe des archives filmées et des archives papier.

4.5 Les archives filmées

Les archives filmées sont constituées d'un nombre important de cassettes, qui sont, pour la plupart, des captations de cours donnés par Lee Strasberg lui-même au Lee Strasberg Institute pendant les dernières années de sa vie. Elles permettent donc de se faire une idée sur la façon dont Lee Strasberg enseignait les exercices qu'il avait développés. Elles sont aussi intéressantes car Lee Strasberg y parle de nombreuses choses en relation avec l'art de l'acteur et avec le théâtre. Une petite partie de ces cassettes a été numérisée, le reste étant dans un état plus ou moins dégradé par le temps, et courant le risque de devenir à terme inutilisable.

4.6 Les archives papier

Les archives papier, quant à elles, regroupent une partie de la correspondance de Lee Strasberg ainsi que de nombreuses photos et de nombreux articles de journaux le concernant. Elles sont conservées à la librairie du Congrès américain à Washington.

5 Conclusion

Lee Strasberg a été une personnalité très influente du théâtre américain du XXe siècle. Il a développé une méthode qui a fait beaucoup parler d'elle. Cette méthode a permis à de nombreux acteurs d'exceller dans leur art sur scène et à l'écran. Elle est encore aujourd'hui la base d'un grand nombre de formations d'acteurs de qualité, notamment aux États-Unis. La recherche passionnante effectuée par Lee Strasberg a laissé derrière elle un certain nombre de traces

et d'archives. Ces archives présentent un grand intérêt à quiconque chercherait à entrer au cœur de cette méthode, et j'espère que cet article donnera à bon nombre de personnes l'envie de s'y intéresser.

6 Bibliographie

- Stella Adler: The Art of Acting*** édité par Howard Kissel. Applause. 2000.
- Stella Adler on Ibsen, Strindberg, and Chekhov*** édité par Barry Paris. Vintage 2000.
- Stella Adler on America's Master Playwrights*** édité par Barry Paris. Vintage. 2013.
- Richard Boleslavsky. ***Acting: The First Six Lessons***. Routledge. 2005.
- Michael Chekhov. ***On The Technique of Acting*** édité par Mel Gordon. Harper. 1991.
- Michael Chekhov. ***To The Actor***. Routledge. 2002.
- Harold Clurman. ***The Fervent Years: The Group Theatre & the 30's***. Da Capo Press. 1983.
- The Collected Works of Harold Clurman*** édité par Marjorie Loggia and Glenn Young. Applause Theatre & Cinema Books. 2000.
- Harold Clurman. ***On Directing***. Fireside. 1997.
- Harold Clurman. ***The Naked Image: Observations on the Modern Theatre***. The Macmillan Company. 1966.
- Harold Clurman. ***Lies Like Truth***. The Macmillan Company. 1958.
- The Lee Strasberg Notes*** édité par Lola Cohen. Routledge. 2010.
- Lola Cohen. ***The Method Acting Exercises Handbook***. Routledge. 2017.
- Elia Kazan. ***Kazan on Directing***. Vintage. 2010.
- Elia Kazan. ***A Life***. Da Capo Press. 1997.
- Elia Kazan. ***An American Odyssey***. Bloomsbury Publishing. 1988.
- Elia Kazan: Interviews*** édité par William Baer. University Press of Mississippi. 2000.
- Robert Lewis. ***Method or Madness?*** Samuel French. 1958.
- Sanford Meisner and Dennis Longwell. ***Sanford Meisner on Acting***. Vintage. 1987.
- Meyerhold on Theatre*** édité par Edward Braun. Methuen. 1978.
- Meyerhold Speaks, Meyerhold Rehearses*** édité par Aleksandr Gladkov. Routledge. 1997.
- Meyerhold. ***Écrits sur le théâtre***. Vol. 1 à 4. L'âge d'homme. 1980-2009.
- Constantin Stanislavski. ***An Actor Prepares***. Routledge. 1989.
- Constantin Stanislavski. ***Building A Character***. Routledge. 1989.
- Constantin Stanislavski. ***Creating A Role***. Routledge. 1989.
- Constantin Stanislavski. ***My Life in Art***. Routledge. 1987.
- Lee Strasberg. ***A Dream of Passion: The Development of the Method***. Plume. 1988.
- Strasberg at the Actors Studio*** édité par Robert H. Hethmon. Theatre Communications Group. 1993.
- The Vakhtangov Sourcebook*** édité par Andrei Malaev-Babel. Routledge. 2011.

**L'acteur et les émotions au théâtre et au cinéma
(2016)**

Ivan Magrin-Chagnolleau. **L'acteur et les émotions au théâtre et au cinéma.** *p-e-r-f-o-r-m-a-n-c-e*, 3, 2016.

11 pages

L'acteur et les émotions au théâtre et au cinéma

Ivan MAGRIN-CHAGNOLLEAU

1 Introduction

En tant qu'acteur¹, je m'intéresse aux émotions depuis longtemps. Je considère en effet qu'il s'agit du matériau brut de l'acteur. Je m'y intéresse également en tant que metteur en scène, puisque mon rôle est alors de faciliter l'émergence de l'émotion chez l'acteur que je mets en scène.

Je vais donc mener dans cet article une discussion sur ce que sont les émotions pour un acteur et un metteur en scène, et la façon de les aborder. Je discuterai également, dans ce contexte, la question de l'artialisation des émotions, c'est-à-dire l'hypothèse selon laquelle l'art contribue à façonner nos émotions.

2 Qu'est-ce que l'émotion ?

Je voudrais partir des définitions proposées par le Grand Robert de la langue française [1]. Le Grand Robert nous donne plusieurs définitions du mot émotion. L'émotion peut être un « mouvement affectant un individu et ayant pour effet de le soustraire à l'état de repos et d'équilibre », ou bien « ce mouvement, considéré dans ses effets physiologiques », ou encore un « mouvement plus complexe, intéressant le cœur autant que la vie organique ». Le Grand Robert nous donne une autre définition plus psychologique : « état de conscience complexe, généralement brusque et momentané, accompagné de troubles physiologiques (pâleur ou rougissement, accélération du pouls, palpitations, sensation de malaise, tremblements, incapacité de bouger ou agitation) ». Et par extension, une « sensation (agréable ou désagréable), considérée du point de vue affectif ».

L'Encyclopædia Britannica nous donne, quant à elle, la définition suivante [17] : « a complex experience of consciousness, bodily sensation, and behaviour that reflects the personal significance of a thing, an event, or a state of affairs² ».

¹J'utiliserai dans cet article les mots « acteur » et « comédien » de façon interchangeable. Tout ce que je dirai s'applique aussi bien au théâtre qu'au cinéma.

²Une expérience complexe de conscience, sensation corporelle et comportement reflétant la signification personnelle d'une chose, d'un événement ou de l'état des choses.

Je voudrais pour ma part partir de l'étymologie du mot é-motion. Il y a dans ce mot la notion de mouvement. C'est également le cas pour le mot é-mouvoir et pour le mot é-meut. Je traduis personnellement cette notion de mouvement de deux façons différentes. Tout d'abord en constatant que l'émotion est un état transitoire. C'est un état énergétique qui est mouvant, donc qui se transforme de quelque chose en quelque chose. L'émotion n'est pas un état permanent. Et il y a aussi le fait que l'émotion peut mettre en mouvement. Elle met en mouvement comme conséquence de cet état transitoire qui agit en nous.

Cela me ramène au fait que l'émotion est une expérience complexe. L'émotion est un phénomène qui comprend plusieurs composantes. Pour moi, l'émotion fait appel à au moins trois composantes différentes : la composante de la conscience, qu'on peut aussi appeler cognitive ; la composante corporelle ; et enfin la composante comportementale. Donc quand nous parlons de l'émotion, nous parlons parfois de la composante cognitive. Parfois, il s'agit plutôt de la composante corporelle. Et parfois, de la composante comportementale. Et si on ne prend pas la peine de clarifier cela, on peut avoir l'impression qu'on ne parle pas de la même chose, alors qu'en réalité on parle de la même chose mais pas des mêmes composantes de cette réalité émotionnelle.

3 Le travail sur les émotions selon Constantin Stanislavski

Je vais m'intéresser dans cet article au travail sur les émotions tel qu'il a été proposé par Constantin Stanislavski [2][4][5][3], puis repris par Lee Strasberg [16][11][12][13]. Il s'agit pour moi de m'appuyer à la fois sur une base théorique, mais aussi sur une base pratique, puisque j'ai été formé en tant qu'acteur à l'Actors Studio, école de comédiens newyorkaise reprenant la technique développée par Lee Strasberg, cette dernière s'appuyant en grande partie sur celle proposée par Stanislavski. Je vais donc parler de cette technique comme quelqu'un qui a vécu ce processus d'apprentissage, et qui s'est en même temps posé de nombreuses questions sur ce processus : à quoi ça sert et qu'est-ce qu'on cherche ? Ce sont donc aussi toutes ces choses-là que je vais essayer de transmettre. J'aborderai ensuite, dans un deuxième temps, la notion d'artialisation des émotions dans ce contexte-là.

Stanislavski est un russe qui a vécu à cheval sur la fin du XIXe siècle et le début du XXe siècle. Il est l'auteur d'une technique de jeu qui a révolutionné le théâtre. Il a été le premier à proposer une théorie du jeu. Il n'a pas été le premier à parler du jeu de l'acteur et de ses questionnements [7][18], mais c'est le premier à avoir pensé le jeu de l'acteur comme une théorie complète. Il est l'un des premiers à avoir dépassé le simple cadre d'une observation. Son problème était un problème de praticien. Il était comédien. Il avait décidé de devenir comédien à l'âge de cinq ans en voyant une performance, non pas d'un acteur, mais de la grande basse Russe Chaliapine. Ce dernier, en plus d'être une très grande basse Russe, était, au dire de Stanislavski, un remarquable acteur.

Et Stanislavski a été tellement frappé, à l'âge de cinq ans, par la vérité de la performance de Chaliapine qu'il a décidé, non seulement, de devenir comédien, mais aussi de passer l'essentiel de sa vie à comprendre comment on pouvait donner une performance avec autant de vérité.

Stanislavski a compris très tôt qu'il y avait une différence entre par exemple la performance de Chaliapine et celle d'un grand nombre de comédiens de l'époque, qui étaient plutôt dans l'imitation et dans le faire. Pour Stanislavski, la différence résidait dans une vérité émotionnelle que l'on retrouvait dans la performance de Chaliapine mais pas dans celle de la plupart des comédiens de l'époque. Stanislavski a donc réfléchi toute sa vie à cette question de la vérité émotionnelle, en tant que comédien d'une part, en tant que metteur en scène d'autre part, et enfin en tant que professeur de théâtre. Le fruit de ce va-et-vient entre théorie et pratique nous a été livré principalement dans trois ouvrages de Stanislavski : *An Actor Prepares*, 1937 ; *Building A Character*, 1950 ; *Creating A Role*, 1961. Mais cette méthode telle que livrée dans ses livres est incomplète. Tout d'abord, parce qu'il n'a rédigé que trois volumes d'une théorie qui devait en comporter davantage. D'autre part, parce que ces trois ouvrages ont été rédigés plusieurs années avant sa mort, et que la dernière partie de sa vie a été très riche en découvertes et avancées diverses. Ces dernières nous ont été transmises essentiellement par ses élèves, principalement Michael Chekhov avec son geste psychologique [14][15], et Meyerhold avec sa biodynamique [19]. Ces deux techniques très physiques découlent directement du travail de Stanislavski, mais bien évidemment approfondies chacune par le travail de leurs auteurs.

L'enjeu de tout ce travail réalisé par Stanislavski est que l'acteur n'a pas un accès direct aux émotions. C'est le grand problème de l'acteur, son plus grand paradoxe. Il ne peut pas se dire : « Je vais avoir peur ! » et avoir peur instantanément. Il ne peut pas se dire : « Je vais être triste ! » et être triste instantanément. Et il n'est pas question non plus pour Stanislavski de faire semblant d'avoir peur ou de faire semblant d'être triste. Mais la grande découverte de praticien de Stanislavski, c'est qu'on peut trouver des déclencheurs de ces émotions. On peut trouver des choses à faire qui vont éventuellement générer une émotion spontanée. Et tout l'enjeu est là.

4 Le travail sur les émotions selon Lee Strasberg

4.1 Mémoire sensorielle

Stanislavski a développé à partir de cette constatation deux familles d'exercices, qui ont été ensuite repris par Lee Strasberg au sein du Group Theater [9], puis au sein de l'Actors Studio [16][11][12][13]. Il s'agit des exercices de mémoire sensorielle et des exercices de mémoire affective. Voici en quelques mots en quoi ces exercices consistent (voir [16][11][12][13] pour des explications plus détaillées).

Une mémoire sensorielle consiste à utiliser ses sens, mais au niveau de la mémoire, c'est-à-dire qu'on va par exemple essayer de reproduire une sensation mais, au lieu de la vivre, on va essayer d'imaginer la vivre. Ça a l'air un peu compliqué comme ça, mais c'est en fait assez simple. Prenons un exemple. Considérons la boisson que l'on prend le matin au petit déjeuner. Si vous voulez créer une mémoire sensorielle de cet événement-là, vous allez commencer par vivre l'événement plusieurs fois. Donc, tous les matins, vous allez prendre, par exemple, votre café. Et vous allez, pendant plusieurs jours, focaliser votre attention dans le moment sur tout ce que vous allez pouvoir noter sensoriellement au cours de cette expérience : la chaleur de la tasse quand on la prend dans sa main, son poids, la couleur du café, son arôme, la chaleur du café, la sensation kinesthésique de porter la tasse aux lèvres, le premier contact au niveau de la lèvre, puis dans la gorge, la deuxième gorgée qui n'a pas le même goût que la première, etc.

On fait cet exercice plusieurs matins de suite. Puis, au bout de quelques jours, on va essayer d'imaginer de toutes pièces cette expérience. On va donc imaginer sensoriellement tous les éléments de cette expérience, mais sans rien utiliser de réel. La vraie tasse sera juste à côté au cas où, mais on fait l'exercice sans rien. Vous allez donc imaginer prendre la tasse de café, sans tasse de café, mais en rappelant la mémoire sensorielle d'avoir tenu une tasse de café dans la main plusieurs matins de suite. Puis vous allez imaginer la porter à vos lèvres, etc. Et comme vous avez fait cette expérience réelle plusieurs fois de suite, alors en faisant appel à votre mémoire sensorielle, vous allez être capable de revivre uniquement par la mémoire certains éléments de cette expérience. Et comme la tasse réelle n'est pas loin, vous pourrez toujours revenir à l'expérience réelle afin de renforcer les éléments de mémoire sensorielle encore incertains. Et au bout d'un moment, vous développerez cette capacité de vivre une expérience très similaire à l'expérience réelle, mais fabriquée uniquement par votre mémoire sensorielle.

Mais à quoi peut bien servir une telle technique ? Cette technique est très utile pour un acteur pour plusieurs raisons. Elle sert, d'une part, à développer une capacité d'attention, et donc d'être dans le moment présent. Tout l'enjeu du métier d'acteur, c'est de développer cette capacité de présence intense dans le moment présent. Et cette technique, quand même elle n'aurait que ce mérite, développe cela. D'autre part, à partir du moment où on devient capable, après de nombreux exercices de mémoire sensorielle, de créer des réalités entières, ça permet à l'acteur qui est sur scène de créer une réalité qui n'existe pas réellement sur scène, mais que lui va faire exister pour lui. Cela permet à l'acteur de se baigner dans cette réalité. Et si il est capable de se baigner dans cette réalité, d'après Stanislavski, il sera capable de transmettre cette réalité, ou tout au moins une partie de cette réalité virtuelle fabriquée, à son public. Ou au moins quelque chose qui aura l'air d'être vraiment vécu.

4.2 Mémoire affective

Il y a un autre intérêt à cette mémoire sensorielle. Elle sert à travailler sur l'autre exercice mentionné, l'exercice de mémoire affective, qui est l'un des autres piliers de cette technique. J'en viens donc maintenant à l'exercice de mémoire affective. Il s'agit d'un des exercices les moins compris. L'exercice consiste à choisir un événement du passé, suffisamment loin dans le passé, et qui a été porteur, au moment où il s'est déroulé, d'une émotion forte. Ça peut être, par exemple, une rupture sentimentale, un accident physique corporel, etc. Il est important que cet événement soit suffisamment loin dans le passé afin que le temps ait pu opérer à une forme de cicatrisation émotionnelle. Il faut néanmoins que cet événement soit resté dans la mémoire comme un événement ayant provoqué une émotion forte. À partir de là, le but de l'exercice n'est pas nécessairement de recréer cette même émotion, mais de recréer par mémoire sensorielle toutes les circonstances de cet événement, ce qui, la plupart du temps, va créer dans le moment présent une émotion également forte, mais qui peut être très différente de l'émotion originale.

J'insiste sur le fait que l'émotion qui surgit dans le moment présent peut être, et généralement est, différente de l'émotion originale. Par exemple, un événement qui a été dramatique dans le passé peut générer du rire à distance. Ou inversement, un moment passé qui a été idyllique peut générer une forme de nostalgie en se le remémorant à distance. On ne sait donc généralement pas à l'avance l'émotion qui va être créée par un exercice de mémoire affective, mais le postulat posé est que si l'événement du passé a généré une émotion forte, alors il sera déclencheur d'une autre émotion forte dans le présent, l'enjeu étant que cette émotion soit suffisamment stable pour qu'elle puisse être utilisée par le comédien dans la durée. J'insiste bien sur le fait que l'émotion déclenchée est généralement différente de l'émotion originale.

D'autre part, l'émotion générée dans le moment présent peut évoluer dans le temps. Un même événement peut générer pendant un certain temps une émotion bien particulière, puis à un moment donné, se mettre à générer une autre émotion. L'enjeu est donc, pour un acteur, de continuer à faire régulièrement ses exercices de mémoire affective, comme les gammes d'un musicien, afin de toujours savoir en permanence quelle mémoire affective déclenche telle émotion. Le comédien aura alors une palette de déclencheurs lui permettant de générer différentes émotions sur commande. À noter cependant que, dans de nombreuses circonstances, le comédien n'a pas besoin de ces exercices. Les circonstances, le contenu de la pièce, le partenaire suffiront à permettre l'émergence chez le comédien de l'émotion nécessaire. Mais quand ça ne sera pas le cas, le comédien aura alors à sa disposition un exercice qui lui permettra d'arriver au même résultat. Comme le dit Stanislavski, les exercices sont là pour tous les moments où un comédien n'a pas la grâce.

Ce que cet exercice de mémoire affective met en évidence, c'est la différence entre l'émotion vécue dans le passé et l'émotion déclenchée dans le présent.

Il y a un lien entre les deux, mais c'est un lien très élu­sif. Mais en revanche, ce qui fonctionne pour Stanislavski et pour tous ceux qui ont travaillé avec ce genre d'exercice, c'est que quelle que soit l'émotion qui va être créée par un exercice de mémoire affective, elle sera vraie. Ce sera une émotion que l'acteur ressentira vraiment dans le moment présent. Cette émotion est liée à l'exercice que l'acteur est en train de réaliser, mais elle est vraie.

5 L'interprétation de l'acteur

On en vient maintenant au débat entre naturalité et culturalité des émotions. Selon moi, la partie culturelle de l'émotion n'est pas l'émotion elle-même, mais ce qui la déclenche, ainsi que la façon de l'exprimer, ou de ne pas l'exprimer. Les émotions, quant à elles, sont d'après moi naturelles. Si je dis « peur », cela renvoie à une expérience émotionnelle qui a toujours existé et qui existera toujours, et dont la nature est universelle. En revanche, ce qui est culturel, et donc qui peut être éventuellement sujet à une artialisation, c'est le fait que, d'une part, ce qui va déclencher la peur n'est pas la même chose en fonction des cultures et, d'autre part, la façon d'exprimer cette peur ou de ne pas l'exprimer sera également différente selon les cultures. Qu'en est-il de l'acteur ? Comment l'acteur se situe-t-il par rapport à l'artialisation des émotions ?

C'est à l'acteur, finalement, de trouver ses propres déclencheurs d'émotions, et un acteur français, par exemple, ne va pas le faire de la même façon qu'un acteur russe, japonais ou américain. Et ça va être lié, non seulement, à sa culture, mais aussi à son vécu, à ses expériences, et à une pratique artistique déjà ancrée dans ses représentations cognitives. Et c'est en cela qu'il peut y avoir une artialisation des émotions chez l'acteur. L'exemple du comédien est très intéressant car un comédien a généralement des modèles, des comédiens et des comédiennes dont il admire le travail, auxquels il se réfère. Il a déjà vu généralement un très grand nombre de films, de spectacles, et il est donc possible qu'il soit influencé par un certain codage des émotions. Et c'est dans ce contexte qu'il peut y avoir une transformation de ce qui déclenche une émotion, de comment l'exprimer, de ce qui est autorisé et ne l'est pas.

Il y a une autre chose intéressante à constater : il y a une différence entre l'émotion de l'acteur et l'émotion du personnage. Mais cette différence est beaucoup plus élu­sive. Si je vais un peu plus loin, je me pose la question suivante : où se trouve l'émotion dans l'écriture dramatique ou scénaristique ? La question se pose lorsque l'on travaille à partir d'un texte, mais elle se pose aussi dans le cas d'une création collective contemporaine par exemple. Donc où se trouve l'émotion dans l'écriture ? Elle se trouve à plusieurs endroits. Elle se trouve dans les didascalies quand il y en a. Et d'ailleurs, l'émotion indiquée dans les didascalies n'est pas toujours la bonne. Ce n'est pas parce que l'auteur a indiqué une émotion particulière dans une didascalie que l'acteur va choisir de travailler avec cette émotion-là à ce moment-là. L'émotion se trouve aussi dans ce qu'on appelle au théâtre le sous-texte, c'est-à-dire dans ce qui n'est pas

dit justement, mais qui est l'enjeu de la scène en question, qui en général se manifeste sous la forme d'un conflit. Et enfin, l'émotion se trouve aussi occasionnellement dans le dialogue lui-même. Il y a des dialogues qui verbalisent certaines émotions, mais rarement toutes.

Donc le choix de l'acteur - et c'est précisément cela que j'appelle le travail d'interprétation de l'acteur - réside dans le choix d'une émotion particulière plutôt qu'une autre à un moment donné du spectacle, en relation avec la scène, les circonstances, un comportement. Et cette émotion est choisie éventuellement en lien avec les didascalies, le sous-texte, et les dialogues s'ils indiquent une émotion particulière.

Mais l'enjeu pour l'acteur, quel que soit son choix, est de créer une émotion vraie. Et c'est encore une fois à ce niveau-là qu'il peut y avoir, d'après moi, une artialisation des émotions. Dans la mesure où, aujourd'hui, on a accès à de nombreux films, captation de spectacles, etc., il y a le danger du cliché, c'est-à-dire de se dire par exemple « quelqu'un qui est amoureux fait comme ça parce que je l'ai vu dans un film ou dans un spectacle », et sous-entendu « ressent ça ». Ou bien « quelqu'un qui est en colère fait comme ça et ressent ça ». Et donc l'artialisation des émotions qui est en jeu ici, du moins dans l'acception que je propose, constitue en fait un obstacle au travail du comédien, dans la mesure où ce que le comédien recherche est plutôt une vérité du moment présent.

Mais ce n'est pas seulement un obstacle. C'est aussi quelque chose qui peut aider. Tout dépend finalement comment on s'en sert. Cela peut en effet être aussi une aide parce que ce dont le comédien s'est imprégné de par son vécu artistique, c'est aussi ce qui a construit ses représentations mentales et émotionnelles, et ce qui va précisément le guider dans la recherche d'une émotion vraie. Il me semble que ces deux arguments sont en faveur d'une artialisation des émotions, au niveau de ce qui déclenche une émotion, et au niveau de la façon de l'exprimer ou non.

6 Différences au niveau du travail sur les émotions entre le théâtre et le cinéma

Avant d'aller plus loin, je voudrais juste mentionner les différences qui existent entre le jeu pour le cinéma et le jeu pour le théâtre. Je pense que les différences sont de trois natures. Et je vous montrerai que ces différences existent, et qu'en même temps elles n'existent pas vraiment. Le rôle du corps n'est pas le même au théâtre et au cinéma, a priori. Et je dis bien a priori. Le rôle du visage n'est pas non plus le même. Et le rôle de l'entendre n'est pas le même.

Au théâtre, le visage, dans la majorité des cas et pour la majorité des spectateurs, n'est pas visible, sauf si le spectateur a des jumelles. Donc si l'émotion n'était exprimée que par le visage, elle serait en grande partie perdue au théâtre.

Au cinéma, en revanche, la caméra peut être beaucoup plus proche. Il y a donc un rôle du visage qui peut devenir prépondérant au cinéma. Et parfois, à l'inverse du théâtre, si la caméra est souvent très proche, le corps est en grande partie perdu, sauf chez des metteurs en scène comme par exemple Pedro Almodovar qui savent très bien mettre en scène les corps et utilisent souvent des plans larges. On voit bien qu'il est difficile de rendre visible à la fois le corps et le visage, que ce soit au théâtre ou au cinéma. Il y a enfin une différence au niveau de l'entendre. Au théâtre, il y a une distance entre le comédien et le spectateur. Au cinéma, les micros qui captent la parole des acteurs ne sont jamais très loin.

Cette différence au niveau de l'entendre a néanmoins tendance à disparaître dans la mesure où de plus en plus de metteurs en scène de théâtre ont recours à des micros HF pour sonoriser leurs comédiens et leur permettre ainsi d'avoir un jeu plus naturel. C'est le cas, par exemple, du grand metteur en scène québécois Robert Lepage. Mais il n'y a pas non plus réellement de différence entre le cinéma et le théâtre en ce qui concerne le corps et le visage dans la mesure où un comédien joue toujours avec son corps et son visage, indépendamment de ce qui est le plus visible. La recherche d'une émotion vraie se fait de la même façon, qu'un acteur soit au théâtre ou au cinéma. Et pour qu'une émotion soit visible sur un visage, il faut qu'elle traverse également le corps, et vice versa. Donc, en réalité, le travail de l'acteur est le même pour le théâtre et le cinéma. Et si l'émotion est vraie, elle se verra à la fois dans le corps du comédien et sur son visage. Et elle s'entendra aussi dans la voix, que celle-là soit sonorisée par un micro ou non.

7 L'interprétation du metteur en scène

De même qu'il y a un paradoxe pour le comédien, il y a aussi un paradoxe pour le metteur en scène. En effet, le metteur en scène travaille à plusieurs niveaux [10]. L'interprétation du metteur en scène est liée, d'une part, au visuel, donc à la façon dont il va composer son espace visuellement, sa mise en scène à proprement parler. Il travaille aussi sur le comportement des comédiens, leurs interactions, leurs déplacements physiques, leurs actions, etc. Il travaille sur le mouvement, tout ce qui bouge sur scène. Et évidemment, il travaille aussi sur les émotions. Mais, comme l'acteur, il ne peut pas travailler directement sur les émotions. Donc le rôle du metteur en scène, au niveau des émotions, sera plutôt de guider le comédien en essayant de l'amener vers une justesse émotionnelle. Et en fonction des comédiens qu'il aura en face de lui, des techniques qu'ils maîtriseront, de leur capacité à être spontanés ou non, il aura plus ou moins de difficultés. Mais l'interprétation du metteur en scène se situe à un niveau plus global que celle de l'acteur, qui lui, est centré sur son personnage. Le comédien a seulement besoin de savoir ce que doit faire son personnage à un moment donné, quelle est l'intention de son personnage à ce moment-là, ce qu'il veut, ce qu'il dit, etc. Tandis que le metteur en scène essaie de composer un ensemble.

8 La relation entre acteur et spectateur

Il y a un troisième problème qui entre en jeu. Et nous reparlons de nouveau d'artialisation des émotions. Il s'agit de la différence – et cela me surprend toujours en tant que metteur en scène, et bien sûr en tant qu'acteur – la différence entre ce qui est exprimé et ce qui est perçu. Et cette différence, elle est énorme. Et c'est là qu'on a une hypothèse qui va dans le sens de l'artialisation des émotions. Le comédien, lui, fait un choix d'interprétation : à ce moment-là, son personnage est par exemple en colère après un autre personnage, et c'est donc une émotion de colère qu'il va essayer de mettre en place dans ce moment-là, de la façon la plus vraie et juste possible. Le metteur en scène a validé cette proposition du comédien, et a rajouté dans sa mise en scène des éléments qui vont dans ce sens. Et le spectateur regarde ce qui est proposé. Il va réagir sensiblement au fait que l'émotion est vraie, mais ce qu'il va percevoir et ressentir peut être quelque chose de totalement différent. Mais finalement, ça n'a pas d'importance. Ce qui est important, c'est que si ce qui existe sur scène est vrai, est juste, et va susciter une réaction chez le spectateur. Il y a à ce niveau-là des travaux très intéressants en cours ainsi que des travaux à initier sur le rôle des neurones miroirs [6] dans cette communication entre acteurs et spectateurs, qui est d'après moi essentiellement de l'ordre de l'empathie [8].

Les résultats des premières expériences semblent montrer que quand il y a des événements saillants sur scène – le mot saillant étant pour l'instant un peu vague mais désignant un moment où il se passe quelque chose d'assez fort émotionnellement sur scène – on observe généralement des réactions cognitives saillantes chez les spectateurs. On peut donc dire qu'il a déjà été établi une corrélation entre ce qui se passe sur scène et ce qui est perçu, mais ce n'est vraisemblablement pas la même expérience émotionnelle qui est vécue. Et c'est là où l'artialisation des émotions est de nouveau une hypothèse intéressante. En effet, le spectateur va, lui aussi, réagir avec des codes reposant sur ses expériences précédentes, et notamment sur la façon dont on exprime ses émotions dans une certaine culture, et dans un certain contexte, par exemple celui d'une salle de théâtre ou d'une salle de cinéma.

9 Conclusion

Pour conclure, je voudrais reformuler quelques propositions. Même si les émotions ne sont pas, d'après moi, un phénomène culturel, ce qui est culturel, selon moi, et qui peut donc faire l'objet d'une artialisation, c'est la façon de raconter des histoires, la façon de mettre en scène une histoire, et la façon d'interpréter et de jouer un personnage. Et en ce qui concerne les émotions, ce qui est culturel et peut donc être sujet à une artialisation est surtout ce qui va les déclencher et la façon de les exprimer.

Et je terminerai avec deux exemples empruntés au cinéma, et qui illustrent cette artialisation des émotions. Si vous prenez le cinéma mexicain, et je pense plus

particulièrement aux films d’Alejandro Iñárritu – ce que je vais dire n’engage que moi et je suis sûr que certains lecteurs pourront dire exactement le contraire, mais cela a surtout valeur d’exemple – il y a toujours un moment où je vais trouver cela beau et fascinant, mais il y a aussi toujours un moment où je vais me demander pourquoi le réalisateur insiste sur ce moment émotionnel, pourquoi il le fait durer, pourquoi il s’y attarde « trop longtemps ». Et je finis par me rendre compte, avec un peu de recul, qu’il y a précisément là quelque chose de culturel, parce qu’on le retrouve dans de nombreux films mexicains. On le retrouve par exemple dans le film *Mother and Child* de Rodrigo Garcia. Ce que je perçois maintenant et que je ne percevais pas avant – j’avais plutôt tendance à penser, avant, avec mon fort biais personnel, que ce n’était pas du très bon cinéma – c’est qu’il s’agit de quelque chose de culturel, et que c’est à moi de faire l’effort de comprendre ces différences culturelles si je veux pouvoir apprécier pleinement ces films. Pour les Mexicains, prendre le temps d’exprimer une émotion au cinéma et le faire avec une certaine intensité est quelque chose de culturel, et donc qui a été artialisé par une pratique cinématographique. Et le public mexicain serait frustré si ce n’était pas le cas. Mais du coup, un public non mexicain, c’est-à-dire ne partageant pas les mêmes codes culturels, et n’ayant donc pas été sujet à la même artialisation des émotions, risque de trouver les films mexicains trop émotionnels, ou les émotions trop délayées, ou les scènes émotionnelles trop longues.

Je voudrais prendre comme deuxième exemple le cinéma japonais. Cette fois-ci, on y exprime les émotions de façon très pudique et très subtile. Les scènes émotionnelles, quand elles existent, sont généralement très courtes. C’est un cinéma qui est généralement assez lent, qui prend le temps, mais pas au niveau de l’expression des émotions. C’est tout au moins ce que l’on pourrait dire de ce cinéma si nous sommes d’une culture différente, où les émotions ne s’expriment pas de la même façon et ne sont pas nécessairement déclenchées par les mêmes facteurs. Il s’agit bien là d’une artialisation des émotions, mais très différente de celle des Mexicains.

Est-ce que cette composante culturelle de l’émotion est véhiculée par l’art, auquel cas nous pourrions définitivement conclure qu’il y a bien une artialisation des émotions, tout du moins en ce qui concerne ce qui les déclenche et la façon de les exprimer ou de ne pas les exprimer, ou est-ce que cette différence culturelle se situe à un autre niveau ?

References

- [1] Rey Alain. *Le Grand Robert de la Langue Française, version électronique, deuxième édition*. Le Robert, 2014.
- [2] Stanislavski Constantin. *An Actor Prepares*. Theatre Arts, Inc., 1937.
- [3] Stanislavski Constantin. *My Life in Art*. Theatre Arts Books, 1948.
- [4] Stanislavski Constantin. *Building A Character*. Max Reinhardt Ltd., 1950.

- [5] Stanislavski Constantin. *Creating A Role*. Theatre Arts Books, 1961.
- [6] Rizzolatti Giacomo Sinigaglia Corrado. *Les Neurones Miroirs*. Odile Jacob, 2011.
- [7] Diderot Denis. *Paradoxe sur le comédien*. Garnier Flammarion, 1993.
- [8] Berthoz Alain Jorland Gérard. *L'Empathie*. Odile Jacob, 2004.
- [9] Clurman Harold. *The Fervent Years*. Da Capo Press, 1945.
- [10] Clurman Harold. *On Directing*. Fireside, 1972.
- [11] Strasberg Lee. *A Dream of Passion*. Plume, 1987.
- [12] Cohen Lola. *The Lee Strasberg Notes*. Routledge, 2010.
- [13] Cohen Lola. *The Method Acting Exercises Handbook*. Routledge, 2017.
- [14] Chekhov Michael. *On The Technique of Acting*. Harper, 1993.
- [15] Chekhov Michael. *To The Actor*. Routledge, 2002.
- [16] Hethmon H. Robert. *Strasberg at the Actors Studio*. Theatre Communications Group, 1965.
- [17] Solomon C. Robert. *Emotion*. Encyclopædia Britannica, 2014.
- [18] Cole Toby. *Actors on Acting*. Read Books, 2011.
- [19] Meyerhold Vsevolod. *Écrits sur le théâtre (Tomes 1 à 4)*. Éditions L'Âge d'Homme, 1990-1992.

La méthode de l'Actors Studio : Genèse, Pratique et Phénoménologie (sous presse)

Ivan Magrin-Chagnolleau. **La méthode de l'Actors Studio : Genèse, Pratique et Phénoménologie.** Dans *Être en jeu, être en scène*, édité par Laurent Berger et Arianna de Sanctis. Presses universitaires de la Méditerranée. Sous presse.

20 pages

La Méthode de l'Actors Studio : Genèse, Pratique et Phénoménologie

Ivan MAGRIN-CHAGNOLLEAU, CNRS
Aix Marseille Univ, CNRS, PRISM, Marseille, France

1. Introduction

1.1. Qui Suis-Je ?

Je voudrais commencer par donner quelques éléments biographiques puisqu'ils sont en lien étroit avec le sujet de cet article. Je me présente généralement comme artiste chercheur. Cela signifie que j'ai une activité artistique ainsi qu'une activité de recherche.

Sur le plan artistique, je travaille régulièrement comme acteur, metteur en scène et auteur pour le théâtre et le cinéma. J'écris aussi de la poésie, je fais de la photographie, de la musique, comme interprète et compositeur, et de la performance.

J'enseigne également le théâtre, le cinéma, la musique, la photographie, l'écriture, l'esthétique, et je coache aussi des artistes.

En recherche, je travaille dans le paradigme de la création recherche, c'est-à-dire que je crois à une recherche en art fondée sur une pratique, ce qui est pour moi la manière la plus riche de conduire ma recherche, dans la mesure où je suis aussi artiste. C'est pour cela que je m'intéresse particulièrement à la dimension phénoménologique de la création, c'est-à-dire à la création comme expérience vécue. Je considère mon travail artistique comme faisant partie intégrale de mon travail de recherche.

1.2. Mon Expertise sur la Méthode de l'Actors Studio

De 2004 à 2007, je me suis formé de manière intensive comme acteur et comme metteur en scène à l'Actors Studio à New York. J'ai donc appris la méthode de l'Actors Studio de professeurs qui avaient été eux-mêmes élèves de Lee Strasberg, le fondateur de la méthode^{1 2}.

Depuis, j'ai régulièrement joué au théâtre et au cinéma en utilisant cette méthode. J'ai fait plusieurs mises en scène de théâtre et réalisé plusieurs courts métrages de fiction, et ainsi dirigé de nombreuses actrices et de nombreux acteurs selon cette même méthode. Je l'ai également enseignée, sous forme de cours, d'ateliers et de masterclasses, et j'ai coaché nombre d'actrices et d'acteurs.

Enfin, étant également chercheur au CNRS, j'ai aussi beaucoup travaillé comme théoricien sur cette méthode^{3 4 5}. En particulier, j'ai accès depuis plusieurs années aux archives de Lee Strasberg au Lee Strasberg Institute, qui contiennent des cassettes VHS et de nombreux documents écrits. J'ai donc 15 ans d'expérience de pratique et de recherche sur cette méthode. Cela m'a conduit naturellement à écrire un livre en français sur cette méthode, livre qui est sur le point d'être publié⁶.

¹ Robert HETHMON, *Strasberg at The Actors Studio*, New York, Theater Communications Group, 1965, 446 p.

² Lee STRASBERG, *A Dream of Passion*, New York, Plume, 1987, 240 p.

³ Ivan MAGRIN-CHAGNOLLEAU, « L'Héritage de Lee Strasberg », *p-e-r-f-o-r-m-a-n-c-e.org*, vol. 2, 2015.

⁴ Ivan MAGRIN-CHAGNOLLEAU, « L'incorporation animale chez l'acteur », *p-e-r-f-o-r-m-a-n-c-e.org*, vol. 2, 2015.

⁵ Ivan MAGRIN-CHAGNOLLEAU, « L'acteur et les émotions au théâtre et au cinéma », *p-e-r-f-o-r-m-a-n-c-e.org*, vol. 3, 2016.

⁶ Ivan MAGRIN-CHAGNOLLEAU. *La Méthode de l'Actors Studio*, à paraître.

1.3. De quoi parle cet Article

En tant que spécialiste de cette méthode, je voudrais détailler ici quelle a été sa genèse, puis en quoi elle consiste, et comment, en tant que praticien, un acteur la fait sienne et continue à la faire évoluer. Je voudrais aussi discuter, en partant de mon expérience et de celle des nombreuses personnes que j'ai dirigées, coachées, enseignées, de la façon dont un acteur fabrique sa propre méthode par la pratique et la réflexion, à partir des enseignements qu'il a reçus, et de sa pratique. Je considérerai notamment la dimension phénoménologique de ce processus, c'est-à-dire l'élaboration d'une méthode personnelle de jeu à partir de l'expérience vécue. Je terminerai en abordant trois questions qui sont fréquemment posées au sujet de la méthode.

2. Genèse de la Méthode

2.1 Stanislavski à New York en 1923

Stanislavski a été une influence majeure sur la genèse de la méthode de l'Actors Studio. Il a été un grand acteur et metteur en scène russe de la fin du XIXème siècle et du début du XXème siècle (1863-1938). Il a développé sa propre méthode de jeu pour l'acteur et a publié plusieurs livres qui détaillent cette méthode^{7 8 9}. Il a également fondé le laboratoire d'acteurs du théâtre d'art de Moscou, qui a ensuite inspiré les laboratoires d'acteurs américains.

En 1923, il a effectué un séjour à New York et a donné plusieurs conférences et masterclasses. C'est à cette occasion que Lee Strasberg a rencontré Stanislavski et s'est enthousiasmé pour son approche du travail de l'acteur.

⁷ Constantin STANISLAVSKI, *An Actor Prepares*, London, Methuen Drama, 1937, 314 p.

⁸ Constantin STANISLAVSKI, *Building A Character*, London, Methuen Drama, 1950, 300 p.

⁹ Constantin STANISLAVSKI, *Creating A Role*, London, Methuen, 1988, 274 p.

À noter aussi une autobiographie rédigée par Stanislavski où il parle également de l'élaboration de son système et du fonctionnement du laboratoire d'acteurs au théâtre d'art de Moscou¹⁰.

2.2 Le Laboratoire d'Acteurs de Richard Boleslavsky et Maria

Ouspenskaya

Fort de cette première expérience avec la méthode de Stanislavski, Lee Strasberg a ensuite suivi les cours de Richard Boleslavsky et Maria Ouspenskaya, deux anciens élèves de Stanislavski qui ont décidé de s'installer à New York et de créer leur propre cours¹¹.

2.3 Deux Laboratoires d'Acteurs : le Group Theater et l'Actors Studio

À la suite de cette expérience du cours et de la façon de travailler de Richard Boleslavsky, Lee Strasberg, avec Harold Clurman, Cheryl Crawford, et d'autres jeunes acteurs, metteurs en scène et auteurs américains fondent un premier laboratoire d'acteurs, le Group Theater^{12 13}.

Interrompu par la Deuxième Guerre mondiale, le laboratoire renaît en 1947 sous l'impulsion d'Elia Kazan et Cheryl Crawford, sous le nom Actors Studio. Ils sollicitent Lee Strasberg dès 1949 pour assurer la direction artistique et prendre en charge le laboratoire.

C'est dans ces deux laboratoires que Lee Strasberg va pouvoir mener ses recherches sur le jeu de l'acteur, tester l'approche de Stanislavski, et mettre

¹⁰ Constantin STANISLAVSKI, *My Life in Art*, New York, Routledge, 1996, 582 p.

¹¹ Richard BOLESLAVSKY, *Acting: The First Six Lessons*, New York, Routledge, 1933, 140 p.

¹² Harold CLURMAN, *The Fervent Years*, New York, Da Capo Press, 1983, 330 p.

¹³ Harold CLURMAN, *The Collected Works of Harold Clurman*, New York, Applause, 2000, 1102 p.

au point ses propres exercices pour compléter l'approche de Stanislavski.

C'est aussi dans ces laboratoires que de grandes pièces du répertoire américain se sont créées, car plusieurs auteurs étaient membres de ces laboratoires et développaient leurs pièces dans cet environnement stimulant, souvent en demandant aux acteurs d'improviser des scènes qui n'étaient pas encore écrites.

2.4 Influences : Stanislavski, Vakhtangov, Meyerhold, Michael Chekhov

Stanislavski est une influence majeure dans la conception de la méthode de Lee Strasberg, mais ce n'est pas la seule.

« La méthode est un amalgame du travail de Stanislavski, Vakhtangov, Meyerhold, et le Group Theater. Observer et analyser leur travail est essentiel pour se former une compréhension pratique d'une approche théorique de notre travail, l'entraînement de l'acteur, et de son utilisation dans diverses formes de production. Les gens feraient une erreur s'ils considéraient une seule de ces sources comme pouvant représenter à elle seule toute l'essence de la méthode. »¹⁴

Ainsi, on voit que Lee Strasberg a puisé dans d'autres sources, par exemple Vakhtangov¹⁵ et Meyerhold¹⁶. Il attribue aussi le développement de sa méthode aux expérimentations qui ont déjà été faites au sein du Group Theater déjà mentionné auparavant.

¹⁴ Traduit par l'auteur à partir d'une citation de Lee Strasberg dans Lola COHEN, *The Lee Strasberg Notes*, New York, Routledge, 2010, 232 p.

¹⁵ Evgueni VAKHTANGOV, *Écrits sur le théâtre*, préface, traduction et notes d'Hélène Henry, postface de Béatrice Picon-Vallin, Lausanne, L'Âge d'Homme, 2000, 391 p.

¹⁶ Vsevolod MEYERHOLD, *Écrits sur le théâtre* : tomes 1 à 4, Lausanne, L'âge d'homme, 1980-1992.

Vakhtangov et Meyerhold ont tous deux été des disciples de Stanislavski, et les approches qu'ils ont développées et sur lesquelles ils ont écrit s'inspirent donc en grande partie du système de Stanislavski. C'est aussi le cas de Michael Chekhov, le neveu d'Anton Chekhov, qui a également été un disciple de Stanislavski, et dont le travail sur le geste psychologique est largement influencé par le travail de Stanislavski¹⁷ ¹⁸. Ces sources sont importantes pour comprendre la méthode de Lee Strasberg, mais aussi pour comprendre le système de Stanislavski, car ce dernier a publié ses livres plus d'une dizaine d'années avant la fin de sa vie, et toute l'évolution de son approche se trouve essentiellement dans les témoignages et écrits de ses disciples.

2.5 Cinquante Ans de Pratique et de Laboratoire

Mais la méthode de Lee Strasberg est aussi le résultat de son travail au sein du Group Theater, et surtout de l'Actors Studio. Il a été régulièrement acteur au théâtre et au cinéma pendant plus de 50 ans, a mis en scène de nombreuses pièces de théâtre, et a enseigné et coaché de nombreux acteurs. C'est ce travail de terrain et de praticien qui lui a permis de développer et de tester de nouveaux exercices, et de trouver ceux qui sont les plus pertinents.

2.6 Résoudre les Problèmes de l'Acteur à Tous les Niveaux

Ce qui intéressait avant tout Lee Strasberg, comme Stanislavski, c'était de résoudre les problèmes de l'acteur. Stanislavski avait déjà pas mal débroussaillé le terrain, mais il restait quelques problèmes qu'il n'avait pas réussi à résoudre, et auxquels Lee Strasberg s'est attaqué. Il fallait aussi adapter certains exercices à la culture américaine.

¹⁷ Michael CHEKHOV, *Être acteur*, Paris, Pygmalion, 1997, 241 p.

¹⁸ Michael CHEKHOV, *L'imagination créatrice de l'acteur*, Paris, Pygmalion, 1997, 249 p.

C'est donc tout ce travail que Lee Strasberg a accompli tout au long de sa vie et de sa carrière. En s'attendant notamment à la question des émotions, à laquelle Stanislavski s'était lui-même beaucoup intéressé. L'une des compétences de l'acteur est de pouvoir proposer quasiment sur commande une émotion vraie. Or, nous n'avons pas d'accès direct à nos émotions. Il est très difficile de décider de ressentir une émotion particulière sur commande de façon consistante et pour toutes les émotions. Il faut donc recourir à des stratagèmes pour cela. Et un stratagème que Stanislavski avait identifié et que Lee Strasberg a développé plus avant est de recourir non pas directement aux émotions mais à une mémoire sensorielle, c'est-à-dire de recréer par les sens une situation déjà vécue ou certains éléments de cette situation. Lorsque la situation vécue est porteuse d'une charge émotionnelle importante, qui peut alors en se la remémorant générer une autre émotion forte, on parle de mémoire affective.

Il y a d'autres exercices que Lee Strasberg a développés pour répondre à des besoins bien spécifiques de l'acteur. Il est important de comprendre que pour Lee Strasberg, un exercice n'a d'intérêt que s'il permet à l'acteur de résoudre un problème, et/ou si l'acteur peut l'utiliser dans la durée de manière consistante.

2.7 Une Extension de Stanislavski

On peut donc dire que la méthode de Lee Strasberg est en quelque sorte une extension de la méthode de Stanislavski, mais enrichie des apports d'autres théoriciens du jeu de l'acteur (Vakhtangov, Meyerhold, Michael Chekhov, ...), mais surtout des 50 ans de pratique, de recherche, de découvertes de Lee Strasberg lui-même.

2.8 Une Méthode Vivante

Pour conclure cette partie dédiée à la genèse de la méthode de l'Actors Studio, je voudrais insister sur le caractère vivant d'une méthode de jeu. La

méthode de l'Actors Studio a été développée par Lee Strasberg, mais elle n'a pas pour vocation de rester figée. Toute sa vie, Lee Strasberg n'a eu de cesse de faire évoluer sa méthode, de remplacer un exercice par une autre forme plus efficace, d'inventer un nouvel exercice pour répondre à un problème précis. Ce serait donc aller à l'encontre même de la méthode que de la considérer comme quelque chose de figé et de gravé dans le marbre.

Chaque comédien, chaque acteur s'approprie une méthode, un exercice, pour les faire siens, les adapter, les modifier, afin que ces exercices remplissent leur fonction première : aider l'acteur à devenir plus consistant, à développer son art du jeu.

3. Pratique de la Méthode

Je voudrais maintenant aborder la méthode au niveau de sa pratique. En tant que praticien de la méthode, qu'est-ce que cela veut dire pour moi ?

3.1 Un apprentissage de la méthode

Tout d'abord, il est nécessaire de prendre des cours pour apprendre cette méthode. Ce n'est pas une méthode que l'on peut apprendre dans les livres. Même si certains livres vont décrire la méthode, ses composantes, les exercices, il est important de les assimiler, et d'être guidé par cela par un professeur ou un coach qui connaît bien cette méthode, qui la pratique depuis longtemps, et qui en a assimilé les subtilités. Bien sûr, chaque professeur ou coach aura sa propre façon de transmettre, et peut-être d'ailleurs que ce qu'il transmettra sera plus ou moins la méthode de l'Actors Studio, mais il est important d'être bien au clair sur ce qui constitue cette méthode, sur les exercices qui en font partie, et sur ce qui représente les idiosyncrasies d'un professeur ou d'un coach.

3.2 Un Ensemble d'Exercices

La méthode de l'Actors Studio est constituée d'un ensemble d'exercices qui ont chacun leur raison d'être. Certains exercices ont été développés pour résoudre un problème bien précis, et n'ont pas pour vocation d'être pratiqués sur la durée (c'est le cas par exemple de l'exercice "Song and Dance"). D'autres exercices sont conçus comme des gammes de l'acteur, et ont donc pour vocation d'être répétés, développés, enrichis tout au long de la carrière de l'acteur. C'est le cas notamment de l'exercice de mémoire sensorielle, qui constitue en outre la base de plusieurs autres exercices (la mémoire affective, le moment privé, l'incorporation animale, etc.)^{19 20}.

3.3 Chaque Exercice a sa Raison d'Être

Je voudrais insister sur le fait que chaque exercice a sa raison d'être. Par exemple, l'exercice "Song and Dance" a pour but d'aider un jeune acteur à surmonter sa timidité lorsqu'il est seul en scène face à un public, et à comprendre également la notion de rythme sonore et corporel lorsqu'il est en scène. Pour un acteur qui n'a pas ces difficultés, cet exercice n'aurait sans doute pas grand intérêt. L'exercice de mémoire sensorielle a pour but d'entraîner la capacité de notre corps à se remémorer des expériences sensorielles (un goût particulier, une odeur particulière, un endroit particulier, etc.) et à les revivre au présent comme s'il y était de nouveau. Cette capacité est essentielle pour pouvoir ensuite réaliser d'autres exercices, et fait partie d'une série d'exercices à pratiquer le plus souvent possible pour progresser en tant qu'acteur.

¹⁹ Lola COHEN, *The Method Acting Exercises Handbook*, New York, Routledge, 2017, 232 p.

²⁰ Ivan MAGRIN-CHAGNOLLEAU. *La Méthode de l'Actors Studio*, à paraître.

3.4 Intégration des Exercices

Un autre point important et qui est souvent passé sous silence est l'intégration des exercices. Quand je parle d'intégration, je parle à la fois de progression (il est nécessaire d'avoir déjà une bonne maîtrise de la mémoire sensorielle avant de pouvoir faire un moment privé, une incorporation animale ou une mémoire affective), mais aussi d'adaptation des exercices au travail quotidien de l'acteur lorsqu'il doit interpréter un rôle.

La plupart des exercices n'ont pas pour vocation d'être simplement des exercices de pratique, mais aussi d'être utilisés concrètement lorsqu'il s'agit de jouer un personnage. Autrement dit, les exercices ne sont pas déconnectés de la pratique professionnelle de l'acteur. L'acteur utilise ces exercices pour faire son travail de construction d'un personnage. Par exemple, un exercice de mémoire sensorielle va être utilisé pour recréer des éléments sensoriels pertinents pour le personnage en construction et dans la situation où il se trouve.

3.5 Un Entraînement Continu

Enfin, cette méthode a été développée de sorte qu'elle puisse constituer un entraînement continu de l'acteur. À quelques exceptions près (quelques exercices qui ont été développés pour des situations bien particulières comme "Song and Dance"), la plupart des exercices ont été conçus pour alimenter un entraînement continu de l'acteur. Plus vous les pratiquerez, plus vous les maîtriserez, et plus vous les maîtriserez, plus ils vous seront utiles.

Le travail continu avec un exercice permet d'améliorer sa maîtrise de l'exercice, et permet de rajouter en complexité. Par exemple, si vous faites un exercice de mémoire sensorielle, vous commencerez peut-être par votre boisson du petit-déjeuner, puis vous irez vers des tâches un peu plus complexes comme la chaleur du soleil sur la peau, et vous finirez par devenir capable de recréer des expériences complexes, par exemple un petit-

déjeuner sur la place Saint-Marc à Venise. En outre, à chaque rôle, les besoins en mémoire sensorielle sont différents et la pratique de cet exercice s'enrichit donc constamment.

4. Phénoménologie de la Méthode

Je voudrais maintenant aborder la dimension phénoménologique de la pratique de la méthode de l'Actors Studio, c'est-à-dire comment cette pratique est vécue par l'acteur. Je m'appuie pour cela sur ma propre pratique et ma propre expérience, et sur de nombreux entretiens que j'ai passés avec des acteurs qui ont été formés à cette méthode et la pratiquent régulièrement.

4.1 Comment l'Acteur Intègre

Tout d'abord, il y a la question de l'intégration, dont j'ai déjà un peu parlé. Les cours intensifs à l'Actors Studio durent trois ans. C'est un enseignement dense, où tous les exercices de la méthode sont enseignés et pratiqués, et où on apprend également à utiliser les exercices dans un travail de personnage. Mais il faut bien sûr bien plus de temps que ça pour intégrer complètement les exercices et en faire une habitude de travail. Lee Strasberg considérait qu'il fallait 7 à 10 ans pour former un acteur. Je pense pour ma part que la formation de l'acteur est continue, et que plus on travaille sur une méthode, plus on la maîtrise et on la fait sienne.

Cette intégration passe par une répétition des exercices, par leur enrichissement, par leur utilisation dans des contextes différents, pour des rôles différents. Et à certains moments, certaines choses se sédimentent, c'est-à-dire qu'on les comprend à un niveau organique, à un niveau intérieur, à un niveau corporel, et cela devient une connaissance incarnée et non plus uniquement intellectuelle.

4.2 Comment l'Acteur Pratique

Cette intégration passe donc par une pratique régulière. Il y a bien sûr la pratique qui se fait dans le cadre des rôles qu'un acteur joue, au théâtre et au cinéma. Il y a aussi la pratique qui se fait dans le cadre des castings. Mais il est important de pratiquer aussi quand il n'y a ni l'un ni l'autre. C'est la pratique qui fait le maître. Ou pour le dire autrement, on est ce qu'on fait tous les jours. Il est important pour un acteur de pratiquer sa technique le plus souvent possible, et si possible tous les jours.

Certains exercices sont comme un muscle, et ce muscle a besoin d'être entraîné tous les jours pour se développer. C'est le cas notamment pour la mémoire sensorielle. Plus on l'utilise, plus elle est facile à utiliser. Et en partant d'exercices de mémoire sensorielle simples, on peut aller vers des exercices de plus en plus complexes.

Cette pratique n'est pas toujours évidente à faire par soi-même et il est possible de créer avec d'autres acteurs un laboratoire d'acteurs où pouvoir s'entraîner régulièrement et bénéficier éventuellement d'un regard extérieur sur son travail et donc d'un retour objectif, à condition bien sûr que ce regard soit suffisamment entraîné et bienveillant.

4.3 Comment l'Acteur Fait Évoluer

Il est aussi important de noter que chaque acteur s'approprie les outils d'une méthode, et les fait évoluer à sa façon, en développant sa propre méthode. Je dis toujours qu'il y a autant de méthodes de jeu que d'acteurs. Deux acteurs qui pratiquent la méthode de l'Actors Studio ne vont pas la pratiquer exactement de la même façon. Chacun a ses exercices préférés, ceux qui marchent bien. Pour ma part, je pratique systématiquement l'exercice de l'incorporation animale quand je travaille un personnage, et également avant d'entrer en scène ou d'aller sur un plateau de tournage. C'est un exercice qui marche bien pour moi, selon mes propres critères d'évaluation, et qui me

permet de sortir de ma tête très facilement. J'ai tellement pratiqué cet exercice, et j'ai guidé tellement d'acteurs dans cet exercice que j'ai sans doute adapté sa pratique à ma propre compréhension de l'exercice et à mes propres besoins, compréhension et besoins qui en outre évoluent dans le temps.

4.4 Comment l'Acteur Mélange

Enfin, il faut noter qu'un acteur, au cours de sa carrière, rencontre de nombreuses approches du jeu. Il le fait par le biais de sa formation initiale, mais aussi à travers les nombreux stages et ateliers, et enfin à travers toutes les rencontres avec des professeurs, des coachs, des metteurs en scène, des réalisateurs, d'autres acteurs, etc. À chaque fois qu'un acteur travaille avec quelqu'un, il apprend une nouvelle méthode de travail, et il va en garder ce qui l'aide, ce qui fonctionne pour lui, ce qui lui permet d'aller plus loin dans son travail. Et même s'il a un entraînement de base avec la méthode de l'Actors Studio, il se peut qu'il intègre dans cette méthode un élément exogène qui y trouvera très bien sa place.

4.6 Une Méthodologie pour Explorer la Phénoménologie de l'Acteur

J'ai le projet de développer une recherche autour de la phénoménologie^{21 22} de l'acteur, afin d'objectiver cette expérience vécue par l'acteur lorsqu'il joue ou pratique son art, et en ne me limitant pas nécessairement aux acteurs qui utilisent la méthode de l'Actors Studio. Pour cela, je souhaiterais utiliser un protocole expérimental qui repose sur des entretiens d'explicitation²³ afin de faire expliciter à des acteurs ce qu'ils font vraiment lorsqu'ils jouent, et pouvoir observer l'expérience vécue des acteurs à un niveau très fin. Cette

²¹ Edmund HUSSERL, *Idées directrices pour une phénoménologie*, Paris, Gallimard, 1985, 624 p.

²² Edmund HUSSERL, *L'idée de la phénoménologie*, Paris, PUF, 1992, 140 p.

²³ Pierre VERMERSCH, *L'entretien d'explicitation*, Issy-les-Moulineaux, ESF Éditeur, 2014, 208 p.

méthode a déjà été utilisée de nombreuses fois avec succès, notamment pour observer des choses aussi fines que l'intuition²⁴.

5. Questions récurrentes sur la Méthode

Je voudrais terminer cet article en répondant à quelques questions qui reviennent souvent au sujet de la méthode de l'Actors Studio.

5.1 Peut-on se faire mal en utilisant la Méthode ?

Cette question m'est souvent posée. Elle concerne, je pense, la peur qu'ont certaines personnes au sujet de l'exercice de la mémoire affective, ce qui fait que la méthode de l'Actors Studio est parfois conçue à contresens comme une méthode psychologisante.

Tout d'abord, je rappelle brièvement en quoi consiste l'exercice de mémoire affective : il s'agit de faire une mémoire sensorielle d'un événement du passé à forte charge émotionnelle, de manière à faire surgir dans le moment présent une émotion forte qui peut être similaire ou non.

À ce sujet, je voudrais faire plusieurs remarques. Tout d'abord, il n'est pas souvent nécessaire de recourir à cet exercice. On ne l'utilise que si on en a besoin et uniquement pour un moment fort émotionnellement dans une pièce ou un film. Ensuite, cet exercice n'est utilisé qu'avec des personnes qui ont des difficultés à générer certaines émotions fortes par de simples exercices de mémoire sensorielle. Quand ces émotions ont été libérées, il n'est parfois pas nécessaire de retourner à un exercice de mémoire affective. D'autre part, la condition pour pouvoir choisir un événement du passé pour cet exercice est précisément qu'il soit dans le passé, et suffisamment loin dans le passé. Lee Strasberg préconisait un minimum de 10 ans afin que l'inconscient ait eu le temps de faire son travail de filtrage.

²⁴ Claire PETITMENGIN, *L'expérience intuitive*, Paris, L'Harmattan, 2003, 386 p.

C'est aussi un exercice qui est pratiqué sous la supervision d'un enseignant, qui ne le fait qu'avec certains étudiants, et qui a appris à superviser cet exercice en toute sécurité psychique. Enfin, avec la pratique, un acteur peut apprendre à pratiquer cet exercice par lui-même, mais seulement une fois qu'il a été guidé plusieurs fois par un enseignant de la méthode.

Pour finir, j'ajoute que le travail sur les émotions, quelle que soit la façon dont l'acteur l'effectue, fait partie de ce qui est demandé à un acteur. Il est donc important qu'un acteur fasse le travail sur soi nécessaire afin d'être à l'aise avec ses émotions, et que ce travail ne constitue plus un obstacle à l'expression de son art. Si cela doit passer par un travail de thérapie, c'est la responsabilité de l'acteur d'entreprendre ce travail. Ce n'est nullement la responsabilité d'un enseignant de théâtre, et il est très important de savoir faire la différence et de fuir tout enseignant de théâtre qui s'ingérerait dans la sphère psychique au nom d'un travail d'acteur sur les émotions. C'est cela qui peut être dangereux pour un acteur, pas de travailler avec la méthode de l'Actors Studio.

5.2 Quelle est la Différence entre Théâtre et Cinéma ?

On me demande aussi régulièrement quelle est la différence entre le jeu au théâtre et le jeu au cinéma, et si la méthode de l'Actors Studio a été développée pour le théâtre ou pour le cinéma. Tout d'abord, en ce qui concerne le travail de l'acteur, il n'y a aucune différence entre théâtre et cinéma. On se prépare de la même manière pour un rôle au théâtre ou pour un rôle au cinéma. On utilise la même procédure, les mêmes exercices, la même approche. La seule différence est liée bien sûr au contexte, qui n'est pas le même. Au théâtre, il faut jouer aussi pour le spectateur du dernier rang et cela demande un engagement du corps et de la voix qui n'est pas le même. Et au cinéma, surtout pour les gros plans, le moindre mouvement prend énormément d'importance et il faut parfois réduire au maximum les expressions corporelles et vocales. Mais ce sont des éléments extérieurs qui vont s'intégrer naturellement dans le travail, à condition d'être guidé par un

metteur en scène compétent. Le travail de l'acteur sur le personnage et sur les scènes, lui, va être le même dans les deux situations.

Quant à la méthode de l'Actors Studio, elle a d'abord été développée pour le théâtre. Mais au moment où le cinéma s'est beaucoup développé, les acteurs formés à l'Actors Studio étaient également très performants au cinéma. C'est d'ailleurs ce qui en a fait sa réputation. Mais cela devrait être le cas de toute méthode de jeu performante, dans la mesure où le travail de personnage de l'acteur est le même, que ce soit au théâtre ou au cinéma.

5.3 Quelle est l'Influence de la Méthode en France ?

C'est la troisième question qui revient le plus souvent au sujet de la méthode de l'Actors Studio. Il y a en effet en France une réticence à cette méthode que je ne m'explique pas. J'ai quelques éléments de réponse mais ils ne sont jamais totalement satisfaisants. Il y a par exemple en France l'importance accordée traditionnellement au texte. Le texte théâtral en France est par tradition très littéraire, ce qui conduit à des jeux généralement plus théâtraux, et qui font moins de place à une interprétation plus réaliste telle que la méthode de l'Actors Studio le permet. C'est aussi le cas du théâtre de boulevard en France, qui est très théâtral et souvent très crié. Il n'y a pas actuellement de place dans ces théâtres pour une approche plus réaliste comme celle de l'Actors Studio. Je crois pour ma part que ces théâtres s'accommoderaient pourtant très bien d'un jeu réaliste, et que les œuvres jouées en sortiraient renforcées.

Il y a peut-être une autre raison qui est liée à la question 1. La méthode de l'Actors Studio fait peur à certaines personnes. Comme c'est une méthode qui repose sur l'idée d'un acteur qui a déverrouillé ses émotions, ce n'est pas une méthode qui convient à tout le monde, et notamment à des personnes qui ont précisément des difficultés émotionnelles. On a peut-être en France une culture qui ne favorise pas l'épanouissement émotionnel. Mais je

laisserai cette question entre les mains des sociologues et des psychologues.

Mais il y a heureusement en France toute une nouvelle génération d'acteurs qui travaillent avec la méthode de l'Actors Studio. Cela a commencé avec des acteurs qui ont des carrières internationales et des coachs de jeu souvent américains ou tout au moins anglo-saxons. En voyant la différence d'intensité au niveau des performances de ces acteurs, la jeune génération a compris qu'il y avait là quelque chose d'intéressant à explorer.

6. Conclusion

6.1 Pour Synthétiser

J'ai commencé, dans cet article, par me présenter succinctement en indiquant quelle était ma relation avec la méthode de l'Actors Studio. Puis j'ai parlé de sa genèse, en mentionnant l'influence importante de Stanislavski, mais aussi de Vakhtangov, Meyerhold et Michael Chekhov, et les deux incubateurs de la méthode qu'ont été le Group Theater entre les deux guerres mondiales, puis l'Actors Studio après la Seconde Guerre mondiale.

J'ai ensuite parlé de la pratique de la méthode, en mentionnant en particulier son apprentissage, la façon dont les exercices interagissent entre eux et dont un acteur les intègre, et l'importance d'un entraînement continu.

Puis j'ai abordé la question de la phénoménologie de la méthode, c'est-à-dire la méthode de l'Actors Studio du point de vue de l'expérience vécue par l'acteur, ce qui recouvre à la fois la façon dont chaque acteur intègre la pratique de la méthode à son travail d'acteur, mais aussi comment il la fait évoluer, et comment il la mélange avec d'autres pratiques. J'ai fini cette partie en proposant une méthodologie pour aborder la phénoménologie de l'acteur plus en profondeur.

J'ai enfin répondu à trois questions fréquentes qui sont posées au sujet de la méthode, concernant un éventuel danger à travailler avec certains exercices, la différence entre le jeu de théâtre et le jeu de cinéma, et l'influence de la méthode de l'Actors Studio en France.

6.2 Pour Aller Plus Loin avec la Méthode

Pour les personnes qui souhaiteraient aller plus loin avec la méthode de l'Actors Studio, je recommande d'abord de prendre des cours ou de faire du coaching avec un spécialiste de la méthode, quelqu'un qui non seulement la pratique depuis de nombreuses années, mais qui a aussi mené une réflexion sur la méthode afin de comprendre ce qui en fait partie et pourquoi.

Je renvoie aussi à la bibliographie de cet article qui contient quelques ouvrages qui ont été écrits sur la méthode de l'Actors Studio et qui présentent la philosophie de la méthode ainsi que les exercices, en particulier COHEN 2017 en anglais et MAGRIN-CHAGNOLLEAU 2020 en français.

6.3 Le Futur de la Recherche sur la Méthode

Il y a à mon avis trois façons d'aborder une recherche sur la méthode de l'Actors Studio aujourd'hui, et je travaille personnellement dans ces trois directions. La première approche est une approche pratique et consiste à pratiquer la méthode pour affiner sa compréhension de cette approche, de chaque exercice, de l'interaction entre les exercices, et de ce que chaque exercice apporte au travail de l'acteur.

La deuxième approche est une approche historique, qui consiste à aller creuser les archives de Lee Strasberg qui sont conservées en grande partie au Lee Strasberg Institute. Je visite régulièrement ces archives et y trouve régulièrement de petites pépites qui me font approfondir ma compréhension de la méthode et du projet de Lee Strasberg.

La troisième approche est une approche phénoménologique et consiste à interroger l'expérience vécue des acteurs qui pratiquent cette méthode, afin de mieux comprendre ce qui se passe pour eux lors de la pratique des différents exercices.

7. Bibliographie

- BOLES LAVSKY, Richard, *Acting: The First Six Lessons*, New York, Routledge, 1933, 140 p.
- CHEKHOV, Michael, *Être acteur*, Paris, Pygmalion, 1997, 241 p.
- CHEKHOV, Michael, *L'imagination créatrice de l'acteur*, Paris, Pygmalion, 1997, 249 p.
- CLURMAN, Harold, *The Fervent Years*, New York, Da Capo Press, 1983, 330 p.
- CLURMAN, Harold, *The Collected Works of Harold Clurman*, New York, Applause, 2000, 1102 p.
- COHEN, Lola, *The Lee Strasberg Notes*, New York, Routledge, 2010, 232 p.
- COHEN, Lola, *The Method Acting Exercises Handbook*, New York, Routledge, 2017, 232 p.
- HETHMON, Robert, *Strasberg at The Actors Studio*, New York, Theater Communications Group, 1965, 446 p.
- HUSSERL, Edmund, *Idées directrices pour une phénoménologie*, Paris, Gallimard, 1985, 624 p.
- HUSSERL, Edmund, *L'idée de la phénoménologie*, Paris, PUF, 1992, 140 p.
- MAGRIN-CHAGNOLLEAU, Ivan, « L'Héritage de Lee Strasberg », *p-e-r-f-o-r-m-a-n-c-e*, vol. 2, 2015.
- MAGRIN-CHAGNOLLEAU, Ivan, « L'incorporation animale chez l'acteur », *p-e-r-f-o-r-m-a-n-c-e*, vol. 2, 2015.
- MAGRIN-CHAGNOLLEAU, Ivan, « L'acteur et les émotions au théâtre et au cinéma », *p-e-r-f-o-r-m-a-n-c-e*, vol. 3, 2016.
- MAGRIN-CHAGNOLLEAU, Ivan, *La Méthode de l'Actors Studio*, à paraître.

- MEYERHOLD, Vsevolod, *Écrits sur le théâtre* : tomes 1 à 4, Lausanne, L'âge d'homme, 1980-1992.
- PETITMENGIN, Claire, *L'expérience intuitive*, Paris, L'Harmattan, 2003, 386 p.
- STANISLAVSKI, Constantin, *An Actor Prepares*, London, Methuen Drama, 1937, 314 p.
- STANISLAVSKI, Constantin, *Building A Character*, London, Methuen Drama, 1950, 300 p.
- STANISLAVSKI, Constantin, *Creating A Role*, London, Methuen, 1988, 274 p.
- STANISLAVSKI, Constantin, *My Life in Art*, New York, Routledge, 1996, 582 p.
- STRASBERG, Lee, *A Dream of Passion*, New York, Plume, 1987, 240 p.
- VAKHTANGOV, Evgueni, *Écrits sur le théâtre*, préface, traduction et notes d'Hélène Henry, postface de Béatrice Picon-Vallin, Lausanne, L'Âge d'Homme, 2000, 391 p.
- VERMERSCH, Pierre, *L'entretien d'explicitation*, Issy-les-Moulineaux, ESF Éditeur, 2014, 208 p.

Actors Studio (2021)

Ivan Magrin-Chagnolleau. **Actors Studio**. *Encyclopædia Universalis*, 2021.

5 pages

Actors Studio

L'Actors Studio a été fondé en 1947 par Elia Kazan, Cheryl Crawford, et Robert Lewis. Après les avoir rejoints en 1948, Lee Strasberg en est devenu le directeur artistique et la figure prédominante jusqu'à sa mort, en 1982.

Il s'agit au départ d'un laboratoire de comédiens, c'est-à-dire d'un lieu où les acteurs professionnels peuvent venir travailler sur des monologues ou des scènes sans aucune pression et dans le seul but de progresser dans leur jeu. Ces monologues ou scènes sont présentés devant des pairs – acteurs, metteurs en scène, dramaturges - en présence d'un modérateur, qui a longtemps été Lee Strasberg. Après la présentation, des retours ont lieu à propos du monologue ou de la scène en question, que les protagonistes peuvent éventuellement proposer à nouveau quelques sessions plus tard.

À l'origine, Stanislavski

L'Actors Studio s'inscrit dans une lignée d'ateliers d'acteurs qui se sont inspirés de celui du Théâtre d'art de Moscou fondé en 1898 par Vladimir Nemirovitch-Dantchenko et Constantin Stanislavski, metteur en scène de plusieurs pièces de Tchekhov. En 1923-24, Stanislavski fait une tournée théâtrale et donne une série de conférences aux États-Unis, notamment à New York. Plusieurs jeunes acteurs et metteurs en scène américains, parmi lesquels Lee Strasberg, Harold Clurman, Stella Adler, Elia Kazan, Sanford Meisner, Robert Lewis, ont pu assister aux représentations et/ou aux conférences, se sont enthousiasmés pour la manière de travailler le jeu de l'acteur qui était la sienne, et certains ont également suivi le cours de théâtre de Richard Boleslavski et Maria Ouspenskaya, des élèves de Stanislavski qui avaient créé leur cours de théâtre à New York. Ces jeunes américains ont ensuite fondé le premier atelier d'acteurs américains, le Group Theater. Par la suite, il allait devenir le lieu de développement de toute une génération de comédiens américains, mais aussi favoriser la création de grandes pièces du répertoire du théâtre contemporain américain –notamment celles de Tennessee Williams ou de Clifford Odets.

Si le Group Theater dut s'interrompre au cours de la Seconde Guerre mondiale, après le conflit la volonté de recréer un atelier d'acteurs était restée très forte. C'est ainsi que naquit l'Actors Studio, où allaient se former nombre d'acteurs américains et où se développèrent également quelques grandes pièces du répertoire - celles d'Arthur Miller notamment.

Lee Strasberg en a donc été l'une des figures-clés, à la fois comme comédien mais aussi comme metteur en scène et surtout comme coach d'acteurs. C'est ainsi qu'il a pu expérimenter de nombreux exercices et mettre petit à petit au point sa méthode.

La méthode de l'Actors Studio

Cette méthode s'inspire directement du travail de Stanislavski, mais aussi de celui d'autres metteurs en scène et théoriciens du théâtre russe tels que Meyerhold, Eugène Vakhtangov et Michael Tchekhov, neveu d'Anton Tchekhov. Il s'agit d'un ensemble d'exercices dont le but est de permettre aux comédiens de résoudre un certain nombre de problèmes qui se posent à eux au cours de l'exercice de leur profession. Ces exercices doivent permettre aux comédiens d'accomplir leur travail de manière consistante, et de produire des performances au théâtre et au cinéma qui soient empreintes de vérité émotionnelle, ce qui est l'un des traits caractéristiques de cette méthode.

La méthode en question est particulièrement adaptée au jeu réaliste, ce qui explique son succès non seulement au théâtre mais aussi au cinéma, où ce type d'interprétation est fortement représenté. En effet, le jeu réaliste au théâtre et au cinéma consiste à développer un personnage dont la construction repose principalement sur les émotions qu'il va ressentir au cours de la pièce ou du film. On parle souvent de vérité émotionnelle. À l'autre bout du spectre, on pourrait parler d'un jeu très théâtral et très formel, où la vérité émotionnelle n'est plus le but principal du travail du metteur en scène et du comédien. Ces deux styles et leurs intermédiaires existent à la fois au théâtre et au cinéma, mais le jeu réaliste reste prépondérant au cinéma, et particulièrement dans le cinéma hollywoodien .

Les exercices de cette méthode ont été proposés par Lee Strasberg dans un ordre et en fonction d'un apprentissage très précis, afin de développer petit à petit la capacité de l'acteur à proposer un jeu ancré dans le moment présent et dans la vérité émotionnelle de ce moment. Certains exercices consistent en des versions améliorées d'exercices proposés par Stanislavski et ses élèves, d'autres sont des créations à part entière de Lee Strasberg.

Mentionnons notamment l'exercice de la mémoire sensorielle, pierre angulaire de la méthode et qui sert de base à de nombreux autres exercices. Il s'agit de recréer par la mémoire une expérience de perception sensorielle. Ce peut être par exemple le moment où l'on boit sa boisson du matin. Lorsqu'on vit l'expérience elle-même, nous prenons cette boisson dans nos mains et nous la buvons. Lorsque nous accomplissons un exercice de mémoire sensorielle, nous revivons par la mémoire cette même expérience mais, cette fois, nous ne sommes plus dans le lieu de l'expérience et n'avons plus réellement la boisson dans les mains. En revanche, nous pouvons faire en sorte de ressentir exactement les mêmes choses.

Il est possible de recréer en nous toutes sortes d'expériences vécues (une douche, une odeur, un goût, un endroit, un objet, etc.), dans la mesure où nous gardons en mémoire l'ensemble de ces expériences, sous une forme ou sous une autre, que nous en soyons conscients ou non. C'est ce que Marcel Proust relate dans l'épisode de la petite madeleine, au début d'*À la recherche du temps perdu*. La grande différence entre l'expérience de Proust et celle d'un exercice de mémoire sensorielle étant que l'expérience évoquée par Proust est involontaire, alors que l'acteur choisit de revivre une expérience sensorielle du passé.

Il faut souligner également l'importance que Lee Strasberg accordait à la relaxation de l'acteur avant tout travail, et ce afin qu'il se rende le plus disponible possible. Aussi chaque séance commence-t-elle toujours par une relaxation, que ce soit en vue d'un entraînement, d'une répétition, d'une représentation ou d'un tournage.

À partir des exercices de base dont fait partie l'exercice de la mémoire sensorielle, on peut alors se tourner vers des exercices

plus complexes, comme celui de la mémoire affective, ou encore l'exercice de l'animal. Leur but est de permettre à un acteur d'utiliser des expériences de son passé afin de créer en lui des émotions qui seront alors interprétées par un public comme étant celles du personnage. Pour l'acteur, l'émotion en question est réelle mais est en lien avec un événement de son passé. Pour le public, elle est également réelle, mais est perçue comme étant celle du personnage dans la situation de la pièce ou du film.

L'ensemble des exercices de cette méthode permet donc à l'acteur de travailler à la fois à la construction de son personnage, mais aussi de créer une vérité émotionnelle qui peut venir habiter une performance entière, ou tout au moins certains moments-clés de cette performance. C'est sans doute là ce qui fait la force de cette méthode : pouvoir proposer une base consistante de travail pour les comédiens, qui puisse s'appliquer aussi bien au théâtre qu'au cinéma, et cela malgré les façons de travailler très différentes qui caractérisent ces deux domaines. Parmi les acteurs qui pratiquent ou ont pratiqué cette méthode, on peut mentionner Paul Newman, James Dean, Marlon Brando, Marilyn Monroe, Al Pacino, Dustin Hoffman, Robert De Niro, ou encore Bradley Cooper.

Cette méthode de jeu fait parfois l'objet de critiques virulentes, souvent de la part de personnes qui ne la pratiquent pas et qui réduisent cette approche de l'Actors Studio, parce que l'acteur travaille en partie avec ses émotions, à une méthode psychologisante. Il faut ajouter que ce travail se fait dans un cadre bien précis et avec des règles claires, en utilisant des exercices très cadrés, et que les dérapages qui pourraient avoir lieu seraient surtout liés aux personnes se déclarant expertes de cette méthode, alors qu'elles ne seraient pas suffisamment prêtes à prodiguer elles-mêmes des cours. Dans ce cas, c'est l'éthique personnelle du coach / professeur / metteur-en-scène / réalisateur qui devrait normalement servir de garde-fou.

Aujourd'hui, son efficacité est si reconnue qu'il existe des modules pour former les comédiens selon cette méthode dans presque toutes les écoles de théâtre et de cinéma aux États-Unis, par exemple à Yale ou à New York University, mais aussi à l'étranger.

Bibliographie

- H. Clurman, *The Fervent Years*, Da Capo Press, New York, 1983.
- H. Clurman, *The Collected Works of Harold Clurman*, édité par M. Loggia et G. Young, Applause Theatre Book, New York, 2000.
- L. Cohen, *The Lee Strasberg Notes*, Routledge, New York, 2010.
- L. Cohen, *The Method Acting Handbook*, Routledge, New York, 2016.
- R. Lewis, *Method or Madness?*, Samuel French, New York, 1958.
- I. Magrin-Chagnolleau, "L'héritage de Lee Strasberg", in *p-e-r-f-o-r-m-a-n-c-e.org*, vol. 2, 2015.
- L. Strasberg, *Le travail à l'Actors Studio*, Cours recueillis et présentés par R. H. Hethmon, NRF, Gallimard, Paris, 1986.
- L. Strasberg, *A Dream of Passion: the Development of The Method*, New American Library, New York, 1990.

3.8 Esthétique du théâtre et du cinéma

The Use of Color in Theater and Film (2013)

Ivan Magrin-Chagnolleau. **The Use of Color in Theater and Film.** *Proceedings of AIC2013.*

4 pages

THE USE OF COLOR IN THEATER AND FILM

Ivan MAGRIN-CHAGNOLLEAU, artist-researcher

Sorbonne Paris 1 & C.N.R.S.

Head of the Aesthetics of Performing and Spectacular Arts Department

ABSTRACT

In this paper, I propose a brief overview on the use of color in theatre and film and point to some references that I consider as seminal on this topic. How is color used in theatre and Film? How is it different and why? What is used on stage and on screen in terms of colors (props, costumes, make ups, sets, lights)? Are there any aesthetic rules and conventions, and how can they be questioned? I relate these rules and conventions to other color theories. In particular, what is the impact of certain colors in terms of meaning and in terms of emotions? I discuss the following questions: how do color choices participate in the aesthetic choices of a stage or screen director? Are these choices conscious or unconscious? How do color impact on the viewer? Is it cultural or universal? I share some of my experience as a stage and screen director in terms of using colors. Finally, I explain how I articulate my own artistic practice with theory, and how my status as an artist-researcher allows me to do so.

1. INTRODUCTION

Colors convey emotions. And they usually convey emotions in a very subtle way, but also in a very uncontrolled manner. Colors are undoubtedly an important aesthetic component in any artistic creation. This component can be consciously used or not, depending on the artist and on the process. But it is always there. Let's have a look at the way colors are used in theatre and film.

In theatre, the audience can see a show because it is lit. A great amount of thinking is usually put into designing the lighting. And it is very common in theatre to put color filters on lights in order to create particular moods. The first aim of lighting is to make things on stage visible. But the second aim, as important as the first, is to create an appropriate mood for each moment of the story. A lot of books have been written about how to light a show (Palmer, 1993) (Setlow & Essig, 2012) (Gillette, 2013) (Reid, 2001) (Shelley, 2009). The work of designing the light for a show is usually done jointly by the director and the light designer. But light is not the only choice connected to colors in the elaboration of a show. Every choice relative to the set, props, costumes also has an impact on the general color perception. And in the end, it is the combination of the two (the elements on stage and the lighting) that provides the audience with an emotional experience, partly impacted by the color perception.

In film, the work on colors is quite different. The lighting is usually much simpler in terms of colors (Alekan, 1998) (Storaro, 2011) (Frost, 2009) (Brown, 2007) (Brown, 2011). Again, the overall color perception is impacted by the choice of sets, props, costumes. But when a director or a cinematographer wants to add a color tone to an image, they put filters on the camera rather than on the lighting itself (there are of course exceptions). And the director also has the option to correct colors with a colorist at the end of the post-production chain. So a lot of choices related to colors are made at the very end of the process, even though it is usually much easier if these considerations about colors have been discussed and planned before the shooting.

To summarize, both theatre and film directors use sets, costumes, and props to impact the colors of their work. But in theatre, the lighting is the important phase, being the one that reveals what is seen. Whereas, in film, the perceived colors are the results of a combination of filters put on the camera and fine-tuning made by the colorist at the end of the post-prod chain.

In part 2, I describe in more details what are the vehicles of color on stage and in film. In part 3, I investigate on the aesthetic rules and choices that can guide the use of color, whether these choices are conscious or not. In part 4, I draw examples from my own experience as a stage and screen director. And finally, part 5 concludes this introductory survey on the use of color in theatre and film.

2. COLOR VEHICLES

Every element that is part of the design of a stage or a set (furniture, costumes, props, make-up, carpets, wallpaper, wall paints, etc.) impacts the overall color perception. A director can choose for instance to have every element in pastel tones or in bright colors, and that will obviously create a very different feeling. Every element has an impact on the overall composition. A lot of inspiration can be drawn in this matter from the paintings of the great masters. Which paintings do a director like the most? How are they composed? What is the palette of their colors? Is it harmonious or not? What mood does it create? Paintings are obviously a great source of inspiration for stage and screen directors. Not just for colors, but also for composition, and for lighting.

Color and light go closely together. Colors are the components of light. So anything that generates light or reflects it or let it go through has an impact on what is seen and on the colors of the scene. Take a moment to look in front of you and notice every color you can see. Which ones catch your eyes first? What is the relationship between color and light? Can you list every element that is defined, among other things, by its color (a book, a wall, a framed picture, a lamp, a floor, a carpet, etc.)? Now, how would you describe your overall feeling in front of this scene? What has the biggest impact on the definition of this feeling? This is the kind of questions stage and film directors ask themselves when preparing a show or a film. As the director, you are the person in charge of the overall aesthetics, and therefore in charge of each decision that contributes to the aesthetics. Are there any rules that guide all the decisions stage or screen directors have to make?

3. AESTHETICS CONSIDERATIONS

The way directors approach colors is very different from one director to another. There are some color theories that can be used as guidelines (Hyman, 2006) (Bleicher, 2011) (Mollica, 2012). Some directors follow them, or at least make their choices being aware of them, and some directors have their own personal approach to color, which may or may not be different from what these color theories propose. The way directors choose colors is based, in my opinion, on several things, and the influence of these things can vary greatly along their careers: what they know about colors, their culture, their experience in using colors, the questions they ask themselves about colors, the possibility of them having a dialog about colors with their lighting designer or their cinematographer, the show or the film they are directing and its needs in terms of color, etc. So there may exist some aesthetic rules about the use of colors, and they may be used by some directors as guidelines, and broken by other directors, or even followed unconsciously by some directors. And some people might argue that the knowledge of these rules can only improve the

artistic abilities of a director. But as with any aesthetic rules and guidelines, this is when directors can transcend them that they truly become artists.

Another interesting question is the way certain colors impact meaning and emotions, and these considerations are also very important in the way directors work with colors. Colors are components of light, and therefore, each color has a different energy. So each color has a different feeling, acts differently upon us, and there are various ways to draw a typology of colors, and these ways have been different across the ages (Finlay, 2003) (Ball, 2003). There are very few books that manage to do a great synthesis on this subject, but two of them are particularly amazing as they draw from history, from various cultures, but also from spiritual traditions (Berton, 2002) (Simpson, 1999).

Finally, there is the question of culture and its impact on the interpretation of colors in terms of meaning and emotions (Gage, 1999). If color is intrinsically linked to a culture, then will an audience in Japan or in Camerouns react the same way to a show or a film? Probably not. And the challenge then becomes to find ways to make a show or a film as universal as possible, in terms of its storytelling and emotional meaning, in spite of the fact that some of its aesthetic components will remain quite identifiable in terms of its origin, and may seem a bit strange or odd to some cultures. But this is in my opinion one of the great advantages of cultural differences: they provide learning experiences about other cultures, and therefore about our own.

4. EXAMPLES

Here are some brief illustrations to the various points I discussed so far, chosen from my experience as a stage and screen director. The first example is my stage direction of the play *The Lesson*, written by Eugène Ionesco, which I directed in 2010. I started by reading the play over and over again, letting images come into my mind, and usually these images came with lighting and colors. At some point in the process, I also played around with the color cards of William Berton (Berton, 2002). I thought of certain aspects that I wanted to emphasize in the play and I let my mind free-associate colors with these aspects. I worked on this show with Olivier Horn, who is an amazing light designer. He captured the essence of what I wanted, and translated it into lighting in a very subtle way. I also started thinking about the set, props, and costumes very early in the process. I thought very early about a small round table covered with a bright red cloth in center stage. And then I wanted the professor to have a dark outfit and the student a bright one. They ended up dark brown and light yellow. I work very intuitively on colors, and the final decisions are always the end result of a fruitful collaboration with the light designer, the costume designer, and the make up artist (I did most of the set and props design on this show).

The second example is the making of the short film *King Cake (La Galette des Rois)*, written by Vinciane Mokry, which I directed in 2012. Very early in the process, I wanted a golden yellow and a royal blue to be the two guiding colors in my creative process. I made choices of props and costumes as well as lighting based on these two dominant colors. And these two colors also served as guidelines for the colorist. At the end, the result is very subtle, and is the consequence of a permanent dialog with the cinematographer, the production designer, and the colorist. But the original intention is there and has an impact on the general tone and mood of the film.

Being an artist-researcher, that is, sharing my time between my art practice and my research, I can articulate the two together. My practice informs my research, and is the basis for it. And my research opens new perspectives to me as an artist and directs some of my artistic choices. For

instance, this investigation on the use of color in theatre and film is based on my practice, but at the same time allows me to go one step further in my thinking, and therefore will have an influence on how I do things next time. There is a permanent interaction and dialog between practice and research.

5. CONCLUSIONS

Colors are perceived in film and theatre as the result of several components: set, props, costumes, make up, lighting, and each of these components has an impact on the overall color perception. Colors have a direct impact on the emotions of the viewer and are usually chosen for that purpose. There are some aesthetic rules that can be used as guidelines and some cultural differences that can be taken into consideration. But in the end, much of the creative process happens in the subconscious mind and is the result of several factors: what directors know about colors, their culture, their experience in using colors, the questions they ask themselves about colors, the possibility of them having a dialog about colors with their lighting designer or their cinematographer, the show or the film they are directing and its needs in terms of color.

REFERENCES

- Alekan, H. 1998. *Des lumières et des ombres*. Editions du Collectionneur.
- Ball, P. 2003. *Bright Earth: Art and the Invention of Color*. University Of Chicago Press.
- Berton, W. 2002. *Couleur énergie : Le Langage des couleurs de la vie*. Colorscope.
- Bleicher, S. 2011. *Contemporary Color: Theory & Use*. Delmar Cengage.
- Brown, B. 2011. *Cinematography Theory and Practice*. Focal Press.
- Brown, B. 2007. *Motion Picture and Video Lighting*. Focal Press.
- Finlay, V. 2003. *Color: A Natural History of the Palette*. Random House.
- Frost, J. B. 2009. *Cinematography for Directors: A Guide for Creative Collaboration*. Michael Wiese Productions.
- Gage, J. 1999. *Color and Culture: Practice and Meaning from Antiquity to Abstraction*. University of California Press.
- Gillette, J. M. 2013. *Designing With Light*. McGraw Hill.
- Hyman, J. 2006. *The Objective Eye: Color, Form, And Reality in the Theory of Art*. University of Chicago Press.
- Mollica, P. 2012. *Color Theory*. Walter Foster.
- Palmer, R. H. 1993. *The Lighting Art: The Aesthetics of Stage Lighting Design*. Pearson.
- Setlow, J., and Essig, L. 2012. *Lighting and the Design Idea*. Cengage.
- Simpson, L. 1999. *Book of Chakra Healing*. Sterling.
- Shelley, S. L. 2009. *A Practical Guide to Stage Lighting*. Focal Press.
- Storaro, V. 2011. *Writing With Light: Colors, and the Elements*. Mondadori Electa.
- Reid, F. 2001. *The Stage Lighting Handbook*. Theatre Arts.

*Address: Ivan MAGRIN-CHAGNOLLEAU, Sorbonne Paris 1 & C.N.R.S.
47 rue des Bergers, 75015 Paris, France
E-mail: ivan.magrין-chagnolleau@cnr.fr*

The Relationship between Photo Retouching and Color Grading (2022)

Ivan Magrin-Chagnolleau. **The Relationship between Photo Retouching and Color Grading.** *Color Culture and Science Journal*, 2022, 14(1), pp. 79-86..

8 pages

The Relationship between Photo Retouching and Color Grading

Ivan Magrin-Chagnolleau ^{1 2}

¹ Aix-Marseille Univ, CNRS, PRISM Laboratory, Marseille, France

² Chapman University, Orange, California

Corresponding author: Ivan Magrin-Chagnolleau (ivan.magrin-chagnolleau@cnrs.fr)

ABSTRACT

This paper deals with photo retouching and color grading. It proposes a simplified workflow for both of them. It also points the commonalities and the differences, and further explore the relationship between the two, including a phenomenological point of view as well as an aesthetic point of view. It also discusses the implications of that relationship in pedagogy.

KEYWORDS

Photography Retouching Workflow, Color Grading Workflow, Color Aesthetics, Phenomenology of the Creative Process, Pedagogy.

1. Introduction

The aim of this paper is to explore the relationship between photo retouching and color grading. The idea of this paper came to life when I started to reflect on my practice as both a photo retoucher and a color grader, and trying to understand if I approached both in the same way, or if I had different modus operandi for each of them.

I am particularly interested in two aspects: (i) why do we make the decisions we make in a creative process, here in retouching a photo or in grading a film, and (ii) what are the aesthetic implications of this relationship between photo retouching and color grading.

I am speaking from the point of view of a scholar dedicated to photography and film philosophy and aesthetics, but also from the point of view of a long-term practitioner as both a photo retoucher and a color grader. The work presented here is therefore the result of a constant dialog between practice and theory.

There are many publications about the aesthetics of photography (see for instance Costello, 2018) and the aesthetics of film (my favorites are Alekan, 2001, and Storaro, 2011; see also for instance Jarvie, 2015). There are a few publications about photo retouching (most of them are mostly guides about how to use the softwares) and color grading (see for instance the excellent Haine, 2019), but they are mostly technical. And this is mostly through online tutorials that ones learns to use photo retouching or color grading softwares. There are very few publications specifically about the aesthetics of photo retouching and the aesthetics of color grading, and these topics are sometimes vaguely touched upon in books about photography aesthetics and film aesthetics, or in books about photo retouching and color grading. But in my knowledge, no one has ever explored the relationship between the two. The main aim of this paper is to touch upon this relationship.

As far as color is concerned, there are a few excellent books about the theory of color (see for instance Mollica, 2012, or Bleicher, 2012).

The outline of the paper is as follows: the first part deals with photo retouching, and tries to answer the question of why we do it in the first place. Then, a simplified workflow for photo retouching is proposed, and a discussion about aesthetic considerations follows.

A second part then deals with color grading, and proposes a simplified workflow for it, as well as a discussion about aesthetic considerations concerning it.

A third part compares the two approaches, points the commonalities as well as the differences, and also shows how the two approaches can feed each other.

And finally a conclusion wraps-up the paper.

2. Photography Retouching and Color

Most professional photographers nowadays use softwares from the adobe suite. One particular program, Lightroom, has been developed with the photographer workflow in mind, and offers various ways to save a lot of time in the process. Even though the topic I discuss here is not related to the use of a particular software, part of how I do things might be.

There are also numerous plug-ins that exist today, whose job is to allow a professional photographer to save time. I am very fond of Color Efex Pro (and Silver Efex Pro for Black&White photography) [1].

2.1. *Why do we retouch a photograph*

Let us first go back to the time before the digital revolution. You did not have that many options to control the final look of an image. You had to carefully think at the moment of taking it (see for instance the visualization technique developed by Ansel Adams in Adams, 1995). And then, the rest of it was a matter of how you developed your film, and then how you printed the chosen picture. This was all chemical, and particularly for color photography, you would generally rely on a strong collaboration with a lab. Occasionally, some photographers would do their own developing and printing jobs, particularly in Black&White photography, but that was not the norm.

Today, retouching a photo is something a photographer can do on their own much more easily. All they need is to open the raw digital file inside a software, and then push buttons and see the corresponding result. This process is non-destructive. It has become entirely part of the creative process. Because there are so many possibilities now, you end up playing with the possibilities and expanding your horizons.

So why do we retouch a photograph? We do it to improve the look of the photograph, in terms of contrast, colors, framing, final look, etc. But we also do it to try new things, to play with the possibilities of the tools we use. Photo retouching is therefore an investigative process.

2.2. *What is the color component of photo retouching*

In this whole process, not everything is directly related to color, though most of it has an impact on how color is perceived. For instance, changing the exposure of a picture impacts the perception of color saturation: making a picture darker tends to make us perceive it as more saturated.

The first steps of a retouching workflow often deals with brightness and contrast. And then we start playing with color. We can do several things at this point. We can check the setting of the white balance, and if we are using raw files, we can choose a different white balance even after the picture has been taken. We can also tell the software where our real whites are and let the software change the settings accordingly. This whole process, called “removing a color cast”, consists in compensating for some technical limitations of the camera we used, or the circumstances of the picture.

There is then the possibility of wanting our picture to look a bit different colorwise. For instance, if you took the picture of a sunset, and you noticed that your camera tended to bring the whole scene a bit more towards yellow than pink or orange, you can then add a tint of pink or orange to make the picture look closer to what you remember or what you intended.

2.3. A possible workflow for photography retouching

Let me now propose a simplified workflow for photo retouching. First of all, as mentioned earlier, I mostly use Lightroom to do my work on pictures. With Lightroom, you can import images, sort them, define all sorts of metadata, retouch them, and then export them. I will deal here only with the part of the workflow that deals with retouching pictures.

I also want to point out a very important principle in art making. There are no immutable laws. Even when you have a workflow in place, occasionally you will choose to deviate from it. Besides, this workflow is also meant to be rethought, and hopefully improved with time.

Photo Retouching Simplified Workflow:

- **test the auto setting:** it consists in using a button that provides automatic settings for the given picture, based on some values that the software extracts from the image. Sometimes, the resulting retouching is close enough and I continue from there, and sometimes the original picture is closer to what I want and I then start with the original picture.
- **adjust the exposure:** it is possible at this stage to modify the white balance inside the software, but I rarely do it. It is because I have usually set the white balance up before taking the picture, and therefore rarely need to change it later. But this is a possibility that is worth mentioning. Personally, I often tend to lower the exposure a bit. I have found that, most of the time, the auto button tends to set an exposure that is a bit too bright for my taste. But occasionally, I will do the opposite, increase the exposure slightly. There are no rules. It is mostly a matter of experience and taste.

- **adjust the contrast:** I tend to prefer more contrasted pictures, because that makes them more pictorial, more graphical, which is usually closer to my personal taste in term of images. For those first steps, auto, exposure, and contrast, I usually use my eyes only. Then, I rely mostly on the histogram, which is a diagram showing the repartition of the values of all the pixels of an image from black to white, and a bit on my eyes.
- **adjust the highlights and the shadows:** the idea behind those two steps is to occupy the whole range of brightness from black to white, in order to have the best possible distribution of pixels. This is not always the best option, but more often than not, it brings the image closer to a pleasing image.
- **adjust the texture, the clarity, the dehazing, and the sharpening:** these four sliders have to do with the granularity of the picture, and how it looks in terms of details. I barely change them, but often just a bit, to increase the crispiness of the picture.
- **adjust the vibrancy and the saturation:** we finally get to the color aspect of retouching pictures. I usually utilize two sliders to do it, the vibrancy slider and the saturation slider. In most cases, I increase both a bit and that gives me a richness in terms of color that I like. Occasionally, I will desaturate slightly, but only for very particular photos.
- **use the color efex pro plug-in:** I will then open the picture in color efex pro, and will play with some of the possibilities of this plug-in to enhance the picture. It is important to notice that part of the settings that I use prior to opening the color efex pro plug-in are decided with the use of the plug-in in mind. From experience, I know how to prepare an image before running the plug-in. If I was not planning on using the plug-in at all, I would make different choices. Once in the plug-in, I have many options. Some of them have to do with some creative techniques effecting contrast, other effecting color, and so on. Some of the looks are also mimicking some vintage looks.

2.4. Aesthetics considerations for photography retouching

I often wonder why I make the choices I make, or why photographers in general make the choices they make, in particular regarding color. One way to answer that question is by saying that a photographer remembers the scene he took a picture of, and consciously or subconsciously tries to match the picture and the memory of that scene.

A perfect example of that is when taking a picture of a sunset. The captured picture will always deviate from what

was really in front of our eyes. And the temptation would be to correct the picture later, to make it as close as possible to what we remember.

But that would be forgetting that a picture is only a representation of reality. The truth is much more complex. When I retouch a picture colorwise, I try find the color treatment that would serve that particular picture best, that would enhance the power of evocation of that picture.

Concretely, there is a part of habit in the process. We tend to do things the way we are used to doing them. But how did we come up with this way of doing? This process is usually incremental. We first learn to do it through practice, on the one hand, and from learning, on the other hand. We can learn from someone, from a book, from an online course or tutorial. All these activities build up our artistic sensitivity, which is also ever evolving.

Then, once we have a basic workflow, we keep experimenting, and we keep learning from courses, books, tutorials. That process will constantly open perspectives,

suggest other ways to do it, sometimes better ways, sometimes not. But as we gain knowledge about the process, be it in practice or by learning, we expand our horizons and the possibilities.

At the same time, we develop a way to adapt to a particular photo our own way of doing things, our own style. Maybe we try to match aesthetics we love, from photographers we admire. Maybe we are also trying to match other media, like painting or film.

In the end, there is something subtle guiding us in this quest for beauty, for the most aesthetically satisfying picture. It has to do with what motivated us to take the picture in the first place, what we want to express through that picture, something that has to do with some universal truth, but seen through the prism of our own experience and our own peculiarities.

Figure 1 is an example of a photograph before retouching (left) and after retouching (right).



Fig. 1. The image on the left is the raw picture, as taken with the camera. The image on the right is a possible version of the picture after retouching it. This picture is part of the series "Calanques".

3. Color Grading and Color

Let us move on to color grading. I got involved in it quite naturally. Because I had done a lot of photo retouching. Because I was making films. And because it made sense to me to do my own color grading on my film projects. And then I started doing it for other people.

3.1. Why do we grade films

There is a slightly different reason why we grade films, compared to photo retouching. In color grading, there is an important motivation which is to save shots. There are situations when the images recorded by the cinematographer present some technical problems (low

light, important color cast, etc.). One of the goals of color grading is then to save the shot, that is, to bring it to a place where it can be used in the film without disrupting the flow of the storytelling.

The rest of the job pretty much resembles that of a photo retoucher, with one big difference: we are now working on moving images, that is, we are not just retouching one still image, or a series of a few images, but 24 or 25 or 30 images per seconds, from the beginning of a clip to the end of a clip, clip after clip, from the beginning of the film to the end of the film. This work is comparatively huge in terms of involvement and time, and there are also considerations about the unity of a shot, of a sequence, of a film.

3.2. What is the color component of color grading

As with photo retouching, the whole process of color grading does not have to do with color, although it can be argued that everything a color grader does impact color perception. But there are mostly two operations that directly have to do with color.

The first one consists in removing a color cast, that is, identifying what is supposed to be white and trying to bring those particular pixels back to white. It can also be done by using some particular scopes, called parade, which display the repartition of luminosity for each of the three sensors (red, green blue), and often allows to detect discrepancies between them, thus indicating a possible color cast.

The second one concerns finding the look of a clip, a sequence, or an entire film. This often includes considerations on whether the mood of that particular clip, sequence, or film should be colder or warmer.

3.3. A possible workflow for color grading

Let me now propose a simplified workflow for color grading. The software I am using is Davinci Resolve, which has imposed itself as the professional tool for color grading. It was first developed only as a color grading tool. Now, Davinci Resolve is an entire post-production suite. The workflow I propose here is not specific to Davinci Resolve though.

Also, as I stated previously for photo retouching, a workflow is a guideline. But there is no law preventing to occasionally deviate from that workflow. And a workflow is also here to be continuously improved.

Color Grading Simplified Workflow:

- **brightness / contrast / highlights / shadows:** for color grading, I prefer using curves, which is a tool I rarely use for photo retouching. The reason why I like curves in color grading is that it allows me to do several operations

all at once, while using at the same time my eyes and the scopes. With some experience, it takes only a few seconds to correct the brightness / contrast / highlights / shadows of a clip all at once. To use the curves tool, I usually set up a point close to white and then move it until I get the expected result. And then I set a point close to black and do the same thing.

- **removing a color cast:** I use again the curves tool to do it, but this time, I use it color by color (red, green, blue). I will mostly use the scopes as a guideline, but I also keep an eye on the image itself.
- **secondary correction:** the first two steps are called primary correction, meaning they are applied to the whole image. A secondary correction is a correction that is only applied to a part of the image, for instance a face. It requires the use of a mask, and then tracking of the mask across the whole clip. This is used only when necessary as it is very time consuming to set it up and adjust it. This is due to the fact that you often have to check the mask frame by frame, and adjust it when it deviates from where it is supposed to be set up, which often happens when the object that is tracked, for instance a face, is moving fast in the shot. It is often used to lighten or darken an area of an image, to make it more or less present. It is also used to adjust a color, for instance a skin color.
- **adjusting between clips and sequences:** this step consists in adjusting consecutive clips and/or sequences in order to smoothen the experience of the viewer. Sometimes, the opposite effect is sought after: increase the transition effect between one clip and another, or between one sequence and another.
- **working on the look of a clip / sequence / film:** the last step consists in working on the look of a clip, sequence, or film. This usually requires applying some additional effects on the images, often through the use of LUTs (Lookup Tables) which are preset tables allowing to apply entire looks and/or filters to clips. Color is often an important component of the look of a film. This is mostly due to the correlation between color and mood.

The last two steps can be inverted. It is sometimes possible to start working on looks before doing all the clips or sequences matching.

3.4. Aesthetics considerations for color grading

Again, the question that comes to mind is: why do we, as color graders, make the choices we make? The answer here is not much different from the one I gave for photo retouching. Habit is part of it, improving a process too, and matching a particular aesthetic we like.

But there is a big difference here. Photography is usually a one-person medium. Film is a collaborative medium and the color grader is not the only person making decisions on the color grading process. The director is also involved, quite often too the director of photography, and sometimes other people too (some producers, some studio executives). It makes the process more complex, and

requires the color grader to be able to argue their choices, and also to be flexible enough to integrate the vision of other people, even when that vision contradicts their own.

Figure 2 is an example of a frame before color grading (left) and after color grading (right).



Fig. 2. The image on the left is the frame as it was captured by the camera. The image on the right is the frame after it has been color corrected. This frame was extracted from the music video "Emmanuelle".

4. A Dialog between Photography Retouching and Color Grading

Let us now investigate on the relationship between photo retouching and color grading. This is rarely done because it is not so common to find people who do both. I happen to do both, and I have consistently and often written accounts of my work as both. This led me to think that a good color grader has to be a good photo retoucher (the opposite is not necessarily needed). It is due to the fact that the building blocks of working on the look of a film is to work on the look of images.

Writing consistently on my work as both a photo retoucher and a color grader also led me to see a dialog between the two, pointing the similarities as well as the differences, and developing an understanding about the tight relationship between the two, and how they could mutually influence each other. This, of course, also has important implications from a pedagogical point of view.

4.1. Similarities between photography retouching and color grading

As you probably noticed, there are similarities in the workflows I presented for photo retouching and color grading. In both cases, we deal with the same first three steps: correcting brightness and contrast, correcting color, and applying a look.

Another similarity has to do with the kinds of tools we find in the softwares that we use to do these jobs. Even though color grading has a few additional tools, the tools we use for the first three steps are very similar. There are sliders, curves, and scopes (mostly a histogram for photo retouching).

Finally, aesthetic considerations about what makes an image beautiful and/or meaningful and/or impactful are also very similar.

4.2. Differences between photography retouching and color grading

Let us now investigate about the major differences between the two. The first major difference has to do with the nature of the medium. Because a film is a rapid succession of still images, there are important additional considerations to take into account when color grading.

First, at the very basic level, we color grade a clip and not a still image. It means we have to select the referent image carefully, if we want the correction to translate in a satisfying way to the other frames of the clip.

It also requires to test the correction we made on the referent image across the whole clip, by watching it entirely to see if there is anything else to adjust. This is particularly difficult when the nature of the material is changing

drastically across the clip, as is the case for a long sequence shot.

In that particular case, it is often possible to have dynamic settings of some color grading parameters, that is, to have some parameters that can vary during the shot according to what happens in the shot.

We also mentioned earlier what is called a secondary correction. Of course, it is not uncommon to correct only part of a photograph when we retouch it. But with film, we are also concerned with tracking, that is, with the fact that this secondary correction has to be applied along a clip and not just to a still image, and to the right area of the image, frame after frame.

There is also the question of clips matching, which does not exist in photo retouching, except when we work with serial photography. When matching clips, it is not just the aesthetics of a still image that we are after, but also the aesthetics of an entire sequence. And that requires different analytical skills and aesthetic knowledge. This is equally true when matching the sequences of a film to give its visual unity to a film.

4.3. *The relationship between photography retouching and color grading*

I would like to expand the previous discussion about similarities and differences between photo retouching and color grading by exploring the relationship between the two a bit further.

Because the nature of the medium is not the same (still images versus moving images) and because the nature of the work is different (solitary work versus collaborative work), it is not uncommon to make some discoveries in one medium and then translating that discovery (when possible) to the other one. Some of my work as a photo retoucher has deeply influenced some of my work as a color grader, and vice versa.

Because the workflow is not entirely the same, and because the tools are not entirely the same, discoveries concerning aesthetics and also the use of tools and the workflow can happen indifferently in one medium or the other. This is also why a photographer can be influenced by a filmmaker and a filmmaker can be influenced by a photographer.

But this relationship goes much deeper. And this is something that is harder to explain since it is mostly phenomenological, that is, based on a lived experience [2]. It is a bit as if those two things were only one thing for me and, as I work on a photo, I mobilize knowledge from both the photo and the film worlds, and from both my experiences as a photo retoucher and a color grader. It is the same when I work on a film as a color grader. This is

also the reason why I keep learning about photography to feed my film work, and why I keep learning about film to feed my photo work.

4.4. *Pedagogical implications of the relationship between photography retouching and color grading*

There are important implications of that relationship between photo retouching and color grading in the area of pedagogy.

I always have students of color grading work on photo retouching first, and I encourage them to keep doing so all over their career. First, because the problem of photo retouching is a simpler one than the problem of color grading, and a building block for it. But also because the questions you ask yourself as a photo retoucher, consciously or unconsciously, can inform to great lengths the questions you ask yourself as a color grader.

I also suggest to students of photography who are particularly interested and fluent in photo retouching to look into color grading as a career, or at least as a strong interest of theirs. This is due to the fact that working as a color grader can inform back on your process as a photo retoucher, and make you a better photo retoucher.

In those two cases, I am not only talking about the tools, and how you can discover various uses of them when going back and forth between photo retouching and color grading. I am also speaking about aesthetics, and the various parameters and aspects that contribute to all the decisions we make while photo retouching or color grading.

Those operations are very complex, and it takes a lot of times to refine the artistry. Any input that can help refine it is always welcome. And working in photo retouching can definitely help refine the artistry of color grading. And the same is true the other way around.

5. Conclusion

In this paper, I have dealt with photo retouching and color grading. I have questioned why we were doing it in the first place. I have then offered a simplified workflow for both.

I have also discussed the commonalities and the differences between the two, exploring the tight relationship they had with each other. I have also explored some implications of that relationship in pedagogy.

The next steps of this research would be to expand even further on the relationship between the two, as well as to develop further the implications for practice and pedagogy. In particular, it could be really interesting to give some more detailed accounts of a phenomenology of photo retouching

and color grading, and see where the creative processes in both activities intersect and diverge.

6. Conflict of interest declaration

The author declares that nothing affected their objectivity or independence and original work. Therefore, no conflict of interest exists.

7. Funding source declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

8. Short biography of the author

Ivan Magrin-Chagnolleau is an artist philosopher who has been involved in art making as well as academic research and teaching for most of his life. He is affiliated with the CNRS in France, and currently works with the PRISM Laboratory in Marseille, a joint-lab between CNRS and Aix-Marseille University investigating on the relationships between art and science. He is also affiliated with Chapman University in California. His interests include: the creative process and its phenomenological dimension, art and philosophy, aesthetics, artificial intelligence and creativity, art and spirituality.

Notes

[1] I owe the discovery of that plug-in to Rick Sammon (2016).

[2] For the readers who are interested in learning more about phenomenology, you can read for instance Husserl, 1985, Husserl, 1992, or Merleau-Ponty, 1976.

References

Adams, A., & Baker, R. (1995). *The Camera* (Reprint Edition). Ansel Adams.

Alekan, H. (2001) *Des lumières et des ombres*. Paris: Éditions du Collectionneur.

Berleant, A. (2004). *Re-thinking Aesthetics: Rogue Essays on Aesthetics and the Arts*. New York: Routledge.

Bleicher, S. (2012). *Contemporary Color: Theory and Use* (2nd Edition). Clifton Park: Delmar, Cengage Learning.

Brown, B. (2016) *Cinematography Theory and Practice, Third Edition*. New York: Focal Press.

Brown, S., Street, S., & Watkins, L. (Eds.). (2012). *Color and the Moving Image: History, Theory, Aesthetics, Archive*. New York: Routledge.

Costello, D. (2018). *On Photography: A Philosophical Inquiry*. New York: Routledge.

Fissoun, D. (2020). *Color Correction with DaVinci Resolve 16*. Blackmagic Design Learning Series.

Frost, J. B. (2009) *Cinematography for Directors: A Guide for Creative Collaboration*. Studio City: Michael Wiese Productions.

Haine, C. (2019) *Color Grading 101*. New York: Routledge.

Hullfish, S. (2012- *The Art and Technique of Digital Color Correction, 2nd Edition*. New York: Routledge.

Hurkman, A. V. (2013). *Color Correction Handbook: Professional Techniques for Video and Cinema* (2nd Edition). Peachpit Press.

Hurkman, A. V. (2013). *Color Correction Look Book: Creative Grading Techniques for Film and Video*. Peachpit Press.

Husserl, E. (1985). *Idées directrices pour une phénoménologie*. Paris: Gallimard.

Husserl, E. (1992). *L'idée de la phénoménologie*. Paris: PUF.

Hyman, J. (2006) *The Objective Eye: Color, Form, and Reality in the Theory of Art*. Chicago: University of Chicago Press.

Jarvie, I. (2015). *Philosophy of the Film: Epistemology, Ontology, Aesthetics*. New York: Routledge.

Magrin-Chagnolleau, I. (2013) 'The Use of Color in Theatre and Film', in *Proceedings of AIC 2013*, pp. 1481-1484.

Merleau-Ponty, M. (1976). *Phénoménologie de la perception*. Paris: Gallimard.

Mollica, P. (2012) *Color Theory*. Mission Viejo: Walter Foster.

Rabiger, M., & Hurbis-Cherrier, M. (2020). *Directing: Film Techniques and Aesthetics*. New York: Routledge.

Sammon, R. (2016) *Evolution of an Image*. New York: Routledge.

Storaro, V. (2011) *Writing With Light, Colors, and the Elements*. Florence: Mondadori Electa.

A Phenomenological Approach to Color Theory and Aesthetics (2022)

Ivan Magrin-Chagnolleau. **A Phenomenological Approach to Color Theory and Aesthetics.** *Proceedings of AIC2022.*

5 pages

A Phenomenological Approach to Color Theory and Aesthetics

Ivan Magrin-Chagnolleau ^{1, 2,*}

¹ Aix-Marseille Univ, CNRS, PRISM, Marseille, France

² Chapman University, California, USA

* Corresponding author: ivan.magrin-chagnolleau@cirs.fr

Abstract

This paper proposes to investigate color theory and aesthetics from a phenomenological point of view, that is, from the point of view of the lived experience. Two angles of study are suggested: one is from the point of view of the artist creating a piece, the other is from the point of view of the audience receiving the work. In each case, it is possible to interrogate the lived experience in order to find some cues about the aesthetic components that contribute to this lived experience, in particular as they relate to color. A tool, the explicitation interview, is proposed as a mean to conduct such investigation.

Keywords: *Phenomenology of Color, A Practice-based Theory of Color, Color and Aesthetics, Color and Emotions, Color as a Lived Experience*

Introduction

For many years, I have approached color and aesthetics from a practical point of view. Being both a scholar and a practitioner, I consider my research in the framework of the creation research paradigm (Magrin-Chagnolleau, 2022a), also called for instance the practice-based research paradigm (Candy, 2018). My practice as an artist being at the foundation of my research on color and aesthetics, it was natural for me to look into phenomenology (Husserl, 1985, 1992), and to approach color and aesthetics from a phenomenological point of view, that is, from the point of view of the lived experience. This was also for me a way to put the creative process at the centre of my research. In other words, I am interested in the creative process as it is experienced by the person who is involved in this process. And I am particularly interested, in this creative process, in the moments when we make decisions that impact the aesthetic of the work at hand, and in particular decisions, conscious or unconscious, that involve color.

Phenomenology is a paradigm that put the lived experience at the centre. But how can we access the lived experience with an approach that is objective enough and can be reproducible? This is where the explicitation interview, developed by Pierre Vermersch, comes into play (Vermersch, 2014). This interview technique allows to access the lived experience of a subject as it really happened, as opposed to how the subject thinks it happened. Developed initially in the framework of experimental psychology, and particularly in the context of pedagogy, this interview technique has been used by people who were interested in phenomenology, and more particularly in a practical approach to phenomenology (Bitbol, 2017). I have been using this interview technique for the past few years in order to learn about the creative process, about the decisions that impacted the aesthetic dimensions of a creative project, and in particular decisions that had to do with color.

Why is color such an important component of aesthetics? Following the theories of embodied cognition (Varela, 1991), I believe this has to do with the fact that color relate to our daily experiences. Colors are everywhere. And because they are, we have a rich web of associations with them. That means that colors have an important impact on us, in particular emotionally. And that

means that artists manipulate colors, consciously or not, in order to create something in the viewer, at an emotional level, but also in more subtly ways. There is a reason why we speak for instance of warm colors and cold colors, in terms of what experience those colors create for the viewer.

I have already written, in the past, about my experience with color as a practitioner (in Film, Theatre, Photography), included for AIC (Magrin-Chagnolleau, 2013, 2021, 2022b). Each time, I have written from the point of view of a practitioner developing a theoretical reflection based on his practice. Here, I would like to go one step further and address the phenomenological dimension of practicing art-making, aesthetics decision-making, and color choices. The ambitious goal of this research would be to develop a theory of color based on practice, that is, based on the lived experience, in other words, a phenomenological approach to color theory and aesthetics. More modestly, in this paper, we will discuss the bases of such an approach to color theory, as well as its feasibility.

A Practical Approach to Phenomenology

What is phenomenology? A phenomenological approach is an approach that deals with the lived experience of a phenomenon. Phenomenology puts experience in the center. Phenomenology tries to get rid of any theory or preconception one can have about a phenomenon, and tries to approach a phenomenon from the point of view of the lived experience, relying only on the lived experience itself to extract any knowledge about the phenomenon (Husserl, 1985, 1992; Heidegger, 1992; Merleau-Ponty, 1976).

What does it mean to have a phenomenological approach of color? It means investigating color as a lived experience, that is, trying to extract knowledge about color as a phenomenon from the experience we have of color. It refers to mostly two broad categories of experiences we can have of color as a phenomenon. One is on the side of reception: how do I experience receiving color? For instance: how do I experience viewing a painting? How do I experience viewing a photograph? How do I experience viewing a film? How do I experience viewing an orange? a tree? etc.

The other is on the side of the creative process: how do I experience making choices about color as I create something? For instance: how do I experience choosing and using a color in a particular area of a painting? how do I experience choosing the color of a piece of cloth for an actor? how do I experience grading color for a particular shot in a film? etc.

Phenomenology can be approached from a theoretical point of view, in which case we deal mostly with the idea of phenomenology. This is sometimes what has been said of Husserl, the founder of phenomenology. But phenomenology can also be approached from a practical point of view. How can I interrogate the lived experience? How can I extract knowledge from the lived experience? How can I use the lived experience to learn something concrete about a phenomenon and the way we experience it?

There is a way to do it through the use of the explicitation interview (Vermersch, 2014). The explicitation interview was created by Pierre Vermersch in the framework of experimental psychology. The interview was developed in order to be able to interrogate a subject on a past experience and have access to what the subject had precisely done during this past experience, as

opposed to what they thought they had done. This is particularly useful in the framework of training, when someone makes a mistake during a process, without being able to pinpoint the moment of that mistake. Through conducting an explicitation interview, it is very easy to find the moment when the person deviates from the learned procedure.

The explicitation interview was then used in the framework of research, in particular as a tool to approach phenomenology from a practical point of view (Depraz, 2011; Vermersch, 2012). What is unique about this interview technique is the fact that it can have access to some pre-reflected knowledge, that is, knowledge that is stored in someone's mind and can be accessed through the interview, but that the subject themselves were not aware about. For instance, when led by a skilled interviewer, this technique gives access to mental operations conducted by the subject while living the experience, such as information taking or decision making.

There is also a version of the explicitation interview called auto-explicitation (Vermersch, 2007, 2008), which consists in administering the explicitation interview to oneself, using in that case particular techniques to be able to be in turn the interviewer and the interviewee.

Now, there is one drawback to the explicitation interview. It takes between 45 and 90 minutes to conduct an explicitation interview, and during such an interview, one can only interrogate a very brief moment of the past experience of the subject. In order to be practical, you usually need to already have a good understanding of the task at hand, and of the particular moment of the task that you want to finely interrogate. Sometimes, you acquire that understanding by first doing some interviews without really knowing what you are after, and then you will conduct some interviews more specifically on a moment that you are interested in.

Relating Phenomenology, Aesthetics, and Color

Now that we have defined phenomenology and how to approach it from a practical point of view, let us have a further look at how we can apply it to aesthetics, and more particularly to color. What does it mean to approach aesthetics from a phenomenological point of view? Let us go back to the way aesthetics can be seen from the point of view of the lived experience. Here again, there are two angles.

The first one is the angle of the artist. As an artist, when creating a piece, I go to a series of decisions, some conscious, some unconscious. Each decision contributes to making me go through a creative process that culminates in the production of the piece. Most of these decisions impact the aesthetic of the final piece. As the artist, this creative process, that is, the succession of all these decisions and actions, are experienced by me. This builds my lived experience of this creative process. By interrogating this process, these decisions, these actions, someone can have access to this process, to this lived experience, and relate all these decisions and actions to the aesthetic of the final piece. This is what I mean by a phenomenological approach of aesthetics. Rather than relying on a pre-conceived theory about aesthetics, it is possible to have access to the lived experience of the artist and correlate some of the decisions made and actions taken to the resulting piece and its aesthetic.

The same can be said about colors. Some of these decisions and these actions directly deal with color. Therefore, if there is a possibility to have a phenomenological approach to aesthetics, then there is a possibility to have a phenomenological approach to color.

The second angle concerns the way a piece is received by an audience. The audience, when receiving the piece, goes through a lived experience that can also be interrogated. And this interrogation can lead to understanding how the person interviewed experienced the piece and its aesthetic. This is a way to understand, from a phenomenological point of view, how the aesthetic of a piece impacts the lived experience of an audience. And again, part of the lived experience of an audience is impacted by colors, and this can also be interrogated.

I would like to point here that color is an important component of aesthetics, mostly because of the particular relationship between color and emotions. There is a reason why we speak of warm or cold colors. This refers to the lived experience that we go through when being exposed to such colors. Color is also a component that is easily manipulated by artists in order, precisely, to create such lived experiences in an audience.

Conclusion

In this paper, I have tried to describe what a phenomenological approach of color theory and aesthetics could look like, and how we could develop it through the use of the explicitation interview. It speaks of the possibility of interrogating the lived experience of an artist creating a piece or of an audience receiving it. In both cases, through this interrogation, we can access some elements that relate to the aesthetic of the piece, in particular as it relates to color. By doing that, we can establish some ground rules concerning aesthetics and colors that are free from any preconceived ideas and theories about aesthetics and colors, but rely directly on the lived experience, therefore allowing to establish a phenomenological approach to color theory and aesthetics.

The future of this work will consist in testing the possibility of developing a phenomenological approach to color theory and aesthetics by conducting some explicitation interviews with artists creating pieces and audience members receiving them, as well as using auto-explicitation.

References

- Bitbol, M., and C. Petitmengin. 2017. Neurophenomenology and the Micro-phenomenological Interview. In S. Schneider and M. Velmans (Eds.), *The Blackwell Companion to Consciousness* (2nd Edition, pp. 726–739). Oxford: Wiley-Blackwell.
- Candy, L., and E. Edmonds. 2018. Practice-Based Research in the Creative Arts. *Leonardo*, 51(1), 63–69.
- Depraz, N., F.J. Varela, and P. Vermersch. 2011. *À l'épreuve de l'expérience*. Bucharest: Zeta Books.
- Heidegger, M. 1992. *Être et Temps*. Paris: Gallimard.
- Husserl, E. 1985. *Idées directrices pour une phénoménologie*. Paris: Gallimard.
- Husserl, E. 1992. *L'idée de la phénoménologie*. Paris: PUF.
- Magrin-Chagnolleau, I. 2013. The Use of Color in Theater and Film. *Proceedings of AIC 2013*.
- Magrin-Chagnolleau, I. 2021. Handling Color in Photography and Film: From Photo Retouching to Color Grading. *Proceedings of AIC 2021*.

- Magrin-Chagnolleau, I. 2022a. Reflections on the impact of creation research, practice-based research, and practice as research, and how to define and maximize that impact. *Media Practice and Education*, 23.
- Magrin-Chagnolleau, I. 2022b. The Relationship between Photo Retouching and Color Grading. *Color Culture and Science Journal*, 14(1), 79–86.
- Merleau-Ponty, M. 1976. *Phénoménologie de la perception*. Paris: Gallimard.
- Varela, F., E. Thomson, and E. Rosch. 1991. *The Embodied Mind*. Cambridge: The MIT Press.
- Vermersch, P. 2007. Bases de l'auto-explicitation (1). *Expliciter*, 69.
- Vermersch, P. 2008. Introspection et auto-explicitation: Bases de l'auto-explicitation (2). *Expliciter*, 73.
- Vermersch, P. 2012. *Explicitation et phénoménologie*. Paris: PUF.
- Vermersch, P. 2014. *L'entretien d'explicitation*. Issy-les-Moulineaux: ESF Éditeur.

3.9 Création Recherche

Éloge de l'artiste chercheur (2014)

Ivan Magrin-Chagnolleau. **Éloge de l'artiste chercheur.** *p-e-r-f-o-r-m-a-n-c-e, 1*, 2014.

7 pages

Éloge de l'artiste chercheur

Ivan MAGRIN-CHAGNOLLEAU

Résumé

Dans cet article, je définis d'abord les termes artiste, chercheur, et artiste chercheur. Puis je donne quelques exemples d'artistes chercheurs : Léonard de Vinci, Constantin Stanislavski, Bertolt Brecht et Lee Strasberg. J'aborde enfin cette question de l'artiste chercheur à partir de ma propre expérience, et en particulier en tant qu'acteur et metteur en scène. Je montre ainsi tout l'intérêt qu'il y a à articuler une pratique artistique avec une recherche plus théorique sur cette pratique, et comment ces deux facettes, art et recherche, se nourrissent mutuellement et contribuent à une créativité épanouie et énaactive.

Mots clés

Artiste, chercheur, artiste chercheur, Léonard de Vinci, Constantin Stanislavski, Bertolt Brecht, Lee Strasberg, créativité, processus créatif, création, acteur, metteur en scène, énaaction, création énaactive, recherche énaactive.

1 Prologue

Je suis un artiste chercheur. Je suis bien sûr d'autres choses. Et en même temps, je ne suis aucune de ces choses-là. Mais si je revendique ici d'être un artiste chercheur, c'est que cela ne va pas de soi dans le milieu de la recherche académique française. Et cela pour plusieurs raisons. Tout d'abord, il semblerait qu'il n'est plus possible aujourd'hui de s'intéresser à la fois à l'art et aux sciences. Il n'y a pourtant pas si longtemps, cela allait de soi. D'autre part, nous sommes à une époque de surspécialisation du savoir. On demande à un chercheur de choisir une discipline et de s'y tenir. Cela permet de le ranger dans la bonne section, dans la bonne case. Je pense pourtant qu'une surspécialisation du savoir conduit à un appauvrissement intellectuel, ainsi qu'à un déclin de la créativité. Il n'y a pas si longtemps, il était courant d'avoir une activité de chercheur tout en pratiquant un art à un très haut niveau, voire un instrument de musique. Je revendique donc d'être un artiste chercheur, et j'en fais même une condition sine qua non pour une créativité épanouie.

2 Qu'est-ce qu'un artiste chercheur

Pour commencer, je voudrais préciser ce que j'entends par artiste chercheur.

2.1 Artiste

Tout d'abord, qu'est-ce qu'un artiste ?

Le Grand Robert de la langue française donne plusieurs définitions pour le mot artiste. Un artiste peut d'abord être une « personne pratiquant un métier, une technique difficile ». Un artiste peut également être une « personne qui se voue à l'expression du beau, pratique les « beaux-arts », l'art ». Un artiste peut être le « créateur d'une œuvre d'art ; spécialement, celui ou celle qui se consacre aux arts plastiques ». La dernière définition que le Grand Robert donne pour le mot artiste est une « personne qui interprète une œuvre musicale, théâtrale ou cinématographique (par opposition à auteur, compositeur, écrivain) ».

Le Grand Robert propose également plusieurs citations pour éclairer ce concept d'artiste. Par exemple, cette citation de Théophile Gautier : « Tout artiste qui se propose autre chose que le beau n'est pas un artiste ». Ou encore cette citation de Flaubert dans ses correspondances : « Où est la limite de l'inspiration à la folie, de la stupidité à l'extase ? Ne faut-il pas pour être artiste voir tout d'une façon différente de celle des autres hommes ? »

La première chose qui ressort de toutes ces définitions est le fait qu'un artiste s'inscrit dans une pratique. Cette pratique peut être une pratique de création ou une pratique d'interprétation, mais dans les deux cas il s'agit d'une pratique artistique. Il y a donc dans les deux cas une expression, que cette expression soit en train d'être créée dans le moment ou bien que la chose exprimée le soit en référence à une œuvre qui a été créée préalablement.

D'après moi, un artiste est une personne qui s'exprime par le biais d'un médium artistique. Toute personne qui s'exprime par le biais d'un médium artistique est selon moi un artiste.

Qu'est-ce qu'un médium artistique ? C'est un médium qui permet d'exprimer quelque chose de ressenti intérieurement autrement que par le langage conventionnel. Voici quelques exemples de médias artistiques : la peinture, la musique, la danse, le théâtre, la performance, la poésie, la littérature, la sculpture, l'opéra, la chanson, la photographie, le cinéma, la vidéo, etc. La littérature et la poésie utilisent le langage mais elles le font de manière non conventionnelle. Ce que j'entends par autrement que le langage conventionnel est lié à la limite intrinsèque du langage pour traduire l'expérience. Dans la mesure où le langage est une convention qui permet une communication entre deux êtres, il contient en lui-même une limite dans sa capacité à exprimer l'expérience. C'est un pis aller, qui est nécessaire, mais qui ne permet pas de tout exprimer. L'artiste va donc rechercher une autre façon d'exprimer son expérience, et en particulier ce qu'il ressent en lien avec son expérience. C'est pour cela qu'il utilise un autre médium que le langage, un médium artistique. Par conséquent, toute personne qui utilise un autre médium que le langage conventionnel pour partager son expérience et exprimer son ressenti intérieur est un artiste.

2.2 Chercheur

Qu'est-ce qu'un chercheur ?

La première définition que nous donne le Grand Robert de la langue française est une définition toute simple : un chercheur est une « personne qui cherche (quelque chose) ». Le Grand Robert nous propose aussi la définition suivante : « personne qui se consacre à la recherche scientifique ». Il y aurait donc deux types de recherche, une recherche dite « scientifique » et une simple recherche, c'est-à-dire une recherche qui n'est pas scientifique. Qu'est-ce qui fait d'une recherche qu'elle est scientifique ? On pourrait proposer qu'une recherche est scientifique à partir du moment où elle suit les étapes de la méthode scientifique. Or la méthode scientifique, telle qu'elle a été proposée notamment par Descartes, ne s'applique pas telle quelle à tous les champs disciplinaires. On a vu, notamment en sciences humaines, de nouvelles méthodologies être proposées. On pourrait dire qu'une recherche est scientifique à partir du moment où quelqu'un se donne une méthode pour effectuer sa recherche, et s'efforce ensuite de suivre cette méthode.

Un chercheur est donc quelqu'un qui cherche, c'est-à-dire quelqu'un qui se pose des questions et qui cherche des réponses. Si l'on veut donner une définition un peu plus scientifique du chercheur, on dira que c'est quelqu'un qui formule une hypothèse et qui essaie ensuite de la valider ou de l'invalider, en se donnant pour cela une méthode. Toute personne qui se pose une question est un chercheur. Toute personne qui se pose une question part à la recherche de sa réponse. Il fait donc de la recherche. Et cette recherche sera scientifique à partir du moment où il se donnera une méthodologie de recherche rigoureuse.

2.3 Artiste chercheur

Il y a donc plusieurs façons d'envisager l'artiste chercheur. Un artiste chercheur peut être un artiste qui se pose des questions, c'est-à-dire quelqu'un qui s'exprime par le biais d'un médium artistique tout en cherchant des réponses à des questions. Cela peut être aussi quelqu'un qui cherche des réponses à des questions par le biais d'une expression artistique. Cela peut également être quelqu'un qui cherche des réponses à des questions, puis les exprime par le biais d'un médium artistique. Cela peut aussi être quelqu'un qui est artiste d'une part, c'est-à-dire qui s'exprime par le biais d'un médium artistique, et qui est chercheur d'autre part, c'est-à-dire qui cherche des réponses à des questions. Pour simplifier, disons qu'un artiste chercheur est une personne qui a une activité artistique et une activité de recherche. C'est cette définition la plus large que j'utiliserai dans la suite de mon article.

3 Quelques exemples d'artiste chercheur

3.1 Léonard de Vinci

Le premier exemple qui me vient à l'esprit est celui de Léonard de Vinci[3]. On connaît bien sûr son grand talent de peintre. On connaît également son enthousiasme et son génie à aligner des inventions en tout genre, par exemple l'hélicoptère, dont il n'est pas l'inventeur mais dont il a amélioré considérablement le prototype. C'était également quelqu'un qui était réputé pour sa capacité à organiser de grandes fêtes sublimes. Il était donc une sorte de metteur en scène de grandes fêtes somptueuses. Et ce ne sont que quelques exemples parmi toutes les différentes activités dans lesquelles Léonard de Vinci pouvait mettre son génie créateur à l'œuvre.

3.2 Constantin Stanislavski

Constantin Stanislavski est un exemple très intéressant d'artiste chercheur. En effet, il a démarré sa carrière en tant qu'acteur, puis il est devenu metteur en scène, puis il a beaucoup réfléchi à la formation de l'acteur, et il a finalement développé toute une théorie du jeu[4][6][7][5]. Mais cette théorie du jeu repose entièrement sur son expérience d'acteur et de metteur en scène. C'est donc sa grande expérience artistique qui lui a permis de mener cette recherche sur le jeu de l'acteur, et de produire une méthode qui est encore une méthode de référence aujourd'hui. Il n'aurait pas pu écrire ce qu'il a écrit s'il n'avait été d'abord acteur et metteur en scène. Il s'agit ici d'expérience et non pas de qualité. Il se trouve que dans le cas de Constantin Stanislavski, il était un acteur et un metteur en scène remarquable. C'est en tout cas ce que l'on sait de tous les témoignages qui ont été écrits à son sujet. Mais quand même il n'aurait pas été un très bon acteur ni un très bon metteur en scène, c'est avant tout le fait qu'il ait une expérience comme acteur et comme metteur en scène qui est important ici. C'est cette expérience importante qui lui a permis de théoriser le jeu de l'acteur et de proposer une réflexion d'une richesse incroyable sur ce que signifie incarner un personnage. Cette recherche a été d'ailleurs tellement pertinente, précisément parce qu'elle s'appuyait sur une pratique artistique intensive, qu'elle fait encore référence aujourd'hui. Les écrits de Constantin Stanislavski ont influencé plusieurs générations d'acteurs et de metteurs en scène, et continuent de les influencer aujourd'hui.

3.3 Bertolt Brecht

Bertolt Brecht est un autre exemple très intéressant d'artiste chercheur. Il a en effet été acteur, puis metteur en scène, avant de commencer à écrire comme théoricien sur le théâtre[2]. Il est bien entendu connu davantage comme metteur en scène que comme acteur, ainsi que comme directeur artistique du Berliner Ensemble Theater, mais tous ses écrits théoriques reposent sur son expérience artistique. Il a davantage écrit sur la mise en scène que sur le jeu de l'acteur, ce qui est finalement assez cohérent avec le fait qu'il était surtout metteur en scène. Il était principalement préoccupé par la possibilité que l'art, et dans son

cas le théâtre, puisse être un outil de changement social. Il a donc développé toute une théorie de la mise en scène comme outil pour rendre la participation du public la plus active possible, et induire ainsi un changement dans les mentalités. C'est ce qu'il a appelé la distanciation. Mais tout ce qu'il a développé au niveau théorique reposait essentiellement sur sa pratique de metteur en scène.

3.4 Lee Strasberg

Je terminerai par l'exemple de Lee Strasberg, qui est un acteur et metteur en scène de théâtre américain, ainsi qu'un grand théoricien du jeu de l'acteur. Lee Strasberg a rencontré Stanislavski lorsque ce dernier est venu à New York. Il a également fait parti de la grande aventure du Group Theater, dont il a été l'un des fondateurs[10]. Après la seconde guerre mondiale, il a été contacté par Elia Kazan afin de devenir l'un des directeurs artistiques de l'Actors Studio. C'est surtout au sein de l'Actors Studio qu'il a pu expérimenter et développer sa théorie du jeu de l'acteur[18][13][14]. Mais cette dernière est également issue de sa grande expérience en tant qu'acteur et metteur en scène de théâtre. Les deux sont inséparables. Sa recherche consistait, d'une part, en son propre travail d'acteur et de metteur en scène, d'autre part, en son expérimentation en tant que coach et directeur d'acteurs.

4 Ma propre expérience d'artiste chercheur

J'en viens maintenant à mon expérience d'artiste chercheur. Je pratique en effet plusieurs activités artistiques en lien avec le théâtre, le cinéma, la musique, la photographie et l'écriture. Par ailleurs, je fais de la recherche en esthétique et en philosophie de l'art. J'essaie donc de combiner ces activités afin qu'elles s'enrichissent mutuellement. Mon travail artistique me sert de terrain pour mes recherches, et mes recherches me permettent d'approfondir mon travail artistique. Il y a donc une osmose parfaite entre mes activités. Chacune d'elle impacte l'autre, dans un processus mutuel d'interaction et de feed-back nommé énaction par Francisco Varela[9].

4.1 Acteur

Je voudrais aborder deux exemples, parmi mes pratiques artistiques, qui illustrent ce va-et-vient entre activité artistique et recherche. Je commencerai par celui d'acteur. J'ai été d'abord formé à Lyon pendant deux ans par Élisabeth Saint-Blancat au théâtre des clochards célestes. Puis j'ai suivi pendant trois ans les cours de l'Actors Studio à New York. À la suite de cette formation, j'ai interprété environ une quinzaine de rôles principaux au théâtre et au cinéma. En parallèle, j'ai lu de nombreux ouvrages sur la théorie du jeu de l'acteur et sur le travail du comédien[4][6][7][5][13][14] etc. J'ai en outre coaché, formé et dirigé de nombreux comédiens. Et j'ai mené tout au long de ce travail une activité de réflexion et d'écriture pour formuler mes propres idées et concevoir mes propres exercices pour me développer comme comédien, et pour aider les autres à se

développer comme comédiens. C'est ce va-et-vient entre pratique artistique, recherche et enseignement qui me permet d'avancer dans ma compréhension du travail de l'acteur et donc de générer de la connaissance pour moi et pour les autres.

4.2 Metteur en scène

Cette même interaction existe dans mon travail de metteur en scène. J'ai également été formé à la mise en scène pendant mes trois années à l'Actors Studio. J'ai réalisé à ce jour une douzaine de mises en scène. J'ai également formé et coaché un certain nombre de metteurs en scène. J'ai lu de nombreux livres sur la mise en scène et le travail de metteur en scène^{[11][19][15][16][17][8][1][12]} etc. Et en parallèle de toutes ces activités, j'ai cherché à formuler ma propre compréhension et ma propre pratique du travail de metteur en scène. Encore une fois, c'est cette interaction permanente entre toutes ces activités qui me permet d'avancer dans ma compréhension et dans ma pratique du travail de metteur en scène, et ainsi de générer de la connaissance pour moi et pour les autres.

5 Épilogue

À travers cet article, j'ai essayé de montrer tout l'intérêt qu'il y avait à être à la fois un praticien et un théoricien dans des domaines comme l'esthétique et la philosophie de l'art. Je ne renie pas du tout le point de vue de certains chercheurs qui n'ont pas d'ancrage dans la pratique artistique et qui sont capables d'articuler des pensées très cohérentes et très utiles. Mais il est également nécessaire, à mon avis, d'entendre des points de vue théoriques émanant de praticiens. Et je pense même qu'il est très utile d'entendre des praticiens s'exprimer sur leur travail, même quand ils n'ont pas d'ancrage théorique avéré. Je prendrai même le risque de dire qu'un point de vue ancré dans une pratique s'avérera généralement plus utile, tout au moins hors d'une communauté composée d'une petite poignée de chercheurs spécialistes.

J'ai également voulu montrer l'intérêt qu'il y avait à s'intéresser à plusieurs domaines à la fois, et même aux frontières entre divers domaines. Je suis convaincu, et c'est mon expérience, que la créativité fleurit lorsqu'elle peut s'exprimer dans plusieurs directions à la fois.

Ces deux points de vue mériteraient d'être étayés davantage, et ils le seront très probablement dans de futures communications. Mais si par cet article j'ai déjà pu contribuer modestement à ce qu'il y ait une plus grande ouverture dans le milieu académique, moins de frontières entre les disciplines, et moins de frontières entre praticiens et théoricien, alors j'aurai accompli mon dessein.

Références

- [1] Bartow Arthur. *The Director's Voice*. Theatre Communications Group, 1988.
- [2] Brecht Bertolt. *Écrits sur le théâtre*. Gallimard (Pléiade), 2000.
- [3] Vecce Carlo. *Leonardo*. Salerno, 1998.
- [4] Stanislavski Constantin. *An Actor Prepares*. Theatre Arts, Inc., 1937.
- [5] Stanislavski Constantin. *My Life in Art*. Theatre Arts Books, 1948.
- [6] Stanislavski Constantin. *Building A Character*. Max Reinhardt Ltd., 1950.
- [7] Stanislavski Constantin. *Creating A Role*. Theatre Arts Books, 1961.
- [8] Jones Richard David. *Great Directors at Work*. University of California Press, 1986.
- [9] Varela J. Francisco Thompson Evan Rosch Eleanor. *The Embodied Mind*. MIT Press, 1991.
- [10] Clurman Harold. *The Fervent Years*. Da Capo Press, 1945.
- [11] Clurman Harold. *On Directing*. Fireside, 1972.
- [12] Féral Josette. *Mise en scène et jeu de l'acteur (Tomes 1 & 2)*. Éditions Jeu, 1997/1998.
- [13] Strasberg Lee. *A Dream of Passion*. Plume, 1987.
- [14] Cohen Lola. *The Lee Strasberg Notes*. Routledge, 2010.
- [15] Brook Peter. *The Empty Space*. Touchstone, 1968.
- [16] Brook Peter. *The Shifting Point*. Theatre Communications Group, 1987.
- [17] Brook Peter. *The Open Door*. Pantheon Books, 1993.
- [18] Hethmon H. Robert. *Strasberg at the Actors Studio*. Theatre Communications Group, 1965.
- [19] Vaughan Stuart. *Directing Plays*. Longman, 1992.

**Reflections on the impact of creation research,
practice-based research, and practice as research, and
how to define and maximize that impact (2022)**

Ivan Magrin-Chagnolleau. **Reflections on the impact of creation research, practice-based research, and practice as research, and how to define and maximize that impact.** *Media Practice and Education*, 2022, 23(2), pp. 105-111.

10 pages

Reflections on the impact of creation research, practice-based research, and practice as research, and how to define and maximize that impact

Ivan MAGRIN-CHAGNOLLEAU^{1,2}

¹*Aix Marseille Univ, CNRS, PRISM, Marseille, France*

²*College of Performing Arts, Chapman University, California, USA*

E-mail: ivan.magrin-chagnolleanu@cnrs.fr

Abstract

In this contribution, I reflect on creation research/practice-based research/practice as research. I mostly explore the notion of the impact of such practices. I underline the fact that impact can mean various things depending on the type of projects: it can be emotional; it can be a way to influence; or it can act more subtly on the subconscious mind. But impact is always about the intention to induce change. I also discuss the difficulty, most of the time, to have a quantitative approach to impact, since it is often difficult to measure. I suggest, instead, to include a reflection on the planning of impact and its maximization before embarking on a project. I illustrate those points with various examples taken from my own practice.

Keywords: impact, inducing change, creation research, practice-based research, practice as research, epistemology

Introduction

I will mostly address the question of the impact of creation research/practice-based research/practice as research by detailing some strategies to define, anticipate, and maximize that impact.

What is at stake here is the impact of creation research, as well as the recognition of that impact by institutions. Impact is not something that can easily be defined, planned, or measured. But it has more to do with intentionality. What is often asked is not so much what the impact of such practice will be, but, mostly, what steps are planned, in order to take into consideration the possible impact of such practice, and to implement strategies to maximize it.

Several examples taken from my practice will be used to illustrate the impact of creation research. What is the impact of art-making with patients of a psychiatric day hospital? What is the impact of producing theatre as a benefit for health research, or for the victims of terror attacks? What is the impact of making short films that encourage empathy and compassion in the viewer?

What is Creation Research?

Let us start by defining creation research / practice-based research / practice as research.

According to the classification developed in several works by Grazia Giacco (Giacco 2018, 2021), we find various names for this paradigm of research that are close to each other but not rigorously equivalent: practice-based research (Candy 2006; Candy and Edmonds 2018), art as research (Macleod and Holdridge 2006), practice as research (Elkins 2009, 2014), practice-led research and research-led practice (Smith and Dean 2009; Easton 2011), art practice as research (Sullivan 2009), artistic research (Borgdorff 2012), research in art (Borgdorff 2012), research-creation (Chapman and Sawchuk 2012), creation as research (Chapman and Sawchuk 2015). In the French language also, Giacco specifies the diversity of names (Giacco 2021): ‘recherche création’ (Gosselin and Le Coguiec 2006; Bruneau and Villeneuve 2007), ‘recherche et création’ (Fourmentraux 2011), ‘recherche-cr ation’ (Giacco 2020, 2018; St evance and Lacasse 2013), and ‘recherche en cr ation’ (Martinez Thomas and Naugrette 2020). I personally advocate, in the French language, for the term ‘cr ation recherche’, which we could translate into English as ‘creation research’.

Despite the polymorphous nature of this concept, which these various names are trying to embrace, we can start to grasp it by saying a few things about it. First, there are various ways of generating new knowledge, and there are various types of new knowledge that can be generated. Through practice, and through art practice, in particular, there is new knowledge to be generated, and art practice is one way to generate new knowledge. What, then, are the conditions for this knowledge to be incorporated, admitted, in the pool of "Scientific Knowledge"? Every scientific community tries to follow some concept of what constitutes science. Every scientific community comes up with its own criteria, and they change from community to community. Therefore, for research to become qualified as research, it needs to federate around it a community that would then set up its own rules for what can enter the field, and what cannot. It is not to say that every art practice has the goal of generating new knowledge to be admitted as scientific knowledge, but

as soon as the word *research* is added, and this art practice is done in the context of the academic world, then this is usually the case.

As a practitioner, I am confronted, through a practice, to this notion of knowledge. Through the decisions I make while practicing, I have to mobilize some knowledge, part of it practical, part of it theoretical. But through the practice, I also generate new ideas, new concepts, new ways of making some decisions: that is, new knowledge. Part of this knowledge might not gain the status of scientific knowledge, but part might. It all depends on the ability of the practitioner to articulate this new knowledge in relation to what has already been accepted as scientific knowledge by a community. And then we come to a paradox, which is that if the knowledge resembles previous knowledge enough, it is more easily accepted by the community; but if it is too disruptive, and, therefore, mostly innovative, the risk of rejection is higher.

To make this concept of creation research clearer, let us give an example and show how practice can inform research. The example I want to use is work I did in a psychiatric day hospital a couple of years ago. In that hospital, they were experimenting with institutional psychotherapy and the therapeutic club, along the lines of work by Tosquelles (2007), Oury (2003, 2016), and Delion (2011, 2018). I met with patients and professionals, who were interested in the idea of creating a video workshop together. We discussed possible activities, and made some collective choices about the activity itself and its content. In just over 12 months, we made three short films, one of about 20 minutes, and two of about 30 minutes. The themes of the films were also chosen collectively, as well as the way to work on them. Everyone participated in the project: patients, professionals, and me. Each person could decide, at each meeting, what they wanted to do that day. How did this practice inform my research? Behind that activity, there is, for me, the intention to understand better how to help patients with psychological problems to feel better. My practice informs my research.

Let us now say a few words about the situation in France concerning creation research. In France, there have always existed two ways of studying art. One way was practical and was the prerogative of the art schools. The other way was theoretical and was the prerogative of the universities. Because of that cleavage, it has always been difficult to articulate practice and research. It is only recently that there has been some effort to bridge this gap, by asking the art schools to start developing research programs, and by asking the universities to start developing practice programs. In some cases, universities and art schools even collaborate to devise research

programs that are strongly based on practice. It has become a bit easier, but still challenging, to address some work handling both the practical and the theoretical side (see Martinez Thomas and Naugrette, 2020, for a good account of the issue).

What is Impact?

Let us now have a look at what *impact* is. There are several components of impact. One is emotional. Something or someone can impact someone or a group of people at the emotional level. Another aspect of impact is influence. Something or someone can influence someone or a group of people through doing or saying something. And there is a third and more subtle dimension of impact which we could call impressionistic. Something or someone creates an impression in someone or in a group of people, that can stay with them and subtly operate on their consciousness.

Ultimately, impact is about change. When we say we want to impact someone, it means we want to induce some change in them. Each creative project is an opportunity to create some change, to modify the representations that people have about something, to provoke some action from them, or to implicate them in an activity whose goal is to produce some positive change in them.

Before dealing with a few examples of creative works and their impacts, I would like to address two last issues concerning impact. The first one is the issue of measurement. Can we measure impact, and how? It is usually quite challenging to measure the impact of a creative activity. Sometimes, some cues can be observed, that someone has been impacted and has possibly started doing something differently. Sometimes, the success of an endeavour can be measured by quantitative data, like money that has been collected, or the number of people who have enrolled. But in a lot of cases, impact can only be approached qualitatively.

If we cannot always easily measure impact, can we at least plan for it? Can we anticipate what kind of impact our work can have, and how to prepare for it, and optimize the chances of having a bigger impact? This is usually the wisest thing to do, and is often asked by institutions offering grants for instance.

Examples

Let us review now a few examples that will help clarify various approaches of impact, its

anticipation, its planning for maximization, and, in some cases, its measurement.

Let us go back, first, to the example mentioned earlier, regarding creation research. When working with patients of a psychiatric day hospital, the goals are multiple. First, the goal is to propose an activity that will be stimulating enough for them to want to invest some time in it. Second, and in my opinion, this is the major goal: we aim to induce some well-being in people whose main experience is either extreme anxiety, or lethargy, when taking medication. Therefore, the main goal is to propose an activity that will allow them to feel better, to have a better image of themselves, and hopefully to make progress towards an autonomous functioning in society. In this approach, the therapeutic goal is first, but it is a shared opinion among the proponents of institutional therapy and therapeutic clubs that an artistic practice contains all the ingredients to reach the therapeutic goal.

A patient's well-being is not so easy to measure. To decide if a patient is doing better after having participated in an activity is even harder. But there are a few things we can plan in that situation, in order to maximize the possibility that the patients will feel better. First, we can involve them in all the decisions concerning the activity: its definition, its content; and, in that case, the themes, the characters, or the role they want to play in the activity. Listening to them carefully and observing them can also help maximize the impact. When something makes them uneasy, it is very important to be able to talk about it, to create a space where they can express themselves in relation to the experience. Finally, by observing their behaviour and listening to what they say, it is often possible to orientate the activity so that they feel even better while being involved in it. But a big part of this ongoing evaluation is highly subjective, and very difficult to quantify. Even though I observed some growing involvement in the activity, and some manifestations of emotions having a positive valence, I could not say for sure that it helped the wellbeing of those patients. In this particular case, I had to rely on the previous work of some pioneers (Tosquelles, Oury, Delion), and on some conviction.

Finally, the impact of such activities is not just on the patients, but also on the professionals and on me. I have learned a lot about the patients, about myself, about what it means to suffer psychologically, and about the daily experiences of the patients. This is also an important impact of this type of work.

I also want to point out here, that the evaluation of the impact of such activities is done at several levels. The first level is the level of the activity itself, and implies the person leading the

activity, that is, myself, the patients participating in its definition and in the activity, and the professionals who are involved in it, including the psychologist in charge of the psychiatric day hospital. There is another level of evaluation at the level of the whole structure, the psychiatric hospital and, in particular, the psychiatrist in charge of the whole sector to which this particular day hospital belongs to. He decides whether to continue such activities or stop them. And he responds to a board, which constitutes a third level of evaluation. Ultimately, it is a mixture of trust between levels of evaluation and of certain indicators about the merits of an approach. One such indicator would be the number of patients that become autonomous in a given psychiatric day hospital, compared to other psychiatric day hospitals of the same sector.

The second example consists in producing shows to raise funds for charity. For instance, I directed a short play, *An Honest Arrangement* by David Wiener, for free and the actors performed in it for free, in order to collect money for the Mercy Ships association. I also directed a friend in *First Love* by Samuel Beckett, as a benefit for the Institute of the Young Blind People in Paris as well as for bone marrow research. Finally, I produced and directed a show, *Choices*, in New York, with actors performing for free and the owner of the performance space letting us use it for free, in order to collect some money for the Life for Paris Association, which supports the victims of the November 2015 terror attacks in Paris.

I must make clear here that in the example of the short play *An Honest Arrangement*, the play had already been directed in a different context, and we were simply asked if we were willing to do it again, this time for a benefit. The artistic gesture had therefore already been done, independently of the goal to collect money for the Mercy Ships, and it was merely a question of redoing it for a different purpose. In this new framework of the benefit, the question of impact then becomes more important.

In the example of the show *Choices*, the idea to do a show and the idea to do a benefit emerged together. But it had a positive effect, as generosity is often a stronger motivator to get something done than the pure will to produce art. It also makes the production of the art a bit easier as it is easier to motivate people for a benefit rather than just for an extra line on a résumé.

In these last examples, the impact is easy to measure: it is the amount of money collected, and what the various beneficiaries of this money could do with it. But there is another impact to these sorts of endeavours, which is on the people doing it. Doing something good and generous opens the heart, increases the capacity for empathy and compassion, and creates strong feelings

that can also bring about some change inside ourselves. Here the process is subordinate to the outcome, and this is subordinate in turn to the goal of raising money. But at the same time, the goal to raise money is a pretext to get some art done.

The financial impact is easy to measure, and is evaluated by the people initiating the benefit. The impact in terms of what the money is used for is measured directly by the structures who receive the money and decide how best to use it. The impact on the well-being of the people involved in doing the benefit or buying tickets for it is harder to measure, and it can only be evaluated by the people themselves.

The last example consists in trying to impact the audience of a work. When developing an artistic work, as an artist, I imagine an audience and how I would like to impact them. For instance, I made a couple of short films, *Party Time!* and *King Cake*, to bring awareness to some situations where we are indifferent to some people's solitude and, sometimes, distress. The intent of these films was to create awareness, but also to develop some empathy and, hopefully, compassion, in the audience.

It is very difficult to know if I have managed to do what I intended. The impact would depend first on the number of people that my films reach. But even if I could somehow measure this number, how could I make sure that they experienced what I intended them to experience? The only approximation of this I can propose is when the film is shown to an audience, and I am there at the end of the film to answer some questions. I can infer, to some degree, from the questions, what people got from the film, what they possibly did not understand, and, also, how they might have been affected. I remember a friend of mine who told me that, since he had seen *King Cake*, he was looking at homeless people differently, with more compassion.

But this is a situation where measuring impact is quite challenging. Also, one could wonder if this kind of impact is enough. I would argue that, even if not very ambitious, there is nothing more important than to open the heart of someone else, even if so little.

In these cases, there is not really an entity responsible for evaluating the impact of the work. In more commercial projects, there is the number of tickets sold and the financial benefit of the film that can at least show if the film reached its intended audience. But the only way to know for sure if the audience was impacted the way it was intended would be to ask an audience to fill in a questionnaire, and to ask questions that address the intended impact.

Conclusion

In this contribution, I have tried to propose a definition of impact in the framework of creation research, with its three proposed dimensions: emotional, influence, impression. Could impact be defined differently? Are there other dimensions of impact?

I have underlined the difficulty to quantify impact prior to the realization of a project, but also sometimes after its realization. I have offered as an alternative to impact measurement the suggestion to plan for it, and for its maximization. How can we plan for impact and for its maximisation?

I have finally given a few examples where I tried to define impact, how I planned for it, and how I reflected on the maximization of this impact.

Is it possible and important to think about the impact of a project before embarking on it? Does it maximize the chances of having a bigger impact when doing it?

Funding

This publication was made possible through a funding from the Excellence Initiative of Aix Marseille University, A*Midex, which is part of the French “Programme Investissements d’Avenir” - Funding #: AMX-19-IET-005.

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Borgdorff, Hendrik Anne (Henk). 2012. *The conflict of the faculties: Perspectives on artistic research and academia*. Leiden: Leiden University Press.
- Bruneau, Monik, & André Villeneuve (Eds.). 2007. *Traiter de recherche création en art : Entre la quête d’un territoire et la singularité des parcours*. Québec: Presses de l’Université du Québec.
- Candy, Linda. 2006. *Practice Based Research: A Guide. Creativity and Cognition Studios Report*. <https://www.creativityandcognition.com/wp-content/uploads/2011/04/PBR-Guide-1.1-2006.pdf> (accessed 31 May 2021).

- Candy, Linda, and Ernest Edmonds. 2018. "Practice-Based Research in the Creative Arts." *Leonardo*, 51(1), 63–69.
- Chapman, Owen, and Kim Sawchuk. 2012. "Research-Creation: Intervention, Analysis and Family Resemblances." *Canadian Journal of Communication*, 37, 5–26.
- Chapman, Owen, and Kim Sawchuk. 2015. "Creation-as-Research: Critical Making in Complex Environments." *RACAR: Revue d'art canadienne*, 40(1), 49–52.
- Delion, Pierre. 2011. *Accueillir et soigner la souffrance psychique de la personne - Introduction à la psychothérapie institutionnelle*. Dunod.
- Delion, Pierre. 2018. *Qu'est-ce que la psychothérapie institutionnelle ?* Éditions d'une.
- Easton, Leonore. 2011. *Rapport sur les méthodes utilisées en recherche artistique dans le domaine des arts de la scène*. Unpublished manuscript. Haute école de théâtre de Suisse romande, Lausanne.
- Elkins, James. 2009. "The Three Configurations of Studio-Art PhDs." In S. R. Riley & L. Hunter (Eds.), *Mapping Landscapes for Performance as Research: Scholarly Acts and Creative Cartographies* (pp. 107–113). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244481_13
- Elkins, James (Ed.). 2014. *Artists with PhDs: On the New Doctoral Degree in Studio Art*. Washington: New Academia Publishing.
- Fourmentraux, Jean-Paul. 2011. *Artistes de laboratoire. Recherche et création à l'ère numérique*. Paris: Hermann.
- Giacco, Grazia. 2018. *Recherche-création et didactique de la création artistique. Le chercheur travaillé par la création*. Louvain-la-Neuve : EME.
- Giacco, Grazia, John Didier, Sabine Chatelain, and Frédéric Verry (Eds.). 2020. *Définir la recherche-création. État des lieux et au-delà*. Louvain-la-Neuve: EME.
- Giacco, Grazia. 2021. "Quelles recherches-créations ? Une cartographie entre singularité et pluralité de la recherche artistique dans les institutions de 3e cycle en France". Communication presented at the monthly Research Seminar of the PRISM Laboratory, Marseille, April 8.
- Gosselin, Pierre, and Éric Le Coguiec (Eds.). 2006. *La recherche création: Pour une compréhension de la recherche en pratique artistique*. Québec: Presses de l'Université du Québec.

- Macleod, Katy, and Lin Holdridge (Eds.). 2006. *Thinking through art: Reflections on art as research*. London: Routledge.
- Martinez Thomas, Monique, and Catherine Naugrette (Eds.). 2020. *Le doctorat et la recherche en création*. Paris : L'Harmattan.
- Oury, Jean. 2003. *Psychiatrie et psychothérapie institutionnelle*. Lecques: Champ Social.
- Oury, Jean. 2016. *La psychothérapie institutionnelle de Saint-Alban à La Borde*. Paris: Editions d'une.
- Smith, Hazel, and Roger T. Dean (Eds.). 2009. *Practice-led research, research-led practice in the creative arts*. Edinburgh: Edinburgh University Press.
- Stévance, Sophie, and Serge Lacasse. 2013. *Les enjeux de la recherche-crédation en musique*. Laval: PUL.
- Sullivan, Graham. 2009. *Art Practice as Research: Inquiry In The Visual Arts*. Thousand Oaks: Sage Publishing.
- Tosquelles, François. 2007. *Education et psychothérapie institutionnelle*. Lecques: Matrice.

3.10 Le processus créatif / Phénoménologie du processus créatif

**La relation réalisateur - compositeur : un exemple de
processus créatif relationnel éactif (2017)**

Ivan Magrin-Chagnolleau. **La relation réalisateur - compositeur : un
exemple de processus créatif relationnel éactif.** *p-e-r-f-o-r-m-a-n-
c-e*, 4, 2017.

9 pages

La relation réalisateur - compositeur : un exemple de processus créatif relationnel énonctif

Ivan Magrin-Chagnolleau

Résumé

Cet article explore la relation qui existe entre un réalisateur et un compositeur. Plusieurs exemples sont étudiés, y compris des exemples tirés de l'expérience personnelle de l'auteur. Ces exemples mettent en lumière les différentes façons de travailler sur la composition d'une musique de film, et notamment les différentes façons pour un réalisateur de collaborer avec un compositeur.

Prologue

Je m'intéresse depuis longtemps au processus créatif. Je m'y intéresse en tant qu'artiste, et en tant que chercheur. Je m'interroge à la fois sur mon propre processus créatif, mais aussi sur le processus créatif d'autres artistes. Et j'observe la façon dont artistes et chercheurs parlent du processus créatif. J'ai voulu m'interroger, dans cet article, sur un aspect particulier du processus créatif, l'aspect relationnel. En effet, il est fréquent pour les artistes de travailler avec d'autres artistes sur certains projets. Il se crée alors une émulation et une collaboration qui transforment le processus créatif et le font passer d'un processus créatif individuel à un processus créatif collectif. C'est particulièrement vrai dans le théâtre et le cinéma, qui sont par nature des arts collaboratifs. Ici, j'explore plus particulièrement la relation qui peut exister entre un réalisateur et un compositeur. Je vais essayer de recenser différentes situations, en étudiant ce qu'ont pu en dire différents réalisateurs et compositeurs, et en m'appuyant également sur ma propre expérience de réalisateur et de compositeur.

Satyajit Ray et la musique

L'expérience de Satyajit Ray est une expérience intéressante. Pour ses trois premiers films, ainsi que pour son cinquième, il a fait appel à Ravi Shankar, le grand musicien bengalais, pour composer ses musiques de films. Pour son quatrième, sixième et huitième, il a eu recours à d'autres musiciens. Mais pour tous ses autres films (le septième et du neuvième au trente-sixième), il a composé lui-même la musique. La musique a en effet été le premier médium artistique qu'il a pratiqué, et ce depuis son plus jeune âge. Lorsqu'il commence à composer lui-même ses musiques de film, il a donc de nombreuses années de pratique assidue de la musique derrière lui. Et ses influences sont nombreuses, aussi bien occidentales qu'indiennes. Il a été plusieurs fois interrogé sur le rôle de la musique dans ses films et sur son utilisation de la musique. Voici par exemple ce qu'il dit dans *American Film*¹ en 1978 :

¹ *American Film* 3, no. 9 (July-August 1978), pp. 39-50.

« J'utilise de moins en moins de musique dans mes films, parce que j'ai toujours eu le sentiment que la musique de fond était un élément qui ne faisait pas partie du cinéma pur. [...] Peut-être fait-on cela par manque de confiance dans le public. [...] Je préfère personnellement une approche plus épurée, mais je reconnais qu'on ne peut pas se passer complètement de musique. Dans la trilogie [d'Apu], je n'ai pas écrit ma propre musique ; j'ai travaillé avec Ravi Shankar. Ces films sans musique auraient semblé encore plus lents. »

Et un peu plus loin :

« Le cinéma lui-même est déjà une transposition de la réalité. Le fait de choisir une histoire avec un schéma dramatique suggère déjà une interprétation de la vie. Et donc il n'y a aucune raison de vouloir éliminer tout ce qui peut aider ce langage cinématographique. La musique en fait partie. Elle aide à magnifier les émotions, permet de créer une unité, sert de pont entre des scènes, et peut aussi aider à corriger certains défauts de la bande son. Cependant, la musique n'est certainement pas indispensable, et certains passages d'un film devraient ne pas en avoir. C'est en travaillant sur le mixage de la bande son qu'on se rend compte s'il faut de la musique ou pas. »

Lorsqu'un réalisateur compose lui-même sa musique, comme dans le cas de Satyajit Ray dans la plupart de ses films, nous ne sommes pas dans un processus relationnel entre le réalisateur et le compositeur, mais plutôt dans une recherche permanente de ce qui est juste, de ce qui fait sens. Nul doute que Satyajit Ray n'était pas la seule personne impliquée dans ce processus décisionnel. Il y avait notamment son monteur, Dulal Dutta, qui a monté tous ses films, et sûrement ses différents producteurs. Mais on peut aussi penser que le fait qu'il ait composé lui-même la plupart de ses musiques de film ait peut-être donné une plus grande unité stylistique à ses films.

Le rapport entre François Truffaut et Maurice Jaubert

Maurice Jaubert a composé des musiques de films essentiellement entre la fin des années 20 et 1940, l'année de sa mort. Il a notamment composé la musique de *Nana* de Jean Renoir, de *Quatorze Juillet* et de *Le dernier milliardaire* de René Clair, de *Zéro de conduite* et de *L'Atalante* de Jean Vigo, de *Drôle de drame*, de *Quai des brumes*, de *Hôtel du Nord* et de *Le jour se lève* de Marcel Carné, ou encore de *Carnet de bal* et de *La fin du jour* de Julien Duvivier. Maurice Jaubert a joué un rôle très important dans l'histoire de la musique de film, et c'est en discutant avec Henri Dutilleul peu de temps avant sa mort que j'ai pu m'en rendre compte. Ce dernier, qui a également abordé un peu la musique de film au cours de sa carrière, a rencontré Maurice Jaubert à plusieurs occasions. Il m'a dit que Maurice Jaubert avait été le premier compositeur de musique de film à envisager une autre formation que l'orchestre symphonique comme médium musical pour un film. Cette trouvaille, aujourd'hui, nous paraît bien sûr évidente, mais à

l'époque, elle était révolutionnaire. Je n'ai pas encore pu vérifier précisément si cette affirmation d'Henri Dutilleux était vraie, mais je n'ai pas encore trouvé de contre-exemples. Il est vraisemblable, cependant, que cette idée est venue à plusieurs compositeurs, à peu près au même moment.

François Truffaut était un grand admirateur de la musique de Maurice Jaubert. Il a utilisé les musiques de Maurice Jaubert dans quatre de ses films, dans *L'Histoire d'Adèle H*, dans *L'argent de poche*, dans *L'homme qui aimait les femmes* et dans *La chambre verte*. Nous sommes ici dans le cas de figure où un réalisateur utilise des musiques préexistantes dans ses films, mais où il va puiser dans le répertoire de la musique de film. Il empreinte donc une musique qui a été composée spécifiquement pour un film et la réutilise dans un autre contexte pour son propre film. Mais dans le cas de François Truffaut, il s'agit principalement des musiques d'un même compositeur, Maurice Jaubert. Voici ce que François Truffaut en dit par exemple dans son texte « Jean Vigo est mort à 29 ans »²:

« André Bazin dans un article sur Vigo a eu un mot très heureux en parlant de son « goût presque obscène de la chair » car il est vrai que personne n'a filmé la peau des gens, la chair de l'homme aussi crûment que Vigo. Rien de ce qu'on a montré depuis 30 ans n'a égalé, dans ce domaine précis, cette image de la main grasse du professeur sur la petite main blanche de l'enfant dans *Zéro de conduite* ou des étreintes de Dia Parlo et Jean Dasté lorsqu'ils vont faire l'amour ou, mieux encore, lorsqu'ils se sont quittés et qu'un montage parallèle nous les montre se retournant chacun dans leur lit, lui dans sa péniche, elle dans une chambre d'hôtel, tous deux en proie au mal d'amour, dans une scène où la prodigieuse partition de Maurice Jaubert joue un rôle de première importance, séquence charnelle et lyrique qui constitue très exactement un accouplement à distance. »

Il semble que François Truffaut ait été particulièrement fasciné par l'osmose qui existait entre Jean Vigo et Maurice Jaubert.

Gabriel Yared et sa conception de la musique de film

Gabriel Yared a donné une master classe sur la musique de film à la cinémathèque française en juin 2012. J'ai eu la chance de me trouver parmi les spectateurs, et aussi de pouvoir échanger avec lui à l'issue de cette master classe. Gabriel Yared a écrit la musique pour de nombreux films, parmi lesquels *Sauve qui peut la vie* de Jean-Luc Godard, *Malevil* de Christian de Chalonge, *La lune dans le caniveau*, *37°2 le matin* et *IP5* de Jean-Jacques Beneix, *La diagonale du fou* et *L'instinct de l'ange* de Richard Dembo, *Adieu Bonaparte* de Youssef Chahine, *Zone rouge* de Robert Enrico, *Désordre* d'Olivier Assayas, *Beyond Therapy* et *Vincent and Theo* de Robert Altman, *Camille Claudel* de Bruno Nuytten, *L'amant* et *Deux frères* de Jean-Jacques Annaud, *Le patient anglais*, *Le talentueux monsieur Ripley*, *Cold*

² François Truffaut. *Les films de ma vie*. Champs Arts, Flammarion, 1975.

Mountain et *Breaking and Entering* d'Anthony Minghella, *La vie des autres* de Florian Henckel von Donnersmarck. Gabriel Yared est autodidacte, mis à part une expérience d'auditeur libre dans une classe d'Henri Dutilleux et quelques cours de contrepoint. C'est un compositeur qui préfère travailler à partir d'un scénario, et surtout de conversations avec le réalisateur. Voici par exemple ce qu'il dit dans un entretien réalisé par la Cinémathèque française peu de temps avant sa master classe³ :

« Je ne suis pas vraiment un homme d'images. Pour moi, quelqu'un qui me parle de son film, ou un scénario que je lis, m'inspirent beaucoup plus que des images qui défilent à 24 ou 25 images par seconde, parce que j'ai toujours pensé que l'image, c'était la fin d'un processus, ce n'est pas un point de départ pour un compositeur, en tout cas pour moi. Ce que j'aime, c'est donc lire ou parler, ou les deux. [...] Et puis ces lectures ou ces conversations sont des déclencheurs, pour moi, d'une inspiration que l'image ne pourrait pas me donner. J'ai besoin des images que moi-même je me fais dans la tête. »

Et un peu plus loin :

« J'ai eu la chance d'avoir fait de belles rencontres, la plus belle restant pour moi cette collaboration et la rencontre avec Anthony Minghella parce que là, vraiment, j'avais quelqu'un qui me connaissait, qui aimait beaucoup ce que j'avais fait pour *37°2 le matin*, pour *L'amant*, qui me comprenait et qui me donnait la permission de risquer, [...] de prendre le temps de commettre des erreurs. [...] Chaque fois qu'on travaillait ensemble, on travaillait avant, parce qu'on voulait installer, comme on installe un décor, comme on installe les acteurs, on voulait installer la musique dans son décor dès le départ. [...] Finalement, c'est avec Walter Murch et Anthony Minghella que j'ai appris à faire de la musique de film, plus en profondeur, de voir toujours les choses dans leur ensemble. [...] Je n'aime pas ce mot, musique de film. Moi, j'écris de la musique pour un projet, pour un homme ou une femme que j'ai rencontré, et avec qui j'ai aimé parler. Donc si j'ai aimé l'entendre, je suis sûr que je vais aimer son œuvre. »

On peut voir dans ces deux extraits l'importance, pour Gabriel Yared, de la rencontre avec le réalisateur. Non seulement cette rencontre est primordiale pour déclencher une envie de travailler ensemble, mais c'est aussi très souvent à travers les mots que le réalisateur utilise pour décrire son projet que Gabriel Yared va puiser ses propres images et son inspiration.

³ <http://www.cinematheque.fr/video/538.html>

Une rencontre SRF - SACEM à Cannes

En mai 2013, j'ai pu participer à une rencontre entre réalisateurs de la SRF⁴ et compositeurs de la SACEM⁵. Cette rencontre était organisée à l'initiative de la Quinzaine des réalisateurs. J'y ai participé en tant que réalisateur membre de la SRF, mais j'y ai aussi donné mon point de vue de compositeur. Cette rencontre était intéressante car elle mettait en contact une dizaine de réalisateurs de la SRF aux profils très variés et une dizaine de compositeurs de la SACEM, eux aussi aux profils très variés. Cette discussion cherchait à cibler les modes de collaboration créative entre réalisateurs et compositeurs. Plusieurs expériences très différentes ont été partagées. Ce qui est ressorti de cette discussion peut se résumer en quelques points :

- La plupart des réalisateurs et des compositeurs ont regretté que le compositeur ne soit pas impliqué plus tôt dans le processus créatif. Cela est généralement dû à des questions de budget, mais aussi parfois au fait que le réalisateur ne se rend pas compte de l'importance de cette collaboration avant qu'il en ait vraiment besoin.
- Il a été aussi débattu du fait d'utiliser une musique préexistante vs. une musique originale. Il y a des réalisateurs qui préfèrent utiliser des musiques préexistantes, arguant du fait qu'ils peuvent choisir parmi des milliers de morceaux dont les qualités évidentes ne sont plus à défendre. D'autres réalisateurs, et bien sûr les compositeurs, défendent l'intérêt de composer une musique originale qui colle à la peau du film et à son atmosphère. Je reviendrai sur ce débat un peu plus loin dans l'article.
- Enfin, il y a des compositeurs qui préfèrent travailler à partir du scénario et d'autres qui préfèrent travailler à partir du film monté. L'idéal est sans doute de faire les deux, de pouvoir arriver tôt dans le processus de façon à proposer des idées en amont du tournage, puis de les affiner au moment du montage.

Quelques expériences personnelles en lien avec la musique de film

J'ai pu, au cours de mon expérience de compositeur et de réalisateur, tester plusieurs configurations dont je fais part dans les paragraphes qui suivent.

***Never Forever* : composer pour une réalisatrice**

Sur le court-métrage *Never Forever*, j'ai été contacté par la réalisatrice Emmanuella Aristil pour composer le thème principal du film. Ce contact a eu lieu a posteriori, pendant la post-production, et une fois que le montage image était quasiment fixé. Cela m'a permis de me faire une idée très précise du film et de son atmosphère visuelle, et j'ai pu faire plusieurs propositions de thèmes et de styles à Emmanuella. Elle a ensuite choisi celui qui lui convenait le mieux, et a fait quelques retours sur cette première version pour

⁴ Société des Réalisateurs de Films

⁵ Société des Auteurs, Compositeurs et Éditeurs de Musique

que je puisse continuer le travail. Ma façon de travailler, dans le cadre de ce travail, qui était entièrement bénévole, a été d'utiliser un logiciel me permettant de composer un morceau, puis de l'écouter avec des instruments midi. J'ai d'abord posé une ligne de basse, jouée en pizzicati à la contrebasse, sur laquelle j'ai rajouté un accompagnement harmonique à la guitare, puis une ligne mélodique très libre à la flûte de pan. Une fois que ce choix d'instrumentation, de tempo, et donc d'atmosphère, a été validé par Emmanuela, j'ai pu faire évoluer cette composition initiale, pour en faire finalement une œuvre musicale qui a trouvé sa place dans la bande-son du film. J'ajoute que pour ce projet, faute de moyens, il n'a pas été possible d'enregistrer le morceau avec des musiciens professionnels. J'ai donc eu recours à une synthèse midi de très haute qualité, et qui a donné un résultat très satisfaisant. Bien sûr, le midi ne remplacera jamais une interprétation par des musiciens professionnels, surtout sur un morceau de ce type.

***C'est la fête !* : travailler avec un compositeur**

Sur le court-métrage *C'est la fête !*, j'étais auteur réalisateur. Le concept de ce film était, dès le départ, une narration musicale. J'ai fait très tôt le choix d'utiliser des musiques préexistantes, surtout parce qu'elles me venaient de façon insistante sur plusieurs séquences du film. J'ai néanmoins sollicité un compositeur professionnel, Florian Mazier, à qui j'ai demandé de composer un morceau pour une séquence du film. Dans ce cas précis, j'ai procédé de la façon suivante : j'ai montré à Florian la séquence image montée, et je lui ai fait écouter deux ou trois morceaux dans le ton de ce que je cherchais. Mais je lui ai ensuite donné carte blanche. Ce qui m'intéressait, dans cette relation, était de pouvoir être surpris par sa proposition. Il m'a proposé trois morceaux, de type musique électronique, et l'un d'entre eux a immédiatement obtenu mon suffrage. Je lui ai fait une ou deux suggestions pour enrichir le morceau, et il est immédiatement reparti au travail. Il m'a finalement rendu un morceau qui correspondait parfaitement à ce passage du film, et à l'atmosphère que je voulais créer. Au final, sur ce film, j'ai utilisé à la fois de la musique préexistante et de la musique composée spécialement pour le film.

***La Galette des Rois* : composer pour son propre film**

Sur le court-métrage *La Galette des Rois*, j'étais réalisateur. Bien que n'ayant pas écrit le scénario du film, j'ai participé à son développement avec la scénariste, dès la deuxième version du scénario. J'ai donc eu très tôt des idées musicales pour le film. J'avais envie d'un morceau, au début du film, qui puisse suggérer le monde de l'enfance, avec à la fois une note joyeuse et une note nostalgique. Ma collaboratrice dans la production du film, Blandine Cantineau, m'a suggéré de ré-écouter les *Scènes d'enfants* de Robert Schuman et j'ai choisi assez vite la scène d'enfant intitulée « Am Kamin » (« Devant la cheminée »). J'ai donc demandé à une amie pianiste de l'enregistrer pour le film. Je ne voulais pas de musique pendant le film, mais je voulais conclure le film de nouveau par un morceau de musique, quelque chose d'un peu mélancolique, de pas trop rapide, et qui permette au spectateur d'entrer en résonance avec le thème du film. Assez vite aussi, un morceau que j'avais composé il y a quelques années s'est imposé à moi. C'était essentiellement une mélodie, pour laquelle je voyais une flûte

traversière, et une ligne de basse au violoncelle pour l'accompagner. J'ai retravaillé ce morceau, notamment au niveau de sa structure, et au niveau d'une deuxième ligne mélodique. Puis j'ai confié la partition à deux amis musiciens, que j'ai ensuite dirigés lors d'une séance d'enregistrement. Cette expérience de choix et de composition de musique pour ce court-métrage *La Galette des Rois* était très intéressante. La musique s'est imposée à moi très tôt, dès la phase de développement, et donc bien avant le tournage. Elle a donc eu un impact direct sur l'esthétique du film, puisque je portais en moi, pendant tout le tournage et la post-production, la résonance émotionnelle de cette musique.

Musique existante vs. musique originale

Je voudrais revenir sur une question qui a été soulevée lors de la rencontre SRF - SACEM pendant le festival de Cannes 2013. Est-il préférable d'utiliser de la musique existante ou de faire composer une musique originale ? Cette question n'a bien sûr pas de réponse définitive. Quentin Tarantino, par exemple, aurait dit qu'il préférerait utiliser de la musique préexistante car cela lui permettait de choisir les meilleures musiques possibles, et de les avoir en tête au fur et à mesure de la préparation du film et de son tournage. C'est un point de vue qui est tout à fait défendable. En même temps, sur son film *Les huit salopards*, il a demandé à Ennio Morricone de lui écrire une partition originale. Il est vrai qu'en sollicitant une musique originale auprès d'un compositeur, le réalisateur peut espérer obtenir une musique qui colle vraiment à la peau de son film. En outre, il devient partie prenante du processus de création en dialoguant continuellement avec le compositeur. On se souvient tous de la musique de Nino Rota pour *Le parrain*, ou encore de la musique d'Ennio Morricone pour *Il était une fois en Amérique*. Je pense que les deux approches ont leurs avantages et inconvénients, et il y a des exemples remarquables pour chacune d'entre elles. En outre, il est tout à fait possible de mélanger les deux approches. Ce choix dépend avant tout de la nature du projet et de la personnalité du réalisateur. Mais même dans le cas où un réalisateur choisirait des musiques préexistantes, je pense qu'il est très intéressant de travailler sur le choix de ses musiques avec une autre personne. En effet, cela permet d'avoir d'autres idées innovantes, et de ne pas rester prisonnier de ses propres biais.

Musique avant ou après le tournage

Une autre question sur laquelle je voudrais revenir est la suivante : faut-il composer/choisir la musique avant ou après le tournage ? Je pense qu'il est très intéressant de pouvoir avoir la musique, ou tout au moins une version préliminaire de la musique, avant le tournage. Cela permet en effet de s'imprégner de cette musique, et si elle est particulièrement adaptée au projet, d'enrichir le processus créatif. En outre, la musique peut aider les acteurs dans leur travail émotionnel. Néanmoins, si cette solution est possible et choisie, il sera probablement nécessaire de retravailler en partie la musique en post-production, notamment pour des raisons de rythme et d'équilibre sonore. Si la musique est composée après le tournage, elle sera

bien sûr composée pour s'adapter parfaitement au montage image. C'est ce qui se fait le plus souvent, essentiellement pour des raisons de coût. Mais, à mon avis, cela nuit au processus créatif. Je pense qu'une collaboration entre un réalisateur et un compositeur est beaucoup plus forte quand le compositeur est sollicité très tôt dans le processus créatif. Il y a aussi une pratique en vogue aux États-Unis, et plus particulièrement à Hollywood, qui fait intervenir le compositeur encore plus tard. Le film est d'abord tourné, puis monté, puis mixé avec des musiques préexistantes temporaires, de sorte de pouvoir avoir une version montrable pour les previews (séances privées où les spectateurs sont invités à remplir un questionnaire à l'issue de la projection, questionnaire qui permet éventuellement de retravailler certains aspects du film, dont son montage, et plus particulièrement la fin). Et c'est seulement une fois que le montage est devenu définitif qu'un compositeur est sollicité pour composer des musiques semblables à celles qui ont été choisies initialement. Je pense que ce processus est très dommageable à la créativité d'un compositeur, mais également à celle d'un réalisateur. Mais c'est malheureusement une pratique qui est encore régulièrement utilisée à Hollywood, car elle permet de réduire le coût de fabrication d'un film.

Un processus créatif relationnel

On voit bien, à travers ces différents exemples, qu'il existe de nombreux cas de figure au sujet de la collaboration entre un réalisateur et un compositeur. Mais quelle que soit la situation, il y a toujours une relation entre les deux, d'une façon ou d'une autre. La relation peut être très minimale, le réalisateur et le compositeur pouvant même ne jamais se rencontrer. Ou bien la relation peut être très riche, avec de nombreuses rencontres et discussions. Cette collaboration peut intervenir très tôt dans le processus créatif, y compris avant le tournage, ou au contraire très tard, vers la fin de la post-production. Mais il est très rare qu'il n'y ait pas de rencontre du tout, et que la seule source d'inspiration du compositeur soit le scénario et les images montées. Je voudrais donc proposer le terme de « processus créatif relationnel »⁶ pour décrire le processus qui est à l'œuvre entre un réalisateur et un compositeur. Le réalisateur influence le processus créatif du compositeur en partageant avec ce dernier l'idée qu'il se fait de son film, de l'histoire, de l'ambiance, de la musique, etc. Mais le compositeur peut également influencer le processus créatif du réalisateur, particulièrement par des musiques qu'il a déjà composées préalablement à la collaboration, ou par des musiques qu'il compose pour ce projet spécifique en amont du tournage. Comme tout processus créatif, celui-ci est influencé par l'environnement au sens très large, mais dans ce cas précis, ce qui prédomine sans doute est la relation entre le réalisateur et le compositeur. Plus la relation entre les deux est riche et plus ils partagent la même vision du film, plus il y a de chances que la musique s'intègre parfaitement au film et en devienne un élément narratif important, au même titre que tous les autres. L'exemple opposé serait un film

⁶ En écho au terme « Esthétique relationnelle » proposé par Nicolas Bourriaud dans son livre du même nom (Les Presses du réel, 1998).

où la musique semble simplement être rajoutée à l'image. Le côté relationnel de ce processus créatif me semble être ici primordial.

Épilogue

Nous avons pu visiter différentes configurations de la relation entre un réalisateur et un compositeur. J'ai essayé de mettre en évidence le côté relationnel de ce processus créatif collaboratif. Je me suis focalisé sur la relation privilégiée qu'il y avait entre le réalisateur et le compositeur. Bien évidemment, il y a de nombreuses autres influences. J'ai proposé dans un article⁷ d'appliquer le concept d'énaction, développé par Francisco Varela⁸, à la création artistique et au processus créatif. Le concept d'énaction est une tentative pour réconcilier le point de vue subjectif et le point de vue objectif. Le sujet est en interaction permanente avec son environnement, et les deux s'influencent mutuellement en permanence. Le sujet influe sur l'environnement, et l'environnement influe sur le sujet. Dans ce cadre, la relation entre le réalisateur et le compositeur peut être vue comme une illustration de ce concept d'énaction : le réalisateur a une idée de son film, qu'il propose au compositeur, qui à son tour propose une musique, qui transforme l'idée que le réalisateur se fait de son film, etc. Mais l'influence de l'environnement ne se limite pas à cette simple relation entre le réalisateur et le compositeur. D'autres facteurs multiples vont bien sûr influencer leur processus créatif.

⁷ Ivan Magrin-Chagnollet. L'énaction dans la création artistique : théâtre, cinéma et performance. À paraître.

⁸ Francisco J. Varela, Evan Thomson, Eleanor Rosch, *The Embodied Mind*, The MIT Press, 1991.

Hasard et Création (2021)

Ivan Magrin-Chagnolleau. **Hasard et Création.** Dans *Le hasard, le calcul et la vie*, édité par Thierry Gaudin, Marie-Christine Maurel, Jean-Charles Pomerol. ISTE Éditions, Collection « Systèmes d'information, web et société », pp. 161-175, 2021. ISBN : 978-1-78405-724-4.

15 pages

Hasard et création

Ivan MAGRIN-CHAGNOLLEAU

Aix Marseille Univ, CNRS, PRISM, Marseille, France

Chapman University, Orange, California

1. Introduction

Je voudrais commencer cet article en disant quelques mots sur mon positionnement actuel, qui est en lien direct avec le contenu de cet article. Je me présente en effet comme artiste chercheur^{1,2}, c'est-à-dire que j'ai une activité de création artistique, et en même temps une activité de recherche au CNRS autour de cette activité artistique. Je travaille au sein du paradigme de la création recherche³, ce qui veut dire que mon travail de création artistique et mon travail de recherche sont intimement liés. Mon travail de recherche est en lien avec mon travail de création artistique, et plus généralement en lien avec l'art, l'esthétique, et la philosophie de l'art. Et mon travail de création artistique est également en lien avec mon travail de recherche, en cela que d'une part il représente le terrain de mes recherches, mais aussi qu'il constitue en lui-même une recherche, qu'elle soit de nature esthétique ou bien plus généralement de nature philosophique. Certaines créations permettent même d'exprimer d'une façon plus pertinente qu'un article le résultat de certaines de mes recherches.

J'ai rejoint au 1er janvier 2019 le laboratoire PRISM à Marseille, une nouvelle UMR⁴ (7061) qui s'est créée précisément pour explorer les liens entre art et science, la création recherche étant une des façons d'explorer ce lien. Dans le cadre de ce paradigme de la création recherche, je m'intéresse plus particulièrement à la phénoménologie⁵ du processus créatif, c'est-à-dire au processus créatif en tant qu'expérience vécue. Comment peut-on décrire le processus créatif tel qu'il est vécu et perçu par la personne qui en est l'acteur ?

La question de l'intervention du hasard dans ce processus créatif est donc une question essentielle, et c'est cette question que je vais traiter dans cet article. Je commencerai par discuter le terme « hasard » et un certain nombre de notions qui lui sont liées. Puis je m'intéresserai au terme « création » et à sa

¹ *L'Artiste Chercheur*, premier numéro du journal *p-e-r-f-o-r-m-a-n-c-e.org*, vol. 1, 2014.

² Ivan Magrin-Chagnolleau, « Éloge de l'artiste chercheur », *p-e-r-f-o-r-m-a-n-c-e.org*, vol. 1, 2014.

³ Également appelée recherche création. Mais je préfère pour ma part mettre le mot création en premier car cet ordre de mots traduit à mon sens davantage le mode opératoire de ce paradigme.

⁴ Unité Mixte de Recherche.

⁵ Edmund Husserl, *L'idée de la phénoménologie*, PUF, 1992.

constellation de notions voisines. Je m'interrogerai ensuite sur l'intervention du hasard dans la création artistique. J'aborderai ensuite la question d'un art du moment présent. Et je refermerai cet article par une conclusion.

2. Le Hasard

J'ai pris l'habitude de démarrer l'exploration sémantique d'un mot à partir des définitions que donne le Grand Robert⁶ de ce mot, car c'est un dictionnaire qui nous fait visiter l'histoire du mot, et donc les différentes notions qui y sont rattachées, ainsi que ses différentes occurrences dans la littérature, chaque occurrence apportant un éclairage différent sur le mot. Et dans le cas du mot hasard, nous allons voir que ses acceptions et ses occurrences sont nombreuses.

Tout d'abord, l'étymologie du mot hasard nous renvoie à un mot arabe, qui est parvenu jusqu'à nous par l'intermédiaire d'un mot espagnol. L'origine du mot arabe est controversée et pourrait signifier ou bien « jeu de dés », ou bien « fleur » ou « fleur d'oranger », les dés ayant porté une fleur sur une face.

La définition la plus ancienne (XIII^e siècle) nous dit : « Jeu de dés en usage au moyen âge. » Et une définition de 1538 nous dit : « Jeu de hasard : jeu où le vainqueur est désigné à l'issue d'une série de « coups » produisant aléatoirement des résultats, et où le calcul, l'habileté... n'ont aucune part. »

Donc, dès le début, la notion de hasard est liée au **jeu**. Il y a des jeux de hasard, où tout le monde a les mêmes chances, et où l'habileté de chacun ne joue aucun rôle, et il y a les autres jeux. Nous voyons aussi immédiatement un lien établi entre hasard et **aléatoire**, et par extension **aléa** (les aléas de la vie).

Au XV^e siècle, nous avons la définition suivante : « Cas, événement fortuit ; concours de circonstances inattendu et inexplicable. » Nous avons donc maintenant l'intrusion des notions de **fortuit**, d'**inattendu** et d'**inexplicable**.

Au milieu du XVI^e siècle, nous avons : « Cause fictive de ce qui arrive sans raison apparente ou explicable. » Là encore, nous avons la notion de ce qui est explicable ou de ce qui ne l'est pas, qui est donc **inexplicable**. Il y a aussi une opposition entre ce qui a une raison d'être et **ce qui n'a pas de raison apparente d'être**.

En philosophie et en science, on a au même moment : « Caractère de ce qui arrive en dehors de normes objectives ou subjectives, de ce qui relève des lois de la probabilité et n'est pas délibéré. » Nous avons ici la première mention de la notion de norme, et le hasard serait ce qui est en dehors des normes, donc **a-normal**. On remarque aussi le lien qui est fait entre hasard et **probabilité**. Et enfin la notion de délibéré, le hasard serait donc le **non-délibéré**.

⁶ Alain Rey, *Le Grand Robert de la Langue Française*, version en ligne, Le Robert, 2017.

Nous avons chez Montaigne en 1580 : « Sans direction déterminée ; sans décider du lieu, de la situation. » Ici, c'est plutôt la **non-décision** qui est mise en avant. On pourrait aussi dire la **non-intention**.

Et chez Fénelon en 1695 : « Sans évolution prévue ni prévisible ; sans modalité déterminée. » Ici, c'est l'évolution qui est **imprévisible**, ou la modalité qui est **indéterminée**.

Nous avons aussi la locution adverbiale « Par hasard » qui peut être traduite par « **Accidentellement, fortuitement**. » Ici, c'est la notion d'**accident** qui apparaît.

Je voudrais terminer cette exploration du Grand Robert concernant le mot « hasard » à l'aide de quelques citations qui vont éclairer quelque peu les propos de cet article. Je commencerai par une maxime de La Rochefoucauld⁷ :

Quoique les hommes se flattent de leurs grandes actions, elles ne sont pas souvent les effets d'un grand dessein, mais des effets du hasard. (François de La Rochefoucauld, *Maximes*, 57)

Nous avons ici une opposition du mot hasard avec le mot dessein, c'est-à-dire avec quelque chose qui est planifié. Le hasard serait donc **non planifié**.

Il y a deux autres citations au sujet du mot hasard que je trouve très intéressantes. La première est de Romain Rolland dans *Jean-Christophe*⁸ :

(...) le hasard sait toujours trouver ceux qui savent s'en servir. (Romain Rolland, *Jean-Christophe*)

La seconde est d'Albert Camus dans *La Peste*⁹ :

Sa seule tâche, en vérité, était de donner des occasions à ce hasard qui, trop souvent, ne se dérange que provoqué. (Albert Camus, *La Peste*)

Ces deux citations établissent un lien entre le hasard et la capacité de chacun et de chacune de s'en servir, de le provoquer. **Le hasard aurait donc besoin d'être provoqué.**

Voltaire, dans son *Dictionnaire philosophique*¹⁰, nous ramène quant à lui au fait que ce que nous appelons hasard est **ce que nous ne comprenons pas**, ou ce que nous ne comprenons pas encore :

⁷ François de La Rochefoucauld, *Maximes*, Garnier Flammarion, 1999.

⁸ Romain Rolland, *Jean-Christophe*, Albin Michel, 2007.

⁹ Albert Camus, *La Peste*, Folio, 1972.

¹⁰ Voltaire, *Dictionnaire philosophique*, Folio, 1994.

Ce que nous appelons hasard n'est et ne peut être que la cause ignorée d'un effet connu. (Voltaire, *Dictionnaire philosophique*, Atomes)

Je voudrais continuer cette exploration sémantique et philosophique par deux citations au sujet du mot « hasard », qui insistent sur la question du point de vue, c'est-à-dire qu'**il n'y a pas de hasard sans sujet concerné par les conséquences de ce hasard.**

La première est d'Henri Bergson dans *Les deux sources de la morale et de la religion*¹¹ :

Une énorme tuile, arrachée par le vent, tombe et assomme un passant. Nous disons que c'est un hasard. Le dirions-nous, si la tuile s'était simplement brisée sur le sol ? Peut-être, mais c'est que nous penserions vaguement alors à un homme qui aurait pu se trouver là, ou parce que, pour une raison ou pour une autre, ce point spécial du trottoir nous intéressait particulièrement, de telle sorte que la tuile semble l'avoir choisi pour y tomber. Dans les deux cas, il n'y a de hasard que parce qu'un intérêt humain est en jeu et parce que les choses se sont passées comme si l'homme avait été pris en considération, soit en vue de lui rendre service, soit plutôt avec l'intention de lui nuire (...) Le hasard est donc le mécanisme se comportant comme s'il avait une intention. (Henri Bergson, *Les deux sources de la morale et de la religion*)

La deuxième est de Paul Valéry dans *Variété*¹² :

Tout le reste – tout ce que nous ne pouvons assigner ni à l'homme pensant, ni à cette Puissance génératrice (la nature) – nous l'offrons au « hasard » – ce qui est une invention de mot excellente. Il est très commode de disposer d'un nom qui permette d'exprimer qu'une chose remarquable (par elle-même ou par ses effets immédiats) est amenée tout comme une autre qui ne l'est pas. Mais dire qu'une chose est remarquable, c'est introduire un homme, une personne qui y soit particulièrement sensible, et c'est elle qui fournit tout le remarquable de l'affaire. Que m'importe, si je n'ai point de billet de la loterie, que tel ou tel numéro sorte de l'urne ? (...) Il n'y a point de hasard pour moi dans le tirage (...) (Paul Valéry, *Variété V*)

Enfin, pour terminer cette exploration du mot « hasard » à l'aide du Grand Robert, je ne peux m'empêcher de terminer par une citation de Jacques Monod dans *Le Hasard et la Nécessité*¹³ :

¹¹ Henri Bergson, *Les deux sources de la morale et de la religion*, Garnier Flammarion, 2012.

¹² Paul Valéry, *Variété III, IV et V*, Folio Essais, 2002.

¹³ Jacques Monod, *Le hasard et la nécessité*, Points Essais, 2014.

Nous disons que ces altérations (du « texte génétique ») sont accidentelles, qu'elles ont lieu au hasard. Et puisqu'elles constituent la seule source possible de modifications du texte génétique, seul dépositaire à son tour des structures héréditaires de l'organisme, il s'ensuit nécessairement que le hasard seul est à la source de toute nouveauté, de toute création dans la biosphère. (Jacques Monod, *Le Hasard et la Nécessité*)

Cette citation, même si son contenu a quelque peu été remis en question récemment par l'épigénétique, est intéressante car elle relie le mot hasard au mot accident, et émet l'hypothèse que toute création (dans la biosphère) et toute nouveauté (dans le texte génétique) seraient nécessairement le fruit du hasard.

3. La Création

Explorons maintenant sémantiquement le mot « création ». Là encore, le Grand Robert nous donne plusieurs définitions. La première définition est plutôt théologique : « Action de donner l'existence, de tirer du néant. » Ici, la « création » est directement liée à une **action**. Ce serait par l'agir que l'on crée. En créant, on donnerait **existence**. Avant d'être créée, la chose créée n'existe pas.

Mais il y a aussi cette notion de « **tirer du néant** ». Qu'est-ce que le néant dans cette définition ? Si l'on rapproche les deux éléments de cette définition, on pourrait dire que le néant est le non-existant. Ou bien le néant est le champ de tous les possibles, mais qui n'ont pas encore été agis, c'est-à-dire en définitive choisis.

Pour illustrer plus avant cette définition théologique, le Grand Robert nous donne plusieurs citations intéressantes. La première nous vient de Montesquieu dans *De L'Esprit des Lois*¹⁴ :

La création qui paraît être un acte arbitraire suppose des règles aussi invariables que la fatalité des athées (...) (Montesquieu, *De L'Esprit des Lois*, I, 1)

Montesquieu nous parle ici des **règles de la création**. La création aurait donc des règles, et qui plus est des règles invariables. Il utilise ici par contraste le terme arbitraire, signifiant qui n'aurait donc pas de règles. Le hasard serait-il donc arbitraire ? Il rapproche aussi ces règles et leur invariabilité de ce qu'il appelle la « fatalité des athées ». On est en plein ici au cœur de la question sur le libre arbitre et la prédestination. Quand il y a création, est-ce que cela vient d'un libre arbitre, d'un choix totalement libre du créateur, ou bien est-ce que tout est prédéterminé, et suit des règles invariables ?

¹⁴ Montesquieu, *De L'Esprit des Lois*, Garnier Flammarion, 1993.

Jean-Jacques Rousseau, dans une lettre à M. de Beaumont¹⁵, nous rappelle combien il est difficile de comprendre, d'appréhender **ce passage de rien à quelque chose** :

Or l'idée de création, l'idée sous laquelle on conçoit que, par un simple acte de volonté, rien devient quelque chose, est, de toutes les idées qui ne sont pas clairement contradictoires, la moins compréhensible à l'esprit humain. (Jean-Jacques Rousseau, Lettre à M. de Beaumont, 18 nov. 1762)

Paul Claudel enfin, dans son *Journal*¹⁶, nous éclaire sur la différence entre **créer à partir de rien** et **créer dans la continuation de ce qui a déjà été créé avant** :

Il y a une différence entre le 1^{er} jour de la Création et les autres. En ce premier jour Dieu crée de rien et parmi rien. Dans les autres Dieu crée au milieu de quelque chose. Son activité créatrice est en quelque sorte conditionnée par les choses qu'il a créées déjà. C'est une matière sur laquelle il agit, qu'il provoque. (Paul Claudel, *Journal*, 13 sept. 1922)

Cette question est essentielle par rapport à la question du lien entre hasard et création, en différenciant une création qui est première d'une création qui viendrait poursuivre une œuvre déjà commencée. Cette question se pose à la fois dans la création artistique et dans la recherche.

Si l'on en vient maintenant à une définition plus commune, faisant davantage référence à l'activité de l'homme plus qu'à celle d'une divinité quelle qu'elle soit, on trouve : « **Action** de faire, d'organiser (une chose qui n'existait pas encore). »

Ici encore, quelques citations viennent éclairer cette nouvelle acception du mot « création ». Commençons par une citation d'Eugène Delacroix tirée de son *Journal*¹⁷ :

Il est bien convenu que ce qu'on appelle création dans les grands artistes n'est qu'une manière particulière à chacun de voir, de coordonner et de rendre la nature. (Eugène Delacroix, *Journal*, 1^{er} mars 1859)

Toute création serait ainsi une manière de **partager sa propre vision du monde**. Car il y a dans la citation de Delacroix à la fois l'action de voir, mais aussi celle de coordonner, c'est-à-dire en quelque sorte de **donner du sens**.

¹⁵ Jean-Jacques Rousseau, *Œuvres complètes*, tomes I à V, La Pléiade Gallimard, 1959-1995.

¹⁶ Paul Claudel, *Journal*, tomes I & II, La Pléiade Gallimard, 1968-1969.

¹⁷ Eugène Delacroix, *Journal*, 1822-1863, Plon, 1996.

Edmond Jaloux, quant à lui, nous parle dans *La Chute d'Icare*¹⁸, du lien entre création et agir, mais aussi de **création continue** :

Agir, c'est une création continue. La nature crée sans arrêt des formes qui n'ont aucune valeur pour elles-mêmes, mais l'ensemble de ces créations infinitésimales est la vie. (Edmond Jaloux, *La Chute d'Icare*, p. 217)

Il est question ici de la création de la nature, mais c'est aussi l'idée d'une **création incrémentale** qui est proposée, c'est-à-dire que dans l'art comme dans la recherche, c'est en ajoutant de petites contributions infinitésimales que se construit une œuvre, une nouvelle théorie, etc.

La prochaine citation que je voudrais partager est celle de Max Jacob dans *Conseils à un jeune poète*¹⁹ :

Ce qui sauve l'art c'est l'invention. Il n'y a création que là où il y a invention. Chaque art a ses inventions. (Max Jacob, *Conseils à un jeune poète*)

Dans cette citation, Max Jacob relie création à **invention**. Il y aurait donc création lorsque quelque chose de nouveau est inventé. Ce qui pose bien sûr la question de la nouveauté. Comment définir le nouveau par exemple dans une création incrémentale ? Ou en recherche, quand il n'y a pas rupture de paradigme, mais seulement une contribution infinitésimale par rapport à l'état de l'art ? Et inversement, comment évaluer l'intérêt, la pertinence, l'esthétique de quelque chose qui est vraiment une rupture de paradigme ? Peut-on comparer une rupture de paradigme en art ou en recherche à une mutation génétique ?

Albert Camus, quant à lui, nous livre, dans *Le Mythe de Sisyphe*²⁰, une pensée plus existentielle :

De toutes les écoles de la patience et de la lucidité, la création est la plus efficace. Elle est aussi le bouleversant témoignage de la seule dignité de l'homme : la révolte tenace contre sa condition, la persévérance dans un effort tenu pour stérile. (Albert Camus, *Le Mythe de Sisyphe*)

La création serait donc pour l'homme une façon de lutter contre l'inévitabilité de sa propre mort, une façon de laisser une **trace**, un **témoignage** de son expérience vécue.

¹⁸ Edmond Jaloux, *La Chute d'Icare*, dans *La Revue des Deux Mondes*, 15 novembre 1935.

¹⁹ Max Jacob, *Conseils à un jeune poète*, NRF, 1945.

²⁰ Albert Camus, *Le Mythe de Sisyphe*, Folio Essais, 1985.

Je voudrais terminer par une citation d'André Malraux dans *Les chênes qu'on abat...*²¹ :

La création m'a toujours intéressé plus que la perfection. (André Malraux, *Les chênes qu'on abat...*)

Cette citation nous rappelle qu'il est peut-être bien plus important d'essayer plutôt que de vouloir être parfait, que ce soit en création artistique ou en recherche. Le plus important est donc de créer quelque chose, et non de viser la perfection.

4. Le hasard dans la création artistique

Après avoir tenté de clarifier quelque peu les termes de « hasard » et de « création », je voudrais maintenant aborder le lien entre les deux. Quel lien y a-t-il entre hasard et création ? Comment le hasard intervient-il dans la création ?

Je m'intéresserai dans cette partie principalement à la création artistique, et je m'appuierai essentiellement sur mon expérience, car pour parler de cette intervention du hasard dans la création, je ne vois pas comment aborder cette question autrement que du point de vue phénoménologique²², c'est-à-dire du point de vue de l'expérience vécue. En effet, seule une expérience vécue peut amener à clarifier ce qui est considéré comme hasard dans une création. Nous l'avons vu chez Henri Bergson et Paul Valéry, le hasard est une notion très subjective, qui dépend d'un point de vue, et donc de l'expérience qui a amené ce point de vue.

Il serait intéressant, pour pousser plus loin cette réflexion, d'envisager un protocole expérimental qui viserait, à l'aide par exemple d'entretiens d'explicitation²³ ou d'entretiens phénoménologiques²⁴, à questionner d'autres artistes, mais aussi des chercheurs, sur des expériences vécues de création où le hasard aurait été perçu comme élément constituant, voire déterminant. On pourrait alors parler de phénoménologie du hasard, ou pour le dire autrement sous forme de question : comment le hasard est-il perçu dans toute forme de création du point de vue de l'expérience vécue ?

Mais revenons-en pour l'instant à ma propre expérience vécue, et à ma relation, dans mon travail de création, à cette notion de hasard. Je voudrais pour cela m'appuyer sur plusieurs exemples, qui illustrent chacun d'une façon différente ce que peut recouvrir le mot « hasard » dans un processus de création.

²¹ André Malraux, *Les chênes qu'on abat...*, NRF, 1971.

²² Edmund Husserl, *L'idée de la phénoménologie*, PUF, 1992.

²³ Pierre Vermersch, *L'Entretien d'explicitation*, ESF Éditeur, 2017.

²⁴ Michel Bitbol & Claire Petitmengin, "Neurophenomenology and the elicitation interview", in *The Blackwell Companion to Consciousness* (2nd edition), Wiley & Sons, 2017.

Le premier est un exemple issu de la photographie, et concerne une série de photos intitulée *Muir Woods Spirits*, qui est régulièrement exposée en galeries et qui a fait l'objet de la publication d'un livre²⁵.

Je m'étais rendu avec un ami au parc national de Muir Woods, au nord de San Francisco et, comme très souvent, j'avais avec moi un appareil photo. Une fois dans ce lieu très majestueux, j'ai donc commencé à prendre quelques photos très classiques, à jouer avec des effets de perspective, tant ces arbres étaient gigantesques.

Mais à un moment, j'ai pris conscience que ce genre de photos classiques ne convenait pas à ce lieu, qui dégageait quelque chose de différent. J'étais plutôt dans une sorte de cathédrale naturelle. Et je me suis mis alors à prendre des photos de ce lieu tout en bougeant mon appareil ou son zoom au moment de la prise de la photo. Ensuite, c'est une question d'expérimentation avec le mouvement, la vitesse d'obturation, et l'ouverture du diaphragme.

Mais ce qui m'intéresse ici est le moment précis où j'ai commencé à expérimenter ce mouvement de caméra. C'est un procédé qui ne faisait pas partie, à cette époque, de mes procédés habituels de photographe. Je l'avais sans doute vu chez d'autres photographes, et peut-être même utilisé une ou deux fois sans y prêter vraiment attention, par exemple lorsque je prends des photos « au hasard », sans regarder dans le viseur. Mais dans le cas présent, je ne me souviens pas d'un moment où j'aurais décidé : « Tiens, maintenant, je vais prendre des photos en bougeant la caméra. » Il se peut que la première photo avec mouvement de caméra ait été prise par hasard, c'est-à-dire que j'ai bougé sans le vouloir, et que j'ai ensuite été captivé par le résultat. Ou bien cette décision de bouger l'appareil en prenant la photo a surgi de mon inconscient, du moment présent, d'une intuition, etc. Dans tous les cas, il y a ici intervention de quelque chose qui n'est pas prémédité, qui n'est pas décidé, qui n'a pas été prévu, qui est accidentel. Et si l'on se réfère à la première partie de cet article, on voit bien qu'on est précisément ici dans le champ sémantique du mot « hasard ».

Le deuxième exemple vient également de la photographie, et concerne une série de photos plus récentes intitulées *Mains Paysages*, qui commence tout juste à tourner en galeries et dont la sortie du livre est imminente²⁶.

Je me suis rendu un été au cœur de l'Auvergne pour suivre une formation. J'ai pris pour cela un petit train régional qui emprunte des chemins à travers des forêts et des canyons magnifiques. Ayant été fasciné par les paysages traversés, je résolu de prendre au retour des photos de ces paysages sublimes. Je sors donc mon téléphone portable et prends une première photo. Et j'observe alors sur cette photo qu'en plus du paysage photographié, je

²⁵ Ivan Magrin-Chagnolleau, *Muir Woods Spirits*, Aloha Edition, 2017.

²⁶ Ivan Magrin-Chagnolleau, *Mains Paysages*, Aloha Edition, 2020.

peux également voir par réflexion sur la vitre du train mes mains et mon téléphone.

Tenté d'abord de trouver le moyen d'éviter cette réflexion, je me rends vite compte qu'il y a quelque chose de fascinant à voir une photo et en même temps l'outil qui a permis de la prendre ainsi que les mains tenant l'outil. Et je prends alors la décision de continuer à prendre des photos avec réflexion apparente.

Encore une fois, ce qui m'intéresse particulièrement ici est le moment où je prends la première photo. Je n'ai pas prémédité de photographier ma réflexion. Je prends juste une photo avec l'intention de capturer un magnifique paysage de la façon dont je le vois. C'est ici le fruit du hasard qui a composé pour moi une photo où l'on voit à la fois un paysage, un téléphone portable et mes mains. Bien sûr, ce n'est pas un hasard si cette réflexion apparaît. Elle est le résultat d'un phénomène physique bien précis. Mais le hasard vient du fait que je n'avais pas du tout anticipé ce phénomène physique, je ne l'avais pas prévu, je ne l'avais pas prémédité, décidé. Cela s'est fait à mon insu. Du point de vue de mon expérience vécue, il s'agit de « hasard ». Cela nous ramène, comme dans le premier cas, aux citations d'Henri Bergson et de Paul Valéry, qui stipulent qu'il n'y a pas de hasard sans sujet.

Je voudrais prendre un troisième exemple qui est encore plus surprenant. Vers la fin des années 90, j'étais à New York pour travailler comme chercheur aux Bell Labs, qui sont ensuite devenus les AT&T Labs. J'habitais alors New York, et je suivais des cours d'aquarelle pour débutants à Cooper Union, l'une des grandes écoles d'art de New York. J'ai réalisé six aquarelles de débutant, qui n'étaient pas particulièrement fantastiques, mais qui ont initié une pratique, et surtout une relation très particulière à ce médium que j'affectionne particulièrement. Les aquarelles ont été rangées dans un dossier cartonné, avec quelques feuilles vierges. À mon retour des États-Unis, ce carton a été rangé dans un grenier dans une maison de campagne familiale.

Une quinzaine d'années plus tard, en rangeant ce grenier, je suis retombé sur ce carton. Je l'ai ouvert et j'ai constaté qu'il avait pris l'humidité, et que la plupart de mes aquarelles de débutant avaient été abîmées par cette humidité. Mais j'ai aussi découvert une chose très surprenante. Grâce à l'humidité, par capillarité, et par sédimentation des pigments, une aquarelle absolument magnifique s'était « composée » sur l'une des feuilles vierges, une sorte de paysage très méditatif²⁷. Encore une fois, on peut expliquer très précisément ce qui s'est passé du point de vue de la physique. Mais qui aurait pu préméditer une chose pareille sans l'avoir d'abord expérimentée « par hasard » ? Du point de vue de mon expérience vécue, c'est vraiment le hasard qui a permis la réalisation de cette aquarelle.

Je pourrais citer de nombreux autres exemples du même type, issus de ma pratique artistique, mais aussi de ma pratique de chercheur, mais il me semble

²⁷ Ivan Magrin-Chagnolleau, *Sedimented Landscape #1*, Aquarelle, 1998-2018.

que ces trois exemples permettent déjà de bien appréhender le sujet de cet article.

Ils illustrent en effet les liens étroits qu'entretiennent hasard et création, et soulignent également, à l'instar d'Henri Bergson et de Paul Valéry, la part phénoménologique de cette relation, c'est-à-dire le fait qu'il y a toujours une subjectivité pour interpréter une expérience vécue comme l'intervention du hasard.

5. Un art du moment présent

Je voudrais terminer cet article en parlant d'une expérience que je mène depuis de nombreuses années autour d'un art du moment présent, c'est-à-dire d'un art qui se décline dans l'instant, sans préméditation, sans prévisibilité, donc un art où le hasard joue un rôle important.

Cette expérience a déjà pris plusieurs formes. Elle se décline pour l'instant autour de deux projets phares.

Le premier s'intitule « Autour du lac ». De 2012 à 2017, j'ai habité à proximité d'un lac. J'ai très vite pris l'habitude d'aller m'y promener, et j'ai aussi très vite pris l'habitude, au cours de ces promenades, de faire le tour de ce lac. J'ai pris rapidement la décision que j'allais profiter de ces promenades autour du lac pour explorer le processus créatif. Cela s'est fait par étapes. J'ai d'abord pris note, au cours de mes premières promenades, de toutes les idées qui me venaient pendant que je marchais autour du lac. Puis certaines de ces idées ont conduit à d'autres formes d'expérimentation. J'ai commencé à prendre des photos au cours de ces promenades. Puis j'ai fait des enregistrements audio de tous les sons que je pouvais entendre. J'ai aussi pris des vidéos courtes de plusieurs moments de ces promenades. J'ai alors pris conscience de l'impact d'une marche régulière sur la créativité. C'est aussi à peu près à ce moment-là que j'ai rejoint le Walking Artists Network (WAN)²⁸, un réseau d'artistes chercheurs qui explorent ce lien entre marche et créativité, et qui font de la marche l'une des pratiques de leur création et de leur recherche.

Au bout de ces cinq années, j'ai amassé une somme de matériaux considérable, une sorte de pâte à laquelle il fallait maintenant que je donne une forme. Le matériau lui-même est déjà une sorte d'œuvre. Mais c'est aussi la mise en forme qui donne du sens à cette œuvre. J'ai donc décidé de produire un certain nombre « d'objets » à partir de ce matériau. Tout d'abord, un livre rassemblant les textes qui ont été produits pendant ces marches, ainsi que quelques photos. Puis un livre contenant, lui, essentiellement des photos. Ces deux livres sont en cours de fabrication et seront sans doute publiés courant 2021. Je voudrais aussi faire un site internet où je pourrais rassembler pêle-mêle une partie des textes, photos, fichiers son et vidéo en un même

²⁸ <https://www.walkingartistsnetwork.org>

endroit. Je voudrais enfin créer une exposition immersive où ces différents matériaux seraient disposés en un tout cohérent, et où le spectateur de l'exposition déambulerait à travers tout cela un peu à la façon où j'ai moi-même déambulé autour de ce lac.

Ce qui est fascinant dans ce genre d'expérience, au-delà de toutes les intuitions profondes qui surgissent pendant des marches régulières, est le rôle prépondérant que joue le hasard dans la fabrication dans l'instant des objets artistiques. Même si, on l'a déjà vu, il faut un sujet pour qu'il y ait hasard, et même si on peut inférer que les pensées qui surgissent, les envies de photos, d'enregistrements de son ou de fichiers vidéo, sont déclenchées par la volition, c'est ce qui se passe à la suite de ce déclenchement qui appartient au hasard.

Par exemple, si je décide, sans toujours savoir pourquoi d'ailleurs, que je vais commencer à enregistrer des sons, ce qui va se passer entre le déclenchement de l'enregistreur et le moment où je l'arrête m'échappe complètement. Même si c'est peut-être le chant d'un oiseau qui a été le moteur de ce déclenchement, je n'ai aucune idée de ce qui va se passer ensuite. Je vais peut-être croiser des enfants qui jouent, ou bien une femme promenant son chien, ou bien des canards qui vont s'envoler à mon approche.

C'est la même chose avec la vidéo. J'ai peut-être une bonne raison pour commencer à filmer, et souvent la bonne raison est juste que j'y pense à ce moment-là, mais même si je filme à partir de là où je suis, et souvent le chemin juste devant moi, ce qui se passe à partir du début de l'enregistrement vidéo m'échappe complètement. Je vais peut-être devoir faire un détour pour éviter un obstacle inhabituel, ou bien je vais être capté par quelque chose qui va me faire dévier de mon chemin.

Le deuxième projet s'appelle « The Haiku Project ». J'écris des haïkus depuis longtemps, mais j'ai eu l'idée de créer un projet entier autour de l'idée du haïku. Je voulais explorer d'autres formes de haïku que le poème, et notamment le haïku photo, le haïku vidéo, et le haïku sonore ou musical. Pour cela, j'ai choisi d'alléger quelque peu les contraintes initiales du haïku et de retenir deux aspects : une forme la plus simple, la plus épurée possible, et l'idée d'un art du moment présent, non prémédité, qui se fabrique à partir de ce qui est ici et maintenant.

Ici encore s'est posée la question des objets fabriqués à partir de tout le matériau accumulé au fil du temps. Comme pour le projet précédent, j'ai choisi de commencer par ce qui me semblait le plus simple à faire, un livre de haïkus poétiques qui contient quelques haïkus photographiques, et un livre contenant essentiellement des haïkus photographiques. Ces deux livres sont également en cours de fabrication et seront sans doute publiés en 2020 ou en 2021. J'ai également l'intention de créer un site internet qui puisse contenir une grande partie du matériau amassé dans le cadre de ce projet. Et je voudrais également proposer une exposition immersive qui puisse non

seulement permettre au spectateur de déambuler au milieu de tout ce matériau, mais aussi d'expérimenter lui-même avec le procédé du haïku sous différentes formes, avec la création dans l'instant.

Ici encore, le hasard joue un rôle considérable dans le processus créatif. Il y a bien sûr toujours cette question du déclencheur. Qu'est-ce qui me pousse à choisir tel moment pour fabriquer un haïku, et qu'est-ce qui me pousse à choisir telle forme de haïku plutôt qu'une autre ? C'est en cela que cette pratique artistique est très phénoménologique, car on peut supposer que c'est quelque chose qui se passe au niveau de l'expérience vécue qui déclenche cette impulsion de faire un haïku et le choix du mode.

Mais depuis quelque temps, j'envisage de pousser cette notion de hasard encore plus loin en mettant en place un petit système qui choisirait aléatoirement les moments des haïkus ainsi que leur type. Il serait alors intéressant de comparer les deux formes de pratique. Je ne suis pas certain néanmoins qu'il soit si intéressant que ça de déconnecter la pratique du haïku de l'expérience phénoménologique qui y est reliée. Encore que dans ce deuxième cas, il y aurait toujours une expérience vécue, mais elle serait d'une autre nature.

6. Conclusion

J'ai donc proposé dans cet article une exploration sémantique du mot « hasard » et du mot « création ». Puis j'ai essayé de montrer, à partir d'exemples issus de ma propre pratique artistique, quelle était la part de hasard dans la création artistique, et à quel point le hasard était ou non un élément constitutif du processus créatif. J'ai ensuite présenté deux expériences artistiques qui explorent un art du moment présent, c'est-à-dire un art où, par définition, le hasard joue un rôle important.

En clôturant cet article, il me vient une idée pour aller plus loin dans cette exploration du lien entre hasard et création, idée que j'ai déjà mentionnée plus tôt. Je mène depuis quelque temps des expériences autour de la phénoménologie²⁹ du processus créatif³⁰. Je fais cela à partir d'entretiens d'explicitation³¹ et d'entretiens phénoménologiques³² que je mène avec plusieurs artistes autour de leur processus créatif. Au cours de ces entretiens, je propose aux artistes de choisir un moment de leur processus créatif, en lien avec une création récente, qui leur paraît particulièrement intéressant. Mais je pourrais aussi leur demander de choisir un moment où il leur semble que le

²⁹ Edmund Husserl, *L'idée de la phénoménologie*, PUF, 1992.

³⁰ Ivan Magrin-Chagnolleau, Ruth Nahoum, Marcus Weisen, "Le projet sipapu : vers une micro-phénoménologie de la création", dans *Les enjeux cognitifs de l'artefact esthétique*, tome 2, sous la direction de Xavier Lambert, L'Harmattan, 2019.

³¹ Pierre Vermersch, *L'Entretien d'explicitation*, ESF Éditeur, 2017.

³² Michel Bitbol & Claire Petitmengin, "Neurophenomenology and the elicitation interview", in *The Blackwell Companion to Consciousness* (2nd edition), Wiley & Sons, 2017.

hasard a joué un rôle important. Cela permettrait notamment de caractériser les différentes notions que différents artistes mettent derrière le mot hasard. C'est en tout cas une piste qui me semble intéressante à explorer.

7. Bibliographie

- L'Artiste Chercheur*, premier numéro du journal *p-e-r-f-o-r-m-a-n-c-e.org*, Vol. 1, 2014.
- Bergson, Henri, *Les deux sources de la morale et de la religion*, Garnier Flammarion, 2012.
- Bitbol, Michel & Petitmengin, Claire, "Neurophenomenology and the elicitation interview", in *The Blackwell Companion to Consciousness* (2nd edition), Wiley & Sons, 2017.
- Camus, Albert, *La Peste*, Folio, 1972.
- Camus, Albert, *Le Mythe de Sisyphe*, Folio Essais, 1885.
- Claudé, Paul, *Journal*, tomes I & II, La Pléiade Gallimard, 1968–1969.
- Delacroix, Eugène, *Journal, 1822-1863*, Plon, 1996.
- Husserl, Edmund, *L'idée de la phénoménologie*, PUF, 1992.
- Jacob, Max, *Conseils à un jeune poète*, NRF, 1945.
- Jaloux, Edmond, *La Chute d'Icare*, dans *La Revue des Deux Mondes*, 15 novembre 1935.
- La Rochefoucauld, François de, *Maximes*, Garnier Flammarion, 1999.
- Magrin-Chagnolleau, Ivan, *Sedimented Landscape #1*, Aquarelle, 1998-2018.
- Magrin-Chagnolleau, Ivan, « Éloge de l'artiste chercheur », *p-e-r-f-o-r-m-a-n-c-e.org*, Vol. 1, 2014.
- Magrin-Chagnolleau, Ivan, *Muir Woods Spirits*, Aloha Edition, 2017.
- Magrin-Chagnolleau, Ivan, Nahoum, Ruth, Weisen, Marcus, "Le projet sipapu : vers une micro-phénoménologie de la création", dans *Les enjeux cognitifs de l'artefact esthétique, tome 2*, sous la direction de Xavier Lambert, L'Harmattan, 2019.
- Magrin-Chagnolleau, Ivan, *Mains Paysages*, Aloha Edition, 2020.
- Malraux, André, *Les chênes qu'on abat...*, NRF, 1971.
- Monod, Jacques, *Le hasard et la nécessité*, Points Essais, 2014.
- Montesquieu, *De L'Esprit des Lois*, Garnier Flammarion, 1993.
- Rey, Alain, *Le Grand Robert de la Langue Française*, version en ligne, Le Robert, 2017.
- Rolland, Romain, *Jean-Christophe*, Albin Michel, 2007.
- Rousseau, Jean-Jacques, *Œuvres complètes*, tomes I à V, La Pléiade Gallimard, 1959-1995.
- Valéry, Paul, *Variété III, IV et V*, Folio Essais, 2002.
- Vermersch, Pierre, *L'Entretien d'explicitation*, ESF Éditeur, 2017.
- Voltaire, *Dictionnaire philosophique*, Folio, 1994.
- Walking Artists Network (WAN) : <https://www.walkingartistsnetwork.org>.

Le projet sipapu : vers une micro-phénoménologie de la création (2019)

Ivan Magrin-Chagnollet, Ruth Nahoum, Marcus Weisen. **Le projet sipapu : vers une micro-phénoménologie de la création.** Dans *Les enjeux cognitifs de l'artefact esthétique : Tome 2*, édité par Xavier Lambert. L'Harmattan, Collection « Ouverture Philosophique », pp. 187-194, 2019. ISBN : 978-2-343-18776-1.

6 pages

LE PROJET SIPAPU : VERS UNE MICRO-PHÉNOMÉNOLOGIE DE LA CRÉATION

**Ivan MAGRIN-CHAGNOLLEAU, Ruth NAHOUM,
Marcus WEISEN**

Introduction :

L'inconscient et le conscient collaborent dans le processus de création. Nombre d'artistes et de penseurs affirment la primauté de l'expérience pré-réfléchie dans la création et dans l'expérience esthétique.

Le projet SIPAPU propose d'éclairer la complexité du processus créatif et ses dimensions pré-réfléchies. Il se propose d'observer le processus créatif d'un point-de-vue phénoménologique, c'est-à-dire au niveau de l'expérience vécue. Il utilise la méthodologie de l'entretien micro-phénoménologique.

Contexte du projet :

La micro-phénoménologie est une nouvelle discipline scientifique qui permet d'explorer très finement notre expérience vécue : que se passe-t-il pour nous lorsque nous vient une idée ? Lorsque tout à coup nous revient un souvenir d'enfance ? Lorsque, lors du processus créatif, une clarté monte à la conscience depuis « les couches les plus souterraines de la pensée sensorielle ? » (Morin 2016)

Des recherches récentes montrent que nous pouvons apprendre à décrire notre expérience vécue de manière très précise et fiable, et en découvrir des dimensions ordinairement inaccessibles. L'entretien micro-phénoménologique est un choix judicieux pour l'étude des processus de création, car il

permet d'amener à la conscience des expériences fugitives pré-réfléchies qui jouent un rôle essentiel dans la création.

L'entretien micro-phénoménologique est inspiré de l'entretien d'explicitation, méthode d'entretien initialement développée par Pierre Vermersch à des fins pédagogiques et d'analyse de pratiques professionnelles. Il a été adapté à la recherche sur l'expérience subjective en sciences cognitives et complété par une méthode d'analyse et de validation des données, pour devenir la micro-phénoménologie.

L'entretien d'explicitation est une technique qui permet d'accéder à la part pré-réfléchie de l'expérience et de la décrire avec un grand degré de précision (Vermersch 2012, 2014).

Les principaux procédés d'explicitation consistent à amener le sujet : 1) à choisir et à explorer une expérience singulière, située dans le temps et dans l'espace ; 2) à évoquer cette expérience passée en l'aidant à retrouver le contexte spatio-temporel et sensoriel de l'expérience, jusqu'à ce que la situation passée soit « rejouée », au point d'être vécue comme plus présente que la situation d'entretien ; 3) à réorienter son attention, au sein de cette expérience évoquée, du « quoi » ou contenu de l'expérience vers le « comment » : par exemple les micro-ajustements du mode d'attention, le ressenti corporel, l'émergence d'émotions et de pensées.

Le travail d'analyse et de modélisation du corpus de descriptions permettra d'identifier la structure de chaque expérience décrite, puis de comparer ces structures afin de détecter une éventuelle structure générique des expériences étudiées.

Cette démarche de recueil et de modélisation de descriptions a déjà reçu deux types de validation, empirique et épistémologique. D'une part, elle a déjà été utilisée pour mettre en évidence des structures dynamiques auparavant inaperçues dans plusieurs domaines : la microgenèse d'une idée scientifique nouvelle (Remillieux 2014), l'émergence d'une intuition

(Petitmengin 2003), l'émergence d'une illusion perceptive (Valenzuela-Moguillansky et al. 2013).

D'autre part, cette démarche a fait l'objet d'un travail de validation épistémologique, qui s'inscrit dans un programme de recherche plus large visant à intégrer l'étude de l'expérience vécue dans le cadre des méthodes d'investigation scientifique de la connaissance (Petitmengin et Bitbol 2009, Bitbol 2014).

Description du projet :

Dans ce projet, nous nous intéressons à la création artistique comme expérience vécue. Nous voudrions en particulier découvrir / approfondir / préciser / éclairer quelles sont les différentes étapes d'un processus de création telles qu'elles sont vécues par ceux qui les vivent. Grâce à l'entretien micro-phénoménologique, nous voudrions identifier des structures dans ce processus créatif, et détecter éventuellement une structure plus générique commune aux différentes expériences vécues décrites par les différents artistes participants au projet.

Objectifs du projet :

Établir une typologie de l'acte créatif, c'est-à-dire expliciter, grâce à l'entretien micro-phénoménologique, des structures génériques de l'acte créatif.

Étapes du projet :

Le projet se décompose en trois étapes complémentaires :

- 1) La première étape consiste à étudier le processus créatif au cours de séances de travail créatif à partir d'une tâche simple et bien identifiée ;
- 2) La deuxième étape consiste à étudier le processus créatif au cours d'une création artistique qui s'étale dans le temps ;

3) La troisième partie consiste à étudier l'expérience esthétique de spectateurs ayant visité une exposition constituée à partir des productions des deux premières étapes.

Déroulé du projet :

- ***Première étape*** : sur environ une année, nous rassemblerons un petit groupe d'artistes (5 à 6) travaillant avec des média différents, à intervalles réguliers (environ tous les deux mois), pour des séances d'environ deux heures (5 à 6 séances). Au cours de chaque séance, l'un des artistes proposera un point de départ (par exemple, travailler à partir d'une matière donnée pour explorer la question de ce que cela signifie pour un artiste de **rencontrer une nouvelle matière** et de **trouver sa « pâte »**) à partir duquel tous les artistes présents exploreront une séquence de création possible. À l'issue de ces deux heures, des entretiens d'explicitation seront conduits avec chaque artiste pour expliciter cette séquence de création. Ces entretiens seront filmés, puis étudiés et analysés dans un second temps.

- ***Deuxième étape*** : sur environ une année, nous proposerons à un petit groupe d'artistes (certains ayant participé à la première étape et d'autres non) de conduire sur cette année un travail de création qui leur est propre, et qui aura un cadre commun issu des pratiques de la première année. Au cours de cette année de création, des entretiens d'explicitation seront menés à intervalles réguliers, et aussi à la demande des artistes, lorsqu'ils auront eu l'impression d'avoir vécu une étape importante pour eux dans leur processus de création. Ces entretiens seront filmés, puis étudiés et analysés dans un second temps.

- ***Troisième étape*** : une exposition sera organisée à partir des productions des deux premières années, accompagnées de certains extraits et analyses des entretiens d'explicitation

menés en parallèle de ces deux premières années, présentés par chaque artiste sous la forme choisie (dispositif scénique, installation, etc.). À partir de cette exposition (qui pourra être itinérante), des entretiens d'explicitation seront conduits avec des spectateurs volontaires afin d'explicitier leur expérience esthétique, c'est-à-dire comment ils ont vécu cette exposition et ce que ça a éveillé en eux.

Délivrables du projet :

- une **exposition** présentant les œuvres et les résultats de recherche ;
- des **entretiens** d'explicitation, étudiés et analysés ;
- un **catalogue** d'exposition avec des contributions artistiques et de recherche ;
- des présentations régulières des avancées du projet à différents **colloques** ;
- des **articles** de recherche.

Conclusion :

Nous avons présenté le Projet SIPAPU, projet de recherche innovant qui s'est donné comme principal objectif d'étudier le processus créatif d'un point de vue phénoménologique, c'est-à-dire au niveau de l'expérience vécue, en recourant à la méthodologie de l'entretien micro-phénoménologique, extension de l'entretien d'explicitation.

Ce projet, d'une durée de trois ans environ, a pour but de faire ressortir des structures génériques réfléchies, mais surtout pré-réfléchies, de l'acte créatif.

Bibliographie :

Bitbol, M. (2014). La conscience a-t-elle une origine ? Des neurosciences à la pleine conscience : une nouvelle approche de l'esprit, Flammarion.

- Depraz, N., F. J. Varela and P. Vermersch (2011). À l'épreuve de l'expérience, Zeta Books.
- Morin, E. (2016). Sur l'esthétique, Robert Laffont.
- Pallasmaa, J. (1996). *The Eyes of the Skin: Architecture and the Senses*, London Academy Editions.
- Pallasmaa, J. (2016). *Six Australian Lectures: Lecture 1- Atmosphere and Mood: Feeling Spaces and Places*.
- Petitmengin, C. (2003). L'expérience intuitive, L'Harmattan.
- Petitmengin, C. (2007). "Towards the source of thought." Journal of Consciousness Studies **14**(3): 54-82.
- Petitmengin, C., and M. Bitbol (2009). "The validity of first-person descriptions as authenticity and coherence." *Journal of Consciousness Studies* 16 (10-12):363-404.
- Remillieux, A. (2014). "Les coulisses d'une invention. ne description expérientielle du processus d'invention technique." *Intellectica* 61:273-310.
- Valenzuela Moguillansky, C., J. K. O'Regan, and C. Petitmengin (2013). "Exploring the subjective experience of the "rubber hand" illusion." *Frontiers in Human Neuroscience*.
- Varela, F. J. (2017). *Le cercle créateur - Écrits (1976-2001)*, Seuil.
- Varela, F. J. and J. Shear (1999). *The View From Within: First-Person Approaches to the Study of Consciousness*, Imprint Academic.
- Varela, F. J., E. Thomson and E. Rosch (1991). The Embodied Mind, The MIT Press.
- Vermersch, P. (2007). "Bases de l'auto-explicitation." Expliciter(69).
- Vermersch, P. (2012). *Explicitation et phénoménologie*, PUF.
- Vermersch, P. (2014). *L'entretien d'explicitation*, ESF Éditeur.

3.11 Énaction

Du chaos à la création : une expérience de pédagogie énaactive pour expérimenter le processus créatif en performance (2015)

Ivan Magrin-Chagnolleau. **Du chaos à la création : une expérience de pédagogie énaactive pour expérimenter le processus créatif en performance.** *p-e-r-f-o-r-m-a-n-c-e*, 2, 2015.

8 pages

Du chaos à la création : une expérience de pédagogie énactive pour expérimenter le processus créatif en performance

Ivan MAGRIN-CHAGNOLLEAU

Résumé

Dans cet article, je décris une expérience pédagogique originale que j'ai menée avec des étudiants de l'université d'Évry au cours de l'année universitaire 2009-2010. J'ai proposé à des étudiants en « administration du spectacle » de créer de toutes pièces une performance pour la fin de l'année. L'objectif de cette proposition pédagogique était de les mettre au contact par l'expérience avec ce qu'est un processus créatif. Je voulais en particulier qu'ils fassent l'expérience des différentes étapes d'un processus de création, y compris les phases où il y a un degré important d'incertitude et d'apparente inefficacité. C'est seulement au prix de l'acceptation de cet inconnu et de ce chaos que peut émerger une œuvre artistique originale et intéressante. C'est cela que je voulais qu'ils découvrent par l'expérience.

Mots clés

Pédagogie, énaction, pédagogie énactive, pédagogie innovante, création, processus créatif.

1 Contexte

En septembre 2009, le département « Arts & Musique » de l'université d'Évry m'a donné carte blanche pour animer un atelier artistique pour les étudiants de leur filière « administration du spectacle ». L'atelier devait s'adresser à des étudiants allant de la première année de licence (L1) à la deuxième année de master (M2). Les seules contraintes que j'avais étaient des contraintes horaires : cinq fois 3 h au premier semestre, et cinq fois 3 h au second semestre. La plupart des ateliers artistiques qui étaient proposés aux étudiants étaient des ateliers de pratique artistique, par exemple pratique théâtrale, chorégraphique ou musicale.

2 Choix d'un processus pédagogique

Il me semblait qu'il était possible d'offrir à des étudiants en université une expérience plus enrichissante que le seul fait de les mettre en contact avec une

pratique artistique. Ces étudiants allaient, pour une partie d'entre eux tout au moins, côtoyer des artistes pendant leur vie professionnelle. Il me semblait donc important qu'ils puissent toucher du doigt ce qu'est un processus créatif, et de le faire de façon expérimentale plutôt que théorique. J'ai donc proposé aux étudiants de créer de toutes pièces une performance pour la fin de l'année. Ce que j'appelle performance dans ce contexte est la réalisation d'un spectacle, créé de toutes pièces, et au cours duquel tous les modes d'expression étaient a priori possibles. Ils pouvaient aussi y avoir des parties plus cadrées que d'autres, et des parties beaucoup plus improvisées. J'ai appelé cet atelier « atelier de création scénique ». J'avais décidé, dès le début, que mon rôle serait le plus minimaliste possible, c'est-à-dire essentiellement celui d'un facilitateur, d'un médiateur. Je voulais en effet que les étudiants fassent l'expérience de ces phases d'incertitude, et de la façon de les dépasser, notamment par des prises de décision. Il s'agissait donc principalement de les mettre en face de résolutions de problèmes, et d'essayer d'intervenir le moins possible dans ces résolutions de problèmes. Ce n'est pas une expérience que les étudiants font couramment dans le système universitaire français, où la grande majorité des décisions sont prises pour eux par le corps enseignant. Or, dans tout processus créatif, il y a un moment où l'artiste est obligé de prendre la responsabilité de ses décisions, et où personne ne peut l'aider et le faire à sa place. Il me semble que c'est là l'une des caractéristiques cruciales de ce qu'est un processus créatif, et aussi précisément la valeur qu'un tel processus peut avoir pour des étudiants en université. J'entends ici par processus créatif le processus qui va permettre à un groupe d'étudiants de créer de toutes pièces pour la fin de l'année une performance publique, et ce sans avoir la moindre idée au départ de ce que va contenir finalement cette performance.

3 Description du processus pédagogique

Dès la première rencontre, j'ai annoncé l'objectif final aux étudiants. Ils devaient, entre maintenant et la fin de l'année, mettre en place une performance, de la durée qu'ils souhaitaient, et qui devait impliquer d'une façon ou d'une autre toute personne du groupe. Ils avaient toute liberté quant au contenu de cette performance, et quant au rôle de chacun. Je m'engageais, de mon côté, à leur trouver un espace de performance pour la fin de l'année, et leur décrivais ce que serait mon rôle dans ce processus, c'est-à-dire essentiellement un rôle de facilitateur, de médiateur. Je leur dis d'emblée que je ne prendrais aucune décision, mais que j'exprimerais de temps à autre mon point de vue, dont ils seraient libres de tenir compte ou non.

Il a fallu également aborder dès le début la question de l'évaluation, puisque le système universitaire impose à toute activité cette évaluation. Pour évacuer cette préoccupation de l'expérience pédagogique que je voulais tenter avec eux, et pour être le plus libre possible vis-à-vis de cette contrainte universitaire, je leur dis qu'ils seraient évalués en fonction de plusieurs critères : leur implication dans l'atelier, leur interaction avec le groupe, et un compte rendu d'une page qu'ils me feraient tous après la performance. Une fois cette question de

l'évaluation réglée, nous pouvions plonger dans le cœur même de cette expérience pédagogique. J'ajoute, et c'est un point qui est très important à mes yeux, que je n'avais aucune idée de ce qui allait se passer, ni de ce que serait véritablement mon rôle dans cette expérience. Mais il me semblait que c'était là la condition même pour qu'une expérience de ce type puisse éventuellement porter ses fruits. Il me semble également que pour pouvoir innover, en pédagogie comme ailleurs, il est nécessaire d'aller vers l'inconnu et, ce faisant, de prendre un risque. L'expérience pédagogique que je tentais avec les étudiants pouvait tout à fait ne mener à rien d'intéressant, mais c'est un risque que j'étais prêt à prendre, la contrepartie positive étant de pouvoir le cas échéant découvrir des choses intéressantes sur la pédagogie d'une part, sur le processus créatif d'autre part, et enfin sur le caractère énonctif d'une expérience pédagogique innovante.

4 Première phase : accumulation de matériel

La première rencontre a consisté principalement à faire connaissance, et à expliquer aux étudiants les contours de l'atelier. Je leur ai ensuite proposé de faire un tour de table où chacune et chacun exprimerait ce qui lui vient à l'esprit en réaction à la proposition que je venais de leur faire. En particulier, je proposais à chaque personne d'imaginer, à chaud, quel pourrait être son rôle dans cette aventure. J'ai ajouté que rien n'était figé, et que tout cela pouvait changer au cours du processus. Il y a donc eu des propositions très diverses : écrire, jouer (en tant qu'acteur ou musicien ou même artiste de cirque), gérer des aspects techniques et organisationnels pour cette performance, etc. Le groupe était composé d'une vingtaine de personnes, et il y avait parmi elles quelques étudiants qui avaient une pratique artistique régulière (danse, cirque, théâtre, musique, etc.). Cela a bien sûr facilité tout le processus. Je leur proposais, afin de faire avancer les choses, d'apporter pour la prochaine rencontre un élément qui pourrait peut-être éventuellement trouver sa place d'une façon ou d'une autre dans la performance finale (par exemple un texte qu'ils avaient écrit ou qu'ils aimaient bien, une image qu'ils avaient créée ou qui leur parlait plus particulièrement, une chanson ou une musique, etc.).

Cette première phase de collecte de matériel a duré plusieurs séances. Elle a permis de voir émerger des envies, des rêves enfouis, des peurs et des inquiétudes, et tous types de réactions par rapport à l'incertitude. Ils se sont tournés plusieurs fois vers moi pour me demander ce que devait être la suite du processus. Je suis bien sûr resté silencieux, car tout l'intérêt de l'expérience était précisément pour eux de trouver eux-mêmes la suite de l'expérience. Ils ont passé plusieurs rencontres à se demander comment ils devaient procéder, ou à attendre de moi des réponses que je ne leur donnais pas. Je voulais qu'ils fassent l'expérience de ne pas savoir, de ne pas réussir à se décider. Et il était important, pour moi, de les plonger au cœur de cet état d'indécision, qui est propre au processus créatif. Il m'arrivait de suggérer une chose ou deux, afin de mettre leurs réflexions en mouvement, ou de susciter des idées de leur part. Mais mon intervention s'arrêtait là. Par exemple, le fait de leur demander ce

qu'ils avaient spontanément envie de faire au cours de cette performance leur a permis de constater que les idées ne manquaient pas, et qu'a priori chaque personne pouvait trouver sa place dans le processus. D'autre part, le fait de leur avoir proposé d'apporter chacune et chacun quelque chose qui pourrait éventuellement trouver sa place dans la performance finale leur a permis aussi d'avoir à leur disposition un certain nombre d'éléments concrets à partir desquels la suite du travail pourrait éventuellement se construire.

5 Deuxième phase : une structure émerge

Il a fallu plusieurs séances pour que les étudiants comprennent que les réponses n'allaient pas venir de moi. C'est à ce moment-là qu'ils ont commencé à se responsabiliser, à s'organiser, à se prendre en main. C'est au même moment qu'une structure pour la performance finale a commencé à émerger. La première structure qui est apparue a été une répartition des rôles au sein du groupe. Il a fallu déterminer qui serait sur scène, et qui ne voulait pas y être. Pour celles et ceux qui ne voulaient pas être sur scène, il a fallu trouver un autre rôle possible. Finalement, il a été décidé de confier l'organisation des activités à deux personnes, qui devaient se tenir informées de ce que chaque groupe faisait et de son état d'avancement. Plusieurs sous-groupes ont émergé à leur tour, en fonction des envies diverses. Par exemple, un groupe tenait à jouer un texte théâtral écrit par l'un d'eux, et voulait donc travailler sur ce texte ensemble. Un autre groupe a voulu articuler cirque et musique, et voulait aussi répéter ensemble. Plusieurs noyaux se sont ainsi constitués, et ils représentaient vraisemblablement de futurs segments de la performance finale. Il n'y avait pas encore de fil conducteur, mais un certain nombre de choses se mettaient déjà en place.

Les étudiants ont aussi pris conscience que ces répétitions par petits groupes devaient se faire en dehors des heures imparties à cet atelier. En effet, il n'y avait pas assez d'heures pour pouvoir inclure en plus des répétitions. Il a donc été convenu que les séances de cet atelier seraient essentiellement consacrées à la mise en place de l'ensemble, et non à la répétition des segments. Chaque petit groupe devait donc trouver des créneaux pour ses répétitions. Il a été convenu, néanmoins, que les dernières séances seraient consacrées à des répétitions collectives afin de mettre l'ensemble en place.

Cette phase a été très intéressante, car elle a soulevé plusieurs questions importantes. Elle a posé d'abord la question de l'organisation du travail, et notamment du choix entre ce qui devait être fait en classe et entre les classes. Elle a aussi posé plusieurs questions au niveau de l'organisation du groupe, en particulier la question du leadership, et celle de la structuration d'un groupe important en sous-groupes plus petits. Enfin, elle a posé la question de la progression, et de comment évaluer cette progression. Mais toutes ces questions ont émergé d'elles-mêmes au cours du processus suivi par les étudiants, et ce sont eux qui ont apporté des réponses à ces questions.

6 Troisième phase : une vue d'ensemble

À ce stade de l'expérience, nous étions déjà au deuxième semestre, et il ne restait plus que quelques séances pour mettre la performance au point. C'est aussi le moment où je leur ai annoncé que je leur avais trouvé un lieu pour leur performance de fin d'année. Il s'agissait du théâtre de l'Agora, à Évry, qui est une scène nationale. Il n'était pas encore clair, à ce point, si nous aurions accès à la grande salle ou à la petite. Les étudiants étaient tous très excités, car ils connaissaient très bien cette salle et ne s'attendaient pas à ce que je l'obtienne pour eux. La condition du théâtre, néanmoins, pour que nous puissions avoir accès à cette salle, était de se regrouper avec deux autres ateliers, afin que nous fassions tous les trois notre restitution le même jour. Cela ne semblait pas être, à première vue, un obstacle. Une fois la surprise du lieu de performance passée, il fallait se remettre au travail.

La question qui se posait maintenant était celle du fil conducteur. Les étudiants étaient conscients qu'ils avaient pour l'instant un patchwork, et qu'il allait peut-être falloir lui donner un peu de sens. Mais ils prenaient également conscience du fait qu'une performance permettait précisément ce genre de juxtaposition, et qu'il n'était pas forcément nécessaire de faire un gros travail de lien. Néanmoins, la question qui se posait était celle, au minimum, de trouver l'ordre le plus efficace pour les différents segments. En cherchant à résoudre ce problème, et à aller un peu plus loin en essayant de donner une cohérence à l'ensemble, ils se sont souvenus d'une phrase que je leur avais dite tout au début de l'expérience, quand je leur avais expliqué que le processus créatif partait toujours de quelque chose de chaotique pour essayer ensuite de lui donner une forme. J'avais même proposé, comme titre provisoire pour cette expérience, « du chaos à la création ». Ils se sont souvenus de cette proposition, ils l'ont adoptée, non seulement comme titre, mais comme idée de fil conducteur. Je les ai aidés à articuler cette proposition, en leur posant des questions pour leur permettre d'aller plus loin dans leur réflexion, et ils ont eu l'idée de démarrer par quelque chose de très désorganisé sur scène, et de faire émerger à chaque fois de ce chaos le segment suivant. Ils ont également eu recours à un fil rouge comique, représenté par un étudiant qui essayait à chaque fois de profiter de ce chaos pour jouer d'un instrument de musique, mais qui était à chaque fois chassé par ce qui allait se passer sur scène, sauf bien sûr à la fin du spectacle, quand il arrivait finalement à jouer son morceau, juste avant la fin du spectacle. Tout s'organisait donc. Les étudiants se sont rendu compte que l'écriture d'un spectacle n'était pas qu'une écriture textuelle. Il fallait également écrire, sous une forme ou sous une autre, le déroulé du spectacle, et faire des choix très précis quant à ce qui allait se passer à un moment bien déterminé sur scène. Ils ont également pensé à intégrer des segments vidéo à leur spectacle, vidéos qui étaient développées par de petits groupes séparément.

7 Quatrième phase : répétitions et mises en place

Une fois la structure du spectacle fixée, il fallait maintenant répéter pour qu'il ressemble à quelque chose. Chaque groupe a donc organisé son planning de répétition en dehors des rencontres du groupe, et chaque séance de travail permettait à toutes et à tous de se retrouver et de tester chaque segment ainsi que les enchaînements entre les segments. Il fallait également régler un certain nombre de problèmes techniques, et les étudiants se sont rendu compte que régler un problème technique faisait partie intégrante du processus créatif, et débouchait non seulement sur un choix technique, mais aussi très souvent sur un choix artistique et parfois esthétique. Les deux dernières séances ont été consacrées à des filages, tout au moins à ce qui pouvait s'en rapprocher le plus compte tenu des circonstances. Et le grand jour est finalement arrivé !

8 Performance

C'est finalement la grande salle du théâtre de l'Agora que nous avons eue à notre disposition. Compte tenu du fait que trois ateliers faisaient leur restitution au même moment, la salle était pleine aux deux tiers, ce qui était très impressionnant pour la plupart des étudiants, qui n'avaient jamais été sur scène.

Les étudiants ont découvert qu'il y avait encore beaucoup de choses à gérer au dernier moment. Par exemple, le groupe qui passait avant avait apporté un vidéoprojecteur, et il avait été décidé que les étudiants pourraient également l'utiliser pour leur performance. Mais au dernier moment, la propriétaire du vidéoprojecteur (qui est pourtant la directrice d'une compagnie de théâtre) a finalement refusé qu'on utilise son vidéoprojecteur. Faisant fi de tout le travail effectué par les étudiants qui avaient préparé les segments vidéo, elle les a obligés à supprimer complètement les séquences vidéo du spectacle. Il a donc fallu trouver très vite de petits réajustements pour que la performance tienne quand même la route. Bien sûr, ce genre d'accident de dernière minute peut arriver à tout moment, et n'est jamais très agréable. Mais d'un autre côté, il peut aussi jouer en faveur du groupe, en le resserrant davantage, et en créant une énergie du groupe encore plus importante. C'est ce qui s'est passé. Cet incident a créé une très forte solidarité entre les étudiants, et leur a permis de s'entraider non seulement pour gérer leur trac, mais aussi pour faire face ensemble à ce qui les attendait sur scène.

La performance finale a duré 35 minutes (au lieu des 45 minutes prévues avec les segments vidéo) et je dois dire que j'ai assisté à une performance d'une qualité remarquable. Il y a beaucoup de performances professionnelles qui n'atteignent pas ce niveau de qualité. Certes, même si mon rôle se limitait à celui d'un facilitateur, d'un modérateur, il est vrai que ma présence a permis d'orienter certains choix, d'aider à certaines prises de décision, d'aider à résoudre certains problèmes, même si mon intervention se limitait à quelques touches très subtiles. Mais les étudiants doivent majoritairement à eux-mêmes la qualité de ce

qu'ils ont créé ensemble, et ils le doivent aussi au fait d'avoir accepté de plonger avec moi dans l'expérience que je leur proposais, c'est-à-dire de plonger avec moi dans l'inconnu. Et cela me conforte dans le fait que, d'un point de vue pédagogique, lorsque l'on propose quelque chose de nouveau et de ludique à des étudiants, ils s'en emparent vigoureusement et le portent bien plus loin que ce à quoi on aurait pu s'attendre.

9 Bilan

Je pense que cette expérience pédagogique a été très enrichissante, à la fois pour moi et pour les étudiants qui y ont participé. Je pense qu'ils ont compris de façon expérimentale ce qu'était un processus créatif, et cela, en soi, est une chose remarquable. Ils ont aussi appris ce qu'était un processus créatif collaboratif, et comment le fait de travailler ensemble de manière collaborative pouvait permettre d'aller bien plus loin que lorsque l'on travaille dans un environnement compétitif. Ils ont également été confrontés à la prise de décision et à la résolution de problèmes. Et le fait qu'ils s'en soient aussi bien sortis a renforcé leur confiance en eux-mêmes. Enfin, ils ont créé quelque chose de toutes pièces, et cette expérience n'a pas de prix.

Ce processus créatif a illustré de très belle manière le concept d'énaction développé par Francisco Varela dans les années 90¹, appliqué au domaine de la création artistique. Il montre en effet qu'il y a une interaction permanente entre les processus cognitifs et l'environnement. J'ai proposé un environnement avec un certain nombre de contraintes aux étudiants, ce qui a orienté leurs processus cognitifs dans une certaine direction, et ces processus cognitifs ont à leur tour influencé l'environnement que je leur avais proposé initialement, et ainsi de suite.

Cette expérience pédagogique a été un succès sur tous les plans, à la fois sur le plan pédagogique lui-même, mais aussi sur le plan artistique, créatif, sur le plan du développement personnel, et au niveau de l'apprentissage du travail collaboratif. Sur le plan pédagogique, car elle a permis de tester un nouveau dispositif pédagogique reposant à la fois sur l'inconnu et sur l'énaction. Et c'est à mon avis un succès sur ce plan, car il me semble, d'après les retours que j'ai collectés par écrit auprès des étudiants après la performance, qu'ils ont appris de nombreuses choses à travers ce dispositif, et bien plus que si je leur avais expliqué moi-même en classe ce qu'était un processus créatif. Sur le plan créatif, car les étudiants ont pris conscience qu'ils avaient la capacité, si on leur offrait les circonstances pour cela, de créer un objet artistique de toutes pièces. Et cette expérience leur a permis d'exercer leur créativité, ce qui a conduit à un très beau spectacle digne d'une compagnie professionnelle tant sur le plan esthétique que sur celui de la fluidité et du contenu. Sur le plan du développement personnel, car chaque étudiant a dû faire un travail sur soi pour être capable de

1. Francisco J. Varela, Evan Thomson, Eleanor Rosch, *The Embodied Mind*, The MIT Press, 1991.

proposer quelque chose au sein du groupe et de dépasser, en partie du moins, certaines barrières. Et enfin sur le plan de l'apprentissage du travail collaboratif, car c'est une expérience à laquelle aucun d'entre eux n'aurait pu faire face seul. Ils ont aussi apparemment compris à quel point le groupe pouvait être porteur d'un dépassement de soi pour chacune et chacun.

**L'énaction dans la création artistique : théâtre,
cinéma et performance (2017)**

Ivan Magrin-Chagnolleau. **L'énaction dans la création artistique : théâtre, cinéma et performance.** Dans *Action, Énaction*, édité par Xavier Lambert. L'Harmattan, Collection « Ouverture Philosophique », pp. 213-224, 2017. ISBN : 978-2-343-11403-3.

12 pages

L'énaction dans la création artistique : théâtre, cinéma et performance

Ivan Magrin-Chagnolleau
CNRS, Institut ACTE (UMR 8218), Paris
ivan.magrin-chagnolleau@cnr.fr

Proposition

En 1993, dans son livre *The Embodied Mind*, Francisco Varela définit le concept d'énaction (*enaction* en anglais) comme une tentative de concilier les approches objectiviste et subjectiviste. « L'organisme donne forme à son environnement en même temps qu'il est façonné par lui. » Ce concept, depuis lors, a eu une influence considérable dans l'évolution des sciences cognitives, aussi bien au niveau des neurosciences, que dans les domaines de l'intelligence artificielle, de la linguistique, de la psychologie cognitive, ou encore de la philosophie. Ce qui est intéressant, lorsque l'on tente d'appliquer le concept d'énaction à la création artistique est de se demander ce qui, dans ce cas-là, joue le rôle de l'organisme. Est-ce l'artiste ou est-ce l'œuvre d'art ? Je suggère d'aborder les deux points de vue en synergie. L'œuvre a besoin de l'artiste pour se manifester, et l'artiste est à la fois canal et co-créateur de l'œuvre, dans la mesure où il donne forme à l'œuvre et lui apporte en même temps ce qu'il est. En ce sens, on peut dire que l'artiste façonne l'œuvre d'art en même temps qu'il est façonné par elle. Mais dans le contexte d'un art collaboratif comme le théâtre, le cinéma ou la performance, on peut aussi dire que le metteur en scène, le réalisateur, ou le performeur donne forme à l'équipe qui travaille sur l'élaboration de l'œuvre et aux contraintes qui vont s'appliquer tout au long de cette élaboration, en même temps que l'œuvre et le metteur en scène, le

réalisateur, ou le performeur sont façonnés par ces interactions avec l'équipe et avec les contraintes. Nous sommes donc vraiment ici au cœur d'un processus créatif qui illustre parfaitement le concept d'énaction tel qu'il est proposé par Varela.

Je me propose, dans cet article, de développer les notions exposées ici, en les illustrant à partir de trois exemples concrets issus de mon expérience personnelle de metteur en scène, de réalisateur, et de pédagogue. Le premier exemple traitera de l'énaction dans le théâtre, et je m'appuierai pour cela sur l'une de mes expériences de metteur en scène. Le second exemple abordera la notion d'énaction dans le cinéma, et je relaterai le processus créatif autour de la réalisation de mon premier court métrage de fiction. Le troisième exemple étudiera l'énaction au cœur d'un processus pédagogique et dans la performance, et l'expérience que j'ai choisie est l'animation d'un atelier de création scénique avec des étudiants de l'université d'Évry.

Intérêt. Enjeux.

Si cette notion d'énaction m'intéresse plus particulièrement, c'est parce qu'elle est d'après moi intimement reliée à la notion de responsabilité. La responsabilité est le devoir de répondre de ses actes, toutes circonstances et toutes conséquences comprises, c'est-à-dire d'en assumer l'énonciation, la réalisation, et toutes les conséquences. En tant qu'artiste, je me suis souvent posé la question de ma responsabilité. Suis-je responsable de l'œuvre produite ? Suis-je responsable de l'idée de départ, de l'étincelle qui donne lieu à la réalisation de cette œuvre ? Suis-je responsable du choix de mes collaborateurs sur chaque projet ? Et finalement, la question qui me paraît la plus intéressante : d'où vient

cet obscur pressentiment dont parle Peter Brook dans *The Empty Space* et qui, d'après lui, mettrait en marche le processus de création ? Suis-je responsable de cet obscur pressentiment, ou bien suis-je responsable des actes que je pose en réponse à cet obscur pressentiment ? Une approche subjectiviste et une approche objectiviste de ces questions donneront des réponses très différentes. Et à travers ce concept d'énaction, qui tend à réconcilier ces deux approches, on peut étudier le lien qui existe entre le sujet, l'artiste dans ce cas, et l'objet, l'œuvre produite, les deux baignant dans un terreau constitué par l'environnement, les contraintes, etc.

L'autre raison pour laquelle cette notion d'énaction m'intéresse particulièrement est qu'elle est une notion de synthèse. Elle tente en effet de réaliser une synthèse entre les approches subjectiviste et objectiviste. Or, j'ai toujours pensé que c'était en cherchant à synthétiser deux points de vue opposés que l'on progressait le plus. Cela permet de construire, de créer une chose nouvelle à partir d'une opposition, d'un conflit, plutôt que de rester prisonnier de cette opposition, de ce conflit.

L'énaction dans le théâtre

Dans le théâtre, il y a souvent débat autour de la question du rôle du metteur en scène dans l'émergence d'un spectacle. Est-il l'unique détenteur de la vision artistique d'une œuvre en élaboration, ou est-il essentiellement un facilitateur, un organisateur de ce qui vient à lui de toutes parts au cours de l'élaboration de cette œuvre ? Cela dépend bien sûr de la personnalité du metteur en scène. Mais même dans le cas d'une personne ayant une vision artistique très forte et très précise, les contraintes extérieures comme l'espace choisi pour le spectacle, ou encore le budget ont une influence non négligeable. Le

concepteur est donc sujet à toutes sortes d'interactions avec son environnement, et c'est de ce mélange entre sa conception initiale et les contraintes environnementales qui s'appliquent sur lui tout au long du processus d'élaboration que découle l'œuvre théâtrale. D'ailleurs, les contraintes que cette personne va rencontrer au cours de son travail sont entièrement dépendantes de son idée de départ. Il me semble que cela illustre parfaitement le concept d'énaction dans la création artistique. Je vais tâcher d'illustrer cela en prenant un exemple bien précis.

En 2010, j'ai effectué une mise en scène de *La leçon* d'Eugène Ionesco. Dans cet exemple, il se trouve que je peux retracer assez précisément la genèse du projet, et cette genèse montrera comment une succession d'évènements extérieurs peut conduire à une prise de décision très personnelle, mais qui découle directement de cette succession d'évènements. J'ai eu la chance de pouvoir assister à une représentation de *La leçon* à New York, à l'Atlantic Theater, en 2004. Il s'agissait d'une nouvelle traduction de Tina Howe. La mise en scène était de Carl Forsman, et le rôle du professeur était joué par Steven Skybell. C'est cette représentation, et plus particulièrement la performance époustouflante de Steven Skybell, qui m'ont vraiment fait découvrir cette pièce. C'est à ce moment que j'ai décidé que je voulais jouer le rôle du professeur dans la pièce. Mais il était clair que je voulais la monter en français. C'est seulement quelques années plus tard que j'ai rencontré à Paris une jeune comédienne, qui venait de lire plusieurs pièces de Ionesco, et qui avait particulièrement aimé *La leçon*. J'ai donc immédiatement pensé à elle pour jouer l'élève. Et j'ai commencé à me plonger dans l'œuvre de Ionesco, à lire des biographies et des interviews. Et plus j'avancais dans ce travail, plus j'aimais cette pièce. C'est aussi à ce moment que j'ai décidé d'en assurer également la mise

en scène. J'avais en effet une idée très précise de la manière dont je voulais la mettre en espace et l'éclairer. À partir de ce moment-là, tout s'est enchaîné très vite. J'ai trouvé un théâtre tout à fait adéquat pour la pièce. J'ai rencontré une maquilleuse-coiffeuse très talentueuse. Et j'ai proposé à mon frère de faire les lumières. Toutes ces décisions sont venues très naturellement. Il y a toujours un moment, dans une aventure créative comme celle-là, où j'ai l'impression que le projet artistique a sa propre existence. C'est un peu comme Michel-Ange lorsqu'il dit que la statue est déjà dans le bloc de marbre et qu'il n'a fait que la libérer. Mais peut-être que la statue est déjà à l'intérieur de lui et qu'il la manifeste dans le bloc de marbre. Il y a quelque chose de très mystérieux dans le processus créatif et il est à mon avis impossible de l'expliquer vraiment totalement, et c'est d'ailleurs d'après moi une bonne chose. Mais j'ai pu observer de façon consistante, projet après projet, que c'était toujours dans des moments de lâcher prise et de jeu que la création artistique se faisait. C'est d'ailleurs le sujet du magnifique livre de Stephen Nachmanovitch, *Free Play*. Au cœur de la création artistique, au cœur du travail de mise en scène de *La leçon* de Ionesco, il y a une osmose totale entre les circonstances extérieures (représentées par le théâtre choisi, les contraintes budgétaires de la production, les contraintes techniques du lieu, les personnes qui travaillent sur le projet, les contraintes d'emploi du temps de chaque personne, etc.) et les décisions créatives que je dois prendre au fur et à mesure que le projet se déroule. Je dois prendre une décision créative parce que je rencontre un obstacle à surmonter. La décision est donc nécessitée par une contrainte extérieure. En même temps, c'est bien moi qui prend cette décision, en fonction de l'obstacle rencontré, mais aussi des idées de départ que j'ai eues à propos de

ce projet, idées qui d'ailleurs le plus souvent se remodèlent au fur et à mesure.

Il existe un couplage structurel entre l'artiste avec son fonctionnement cognitif (idée de départ, idée de réalisation, idées créatives pour répondre à des contraintes, etc.) et l'environnement (son éducation, ses expériences, les contraintes extérieures, les personnes avec lesquelles il interagit, etc.). Et j'irai même un peu plus loin : plus l'artiste sera dans le lâcher prise et dans le jeu, plus ce couplage structurel sera osmotique, c'est-à-dire moins il y aura de résistance. Par résistance, j'entends à la fois ce que l'artiste perçoit comme des résistances extérieures (des choses qui semblent aller à l'encontre de ses idées) et ce qui constitue ses résistances internes (par exemple, la difficulté à reconnaître qu'un obstacle extérieur lui montre peut-être en fait un chemin nouveau et plus intéressant à explorer).

L'énaction dans le cinéma

La même question se pose lors de l'élaboration d'un film, à propos du rôle du réalisateur. Le réalisateur est-il seul détenteur de la vision du film, ou bien est-il là pour permettre aux diverses interactions successives avec ses collaborateurs de donner forme à une œuvre, qui se construit tout au long du processus créatif, et n'est vraiment terminée que lorsqu'elle sort en salles ? Une fois encore, les contraintes rencontrées ne sont-elles pas intrinsèquement liées à l'idée du projet initial ? Voici un autre exemple qui illustre encore à merveille le concept d'énaction dans la création artistique.

Entre 2008 et 2013, j'ai écrit et réalisé mon premier court métrage de fiction, *C'est la fête !* J'avais écrit de nombreuses idées de films, et plusieurs scénarios de

courts métrages. J'ai choisi ce projet d'une part parce qu'il était l'un des plus simples à réaliser comme premier projet de court métrage de fiction, et aussi parce que son thème, la solitude, était l'un de ceux sur lesquels je voulais à ce moment-là le plus m'exprimer. J'avais écrit au départ une histoire avec un personnage principal masculin, que je comptais jouer moi-même, et un personnage secondaire féminin, que je pensais confier à une amie comédienne. J'avais aussi en tête de demander à un chef opérateur - cadreur de me seconder pour les aspects techniques du tournage. Or, il se trouve que mon amie comédienne et le chef opérateur - cadreur auquel je pensais étaient très peu disponibles, et surtout presque jamais en même temps. Le tournage était sans arrêt repoussé. Je cherchais aussi quelques financements, et je n'en trouvais pas. Je me suis finalement demandé ce qui m'empêchait vraiment de tourner ce film. Et cela m'a conduit à prendre quelques décisions qui m'ont finalement permis d'avancer. J'ai d'abord réécrit le scénario pour qu'il n'y ait plus de personnages secondaires. Et j'ai modifié le découpage technique pour qu'il n'y ait plus que des plans fixes. J'ai aussi opté pour le noir et blanc afin de pouvoir faire un éclairage de qualité avec très peu de moyens. Toutes ces décisions ont été dictées par des contraintes extérieures, mais sont finalement devenues des décisions artistiques et esthétiques. J'ai donc pu tourner en trois jours mon premier court métrage de fiction, en l'ayant écrit et produit, et en assurant sur le tournage à la fois les rôles de directeur de la photographie, de cadreur, d'ingénieur du son, et de comédien. C'est à ce moment que j'ai eu l'idée, par jeu, de réaliser un film sur lequel j'aurais tout fait. Et j'ai donc décidé d'en faire également le montage et le mixage. Tout cela a bien sûr été possible parce que j'avais déjà quelques compétences sur chacun de ces postes techniques, et aussi parce que j'étais le seul à

devoir assumer toutes les conséquences de ces choix. Mais ces décisions ont aussi été le fruit des circonstances extérieures. Et même si l'envie de départ de réaliser un premier court métrage de fiction et que le choix du projet à réaliser étaient bien les miens, le chemin suivi a découlé entièrement des contraintes extérieures, y compris certaines décisions artistiques et esthétiques. Il y a donc bien eu un couplage structurel entre l'organisme (mes processus cognitifs, l'œuvre) et l'environnement.

L'énaction dans la pédagogie et la performance : du chaos à la création

La question de l'énaction se pose aussi dans le cadre pédagogique. Et plus particulièrement quand ce cadre pédagogique concerne la création artistique. Est-ce qu'un savoir peut être transmis indépendamment de la personne qui transmet ce savoir et de son audience ? Est-ce qu'un savoir peut être transmis autrement que par l'expérience ?

J'ai tenté de répondre à ces questions à travers une expérience pédagogique que j'ai imaginée et menée avec des étudiants de l'université d'Évry. Au cours de l'année scolaire 2009-2010, le département arts et spectacles de l'université d'Évry m'a demandé de mettre en place un atelier pour lequel j'avais carte blanche. Il s'agissait d'un atelier s'étalant sur toute l'année, comprenant cinq séances de trois heures au premier semestre et cinq séances de trois heures au second semestre, et pour lequel le but était de mettre les étudiants en contact avec une pratique artistique. Plutôt que de faire un autre atelier théâtre ou danse ou musique, ce qu'offraient la plupart des autres ateliers, j'ai conçu le projet ambitieux et novateur de proposer un atelier de création scénique au

cours duquel les étudiants seraient confrontés par l'expérience aux différentes étapes de la création d'une performance scénique. Le cadre que j'ai donné à cet atelier était le suivant : en partant de rien, nous devons arriver à une performance à la fin de l'année ; je ne jouerais qu'un rôle de facilitateur mais ce seraient les étudiants qui prendraient en charge tous les rôles (écriture au sens large, mise en place, en espace, en scène, interprétation, lumières, son, etc.) ; je m'engageais à leur trouver un espace de performance pour la fin de l'année. Je n'avais donc pas d'idée précise sur ce que serait cette performance scénique de fin d'année car je voulais que ce soient les étudiants qui prennent toutes les décisions créatives, et qu'ils soient confrontés aux différentes étapes que les artistes traversent lorsqu'ils sont dans un processus créatif.

Je dois dire que ce projet a dépassé toutes mes espérances. Les premières séances ont permis aux étudiants d'apprendre à se connaître un peu car ils venaient d'années différentes. J'ai aussi proposé que chacun d'entre eux apporte quelque chose de créatif (un texte, un dessin, une musique, une danse, etc.). J'ai suggéré que chacun d'entre eux puisse participer en fonction de ses compétences (artistiques ou techniques) et de ses envies. Le début du processus a été très chaotique, ce qui est généralement le cas au cours d'un processus créatif. Mais je suis intervenu le moins possible pour laisser ce chaos s'auto-organiser en fonction des idées que les étudiants avaient, et des contraintes qu'ils rencontraient. Et petit à petit, c'est effectivement ce qui s'est passé. Plusieurs textes écrits par certains étudiants ont émergé. Des envies de danser ou de faire de la musique se sont également exprimées. Des petits groupes se sont constitués autour d'un morceau de ce qui serait à la fin de l'année une

performance. Et les groupes ont commencé à travailler séparément sur leurs morceaux. Quelques personnes se sont proposées pour organiser l'ensemble et ont tenté, avec l'aide de tout le groupe, de donner une cohérence à l'ensemble. Un nom a été trouvé pour la performance : « Du chaos à la création ». Et le principe de base de cette performance s'est peu à peu dessiné, à savoir le retour entre chaque morceau à un état un peu chaotique et à l'émergence du morceau suivant à partir de ce chaos. Un fil rouge comique s'est aussi mis en place, constitué d'un étudiant qui tenterait à plusieurs reprises de jouer de la flûte mais serait systématiquement contrarié dans ses tentatives. Et pour finir, quelques personnes ont pris en charge les aspects techniques de la performance (lumières, son, vidéo, etc.). J'ai obtenu la grande scène du théâtre de l'Agora, scène nationale d'Évry, une salle de huit cents places. Et à la fin de l'année, les étudiants sont parvenus à proposer une performance de quarante minutes devant environ trois cents personnes, et cette performance était de très grande qualité, en dépit de (ou grâce à) certaines difficultés rencontrées au dernier moment. Les étudiants ont vécu tout au long de l'année une aventure très riche, qui leur a permis d'apprendre de nombreuses choses sur la création scénique par l'expérience.

Ce projet pédagogique illustre parfaitement le concept d'énaction : à la fois par rapport à l'émergence de l'œuvre elle-même, la performance, qui a trouvé son existence propre, après avoir été imaginée conjointement par plusieurs personnes ; par rapport à mon rôle dans ce projet, qui s'est inventé au fur et à mesure, en réaction à ce qui se passait dans le groupe, et fort de mon expérience passée de metteur en scène et de pédagogue ; par rapport aux étudiants qui se sont inventés artistes par la force des choses. Plusieurs

facteurs ont permis à cette performance d'exister, et chacun de ces facteurs a été nécessaire.

Conclusion

J'ai proposé dans cet article une réflexion préliminaire sur le concept d'énaction dans la création artistique, et plus particulièrement dans le théâtre, le cinéma, et la performance. L'énaction a été définie par Francisco Varela dans son livre *The Embodied Mind*. Il résume le concept par cette phrase : « L'organisme donne forme à son environnement en même temps qu'il est façonné par lui. »

J'ai choisi des exemples dans les trois domaines du théâtre, du cinéma et de la performance précisément parce que l'on parle, à propos d'eux, d'arts collaboratifs. Contrairement aux autres arts, qui sont souvent le fait d'une seule personne, même si elle est bien sûr en contact avec son environnement pendant la création de son œuvre, un spectacle théâtral, un film ou une performance ne peut être le fait d'une seule personne, à de très rares exceptions près. Nous sommes donc en présence de l'un des systèmes les plus complexes de création artistique. Il est donc relativement facile de mettre en évidence des interactions entre organisme et environnement, et nous avons vu que nous pouvions définir de plusieurs façons la paire organisme/ environnement.

Le rôle de l'organisme peut être joué par l'artiste ou par l'œuvre, selon le point de vue, mais il est parfois difficile de différencier vraiment les deux, et c'est peut-être la paire artiste-œuvre qui joue le rôle de l'organisme. Quant à l'environnement, il comprend tout ce qui est extérieur à l'organisme artiste-œuvre, c'est-à-dire le lieu, les

contraintes budgétaires, les autres personnes travaillant sur le projet, les contraintes de temps, les contraintes techniques, etc. Les contraintes sont souvent source de créativité : elles la stimulent. C'est une autre idée qui est développée dans *Free Play* de Stephen Nachmanovitch, et à laquelle j'adhère complètement.

Et je proposerai, pour aller plus loin, que cette énonciation devient véritablement osmotique dans un état de lâcher prise et de jeu, état qui me paraît être, à l'instar de Stephen Nachmanovitch, le seul état possible pour une véritable création artistique.

Bibliographie

Francisco J. Varela, Evan Thomson, Eleanor Rosch, *The Embodied Mind*, The MIT Press, 1991.

Peter Brook, *The Empty Space*, Touchstone, 1968.

Eugène Ionesco, *La leçon*, Gallimard, 1954.

Stephen Nachmanovitch, *Free Play*, Tarcher/Putman, 1990.

Ivan Magrin-Chagnolleau. Du chaos à la création : une expérience de pédagogie énonciative pour expérimenter le processus créatif en performance. *p-e-r-f-o-r-m-a-n-c-e.org*, 2 (1-2). 2015.

Filmographie

C'est la fête !, court métrage écrit et réalisé par Ivan Magrin-Chagnolleau, Aloha Films, 2013.

**Expérience phénoménologique et énaïve d'une
oeuvre de poésie spatialisée visuelle et sonore (sous
presse)**

Ivan Magrin-Chagnolleau. **Expérience phénoménologique et énaïve
d'une oeuvre de poésie spatialisée visuelle et sonore.** Dans *Corpo-
rité et construction du sens dans la poésie visuelle*, sous presse.

10 pages

Expérience phénoménologique et énaactive d'une œuvre de poésie spatialisée visuelle et sonore

Ivan MAGRIN-CHAGNOLLEAU, CNRS
Aix-Marseille Univ, CNRS, PRISM, Marseille, France
Chapman University, California, USA

Résumé

Dans ce texte, je propose un dispositif permettant de faire l'expérience du caractère énaactif de la réception d'une œuvre artistique, et de prendre conscience de ce caractère énaactif. Pour cela, j'ai construit une œuvre de poésie spatiale visuelle et sonore de 5 minutes et 32 secondes. J'ai proposé à des sujets de prendre connaissance en ligne de cette œuvre, puis de répondre immédiatement à une question, dont le but était de faire expliciter par le sujet un moment particulièrement intéressant dans l'expérience qui venait d'être vécue. Cette question a été construite selon le modèle de l'entretien d'explicitation. Il s'agit donc d'une approche phénoménologique, qui s'intéresse à l'expérience vécue du sujet, et dont le but est de mettre en évidence le caractère énaactif de cette expérience. Parmi les 20 sujets qui ont répondu à la question, 14 ont accepté de jouer le jeu du contrat posé par la question, et ont répondu en lien avec la question posée. Les réponses données par ces sujets sont conformes à l'hypothèse énaactive de la cognition.

Mots-clés

Énaaction, hypothèse énaactive, phénoménologie, cognition, entretien d'explicitation, auto-explicitation, poésie spatiale, poésie visuelle, poésie sonore.

Une approche phénoménologique de la création artistique et de la réception d'une œuvre

Avant d'entrer dans le vif du sujet, je voudrais poser, comme contexte à ce travail, deux fils conducteurs de ma recherche en ce moment. Ces fils sous-tendent le travail présenté ici. Le premier concerne une approche phénoménologique de la création artistique et de la réception d'une œuvre. Par approche phénoménologique, j'entends une approche reposant sur l'expérience vécue d'un artiste en train de créer une œuvre, ou bien l'expérience vécue d'une personne qui en fait l'expérience (voir par exemple

Magrin-Chagnolleau *et al.*, 2019). Je m'appuie, bien sûr, sur quelques textes fondateurs de la phénoménologie (Husserl, 1985, 1992 ; Heidegger, 1992 ; Merleau-Ponty, 1976 ; etc.), mais j'essaie d'adopter une approche plus pratique que théorique de la phénoménologie. J'essaie d'adopter une approche expérientielle de la phénoménologie, ce qui pour moi fait sens puisque la phénoménologie s'intéresse précisément à l'expérience vécue, mais ce qui ne va pas toujours de soi dans les textes théoriques fondateurs sur la phénoménologie, où il s'agit parfois davantage d'une idée de la phénoménologie plutôt que de sa mise en pratique.

L'hypothèse éactive appliquée à la création artistique et à la réception d'une œuvre

L'autre fil conducteur qui traverse ma recherche depuis quelque temps est l'hypothèse éactive de la cognition, telle qu'elle est présentée par Francisco Varela (Varela *et al.*, 1991). J'ai rencontré l'hypothèse éactive par le biais d'une participation à un ouvrage collectif (Magrin-Chagnolleau, 2017), et je tente de mettre en lumière l'hypothèse éactive dans des activités cognitives ayant un lien avec la création artistique ou bien avec la réception d'une œuvre. Pour moi, l'hypothèse éactive traduit principalement le lien étroit qui existe entre cognition, perception et action. Pour le dire autrement, l'hypothèse éactive traduit le fait que notre cognition est incarnée, c'est-à-dire qu'elle se développe sur la base des expériences perceptives, sensorielles, corporelles que nous faisons de notre environnement au sens très large.

Faire comprendre par l'expérience l'hypothèse éactive

Ce qui m'intéresse en premier lieu avec l'éaction, c'est d'en faire faire l'expérience par le biais de propositions artistiques et/ou performatives, ce que je m'attache généralement à faire dans les colloques où je participe. C'est aussi ce que j'ai proposé ici : approcher l'hypothèse éactive en faisant faire l'expérience d'une poésie spatialisée visuelle et sonore. Je vais détailler cette intention et ce dispositif un peu plus loin.

Un outil pour explorer ces dimensions : l'entretien d'explicitation, et l'auto-explicitation

Je voudrais enfin expliquer la méthode que j'utilise pour explorer l'expérience vécue, et dans ce cas-là, l'hypothèse éactive. J'utilise l'entretien d'explicitation développé par Pierre Vermersch (Vermersch, 2012, 2014). Cet entretien consiste à interroger un sujet

sur une expérience vécue passée à l'aide d'une technique très précise. Il s'agit de faire revisiter à un sujet un moment très spécifique d'une expérience vécue, afin de pouvoir accéder par ce biais à des informations très précises sur le déroulé de l'action lors de ce moment spécifique, y compris des informations pré-réfléchies sur certaines opérations cognitives comme des prises d'information ou des prises de décision. Cette technique d'entretien permet d'accéder très précisément à ce qu'un sujet a réellement fait, et non à ce que le sujet croit qu'il a fait, ce qui n'est pas toujours la même chose. Ce type d'entretien peut conduire à des prises de conscience tout à fait intéressantes, chez le sujet autant que chez la personne qui conduit l'entretien.

Il faut noter ici que cette technique d'entretien peut être très coûteuse en temps puisqu'il faut généralement compter 30 à 60 minutes pour faire passer un entretien sur des moments spécifiques qui sont parfois de l'ordre de quelques secondes seulement, voire moins. Il est donc préférable de s'assurer d'avoir déjà une bonne compréhension de la tâche à explorer, et du moment spécifique visé.

J'utilise aussi, dans ma recherche, et c'est le cas ici, une version réflexive de l'entretien d'explicitation qui s'appelle l'auto-explicitation (Vermersch, 2007, 2008). Dans cette version de l'entretien d'explicitation, on se fait passer l'entretien à soi-même, en utilisant ou bien un enregistreur, ou bien l'écrit. Il faut noter que cela requiert de faire des va-et-vient entre la position de sujet et celle d'intervieweur, ce qui nécessite déjà une très bonne maîtrise de l'entretien d'explicitation, et d'avoir joué régulièrement le rôle de sujet, et celui d'intervieweur.

L'expérience proposée : une poésie spatialisée visuelle et sonore

L'intention : faire expérimenter l'énaction

Comme je l'ai déjà noté, une de mes intentions est d'arriver à faire comprendre l'énaction par l'expérience, en créant des situations qui font faire l'expérience d'un processus cognitif et de sa nature énaactive. En tant qu'artiste, je questionne souvent mes propres processus créatifs, et les processus cognitifs sous-jacents, notamment à l'aide de l'auto-explicitation (Vermersch, 2007, 2008). Je questionne également les processus créatifs d'autres artistes en les interrogeant à leurs sujets à l'aide de l'entretien d'explicitation (Vermersch, 2012, 2014). J'ai pu à plusieurs occasions mettre en évidence le caractère énaactif de ce processus créatif de l'artiste, et réfléchir à la nature énaactive de ce processus (Magrin-Chagnolleau, 2017). J'ai également commencé à explorer l'hypothèse énaactive au niveau de la réception d'une œuvre. Et

je me suis intéressé aux différentes façons de faire expérimenter cette nature énaactive de la réception d'une œuvre en proposant divers dispositifs expérimentaux, notamment à un colloque sur Langage et Énaction à Clermont-Ferrand. J'ai donc voulu renouveler cette proposition ici, mais cette fois-ci en m'appuyant sur une œuvre de poésie spatiale visuelle et sonore, à la fois pour me rapprocher du thème de la journée d'étude, mais aussi parce qu'une œuvre de poésie spatiale visuelle et sonore me paraît particulièrement pertinente pour ce genre d'expérimentation.

Pourquoi la poésie spatialisée est intéressante

En effet, en utilisant la spatialisation, à la fois sur le plan visuel et sur le plan sonore, cela permet de créer une situation cognitivement complexe, dans laquelle un sujet reçoit à la fois des informations visuelles dans l'espace, mais aussi des informations sonores dans l'espace, le tout à travers un geste poétique qui demande également au sujet de faire une interprétation à partir du symbolique. Il y a donc une saturation cognitive pour le sujet qui conduit le plus souvent à un lâcher prise de sa part, et donc à une disponibilité plus grande vis-à-vis de l'expérience qui lui est proposée. Cela oblige alors au sujet à faire des choix, à la fois attentionnels, mais aussi de fabrication du sens. C'est dans ce genre de situations qu'il est favorable de mettre en évidence le caractère énaactif de la cognition.

Construction et description de l'œuvre

Pour construire cette œuvre, je me suis donné comme contrainte d'explorer un champ sémantique autour de l'énaaction et de la représentation que je m'en faisais, tout en laissant venir d'autres champs sémantiques par libre association. Par exemple, l'idée de va-et-vient est très proche du concept d'énaaction, et il a conduit par libre association à la mer, et aux vagues, d'où le champ sémantique marin également présent dans l'œuvre.

J'ai aussi voulu créer une sorte de parallélisme entre l'écrit et le sonore, mais en créant à chaque fois des décalages temporels entre les deux, et parfois aussi une certaine variabilité, ce qui fait que le sonore, qui vient généralement après le visuel, n'est jamais exactement une répétition du visuel, et est presque toujours en décalage avec ce dernier. Cette idée est venue en créant l'œuvre, comme un dispositif que j'associais avec l'idée de faire expérimenter l'énaaction.

J'ai donc finalement construit une œuvre audio-visuelle de 5 minutes et 32 secondes, comprenant des mots qui apparaissent au fil de l'œuvre dans différentes zones de l'écran, ainsi qu'une bande sonore qui est constituée de ma voix disant également des mots, mais que l'on entend tantôt à droite, tantôt au milieu, tantôt à gauche.

Une utilisation un peu particulière de l'explicitation

J'ai également voulu tester, dans le cadre de cette expérimentation, une façon un peu différente d'utiliser l'explicitation. En effet, comme il était peu réaliste d'envisager de faire passer un entretien d'explicitation à tous les participants du colloque, ce qui aurait pris beaucoup de temps, j'ai eu l'idée de proposer aux personnes qui avaient regardé et écouté l'œuvre de répondre immédiatement après à une question en ligne. La question était posée de telle façon qu'elle pouvait permettre la mise en évocation du sujet (voir Vermersch, 2014), mais bien sûr ne la garantissait pas puisque nous n'étions pas à proprement parler dans le cadre d'un entretien d'explicitation. Il s'agissait plutôt d'une incitation à l'auto-explicitation (Vermersch, 2007, 2008). La question posée était : « Vous venez de regarder et d'écouter une vidéo poésie. Je vous propose, si vous le voulez bien, de prendre le temps de laisser revenir un moment qui vous intéresse particulièrement dans l'expérience que vous venez de vivre. Quand vous avez ce moment, je vous demande de le décrire aussi finement que possible. » J'ai également proposé à d'autres personnes ensuite de regarder l'œuvre en ligne et de répondre également à la question.

Analyse des résultats

J'ai eu au total 20 réponses. Ce qui est intéressant à observer en premier lieu, c'est le fait qu'on peut observer deux sortes de réponses. Une partie des réponses ne donne absolument aucune information sur l'expérience que le sujet vient de vivre. Ce sont des sujets qui n'ont pas joué le jeu de la question, c'est-à-dire qui n'ont pas accepté le contrat proposé. Et c'était tout à fait prévisible. Dans la mesure où je n'étais pas là en personne pour proposer un entretien d'explicitation au sujet, et pour m'assurer que le sujet était d'accord pour le faire, je m'attendais obligatoirement à ce qu'un certain nombre de sujets ne souhaitent pas entrer dans la proposition qui leur était faite. Ce type de réponse représente 10% de l'ensemble des réponses, soit 2 réponses. Il n'y a pas grand chose d'intéressant à tirer de ces réponses par rapport au propos qui nous intéresse ici. On peut noter simplement que les raisons de rejeter la proposition qui était faite aux sujets sont principalement liées au fait que ces sujets étaient déjà très

familiers avec l'entretien d'explicitation et ne semblait pas en mesure de pouvoir s'adapter à cette nouvelle proposition.

C'est dans l'autre catégorie de réponses qu'on trouve les choses les plus intéressantes, et en lien avec le sujet de l'article et de la journée d'étude. Cela concerne donc 90% des réponses, c'est-à-dire 18 sujets. Ce sont des sujets qui ont joué le jeu, c'est-à-dire qui ont accepté la proposition de se remémorer quelque chose de l'expérience passée et d'en faire part. La question était posée de manière à ne pas influencer le sujet sur le moment choisi de son expérience. Cela m'intéressait en effet de savoir quel moment le sujet allait choisir. Nous sommes dans le cas de figure où le moment choisi n'est pas spécifié par la question mais laissé au choix du sujet. C'est généralement la procédure que l'on suit lorsqu'on explore une tâche (ici, vivre l'expérience de cette poésie spatiale visuelle et sonore) sans avoir d'idée préconçue sur la typologie de la tâche (c'est-à-dire son organisation) ou de ce qui fait le plus de sens au regard de la question de recherche posée.

En analysant les réponses, on remarque que 8 sujets ont évoqué plutôt le début de l'expérience, tandis que 6 sujets ont choisi un moment un peu plus loin dans l'œuvre. Les 4 autres sujets ont plutôt fait une réponse prenant en compte la globalité de l'expérience, c'est-à-dire sans s'attacher à un moment précis, donc n'ayant pas intégré l'ensemble de la consigne. C'est assez cohérent. En effet, les sujets, quand on leur laisse le choix d'un moment spécifié, vont choisir le moment le plus saillant pour eux, qui est souvent le moment le plus disruptif, dans ce cas, souvent le début car on ne sait pas à quoi s'attendre et on est surpris. Les autres moments correspondent généralement pour les sujets qui les ont choisi à des moments de plus grande surprise que le début.

On peut également classer les réponses en fonction du fait que le sujet a choisi de « revivre » un moment spécifique pour répondre à la question, c'est-à-dire de se mettre soi-même en évocation, ou non. En me basant sur les libellés des réponses, j'ai identifié 14 sujets qui semblent s'être mis en évocation pour répondre à la question, ce qui correspond également aux sujets qui ont choisi un moment spécifique.

Voici quelques extraits de réponses :

« Ce qui m'est arrivé au début, dès l'entrée de la pièce, est la crainte liée à l'écoute stéréo, entièrement située à l'oreille droite, d'avoir un problème technique avec mon casque. Le texte ainsi que la bascule du son au côté

gauche m'ont détrompé assez vite mais j'ai pendant un moment été déstabilisé. »

« J'ai été marquée par le tout début de la vidéo. Les mots sont posés sur la surface de l'écran comme si l'écran était de l'eau, notamment les prépositions "sur" (sur la mer) ou "sous" (sous la mer) comme si je me situais exactement à la surface (l'eau au niveau des yeux ou du nez). Cela inverse le dessus et le dessous, puis l'intérieur et l'extérieur (comme si on faisait une pirouette dans l'eau). »

« Ce décalage est venu chez moi créer dans mon corps ce mal de mer. »

« Il y a eu une attente que les mots parlés correspondent aux mots écrits, puis une déception que ça n'arrive pas vraiment. »

« Les mots amenaient mon esprit à vagabonder. »

« J'ai essayé de comprendre comment les rapports son-texte étaient imbriqués, s'ils se rapprochaient (ça m'a semblé être le cas par moment) ou si au contraire ils s'éloignaient. »

« Le moment le plus marquant de la vidéo est pour moi le passage sur l'opposition entre le monde intérieur et extérieur. Dans cette séquence, des mots associés au monde intérieur (intérieurité, monde intérieur, intérieur) et au monde extérieur (extériorité, environnement, extérieur) sont affichés un par un et sont lus un par un dans un ordre différent de l'affichage. »

« Je revois le moment où les mots "va et vient" sont apparus sur fond noir en même temps et avec un temps de décalage par rapport aux paroles prononcées. »

« La surprise de la spatialisation du son, au tout début, qui passe de l'oreille droite à l'oreille gauche, et en même temps je pense au titre et à la marée, et je me demande à ce moment-là en voyant le mot "surface", si la surface est au niveau de la surface de la mer d'un point de vue spatial, c'est-à-dire au milieu de l'écran, et si je vais voir apparaître, dans les secondes à venir, le ciel en haut de l'écran, et les profondeurs en bas. »

« L'élément aquatique est pour moi évocateur. Je me sens rapidement immergée dans le noir de l'écran mais le décalage entre les mots et la voix me perturbe légèrement. »

« Le début m'a fortement déconcertée. Le décalage entre la voix et l'affichage du texte m'a rendue impatiente d'attendre le moment quand et où le texte apparaît. »

« Au début, je ne sais plus si je suis sous l'eau ou sur la vague, si je flotte ou si je coule, si je lis ou si j'écoute. Des injonctions contradictoires me manipulent. Malaise. Effervescence électrique dans mon cerveau, palpitations cardiaques. »

« L'attente de découvrir si tout ce qu'on écoute apparaît écrit sur l'écran. »

« Expérience extérieure, expérience intérieure : extériorité, introspection. »

« Sous la mer le poisson nage : je plonge sous la mer et cherche le poisson. »

« Va et vient, répété : je vois des "tildes" espagnols. »

Je voudrais d'abord faire deux remarques préliminaires. Tout d'abord, la mise en évocation par soi-même des sujets ne peut être formulée que de manière hypothétique dans la mesure où c'est le sujet lui-même, juste après avoir fait l'expérience de la poésie spatiale visuelle et sonore, qui choisit comment répondre à la question posée, et que personne ne pose vraiment cette question en personne et en présence, de manière à guider ensuite la réponse en utilisant les techniques de l'entretien d'explicitation et en observant certains signes qui peuvent confirmer ou non la mise en évocation. Ce qu'on peut dire à partir de cette expérience, c'est que 14 sujets ont accepté de parler d'un moment de l'expérience que ces sujets venaient de vivre, et que ces sujets le font de manière spécifique.

La deuxième remarque est que cette expérience ne concerne que 20 sujets, et que seulement 14 d'entre eux répondent effectivement à la question posée en choisissant de parler d'un moment spécifique de l'expérience vécue, donc c'est un nombre de sujets relativement réduit pour pouvoir en tirer des conclusions générales.

Ceci étant dit, en regardant le contenu des réponses ci-dessus, on peut néanmoins faire quelques d'observations. Tout d'abord, on peut observer à quel point les

expériences relatées sont différentes d'un sujet à l'autre. Cela n'est bien sûr pas surprenant, mais correspond assez bien à l'hypothèse énaactive de la cognition. En effet, même si les sujets sont tous exposés à la même expérience, c'est-à-dire aux mêmes stimuli de l'environnement extérieur (visuel et sonore), ces stimuli viennent rencontrer à chaque fois une cognition singulière, et vont donc faire sens de manière totalement différente d'un sujet à l'autre. Ceci est tout à fait conforme à l'hypothèse énaactive.

Notamment, on peut noter dans certaines réponses des éléments qui font référence à une certaine anticipation par rapport à l'expérience en train de se vivre, et cette anticipation, lorsqu'elle est mentionnée, se fait sur des éléments très différents selon les sujets. Il me semble que cette idée d'anticipation (cognitive) est aussi tout à fait conforme à l'hypothèse énaactive, dans la mesure où elle se construit sur la base des expériences passées du sujet, et donc sur la base de la façon dont sa cognition s'est organisée à la suite de ces expériences passées. Il semble donc bien y avoir un va-et-vient entre la façon dont la cognition du sujet est déjà organisée avant de vivre l'expérience, et les stimuli que cette expérience propose, et comment ils vont pouvoir s'agencer ou non avec la cognition qui résulte des expériences passées du sujet.

Conclusion et perspectives

J'ai proposé dans cet article une réflexion autour des dimensions phénoménologique et énaactive d'une expérience d'une poésie spatiale visuelle et sonore. Mon ambition était de faire faire l'expérience de la nature énaactive de la cognition par le biais d'une œuvre de poésie spatiale visuelle et sonore. Pour cela, outre la fabrication de l'œuvre elle-même, qui mobilisait déjà cette dimension énaactive, j'ai proposé aux personnes ayant fait l'expérience de l'œuvre de répondre à une question qui était posée selon la technique de l'entretien d'explicitation développée par Pierre Vermersch.

En dépouillant ensuite le questionnaire, il apparaît assez clairement dans certaines réponses que la cognition mobilisée par les spectateurs de cette œuvre est de nature énaactive, ou en tout cas, a minima, ne contredit pas cette hypothèse et va plutôt dans son sens.

Il y a plusieurs pistes possible pour aller plus loin dans ce travail. Tout d'abord, on pourrait proposer cette expérience à un nombre plus grand de personnes afin de vérifier ces premières tendances sur un nombre plus important de sujets.

Une autre possibilité serait de faire passer des entretiens d'explicitations en bonne et due forme à des sujets venant de faire l'expérience de l'œuvre de poésie spatiale visuelle et sonore et de vérifier si on obtient des résultats similaires à ceux décrits dans l'article, en étant sûr cette fois-ci que les sujets sont bien en évocation d'un moment spécifique de leur expérience passée.

Cela permettrait en outre de valider le protocole expérimental proposé dans cet article, dans un cadre où des entretiens d'explicitations ne seraient pas, ou pas immédiatement, possible. A minima, le protocole proposé dans l'article permet de faire une première analyse de la tâche et de tirer des premières conclusions de manière à pouvoir ensuite mieux cibler les moments spécifiés lors d'entretiens d'explicitation.

Bibliographie

Heidegger, M. (1992). *Être et Temps*. Gallimard.

Husserl, E. (1985). *Idées directrices pour une phénoménologie*. Gallimard.

Husserl, E. (1992). *L'idée de la phénoménologie*. PUF.

Magrin-Chagnolleau, I. (2017). L'énaction dans la création artistique: Théâtre, cinéma et performance. In *Action, Énaction* (pp. 213–224). L'Harmattan, Collection « Ouverture Philosophique ».

Magrin-Chagnolleau, I., Nahoum, R., & Weisen, M. (2019). Le projet Sipapu: Vers une micro-phénoménologie de la création. In X. Lambert (Ed.), *Les Enjeux Cognitifs de l'Artefact Esthétique—Tome 2* (pp. 187–194). L'Harmattan.

Merleau-Ponty, M. (1976). *Phénoménologie de la perception*. Gallimard.

Varela, F., Thomson, E., & Rosch, E. (1991). *The Embodied Mind*. The MIT Press.

Vermersch, P. (2007). Bases de l'auto-explicitation (1). *Expliciter*, 69.

Vermersch, P. (2008). Introspection et auto-explicitation: Bases de l'auto-explicitation (2). *Expliciter*, 73.

Vermersch, P. (2012). *Explicitation et phénoménologie*. PUF.

Vermersch, P. (2014). *L'entretien d'explicitation*. ESF Éditeur.

3.12 Entretien d'explicitation

Pratiquer l'entretien d'explicitation à distance : petit ABC (2021)

Jocelyne Leblanc, Anne Philippon, Ivan Magrin-Chagnolleau, Marie-Hélène Lachaud, Magali Boutrais. **Pratiquer l'entretien d'explicitation à distance : petit ABC**. *Explicititer*, 2021, 130, pp. 51-60.

9 pages

Pratiquer l'entretien d'explicitation à distance : petit ABC

Jocelyne LeBlanc, Anne Philippon, Ivan Magrin-Chagnolleau,
Marie-Hélène Lachaud, Magali Boutrais

INTRODUCTION

Le 28 novembre 2020, nous avons participé à un premier atelier de pratique des techniques d'explicitation à distance, c'est-à-dire à travers une plateforme de visioconférence, animé par Marine Bonduelle. À la suite de trois heures de pratique, plusieurs d'entre nous ont eu envie de réfléchir à l'impact d'un dispositif de visioconférence sur la manière de mener l'entretien d'explicitation. Nous avons donc formé un groupe de travail, et cet article est le fruit de notre réflexion. Deux d'entre nous ont participé aussi au 2^e atelier de pratique du 30 janvier 2021. Nous souhaitons, à travers cet article, aborder les questions suivantes : comment mener un entretien d'explicitation à distance et que devons-nous mettre en place pour qu'il soit le plus harmonieux et efficace possible ? Nous vous partageons donc notre courte mais éclairante expérience qui se divise ainsi dans cet article : 1) l'installation technique ; 2) quelques moments d'expérimentation ; 3) un tableau récapitulatif de ce qui est commun et différent entre le présentiel et le distanciel ; 4) quelques constats.

Dès le départ, nous avons toutes et tous été plus ou moins déstabilisés par la pratique des entretiens à distance, mais il semble possible de trouver là une belle occasion d'innover et de nous laisser surprendre par ce que permet la visioconférence. La meilleure stratégie n'est peut-être pas de chercher à faire à tout prix exactement de la même façon en visioconférence qu'en présentiel, mais plutôt de trouver comment exploiter au mieux cette condition dans le cadre d'un entretien d'explicitation, voire d'inventer de nouvelles choses.

Rappelons également que dans le cas des ateliers de pratique du Grex, les A¹ sont souvent expert.e.s, ce qui n'est que rarement le cas dans une situation professionnelle (éducative, de recherche, etc.). Dans ce cas-là, les A sont plutôt novices. La principale différence entre les deux est la capacité du A expert à entrer plus facilement en évocation et à y retourner. Inversement, le A novice éprouvera fort probablement une difficulté réelle à entrer en évocation et une grande facilité à en sortir.

L'INSTALLATION TECHNIQUE OU *LES PETITS COUSSINS DE MARINE*

Lors de la journée de pratique du 30 janvier dernier, Marine a partagé qu'en tant que formatrice elle apportait toujours des petits coussins pour ses participant.e.s, car les chaises sont toujours trop dures. Nous avons trouvé que cette expression s'appliquait bien à la section technique de la pratique à distance, et que c'était une jolie métaphore. Les *petits coussins* auxquels il faut penser pour le volet technique sont nombreux dans une pratique à distance contrairement à celle en présentiel. Tout comme un EDE² « normal » (c'est-à-dire en présence physique) demande une logistique et une installation particulières, celui fait à distance, exige, lui aussi, des manières de faire spécifiques. Ainsi, la disposition des deux chaises à 45 degrés pour A et pour B lorsque l'on est dans le même

¹ Dans la suite de l'article, nous appellerons **A** la personne interrogée, **B** la personne qui mène l'entretien, et **C**, la personne qui observe éventuellement l'entretien, dans le cas notamment d'un atelier de pratique.

² EDE = Entretien **D**'Explicitation

espace physique, et qui est notre principal souci, devient en distancié, une panoplie de détails techniques à prendre en compte.

Voici donc certains points à considérer et à intégrer dans cette nouvelle pratique à distance de l'entretien. Les points suivants, une fois abordés et réglés, permettront de se consacrer entièrement à l'EDE et de ne plus être soucieux ou inquiet d'un mauvais fonctionnement ou autre. Cela demande du temps, mais nous avons constaté qu'il est nécessaire de le prendre !

a) Avant l'entretien ou l'ante-début

- Le réseau internet : Joindre le A afin de vérifier la connexion internet. Si celle-ci n'est pas bonne, il est possible que l'entretien ne puisse se faire.
- Un endroit calme : S'assurer que notre A sera dans un endroit calme où les distractions extérieures seront réduites au maximum (enfants, animaux, bruits ambiants, etc.) et si besoin, dans lequel la confidentialité pourra être garantie. L'idée de s'installer dans un endroit familier pourra favoriser la concentration si la personne est habituée à y travailler par exemple. Il faut s'assurer que la personne ne sera pas dérangée pour toute la durée de l'entretien (contrat entre A et B sur la durée approximative nécessaire).
- Le silence : Tout comme en présentiel, il est aussi important de désactiver tout autre appareil et d'éteindre tout autre écran. Autrement dit, il importe d'éliminer toute source d'interruption possible lors de l'entretien.

L'ordinateur et ses ajustements :

- La caméra : Penser à éviter d'installer la caméra à contre-jour, afin de pouvoir bien distinguer les traits du visage.

En visio, le regard n'est pas toujours évident à bien saisir selon que l'on regarde directement ou non la caméra. On s'assurera donc que la caméra est au plus près de la fenêtre d'interaction à l'écran afin de créer la plus grande proximité de regard entre A et B. Par exemple, une caméra placée sur le côté de l'écran de l'ordinateur amènera un point focal décalé entre celle-ci et la fenêtre du dispositif de visioconférence.

- Le cadrage : Prévoir le cadrage de l'entretien. Pour B, il est nécessaire de voir les mains et le haut du corps de A, et bien sûr le visage, ainsi qu'éventuellement une partie des jambes. Pour A, il est probablement important d'avoir le moins de distraction visuelle dans l'espace du B, mais aussi de voir à peu près la même chose que ce que B voit de A, afin de pouvoir mettre en œuvre *a minima* un accord postural. Dans tous les cas, le mode plein écran est recommandé, c'est-à-dire que A ne voit que B à son écran et vice-versa.

Il est avantageux de régler ces dimensions techniques lors d'une visio précédente afin de pouvoir tester le dispositif et ses modalités, micros, casques, écrans, etc. Ainsi, le jour de l'entretien, le risque que ces éléments viennent interférer avec le bon déroulé est amoindri.

b) Durant l'entretien

- Réglage de l'affichage à l'écran : Lors d'un atelier de pratique ou d'une situation avec observateur(s) ou observatrice(s), nous recommandons d'éteindre les caméras pour les C et de régler l'affichage pour que seuls apparaissent à l'écran le A et le B. A peut même faire disparaître son image de l'écran en la faisant glisser sur le côté. Ainsi, A n'est pas perturbé.e. B peut faire de même de son côté.
- L'utilisation du *Chat*³ : En présence d'un C, le *Chat* est un moyen facile et discret à utiliser entre B et C durant l'entretien afin de communiquer si nécessaire. Moins invasive que l'intervention en présentiel, cette manœuvre est cependant susceptible de déconcentrer B comme on a pu s'en rendre compte lors de l'entretien de A par Jocelyne analysé plus loin. Il est possible de l'activer entre B et C seulement et de s'entendre au préalable sur cette manière de faire.
- Établir également un contrat entre B et C : B souhaite que C utilise le *Chat* ou bien préfère éviter toute perturbation ? Si oui, de quelle manière ? B et C peuvent convenir d'un signe visuel pour indiquer une demande d'aide de C (en particulier pour un.e B novice, ou dans le cas d'un relais éventuel pris par C pour devenir B, si l'occasion se présente).
- L'enregistrement : L'utilisation d'un outil de visioconférence facilite l'enregistrement audio et/ou vidéo de l'entretien à des fins de recherche, vu qu'il suffit d'appuyer sur un bouton pour démarrer cet enregistrement, dont la possibilité est offerte par à peu près tous les outils de visioconférence. Comme dans le cas d'un entretien en présence, il est bien sûr requis de demander l'autorisation verbale d'enregistrer cet entretien, en expliquant le pourquoi de cet enregistrement, et en rappelant l'anonymisation de tout résultat qui serait inféré à partir de l'entretien. Il peut d'ailleurs être souhaitable de redemander cette autorisation une fois l'enregistrement démarré, afin que cette autorisation verbale soit également enregistrée.
- Le casque d'écoute : Il y a plusieurs possibilités pour entendre l'autre à travers un dispositif de visioconférence : un casque englobant, des écouteurs moins immersifs, ou bien des haut-parleurs. Un casque immersif peut mettre certain.e.s A dans une situation d'inconfort. Il semble préférable de demander à A sa préférence. La même question se pose au niveau de B. Il s'agit dans ce cas de favoriser la meilleure compréhension possible par B de ce qui est dit par A. Dans les deux cas, on prêtera particulièrement attention au risque d'écho quand on utilise des haut-parleurs d'un côté et d'un autre, en privilégiant autant que possible l'usage de casques. Pour B, une possibilité est par exemple de mettre un seul côté du casque, afin d'avoir un retour direct de sa propre voix. Pour A, cela n'est pas forcément nécessaire.
- Le micro : Il est important de s'assurer de très bien s'entendre et de se comprendre de part et d'autre du dispositif. En présence, cette compréhension est implicite et n'a pas besoin d'être discutée, mais en visioconférence, ceci doit être testé en amont de l'entretien pour éviter l'interruption et la sortie d'évocation qu'une difficulté de compréhension pourrait susciter au cours de l'entretien.

³ Le *Chat* est un outil de conversation par écrit inclus dans les modalités de l'outil numérique.

QUELQUES MOMENTS D'EXPÉRIMENTATION SELON LES POSTURES

En tant que A

- On a entendu, lors de notre journée de pratique du 30 janvier, une A être complètement rassurée par cette pratique à distance. Dans le même souffle, elle disait être “un peu plus rassurée” pour sa prochaine formation à distance à venir.
- Observations des perceptions sensorielles du côté de Marie-Hélène

Ce qui me revient de ma première pratique de l'EDE à distance avec le groupe de pair.e.s lyonnais en juillet 2020 en tant que A : la voix de Marine m'aide à élucider, elle me guide mais mon corps ne sent pas sa présence. C'est comme si mes capteurs sensoriels étaient ouverts mais qu'ils ne percevaient rien. Au moment où j'ai partagé cela, il m'est revenu que j'avais perçu la même chose quand j'ai animé les premières formations à distance, au mois de mai 2020. Cette sensation m'a déstabilisée, j'ai essayé d'y remédier et il me semble qu'aujourd'hui, c'est plus facile. Je serais curieuse de savoir comment j'ai fait, ce que j'ai modifié dans ma pratique d'animation à distance pour me sentir plus connectée aux participant.e.s.

Au cours de la pratique à distance du 28 novembre, j'ai découvert en tant que A ce que je peux décrire comme étant dissociée. Il me revient la joie associée à cette découverte, je (mon observatrice) me revois observer MH qui explicite devant l'écran de l'ordinateur. Mon observatrice se trouve au-dessus de moi, sur la droite. De là, je peux me voir, je suis assise face à mon bureau, la lampe éclaire mon visage, je parle, j'ai les yeux fermés, je suis calme.

En tant que B

- Observation des perceptions sensorielles du côté de Marie-Hélène (suite)

Au cours de la pratique de juillet, en tant que B, j'ai noté : “je me charge de mon corps, je me sens reliée à A quand je le questionne, je vois qu'il part en évocation, ses yeux sont fermés, son débit de parole est ralenti”. Au moment du débriefing, A constate qu'il est en auto-explicitation (“comme si je m'étais chargé de mon corps et j'avais ta voix, je lâche moins prise sur le contenu, les questions de l'EDE”). J'ai noté qu'il n'a pas fait de différence avec le présentiel. Ce même jour, j'ai découvert que je ralentis le débit de A en maintenant les mains à plat sur mon bureau. Je perçois la pression de mes mains à plat sur la surface de mon bureau, elle est lisse, c'est un peu chaud. Ce geste m'ancre, me met en contact avec A (comment ?). Il me revient aussi que mes pieds sont à plat sur le sol. Ces deux gestes m'aident à rester connectée, à voir les images de ce que décrit A, j'entends sa voix, je ferme les yeux et les ouvre de temps en temps.

Comme en présentiel, je découvre aussi que j'ai plus de difficulté à mémoriser les mots de A et à visualiser le film de ses actions si je ne ferme pas les yeux.

A la fin d'une autre pratique au mois de janvier, j'étais B et me suis sentie connectée à A mais cette fois, je faisais les mêmes gestes qu'elle, bien qu'elle fermait les yeux. Je la regarde, je sens le contact de mes pieds à plat sur le sol, je regarde l'écran et quand je ne vois plus le film de A qui fait ce qu'elle décrit, je ferme les yeux et là, je vois ce qu'il manque, je vois si c'est clair ou pas.

- L'accord postural du côté de Jocelyne

Tout récemment, lorsque j'ai pratiqué en tant que B, j'ai repris le geste de A qui se faisait de son côté droit et où elle le mentionnait aussi verbalement : ça se passe de mon côté droit, comme ça. Je me suis donc appliqué à reprendre le geste du même côté qu'elle, soit celui de mon côté droit. J'ai alors vu à l'écran que mon geste était reflété du côté gauche pour elle. Un peu plus tard dans l'entretien et jusqu'à la toute fin, j'ai expérimenté de faire plutôt l'effet miroir. Son côté droit devenait le gauche pour moi et vice versa. Après l'entretien, je lui ai posé la question : Comment as-tu vécu le fait que je fasse le geste à l'opposé du tien ? Elle m'a répondu qu'elle avait dû interpréter et comprendre que je faisais bien le geste du côté droit, mais que l'ordinateur le lui transmettait à l'opposé. Elle disait qu'une fois cela décodé et compris, elle n'avait plus été dérangée.

- Une belle surprise du côté de Jocelyne

Lors de notre pratique de fin novembre, je me souviens que j'avais peur que mon A ne me sente pas assez présente. Afin d'éviter cette situation, j'avais tendance à me pencher vers l'écran afin de lui faire comprendre, par mon non-verbal, que j'étais bien présente à elle. Je n'ai jamais pensé pratiquer l'accord postural à ce moment. Je ne voyais que le vide entre nous deux créé par la distance virtuelle. Et pourtant, à ma grande surprise, dans le feed back après l'entretien, j'avais été étonnée par le fait que j'avais réellement senti une « bulle » entre nous deux. Par « bulle », j'entends le sentiment que je pouvais bien saisir ce qu'elle me disait, ce qu'elle ne disait pas, et surtout, que toute mon attention était dirigée vers elle (même si un C était aussi présent !). Dans la pratique du 30 janvier, j'ai à nouveau revécu ce sentiment de « bulle ». Je planifie de faire un entretien de cet entretien afin d'en comprendre la teneur. C'est à suivre !

- Un manque de sensations du côté d'Anne

En novembre puis en janvier, j'ai fait l'expérience d'être B à distance. Au-delà de ma difficulté à accompagner un A expert (ce dont je n'ai pas l'habitude), j'ai l'impression que je n'ai pas réussi à retrouver les sensations que j'éprouve lorsque j'accompagne A en présentiel. Il me semble que toute mon attention est portée sur le visuel et l'auditif que j'analyse intellectuellement, alors qu'en présentiel, je me laisse davantage porter par la dimension kinesthésique, plus intuitive. Lorsque A est près de moi, je ressens une énergie qui nous relie, comme des fils tissés et animés et qui me permettent, je crois, d'accompagner A dans son évocation. À distance, je n'ai pas retrouvé cette sensation, j'ai l'impression d'avoir été davantage guidée par ma réflexion.

En tant que C du côté d'Anne et Jocelyne

Anne : Ce samedi matin du 30 janvier 2021, nous sommes 4 dans le sous-groupe et je suis le C de Jocelyne qui est B. Jocelyne et moi avons ce projet d'écriture d'un article sur l'EDE à distance. C'est donc avec l'intention de tester le rôle d'un C qui accompagne B à distance que je débute l'écoute de cet EDE. A souhaite comprendre ce qui lui a permis dans un moment spécifié, d'écrire facilement, alors que d'habitude elle a du mal à s'y mettre. En tant que C, je me sens très libre puisque je n'ai pas la responsabilité de l'accompagnement de A, je suis "invisible" pour A. Dans cette exploration, les deux C ont convenu d'essayer l'utilisation du *Chat* pour communiquer avec B.

Jocelyne : Pour vous situer dans le temps, cet entretien a duré 20 minutes précisément. Il a été suivi par un échange entre A et moi (B) alors que les deux C sont demeurées invisibles tant au niveau

sonore que visuel. Toutefois, elles entendaient et voyaient tout. À la 31^e minute, elles se sont jointes à nous en réactivant leur micro et leur caméra afin d'échanger à 4 sur cet entretien. Cet échange post-EDE a duré 30 minutes. En résumé, l'entretien, l'échange sur l'entretien entre A et B et l'échange entre A, B et les deux C, tout cela a duré 1 heure.

L'EDE débute :

(A4) A : *Je suis assise à mon bureau, j'avais pas mal de choses à faire et puis cet article à écrire, ... j'étais d'humeur assez fraîche, j'étais guillerette en fait, je voyais bien que je pouvais rassembler tous les éléments, des kilos et des kilos de documents ; de vidéo etc; et puis à la place, j'avais mon bloc, mon crayon , et puis j'ai ... tout de suite j'ai , je suis partie sans technique, voilà, sans technique d'écriture.*

(B5) Jocelyne (B) : *D'accord.*

(A5) A : *et j'ai écrit l'intro, en me disant tant pis, j'ai dû faire 3-4 paragraphes, des textes courts en tout cas.*

Anne : Dès la fin de A4, je relève l'expression "je suis partie sans technique" et j'ai la sensation que c'est le moment à creuser. Je décide donc, assez spontanément, de partager cette sensation à Jocelyne et j'écris sur le *Chat* pendant A5, je formule mon hypothèse :

10:59:42 De Anne à Jocelyne (Message direct) : tout de suite je suis partie sans technique : peut-être le moment clé à creuser.

Et tout en écoutant je poursuis cette fois avec une proposition de relance :

10:59:59 De Anne à Jocelyne (Message direct) : c'est comment « sans technique » ?

Anne : À ce moment-là, c'est comme si j'étais B, je ne me pose pas la question de ce que B peut faire de cela, je ne m'interroge pas sur la perturbation que cela peut engendrer, alors que Jocelyne est au début de l'EDE, et qu'elle accompagne A dans l'évocation.

Jocelyne : Je n'ai jamais vu ses messages apparaître sur mon écran. J'étais tellement concentrée à suivre A que cette petite fenêtre du *Chat* n'a pas attiré mon attention. Ni à 10 : 59 : 42, ni à 10 : 59 : 59.

Anne : L'EDE se poursuit et je n'interviens plus. Je me dis que j'ai testé l'utilisation du *Chat* et je constate que B n'a pas besoin de C. Cinq minutes plus tard dans l'entretien, A développe un long récit du moment du séminaire sur lequel elle devait écrire. J'analyse qu'elle est souvent en commentaires et qu'elle explique le contenu du séminaire. Par 3 fois, j'ai l'impression que Jocelyne veut la faire revenir sur le moment d'écriture, elle dit "d'accord" mais ne parvient pas à poursuivre, car A poursuit son récit.

C'est à ce moment-là que je partage sur le *Chat* avec B :

11 : 07 : 18 De Anne à Jocelyne (Message direct) : il faudrait peut-être la recentrer sur le moment d'écriture : "et tout ça, ce que tu as ressenti, ça s'est donné à toi comment au moment où tu écris ?"

Jocelyne : Au tout début de l'entretien, le second C a envoyé un *Chat* pour vérifier si ça fonctionnait bien, ce système de communication fermé entre chacun des C et moi, en B. Je lui avais alors répondu immédiatement : oui. Je n'ai plus porté attention au *Chat* par la suite. Le *Chat* de 11 : 07 : 18 est le premier que j'ai aperçu. Je me souviens de le voir et d'être surprise. Comme il apparaissait dans une petite fenêtre où on ne pouvait lire seulement que : *Il faudrait peut-être la recentrer sur le moment d'écriture*, je tentais d'en comprendre le sens tout en écoutant mon A. C'était difficile, car je tentais de relier cette suggestion au fil de l'entretien qui continuait de se dérouler. J'ai laissé tomber pour finalement me concentrer sur A. A me décrivait alors la journée de la conférence et je me demandais comment la ramener, et ce, même si ce qu'elle disait m'aidait à mieux comprendre la chronologie et la teneur de ce qu'elle avait vécu. Je continuais malgré tout à me questionner sur cette proposition d'Anne (11 : 07 : 18). C'est à ce moment que j'ai cliqué sur la fenêtre et que j'ai pu voir l'entièreté de la phrase. Tout ça s'est déroulé rapidement, pas plus de deux minutes entre le moment où je le vois et le moment où je clique sur la fenêtre. Dans ce laps de temps, l'entretien s'était donc poursuivi et la suggestion n'était plus à propos. Il y a donc eu ici un enjeu lié à la non-connaissance de cet outil de visioconférence. Par la suite, la prise de conscience du A en lien avec cette écriture qui s'était faite facilement a été très belle. Je n'ai plus vu aucun des *Chats* de mes deux C apparaître. Il y en a eu trois autres.

Ce que cela nous apprend du rôle de C dans un EDE à distance

Il est difficile pour B d'accompagner A et, en même temps, d'avoir son attention sur le *Chat*. C devrait intervenir seulement si B est en difficulté. C'est pourquoi nous proposons que C et B conviennent *a priori* d'un signe visuel qui permettrait à C d'intervenir à la demande de B. Toutefois, cette possibilité de soutenir le B sans que A ne soit témoin de rien est vraiment intéressante, en particulier lorsque B est novice.

Du côté des sous-groupes et la gestion du temps

Lors des pratiques à distance avec plusieurs personnes, l'organisation des sous-groupes fonctionne bien, mais l'expérience nous conduit à évoquer un point de vigilance autour de la gestion du temps. En effet, à la fin du temps imparti, le basculement du sous-groupe au groupe entier se fait manuellement par l'hôte de la visioconférence. Si la durée prévue pour les temps de pratique n'a pas été respectée, il peut être brutal de se retrouver dans le groupe entier. Cette situation nous est arrivée alors qu'un.e B, yeux fermés et en pleine évocation, s'est retrouvé.e de retour parmi tous sans même en être conscient.e. Marine, l'animatrice, avait alors réagi prestement pour informer, par *Chat*, les participant.e.s de l'autre groupe de prendre une pause de quelques minutes.

Il est important de bien mentionner la durée du temps de pratique avant de partir en sous-groupes, et aussi de le noter dans les sous-groupes ! A, avec les yeux fermés et dans sa « bulle », peut ne pas se rendre compte de son retour dans la « salle virtuelle principale », et cela peut être un choc.

D'autre part, se retrouver écouté.e par le groupe entier lorsque ce n'est pas dans le contrat de départ pose un problème éthique. Pour éviter cela, B doit veiller impérativement à mettre fin à l'entretien avant le temps imparti, le C peut aussi y veiller et intervenir dans le *Chat* pour alerter B, et cela doit faire l'objet du contrat de départ. On pourrait aussi imaginer que l'animateur.trice attende que les sous-groupes signalent sur le *Chat* qu'ils sont prêts à revenir en grand groupe.

PRATIQUER L'EXPLICITATION EN VISIO OU EN PRÉSENCE :
TABLEAU RÉCAPITULATIF DE CE QUI EST COMMUN
ET DE CE QUI EST DIFFÉRENT
(à partir des ateliers de pratique)

| EN VISIO | EN PRÉSENCE |
|--|--|
| MATÉRIEL | |
| Ordinateur Outil de visioconférence Connexion Internet adaptée à la visio Casques d'écoute avec micro intégré | Chaises Coussins (ceux de Marine !) |
| Enregistreur numérique : son, image et <i>Chat</i> de l'outil de visioconférence utilisé | Enregistreur numérique |
| ESPACE | |
| Lieu favorable à la concentration Éteindre téléphone et autres sources de distraction | |
| Cadrage de la caméra de l'ordinateur : B voit les mains et le haut du corps de A et place sa caméra face à son regard C éteint sa caméra : seuls A et B apparaissent à l'écran B affiche A en mode plein écran | Placement des chaises Accord postural |
| Lors d'un atelier de pratique, établir un contrat entre B et C B est-il d'accord pour que C prenne le relais en tant que B ? | |
| B souhaite-t-il que C utilise le <i>Chat</i> pour faire quelques suggestions de relances/questions ou bien préfère-t-il éviter toute perturbation ? | Comment B souhaite-t-il communiquer son besoin de soutien au C ? Par une tape sur l'épaule, un regard ? |
| TEMPS SUPPLÉMENTAIRE À PRÉVOIR | |
| Essai technique en amont de l'entretien Durée de connexion : à noter pour les participants avant la répartition en sous-groupes (horaire du retour en grand groupe) | Ne s'applique pas |
| RELATION HUMAINE | |
| Moins d'accès aux détails du non verbal : mimiques du visage Perception sensorielle réduite Les échanges informels entre deux personnes ne peuvent se faire que par le <i>Chat</i> | Les échanges informels ont lieu et servent bien souvent à détendre l'atmosphère Souplesse du moment de fin des entretiens en sous-groupes |

CONCLUSION

Voici en guise de conclusion quelques constats tirés des ateliers de pratique. Selon les personnes, la visioconférence est plus ou moins gênante pour l'entretien et cela varie en fonction des modalités sensorielles dominantes de chaque individu (auditive, visuelle, kinesthésique...). Pour certain.e.s le son de la parole suffit, A ferme les yeux, B est absorbé.e par les paroles de A, il n'y a pas de différence entre le présentiel et la distance. Pour d'autres, la présence des corps et la sensation d'énergie qui s'en dégage peut manquer lors de la visioconférence. L'accord postural n'est pas le même et la reprise de gestes peut être perturbée.

En présence comme à distance, il apparaît qu'un accord entre B et C est nécessaire : Qu'est-ce que C observe et/ou note ? C peut-il à un moment de l'entretien prendre le relais de B ?

La connaissance de la plateforme de visioconférence n'était pas égale pour tous. Nous avons mis en commun nos connaissances et habiletés à chaque instant de ces moments de pratique. L'idéal est de prendre le temps de se familiariser avec les fonctions de la plateforme avant l'entretien.

Dans les ateliers de pratique du GREX2, nous explorons, nous essayons, nous expérimentons, que ce soit en présence ou à distance. Nous avons encore des choses à apprendre sur les techniques d'aide à l'explicitation.

La pandémie de Coronavirus a entraîné l'utilisation d'outils de visioconférence, ce qui a donné une dimension internationale (intercontinentale !) aux ateliers de pratique du samedi et aux séminaires du vendredi. Vive les levers tôt pour ceux et celles du Québec ! Avant d'y être contraint.e.s par cette pandémie, nous n'avions pas envisagé de « rapprocher ainsi les continents ». Malgré certaines limites pour les entretiens, nous nous adaptons et apprenons de nouvelles manières de faire. D'autres expériences restent à partager, sans doute, et de nouvelles explorations nous attendent...

Mise en évocation : Entretien d'explicitation et méthode de l'Actors Studio (2023)

Ivan Magrin-Chagnolleau. **Mise en évocation : Entretien d'explicitation et méthode de l'Actors Studio**. *Explicititer*, 2023, Hors-série (1), pp.53-61.

9 pages

Mise en évocation :

Entretien d'explicitation et méthode de l'Actors Studio

Ivan MAGRIN-CHAGNOLLEAU, CNRS

Aix-Marseille Univ, CNRS, PRISM, Marseille, France
Chapman University, Orange, California

Introduction

Je vais vous parler dans cet article d'évocation. J'ai eu en effet deux sujets de discussion de prédilection avec Pierre Vermersch, qui sont d'une part l'évocation, et notamment le lien avec le travail fait à l'Actors Studio, et d'autre part l'auto-explicitation. C'est le premier sujet que j'aborde ici.

Pourquoi parler de la méthode de l'Actors Studio en lien avec l'évocation ? La première raison, c'est que quand j'ai découvert Pierre Vermersch à travers son livre sur l'entretien d'explicitation¹, j'ai remarqué qu'il parlait de l'Actors Studio à un moment dans le livre. Et il se trouve que j'ai fait l'Actors Studio, j'ai été formé pendant trois ans là-bas en tant que comédien. Et du coup, ne le sachant pas au début, je suis allé dans l'explicitation en étant un expert de l'évocation, de la mise en évocation, et de la mise en auto-évocation. Donc j'en ai beaucoup discuté avec lui.

La mise en évocation dans l'entretien d'explicitation

Pour cadrer un peu les choses, je voudrais d'abord reprendre la définition de l'évocation dans le glossaire du livre :

« Le fait qu'un acte mental s'accompagne d'un contenu représenté de façon quasi sensorielle.

Tous les actes cognitifs ne s'accompagnent pas nécessairement d'évocation. Toute évocation peut être décrite par son contenu (qu'est ce qui est évoqué), par la texture sensorielle en laquelle elle représente ce contenu (quelle est la ou les modalités

¹ Vermersch, P. (2014). *L'entretien d'explicitation*. Issy-les-Moulineaux : ESF Éditeur, 208 p.

sensorielles qui supportent cet évoqué, visuel, auditif, etc.) et par le déroulement de l'acte d'évoquer en tant que tel. » (page 176)

Et puis si on reprend l'endroit du livre où il parle de l'Actors Studio, c'est dans le chapitre « Mémoire concrète et questionnement sensoriel » pages 82 et 83. Il cite deux domaines professionnels ayant développé l'utilisation délibérée de la mémoire concrète, le théâtre et la psychothérapie. Et donc pour le théâtre, il parle des techniques de formation à l'Actors Studio. Il fait référence à plusieurs choses. Alors il y a à mon sens des petites confusions liées au fait qu'il y a deux outils différents qui sont utilisés dans les techniques de l'acteur, qui sont d'une part la mémoire sensorielle, et d'autre part la mémoire affective, qui sont deux choses légèrement différentes même si elles induisent un peu les mêmes processus. J'y reviendrai plus loin.

Ce qui est intéressant, c'est qu'il parle aussi de déclencheurs, donc il en parle à la fois côté évocation, et je vais citer à ce sujet quelques passages des premiers numéros d'*Expliciter*. Et il se trouve que le mot déclencheur est aussi utilisé dans le vocabulaire du comédien. J'y reviendrai aussi. Et puis il y a aussi l'idée de ce que Pierre Vermersch appelle la stabilité. Je vais également parler de ça, parce que la stabilité dont il parle n'est pas tout à fait la même que la stabilité dont on parle du point de vue du comédien.

Pour revenir un peu d'abord du côté « évocation » avant d'aller du côté de ce que je vais appeler « mémoire sensorielle », je vous propose, si vous en êtes d'accord, de fermer les yeux et de prendre le temps de laisser revenir le moment où vous avez commencé la lecture de cet article. Donc prenez le temps de laisser revenir ce moment. Où êtes-vous ? Est-ce que vous ressentez une certaine température, du vent, de l'air ? Est-ce que vous sentez des odeurs particulières ? ... Je vous propose de revenir tranquillement ici et maintenant, à votre rythme. Je ne peux pas certifier que vous étiez en évocation. Il n'y a que vous qui puissiez le dire. Mais si vous y étiez, et si vous avez ressenti de nouveau les éléments de votre expérience de quand vous avez commencé la lecture de cet article, il y a des chances que vous ayez effectivement été en évocation. Donc l'évocation, c'est ça. Évidemment, on peut aller beaucoup plus loin dans la qualité d'évocation.

Il se trouve que cet été, à l'université d'été du GREX², on a travaillé justement sur le remplissement intuitif, c'est-à-dire sur comment on passe d'un état où on n'est pas en évocation à un état où on l'est de plus en plus, jusqu'à ce que le remplissement intuitif soit

² Maurel, M. (2022). Université d'Été de Goutelas 2022 : Décrire le remplissement intuitif typique de l'évocation. In *Expliciter*, 135, p. 1-21.

complet, qu'on puisse dire : « Ça y est, j'y suis ! » Ceci pour dire que finalement, ce thème de l'évocation, c'est toujours un sujet de recherche actif et c'est toujours quelque chose sur lequel on se pose pas mal de questions.

Ce que je voudrais faire, maintenant, c'est revenir aux premiers numéros d'explicitation pour voir un peu comment, dès le début, Pierre Vermersch parle déjà d'évocation comme sujet de recherche. On commence par le numéro 6 de septembre 94, qui est un numéro court de quatre pages. Sur la quatrième page, on trouve un texte intitulé : « Projet pour une analyse phénoménologique de la conduite d'évocation³. » Je voudrais en citer un extrait :

« Un des éléments essentiels de l'explicitation est l'évocation d'un vécu spécifique passé. On peut évoquer autre chose que du vécu, et si c'est du vécu, pour aller vers la dimension procédurale de l'action, il faut un guidage de l'intervieweur, autrement dit une conscience du contenu des répliques et une formulation judicieuse des relances. Dans la grande majorité des cas, cet accès à une situation spécifique se fait de façon aisée. L'évocation de ce vécu est possible et stable. Bref, dans tous ces cas, nous produisons un entretien efficace et évident dans sa conduite. Au total, cette évidence ne nous apprend rien sur ce qu'est la conduite d'évocation. Nous sommes les utilisateurs avertis d'un mécanisme psychologique suffisamment familier pour être opaque à notre compréhension. Une manière d'avancer est de rechercher activement à décrire ce qui se passe quand cette conduite est perturbée, qu'elle ne se met pas en place. Les difficultés permettent de rendre apparent des aspects d'une conduite qui, autrement, est rendue opaque par son caractère habituel, rapide, bien rôdé. » (page 4)

On voit déjà là plusieurs éléments de la méthode de travail de Pierre Vermersch, qui sont déjà esquissés ici. Et on peut aussi noter quelque chose d'intéressant, c'est que Pierre Vermersch semble suggérer que cette capacité d'évocation, elle est finalement assez naturelle. Et dans son livre sur l'entretien d'explicitation, et dans l'autre article dont je vais parler, il se réfère à la madeleine de Proust, et donc à cette capacité qu'on aurait de se remettre facilement dans une mémoire d'une expérience passée au point de ressentir un certain nombre d'éléments de cette mémoire.

Dans le numéro 8 de janvier 95, qui est un numéro de huit pages, il y a un article de quatre pages intitulé : « L'évocation, un objet d'étude ?⁴ » C'est un article vraiment intéressant quand

³ Vermersch, P. (1994). Projet pour une analyse phénoménologique de la conduite d'évocation. In *Explicitation*, 6, p. 4.

⁴ Vermersch, P. (1995). L'évocation, un objet d'étude ? In *Explicitation*, 8, p. 4-8.

on s'interroge sur l'évocation, parce qu'il reprend plusieurs éléments qui vont être ensuite développés pendant des années et des années. C'est quasiment tout le programme de recherche de Pierre Vermersch autour de l'évocation qui est déjà en filigrane dans cet article.

La première question qu'il pose, c'est comment étudier l'évocation. Il liste un certain nombre de méthodes possibles pour s'intéresser à l'évocation et arriver à l'étudier. Par exemple, on peut faire une première étude sous forme d'études de cas, donc en particulier les cas où ça ne marche pas, pour comprendre quand ça ne marche pas, pourquoi ça ne marche pas, et ce qu'il faudrait éventuellement faire pour que ça marche. Il parle aussi de la possibilité de combiner des méthodes classiques reposant sur des expérimentations, et des données qu'on peut recueillir sous forme d'entretiens, c'est-à-dire croiser des données recueillies par l'entretien d'explicitation avec des données qui viendraient d'autres sources que l'entretien.

Ensuite, ce qui est vraiment intéressant dans cet article, c'est la classification qu'il fait en s'appuyant sur Piaget, sur les différents types de prises de conscience possibles. Il précise également qu'on peut s'intéresser à l'objet d'évocation, donc le contenu de l'évocation. On peut s'intéresser au sujet qui évoque, donc la personne qui est interviewé. Ou bien on peut s'intéresser à l'acte lui-même d'évocation.

Dans cet article, ce qui est particulièrement intéressant, c'est que Pierre Vermersch pose déjà d'emblée les catégories descriptives de l'évocation. Et là encore, il y a toute une cartographie de travail. Il parle des différentes phases de l'évocation : de l'ante-début de l'acte d'évocation et du début de l'acte d'évocation. C'est aussi dans cet article qu'il s'interroge sur le déclencheur de l'évocation. Il discute aussi des sous-modalités. Ensuite, il parle de l'acte d'évocation installé, en proposant la notion de stabilité de l'évocation, sa précision, sa vivacité. Et il parle ensuite de la fin de l'acte d'évocation, et finalement de la post-action de l'acte d'évocation. Il pose en particulier les questions suivantes : Est-ce qu'il y a une rémanence de l'évoqué ? Est-ce qu'il y a un changement d'état ? Si changement d'état il y a, de quelle nature est-il ? Et est-ce que le fait d'avoir évoqué un passé vécu rend sa ré-évocation ultérieure plus facile ou non ? J'y reviendrai quand je vous parlerai de la méthode de l'Actors Studio et de la mémoire sensorielle. Donc je voulais revenir un peu sur ces fondamentaux-là, parce qu'ils vont nous aider à comprendre ce que je vais dire maintenant de la mémoire sensorielle et de la méthode de l'Actors Studio.

Ce qu'il faut bien comprendre dans tout ce dont je viens de parler, c'est qu'on retrouve un certain nombre de questionnements autour de l'évocation, notamment de ses déclencheurs. Et on retrouve aussi un certain nombre de questions qu'on va retrouver côté comédien, qui sont des questions notamment de stabilité et d'entraînement et de facilité ou non de ré-évoquer un

passé une fois qu'on l'a déjà évoqué une fois, deux fois, etc. Alors, avant de passer à la méthode de l'Actors Studio et à la mémoire sensorielle, je voudrais juste rappeler une chose fondamentale, qui est que, bien évidemment, le but visé n'est pas le même. La méthode est plus ou moins la même, au moins au début, mais le but visé n'est bien sûr pas le même. Dans l'explicitation, le but visé, c'est vraiment la description de l'action. Du côté du comédien, le but visé, c'est le surgissement de l'émotion.

La mise en évocation dans la méthode de l'Actors Studio : mémoire sensorielle

La méthode de l'Actors Studio a été développée par Lee Strasberg. Elle s'inscrit dans la lignée de quelques théoriciens russes, notamment Constantin Stanislavski, qui est venu faire des conférences et des stages aux États-Unis à New York dans les années 30, et auxquels ont assisté Lee Strasberg, Elia Kazan, Stella Adler, Sanford Meisner, etc. La question centrale pour Stanislavski, et donc aussi pour Lee Strasberg, c'est comment peut-on faire pour pouvoir accéder à l'émotion sachant qu'on ne peut pas y accéder de manière directe ? Il faut trouver alors une technique qui permette à l'acteur d'être consistant et stable dans son travail d'acteur, notamment dans sa capacité à ressentir des émotions vraies pour pouvoir les jouer, et donc les faire exister sur scène ou devant une caméra.

La technique que propose Stanislavski, et qui a ensuite été développée encore plus loin par Lee Strasberg, c'est ce qu'il appelle la mémoire sensorielle, c'est-à-dire une technique permettant de revivre un moment du passé, mais en se focalisant sur tous les éléments sensoriels de cette situation du passé. Ce qui revient finalement à faire une évocation d'un moment du passé, mais de choisir ces moments du passé sur d'autres facteurs que ceux qu'on va utiliser pour l'explicitation. Dans l'explicitation, on recherche la description de l'action, on va donc s'intéresser à des moments qui présentent un intérêt du point de vue de l'action. Alors que pour un comédien, on va chercher des moments qui présentent un intérêt du point de vue de l'émotion.

Je vous propose un autre exercice. Je vais juste vous proposer de fermer les yeux et de prendre deux ou trois respirations profondes. Je vous propose de choisir maintenant un moment dans votre passé récent ou lointain, un moment de grande joie. Donc vous choisissez ce moment et puis vous essayez de vous mettre en contact avec ce qui se passe à ce moment-là, ce que vous ressentez au niveau de vos sensations. Est-ce que vous voyez quelque chose ? Est-ce que vous entendez quelque chose ? Est-ce que vous sentez quelque chose ? Est-ce qu'il y a des gens autour de vous que vous pouvez voir, que vous pouvez entendre ? Dont vous pouvez entendre la voix ? Quel temps fait-il ? Quelle température fait-il ? Comment êtes-vous

habillé ? Etc. Tranquillement, à votre rythme, je vous propose de revenir dans l'ici et maintenant et de rouvrir les yeux.

Vous pouvez constater qu'il n'y a pas une très grande différence entre les deux exercices, si ce n'est le cadre bien sûr. Et si ce n'est que dans le premier, en explicitation, l'étape de sensorialisation est juste un passage pour amener à l'évocation, pour ensuite permettre une description fine de l'action, de l'action physique ou de l'action mentale, de l'action cognitive, etc. Dans le deuxième cas, la mémoire sensorielle, on va rester avec la sensorialité aussi longtemps que nécessaire pour qu'émerge éventuellement une émotion. Et c'est ça qui est visé par un exercice de mémoire sensorielle.

Il y a plusieurs choses à noter en termes de différence. La première chose, c'est que quand on travaille sur la mémoire sensorielle, on commence généralement à travailler avec un professeur, donc on est guidé comme je vous l'ai fait là, mais on passe assez vite de l'évocation à l'auto-évocation. C'est à dire qu'on va développer très vite une expertise pour pouvoir le faire soi-même sans avoir besoin d'être guidé. Et donc en ce sens, on se rapprocherait presque plus de l'auto-explicitation et de l'auto-évocation qu'on fait en auto-explicitation plutôt que d'évocation guidée par quelqu'un d'extérieur. Donc ça nécessite une expertise d'évocation qui a été entraînée d'abord par le guidage extérieur, puis par l'auto-guidage. La pratique de l'évocation permet clairement de développer une expertise de l'évocation et donc d'acquérir une capacité à s'y mettre soi-même d'une part, mais aussi à s'y mettre plus rapidement.

Il y a une autre chose qui est intéressante, c'est que plus on travaille sur l'évocation d'un passé précis, d'une expérience vécue précise, plus il va être facile de l'évoquer. Donc il y a bien quelque chose de l'ordre de la stabilité de la mémoire d'un vécu, et d'une stabilité qui devient plus grande avec la pratique. Donc plus un comédien pratique la mémoire sensorielle, plus c'est facile de la pratiquer. Et plus il la pratique sur un vécu particulier, et plus c'est rapide de ré-évoquer ce vécu.

Là en revanche où il y a une différence importante, c'est par rapport à ce qui est visé, c'est-à-dire l'émotion qui surgit au présent dans le cas d'une mémoire sensorielle. Et cette émotion qui surgit au présent, elle, elle n'est pas nécessairement stable dans le temps. Par exemple, si vous faites l'évocation d'un moment joyeux du passé, peut-être que vous allez l'évoquer et ressentir de la joie. Mais peut-être qu'en l'évoquant, vous allez ressentir une certaine nostalgie, parce que le moment, maintenant, appartient au passé, et peut-être que l'émotion dominante qui va surgir, c'est plutôt de la nostalgie. Inversement, s'il vous est arrivé une grosse contrariété qui vous a mis dans un état de stress intense, voire traumatique, le fait de ré-

évoquer ce passé va peut-être vous faire sourire ou rire avec du recul. Donc d'une part, l'émotion qui va surgir, vous ne pouvez pas la contrôler, et d'autre part elle ne sera peut-être pas de même nature que l'émotion que vous avez ressentie au moment où vous avez vécu ce moment.

Mais en outre, à force de refaire une mémoire sensorielle donnée, il se peut qu'à un moment l'émotion change. C'est-à-dire qu'elle peut s'atténuer, ou vous pouvez l'avoir en quelque sorte épuisée. Et donc ce qui vous a servi de déclencheur, par exemple, pour une émotion de joie, à un moment peut ne plus fonctionner. Il se peut même qu'avec la même mémoire sensorielle, une autre émotion se mette à surgir au bout d'un moment. Et c'est en ça qu'on dit souvent dans cette méthode que les exercices de mémoire sensorielle sont un peu les gammes de l'acteur, c'est-à-dire que c'est un exercice qu'il faut faire régulièrement, revivre des expériences différentes, voir quelles émotions elles font surgir, faire une sorte de petit catalogue ou d'inventaire. Si j'ai besoin de telle émotion, je peux utiliser tel vécu. Mais sans oublier que ce n'est pas acquis et que ça peut changer, et qu'il faut donc le revisiter régulièrement.

Les gammes de l'acteur, c'est de trouver ces différents déclencheurs. Et il y en a potentiellement plein. Mais tout en sachant que dans ce cas-là, ce qu'on vise, c'est vraiment l'émotion, et donc ce qui va caractériser la qualité d'un moment passé, vécu, c'est plutôt sa capacité à faire surgir une émotion qu'on va pouvoir utiliser en tant que comédien. Ce ne sont donc pas du tout les mêmes critères que l'on va utiliser pour choisir les moments du passé. On est toujours sur des moments spécifiés dans le sens où on n'est pas dans le générique, dans le général, sinon il y a peu de chance qu'on fasse surgir l'émotion, mais en se rappelant que le but visé, c'est vraiment l'émotion.

Pour finir, ce qu'on appelle la mémoire affective est un exercice de mémoire sensorielle, mais qu'on va faire en choisissant d'emblée un moment d'un vécu du passé qui a une forte charge émotionnelle, donc qui peut être traumatique par exemple. Mais il y a une précaution importante, quand on effectue cet exercice, c'est que d'une part on ne le fait pas avec tout le monde, c'est-à-dire qu'un professeur aguerri va choisir délibérément avec qui le faire pour des raisons bien particulières. Et l'autre précaution sur laquelle Lee Strasberg insistait, c'est que dans ce cas-là, on va choisir un moment d'un passé plus lointain parce qu'il faut avoir fait le travail d'élaboration pour pouvoir en dépasser justement le côté traumatique, mais avoir encore ce côté émotionnel. Lee Strasberg disait de choisir un moment du passé ancien d'au moins dix ans. En général, on effectue des exercices de mémoire affective avec des comédiens qui ont un accès peut-être plus difficile à leurs émotions. Et du coup, il faut aller chercher des vécus un peu plus chargés pour faire surgir l'émotion. Mais pour des comédiens pour qui les émotions arrivent facilement par le biais de simples exercices de mémoire

sensorielle, on va rarement aller chercher ce genre de mémoire affective parce qu'elles sont évidemment beaucoup plus chargées.

En guise de conclusion : l'importance du sensoriel pour entrer en évocation

En conclusion, ce qui est important, je pense, c'est l'importance du sensoriel comme voie d'accès à l'évocation, que ce soit pour faire de l'explicitation, ou pour faire surgir une émotion pour le comédien. Démarrer un entretien d'explicitation en étant vraiment dans le sensoriel permet donc de faciliter grandement la mise en évocation. Mais dans le cas d'un entretien d'explicitation, on souhaitera passer dès que possible du sensoriel à la description de l'action, puisque c'est le but visé d'une part, mais aussi pour éviter le surgissement éventuel de l'émotion. Dans le cas d'un exercice de mémoire sensorielle pour le comédien, on va au contraire rester le plus longtemps possible dans le sensoriel, en espérant que justement l'émotion surgisse. Et puis, dans ce dernier cas, on aura en outre choisi délibérément un moment du passé à charge émotionnelle plus forte, pour favoriser justement le plus possible ce surgissement de l'émotion.

Financement

Cette publication a été rendue possible grâce à un financement de l'Initiative d'excellence d'Aix-Marseille Université, A*Midex, qui fait partie du Programme d'Investissements d'Avenir - Financement No. : AMX-19-IET-005.

Bibliographie

Cohen, L. (2010). *The Lee Strasberg Notes*. New York: Routledge, 201 p.

Cohen, L. (2017). *The Method Acting Exercises Handbook*. New York: Routledge, 222 p.

Magrin-Chagnolleau, I. La Méthode de l'Actors Studio : Genèse, Pratique et Phénoménologie. In *Être en scène, être en jeu. Statuts, esthétiques, processus et techniques pour l'interprète et le performeur*. In Press. <https://hal.archives-ouvertes.fr/hal-02342281>

Magrin-Chagnolleau, I. (2021). Actors Studio. In *Encyclopædia Universalis*.

Maurel, M. (2022). Université d'Été de Goutelas 2022 : Décrire le remplissage intuitif typique de l'évocation. In *Expliciter*, 135, p. 1-21.

Strasberg, L. (1987). *A Dream of Passion*. New York: Plume, 202 p.

Strasberg, L., and Hethmon, R. H. (1965). *Strasberg at The Actors Studio*. New York: Theater Communications Group, 410 p.

Vermersch, P. (1994). Projet pour une analyse phénoménologique de la conduite d'évocation. In *Expliciter*, 6, p. 4.

Vermersch, P. (1995). L'évocation, un objet d'étude ? In *Expliciter*, 8, p. 4-8.

Vermersch, P. (2007). Bases de l'auto-explicitation (1). In *Expliciter*, 69, p. 1-31.

Vermersch, P. (2008). Introspection et auto-explicitation : Bases de l'auto-explicitation (2). In *Expliciter*, 73, p. 42-56.

Vermersch, P. (2012). *Explicitation et phénoménologie*. Paris : PUF, 472 p.

Vermersch, P. (2014). *L'entretien d'explicitation*. Issy-les-Moulineaux : ESF Éditeur, 208 p.

Le rôle de l'auto-explicitation en recherche (2023)

Ivan Magrin-Chagnollet. **Le rôle de l'auto-explicitation en recherche.** *Expliciter*, 2023, Hors-série (1), pp.105-111.

7 pages

Le rôle de l'auto-explicitation en recherche

Ivan MAGRIN-CHAGNOLLEAU, CNRS

Aix-Marseille Univ, CNRS, PRISM, Marseille, France
Chapman University, Orange, California

Introduction

Dans cet article, je vais vous parler d'auto-explicitation et de son utilisation en recherche.

Ce que j'ai observé en discutant avec plusieurs personnes sur comment elles font de l'auto-explicitation, c'est qu'avec le temps, on développe chacun ses propres stratégies, ses propres façons de faire, et on découvre au fur et à mesure ce qui fonctionne pour nous, ce qui ne fonctionne pas, etc.

Pour démarrer et entrer vraiment dans le vif du sujet, il y a deux articles de base. Ils s'appellent d'ailleurs « Bases de l'auto-explicitation ». Ce sont des articles à lire et à relire quand on travaille sur l'auto-explicitation.

Un premier article de base

Le premier est dans le numéro 69 de la revue *Expliciter* de mars 2007¹ et fait 31 pages, donc il y a de quoi lire. Il est vraiment intéressant parce que, bien que ça semble être un article sur l'auto-explicitation, en fait, c'est un article sur plein d'autres choses, et notamment sur tout ce qui a conduit Pierre Vermersch à développer cette technique d'auto-explicitation. Je voudrais citer quelques passages de cet article qui sont pour moi essentiels, notamment par rapport à ce que je vais dire dans la suite de cet article.

La différence essentielle entre les deux techniques, donc à savoir l'explicitation et l'auto-explicitation, va se situer entre la conduite de guidage et de l'auto-guidage, avec les conséquences pratiques qui en découlent. (page 19)

¹ Vermersch, P. (2007). Bases de l'auto-explicitation (1). In *Expliciter*, 69, p. 1-31.

Et c'est un point essentiel. Dans l'auto-explicitation, on se guide soi-même, donc on n'est plus un A² qui se laisse guider ou un B qui guide un A, mais on est à la fois le « A qui sait »³, et on est aussi celui qui guide A. Et donc tout l'enjeu de l'auto-explicitation, c'est de trouver cet équilibre, de comprendre à quel moment on a besoin de se reguider et de savoir aussi se remettre en évocation. Et donc Pierre Vermersch distingue deux types de guidage, ce qu'il appelle le « guidage relationnel » et ce qu'il appelle le « guidage cognitif ».

Au niveau du guidage relationnel, il dit quelque chose qui pour moi est important :

Dans l'auto-explicitation en revanche, l'informateur se trouve seul, mais pour autant toutes les attentions relationnelles qui sont mobilisées par un intervieweur doivent aussi être présentes. Commencer une auto-explicitation sans avoir vérifié que l'on est bien d'accord avec soi-même pour le faire, que l'on est au clair avec le but poursuivi ou avec la décision de ne pas s'en donner immédiatement, renouveler cet accord avec soi-même quand les difficultés surviennent etc. tout cela doit faire l'objet d'une négociation avec soi-même, au risque de rester coincé si ce n'est pas fait, ou si c'est mal fait. (Pages 19-20)

Pour celles et ceux qui pratiquent l'auto-explicitation régulièrement, on a tous des expériences où on a complètement oublié de se donner l'accord à soi-même et où ça ne marche pas du tout. Et on ne comprend pas pourquoi jusqu'à ce que tout à coup on se dise : « Ah mais oui, peut-être que je ne suis pas d'accord pour faire cette auto-explicitation. » C'est vraiment un point important.

Ensuite, au niveau du guidage cognitif, il revient sur cette question de l'expertise qui est nécessaire quand on fait de l'auto-explicitation :

Rappelons qu'il faut acquérir une double expertise, celle de l'introspection rétrospective dans son aspect élucidant, descriptif, analyseur, et celle de l'introspection actuelle. La seconde sans la première ne donne pas une compétence de chercheur, la première sans la seconde ne permet pas la pratique de l'auto-explicitation. (page 24)

C'est vraiment quelque chose de crucial parce que ça pose vraiment le cadre de quand est-ce que c'est de l'auto-explicitation et quand est-ce que c'est de l'auto-explicitation à visée de recherche. Cette façon de faire de l'introspection est double : elle a besoin d'être élucidante et analysante pour qu'on puisse parler d'une situation de recherche, mais elle a aussi besoin d'être actuelle, c'est-à-dire de questionner l'actualité de l'expérience vécue pour qu'il s'agisse d'auto-

² Dans la nomenclature de l'explicitation, A désigne la personne qui est interviewée, et B désigne la personne qui conduit l'entretien.

³ Cette formule renvoie au fait que dans un entretien d'explicitation, la personne interviewée, A, est celle qui est experte de son passé vécu.

explicitation, et pas simplement d'introspection. Autrement dit, l'auto-explicitation n'est pas une introspection comme les autres, elle suit un protocole bien précis, qui donne à la fois accès à l'expérience vécue du sujet qui conduit cette auto-explicitation, mais d'une manière éclairante, descriptive, analysante.

Au sujet de cette expertise, il dit aussi :

La pratique du dessin de vécu (Vermersch 2005) comme j'aime nommer l'auto-explicitation est une pratique experte qui s'apprend, qui s'exerce, qui se développe et se perfectionne et qui se transmet. (page 24)

Cette question de l'expertise est vraiment essentielle pour comprendre le rôle de l'auto-explicitation en recherche. Pour conduire une auto-explicitation à visée de recherche, la personne qui fait cette auto-explicitation doit d'abord avoir une expertise en auto-explicitation, c'est-à-dire à la fois une expertise en tant que A (personne interviewée), une expertise en tant que B (personne interviewante), mais aussi une expertise de passage de l'un à l'autre, c'est-à-dire la capacité à aller en évocation en tant que A, puis à diagnostiquer en temps réel le besoin d'en ressortir pour jouer le rôle de B, dans le but d'y retourner pour être de nouveau A en évocation, etc. Mais il faut également une expertise au niveau de la tâche qui est étudiée par le dispositif de recherche. Je vais donner un exemple un peu plus loin.

Un deuxième article de base

L'autre article, qui est la deuxième partie, est dans le numéro 73 de février 2008⁴. C'est un article plus court. Je voudrais citer le préambule qui est très éclairant :

Ce texte avait pour idée initiale d'essayer de mettre de l'ordre dans les rapports entre le thème du « sens se faisant », et la méthode de production de données de verbalisation désignée par le terme « auto-explicitation ». En chemin, il a bifurqué et il est devenu un retour sur la notion même d'introspection et les différentes pratiques qui peuvent être nommées ainsi, renvoyant à plus tard la clarification de l'intention initiale relative au « sens se faisant » et la publication du texte déjà existant. (page 42)

Ça, c'est aussi vraiment important parce que du coup, évidemment, le cœur de l'auto-explicitation, c'est l'introspection, mais c'est une certaine forme d'introspection et avec des visées particulières.

⁴ Vermersch, P. (2008). Introspection et auto-explicitation : Bases de l'auto-explicitation (2). In *Expliciter*, 73, p. 42-56.

Il est aussi question ici du « sens se faisant », qui est aussi une spécificité intéressante de l'auto-explicitation. Pour le dire autrement, lorsqu'on fait une auto-explicitation, on découvre au fur et à mesure l'intérêt d'une telle pratique au moment même où on la fait, ce qui nécessite un va-et-vient permanent entre une posture de lâcher-prise de A qui s'en remet à B pour être guidé, et en même temps une interrogation continue de ce qu'on recherche afin de pouvoir jouer le rôle de B et de décider si et comment réorienter l'entretien quand nécessaire, ainsi que changer de moment spécifié, ou faire de la fragmentation, etc.

Une découverte de l'auto-explicitation

Je voudrais maintenant vous rappeler brièvement comment je suis arrivé à l'auto-explicitation. Je pense que, comme la plupart d'entre vous qui avez fait un stage de base, vous avez dû commencer à toucher à l'auto-explicitation pendant ce stage de base. Ça a été aussi mon cas.

Et puis l'année d'après, comme j'ai vraiment voulu creuser cet outil d'auto-explicitation, j'ai fait un stage d'auto-explicitation avec Pierre Vermersch, que j'avais déjà rencontré à quelques reprises lors de séminaires du GREX2, mais avec qui je n'avais pas vraiment eu d'échanges soutenus et de vraies rencontres. C'est vraiment au cours de ce stage que ces rencontres se sont passées et c'est vraiment au cours de ce stage que j'ai vraiment commencé à comprendre ce que c'était vraiment que l'auto-explicitation, comment ça pouvait être utilisé, et que j'ai commencé à jouer un peu sur les liens entre entretien d'explicitation, auto-explicitation, qu'est-ce qu'on peut faire avec l'un, qu'est-ce qu'on fait avec l'autre. Qu'est-ce qu'on fait de toutes ces données qualitatives, comment on les utilise, comment on peut publier sur ce genre de travaux ? Qu'est-ce qu'on fait des données recueillies, comment on recueille les données et qu'est ce qui est pertinent dans les données ? etc.

Pourquoi utiliser l'auto-explicitation en recherche

Je voudrais maintenant parler de trois choses qui éclairent le rôle de l'auto-explicitation en recherche. La première, c'est que pour moi, le gros avantage de l'auto-explicitation, surtout dans un contexte où on est soi-même expert de la tâche qu'on veut interroger, c'est la réduction du coût. En effet, ce qui est compliqué avec un entretien d'explicitation, c'est son coût. Ça coûte en temps, ça coûte en temps de l'autre, parce que du coup on interroge un A mais il faut que le A soit d'accord. Et puis si on n'est pas tout à fait sur le bon moment, il faut refaire un deuxième entretien derrière. Et puis si on n'a pas assez fragmenté, il faut refaire un troisième entretien derrière, etc. Donc à moins d'avoir des A extrêmement complaisants et qui n'ont que ça à faire, c'est un peu compliqué. L'intérêt d'une auto-explicitation dans ce contexte, encore une fois pour une tâche où on peut soi-même être un expert, ou en tout cas suffisamment expert, c'est de

pouvoir un peu défricher la tâche, de pouvoir commencer aussi à réfléchir à la typologie de la tâche. C'est quelque chose d'important pour commencer à comprendre comment elle se déroule. Et puis ça permet aussi de commencer à intuiter quels peuvent être les moments spécifiés intéressants par rapport à l'objet d'étude. Donc ça réduit le risque de faire des entretiens dont on ne pourrait rien faire parce qu'on ne serait pas allé au bon endroit.

Ceci étant dit, et là on revient à « C'est A qui sait », c'est aussi intéressant de faire quelques entretiens sans savoir, pour justement laisser A et son expertise nous guider vers les moments qui lui semblent intéressants et qui peut-être le seront aussi pour notre objet de recherche.

Finalement, c'est un peu un jeu entre les deux. À mon avis, il ne faut pas en savoir trop avant de commencer à faire les premiers entretiens d'explicitation, sinon on risque de projeter des idées qu'on a sur le sujet. Et du coup la recherche est verrouillée parce qu'elle est déjà contrainte par ces idées. Mais inversement, on ne peut pas faire des tas et des tas d'entretiens, d'abord parce que les A ne s'y prêteraient pas forcément plusieurs fois de suite, et aussi parce que plus on fait des entretiens, plus il y a des données à analyser et plus ça prend du temps. Donc c'est vraiment un jeu entre les deux. Et encore une fois, c'est à condition uniquement de pouvoir être soi-même un A relativement expert par rapport à la tâche visée.

Une typologie de l'auto-explicitation

Le deuxième point, c'est le travail qui est fait en université d'été. J'ai eu la chance d'y participer cette année pour la première fois et du coup de vraiment bien comprendre comment ça se passe. Cela m'a permis de mettre en évidence trois types d'auto-explicitation, si on fait une typologie de l'auto-explicitation. Il y a les auto-explicitations qu'on fait avant pour avoir une idée de la tâche (ce dont je viens de parler dans le paragraphe précédent), pour voir comment ça va se passer. Il y a les auto-explicitations qu'on fait pendant, c'est-à-dire des sortes de va-et-vient. On fait de l'explicitation, on fait un peu d'auto-explicitation, on revient à l'explicitation. Et puis il y a les auto-explicitations qu'on peut faire après, pour creuser un entretien d'explicitation lorsqu'on a été A et qu'on analyse l'entretien. Ces trois types d'auto-explicitations ont des natures différentes. Elles n'interrogent pas les mêmes choses et du coup elles ne se mènent pas tout à fait de la même façon.

Un exemple d'utilisation de l'auto-explicitation en recherche

Pour donner un exemple concret, je vais expliquer comment j'utilise l'auto-explicitation dans mon propre travail de recherche. Je travaille essentiellement sur le processus créatif, le processus de création. Et donc, étant artiste moi-même, je peux du coup être aussi mon propre A. Mais je ne m'en tiens pas uniquement à ce que je peux faire en auto-explicitation. Je vais aussi faire des

entretiens d'explicitation avec d'autres artistes sur leur processus de création. Et donc c'est vraiment une histoire de va-et-vient, c'est à dire c'est un peu ce que je disais ci-dessus : c'est à la fois défricher en travaillant en auto-explicitation, en allant voir déjà comment je fonctionne, comment je travaille, quelles sont les différentes étapes qui m'intéressent. Et là, ce qui est bien, c'est que je peux y consacrer le temps que je veux, je peux y revenir, je peux aller sur différents moments, je peux fragmenter, je peux faire à peu près tout ce que je veux. Et ensuite, j'interroge d'autres artistes.

En fait, la recherche se mène en cheminant avec ces deux outils. Je peux avoir une idée, je peux la tester sur moi, imaginer quels peuvent être les moments intéressants pour creuser cette idée, tester avec un artiste en lui proposant de faire un entretien. En faisant l'entretien avec l'artiste, qui lui a une autre façon de fonctionner, je vais découvrir des choses auxquelles je n'ai pas pensé. Je vais aller les interroger en auto-explicitation, ce qui va me donner d'autres idées. Je vais repartir en entretien d'explicitation, ou bien avec le même artiste, ou bien avec un autre artiste, et ainsi de suite.

Ce qui est intéressant dans cette démarche itérative, c'est qu'au fur et à mesure, le propos s'affine, les moments spécifiés s'affinent, le type de fragmentation dont j'ai besoin s'affine, le type d'informations auxquelles je peux accéder s'affine, et c'est comme ça que se construit la recherche. Il arrive un moment finalement où l'analyse des données devient en quelque sorte plus simple, parce que je commence à avoir des entretiens qui sont mieux ciblés, qui sont mieux conduits, mieux pas au sens de l'expertise de B, mais au sens du sujet de recherche choisi et visé. Ça permet petit à petit d'avancer et de s'approcher au plus près de mon sujet.

Pour conclure

Il y a toujours cette idée du contrat qu'on passe, ou bien avec l'autre, ou bien avec soi-même quand on fait de l'auto-explicitation. Et dans les deux cas, ça me paraît important d'être toujours dans une posture d'accueil et d'accepter de se laisser détourner - par les informations qu'on acquiert - de l'idée qu'on se faisait de son sujet de recherche. Parce que sinon on perd un peu la force et l'intérêt de ce genre d'approches.

Financement

Cette publication a été rendue possible grâce à un financement de l'Initiative d'excellence d'Aix-Marseille Université, A*Midex, qui fait partie du Programme d'Investissements d'Avenir - Financement No. : AMX-19-IET-005.

Bibliographie

- Balas-Chanel, A. (2007). Mes premières expériences d'auto-explicitation. In *Expliciter*, 69, p. 55-67.
- Berger, È. (2007). Auto-explicitation du vécu du corps sensible : Éléments méthodologiques et expérientiels pour la description de l'expérience corporelle interne à visée de recherche. In *Expliciter*, 69, p. 32-54.
- Crozier, J. (2008). Mon expérience du stage d'auto-explicitation. In *Expliciter*, 74, p. 34-37.
- Hatier, C. (2010). Le stage de formation à l'auto-explicitation de janvier 2010 à Paris. In *Expliciter*, 85, p. 33-35.
- Magrin-Chagnolleau, I., Nahoum, R., Weisen, M. (2019). Le projet Sipapu : vers une micro-phénoménologie de la création. In *Les enjeux cognitifs de l'artefact esthétique : tome 2*, Paris : L'Harmattan, pp.187-194. <https://hal.archives-ouvertes.fr/hal-03511242/document>
- Maurel, M. (2007). Petites histoires d'auto explicitation. In *Expliciter*, 68, p. 22-29.
- Snoeckx, M. (2007). Les puissances dormantes : Témoignage du stage d'auto-explicitation. In *Expliciter*, 70, p. 10-14.
- Vermersch, P. (2005). Éléments pour une méthode de « dessin de vécu » en psychophénoménologie. In *Expliciter*, 62, p. 47-57.
- Vermersch, P. (2007). Bases de l'auto-explicitation (1). In *Expliciter*, 69, p. 1-31.
- Vermersch, P. (2008). Introspection et auto-explicitation : Bases de l'auto-explicitation (2). In *Expliciter*, 73, p. 42-56.
- Vermersch, P. (2008). Auto-explicitation d'un exemple d'émergence de sens par mobilisation du « sens corporel » (focusing). In *Expliciter*, 76, p. 22-31.
- Vermersch, P. (2008). Contenu de la formation à l'auto-explicitation et à l'introspection descriptive. In *Expliciter*, 76, p. 32-33.
- Vermersch, P. (2012). *Explicitation et phénoménologie*. Paris : PUF, 472 p.
- Vermersch, P. (2014). *L'entretien d'explicitation*. Issy-les-Moulineaux : ESF Éditeur, 208 p.

Le rôle du corps dans la mise en évocation (2023)

Magali Boutrais, Ivan Magrin-Chagnolleau, Maryse Maurel, Patricia Rottement. **Le rôle du corps dans la mise en évocation, le remplissement intuitif et la présentification dans le cadre d'un entretien d'explicitation.** *Expliciter*, 2023, 137, pp. 37-41.

5 pages

Le rôle du corps dans la mise en évocation, le remplissement intuitif et la présentification dans le cadre d'un entretien d'explicitation

Magali Boutrais ⁽¹⁾⁽²⁾, Ivan Magrin-Chagnolleau ⁽¹⁾⁽³⁾⁽⁴⁾, Maryse Maurel ⁽¹⁾, Patricia Rottement ⁽¹⁾

⁽¹⁾ GREX2, Paris, France

⁽²⁾ Université de Picardie, Amiens, France

⁽³⁾ Aix-Marseille Univ, CNRS, PRISM, Marseille, France

⁽⁴⁾ Chapman University, Orange, California

Résumé

Dans cet article, nous nous intéressons au moment où, dans un entretien d'explicitation, le sujet de l'entretien (A) est en mesure de dire qu'il a trouvé un moment spécifié en réponse à l'intention éveillante que l'intervieweur (B) lui a proposé. Qu'est-ce qui permet à A de pouvoir dire que c'est bon ? Y a-t-il des corrélats à cette prise de conscience ? Nous proposons dans cet article la notion d'ancrage corporel, c'est-à-dire que l'un des corrélats serait que A s'ancre tout d'un coup corporellement dans son revécu.

Mots clés

Évocation, remplissement intuitif, présentification, ancrage corporel, explicitation.

Introduction

Le travail présenté dans cet article est le résultat d'expériences réalisées au cours de l'université du GREX2¹ pendant l'été 2022. Le sujet d'étude de cette université d'été était le remplissement intuitif, c'est-à-dire la façon dont le sujet d'un entretien d'explicitation² (A)³ entre en évocation. Nous nous intéressons particulièrement à la façon dont ce remplissement menait à une présentification, c'est-à-dire à la conscience d'être vraiment présent dans la situation que l'on est en train d'évoquer.

Notre groupe de travail était composé de Magali Boutrais, Ivan Magrin-Chagnolleau, Maryse Maurel et Patricia Rottement. Notre protocole, qui a été décrit précédemment dans un article par Maryse Maurel⁴ était le suivant : nous proposons à A de se remémorer un moment de son passé

¹ Groupe de Recherche sur l'EXplicitation 2.

² Vermersch, P. (2014). *L'entretien d'explicitation*. ESF Éditeur.

³ On rappelle que dans la terminologie de l'explicitation, A est la personne interviewée, B la personne qui conduit l'entretien, et C la personne qui observe l'entretien en train de se faire et peut éventuellement noter des observations au sujet de cet entretien.

⁴ Maryse Maurel. (2022). Université d'été de Goutelas 2022 : Décrire le remplissement intuitif typique de l'évocation. *Expliciter*, 135.

en lui proposant une consigne que nous souhaitons surprenante, et nous faisons donc un premier entretien d'explicitation avec A sur ce premier vécu de référence, le V1. Puis dans un deuxième temps, nous faisons un second entretien d'explicitation avec le même A sur le vécu de ce premier entretien, le V2. Comme nous étions un groupe de quatre, il y avait un A et généralement plusieurs B à tour de rôle, les B n'étant pas en train de questionner pouvant prendre temporairement un rôle de C.

La particularité de notre groupe, c'est d'avoir décidé de prendre tous les quatre à tour de rôle le rôle de A. Nous avons donc réalisé quatre entretiens sur un V1 et quatre entretiens sur un V2. Nous avons donc un matériau très intéressant pour pouvoir faire des études comparatives. C'est cette approche comparative que nous avons choisie pour cet article.

Nous avons choisi ici de nous focaliser sur le rôle du corps au moment précis où le A annonce qu'il a trouvé le moment spécifié proposé par l'intention éveillante de B.

Le rôle du corps

Pourquoi s'intéresser au rôle du corps dans le remplissage intuitif, et de quel corps parle-t-on ? Nous avons choisi de nous intéresser au rôle du corps car nous avons noté dans les quatre entretiens sur les V2 qu'il y avait quelque chose de l'ordre d'un **ancrage corporel** au moment du « Je l'ai ».

Cela a commencé avec la prise de conscience d'Ivan qu'au moment de son « Je l'ai », il y avait pour lui quelque chose de l'ordre d'un « posement » de son corps devant le carton en même temps que le carton se posait⁵. Nous avons donc été alors attentif au rôle du corps lors des autres entretiens. Nous avons pu noter que cet ancrage corporel s'était manifesté de quatre façons différentes, mais ce sont toutes des expressions qui réfèrent d'une façon ou d'une autre à un ancrage corporel. Nous allons y revenir plus en détails.

Mais avant, nous souhaiterions préciser de quel corps il s'agit. Il s'agit du corps de A qui se déplace dans son vécu de référence V1 lors du premier entretien, donc du corps de A qui revisite son passé, à la recherche du moment spécifié induit par la consigne de B.

Et cela est interrogé ensuite au cours de l'entretien sur le V2, c'est-à-dire le deuxième entretien qui cherche à explorer comment le sujet A s'est mis (ou a été mis) progressivement en évocation et a entrepris son remplissage intuitif sous la conduite de B, et ce jusqu'à la présentification.

Explorons maintenant cette thématique dans chaque entretien réalisé :

- A1 : Ivan (V2 du V1) – B1 : Magali.
A1' : Ivan (V3 du V2) – B1' : Maryse, puis B1'' : Patricia.
Consigne : « ... le moment où tu as ouvert le premier carton dans ton nouvel appartement. »
- A2 : Patricia (V2 du V1) – B2 : Maryse.
A2' : Patricia (V3 du V2) – B2' : Magali, puis B2'' : Ivan.
Consigne : « ... le moment de ton arrivée au château où pour la première fois tu as dit bonjour à quelqu'un du GREX. »

⁵ Voir les extraits de l'entretien un peu plus loin dans l'article.

- A3 : Maryse (V2 du V1) – B3 : Patricia.
A3' : Maryse (V3 du V2) – B3' : Magali.
Consigne : « ... le moment où tu as planté dans ton jardin le dernier légume que tu as pu manger. »
- A4 : Magali (V2 du V1) – B4 : Ivan.
A4' : Magali (V3 du V2) – B4' : Maryse.
Consigne : « ... le moment où tu as joué de ton nouveau traverso en bois, pour la première fois. »

Nous nous intéressons ici plus particulièrement aux entretiens notés A', c'est-à-dire sur le V3 du V2. Il s'agit donc de repérer dans chacun de ces quatre entretiens le ou les moments où nous apprenons quelque chose au sujet du corps de A tel qu'il le perçoit dans son revécu de l'entretien de référence V2 du V1.

Commençons par le premier entretien, celui où Ivan est A. Voici les passages pertinents par rapport à notre sujet d'étude :

49. I. C'est, c'est le moment où Magali commence à me questionner et là, je me vois en fait devant le carton et je me vois prendre le cutter dans ma poche droite et je me vois m'accroupir pour couper le premier morceau. Et du coup, là, il y a vraiment quelque chose de...

50. Ma. Tu y es.

51. I. J'y suis. C'est physique, il y a ... Mon corps est dedans. Avant, mon corps il est pas dedans. C'est plus quelque chose qui se promène, qui se balade, même quand le carton il est stable.

52. Ma. Qu'est-ce que ça fait comme différence entre le moment où le corps, il est pas dedans, et là tu y es ? Qu'est-ce que ... Qu'est ce qui fait la différence ?

53. I. Bah je pense que... Ça serait à vérifier. On les a pas filmé les entretiens, mais je pense que c'est le moment où mon corps commence à être sollicité dans l'entretien en fait, c'est-à-dire que ça devient corporel, je sais pas, j'ai l'impression de...

67. I. Il y a un ancrage là et l'ancrage je pense qu'il commence avec le carton qui trouve sa bonne place sur le sol. Donc il y a une certaine stabilité mais dans le même moment, du coup moi je suis aussi en face du carton donc je suis moi-même stable.

71. I. Mais je pense que la présentification, elle commence vraiment quand le carton trouve sa place et dans la même fraction de seconde, du coup, le corps y est aussi comme si les deux se posaient en même temps.

73. I. Ah oui, en fait, il y a pour moi, il y a vraiment une différence entre les deux moments que j'ai décrits, c'est à dire il y a le moment où, où je sais quel est le premier carton que j'ai ouvert, mais je sais aussi que j'y suis pas encore, c'est à dire que je l'ai retrouvé, mais c'est pas encore présentifié, c'est pas encore, je suis pas encore moi, là, au moment de mon revécu où ça va, où ça se passe, quoi.

77. I. Ben c'est pas incarné, c'est-à-dire c'est, ça reste quelque chose qui est plus en fait, c'est un peu comme si j'avais la capacité à ce moment-là toujours, d'avoir une sorte de vue d'oiseau, la *bird's eye*, et que je peux aller un peu où je veux, me balader, tout ça. Je suis encore un peu dans ce, dans ce truc flottant. Et du coup, la différence

c'est que quand je, quand je viens à l'endroit du carton, là je sais que je suis à l'endroit, même si je ne suis pas encore ancré. Et quand le carton se pose, là, du coup le corps se pose, c'est-à-dire là j'ai plus cette capacité de, je l'ai toujours mais en fait je la ressens plus. C'est à dire que là...

78. Mg. Quand tu la ressens, tu la ressens où, comment ? Peut-être dans ton corps, peut être autrement ?

79. I. Ouais, c'est un truc un peu aérien, un peu éthéré quoi, un peu, ouais, par là plutôt [montre vers la tête]. Alors que, alors que quand je suis posé, que le carton est posé, c'est tout le corps. C'est pas juste cette partie.

On voit à travers ces extraits d'entretien qu'il se joue quelque chose de l'ordre de l'ancrage corporel. A perçoit qu'il y est au moment où son corps se pose en même temps que le carton, devant le carton, au moment où il y a un ancrage corporel. Voyons voir si nous pouvons retrouver quelque chose de similaire dans les trois autres entretiens.

Dans l'entretien suivant, c'est Patricia qui est A. Voici les passages pertinents que nous avons relevés :

| | |
|-----|--|
| A7 | Eh bien j'étais dans une autre dimension, avec quelque chose de physique. Je me sentais corporellement. Je me sentais connectée. Nicolas...Bonjour...Vous avez remarqué...la voiture, les étapes, je faisais beaucoup de choses en même temps. Ça y est, des choses me reviennent. Donc j'ai quitté la cour. Waouf une sorte de marche arrière. Et je me suis posée comme j'étais, sur le parking. |
| A8 | Quand j'ai eu le gros plan du visage de Nicolas, son sourire. On s'est regardés. C'est à ce moment-là que j'ai dit que c'était bon |
| A11 | Une tension corporelle, debout tendu et puis de se dire bonjour d'accueillir, j'ai la tension, et le contact c'est ça qui était complètement vrai je ne l'ai pas vu de l'extérieur, je l'ai senti de l'intérieur de mon corps. |
| B12 | Ça te paraît stable à ce moment-là ? |
| A12 | Oui |
| B13 | Comment tu sais que c'est stable ? |
| A13 | Je peux le vérifier dans mon corps |
| B14 | Quand tu vérifies, comment tu fais ? |
| A14 | J'ai une conscience de mon corps |
| B15 | Un endroit particulier ? |
| A15 | Une tension verticale, la tension des muscles, la tension debout. Une ouverture comme ça (épaules-thorax) un ressenti de l'intérieur. Et quand j'ai ça, je sais que c'est vrai. Comme je pratique beaucoup, c'est mon accès... |

Il y a aussi quelque chose de l'ordre du corporel qui se joue ici. A utilise l'expression « Je me suis posée », ce qui est très similaire au fait de se poser en même temps que le carton du premier entretien. A dit aussi : « Je peux le vérifier dans mon corps », ce qui indique également clairement quelque chose de corporel.

Prenons maintenant l'entretien suivant. Pour celui-ci, nous n'avons pas conduit un entretien V3 du V2 à proprement parler, mais plutôt sous forme de « mine de rien », de relances et de commentaires.

Mais on a pu relever quand même quelques énoncés très éclairants sur le rôle du corps :

Je bougeais les pieds parce que j'allais d'une scène à l'autre, je n'arrivais pas à me stabiliser.

Je vois la scène, je suis fixée, mes pieds se stabilisent, je ne suis plus en recherche, je ne flotte plus.

En même temps que j'ai réintégré mon corps, et en même temps, j'ai perçu la scène comme si j'y étais.

Nous pouvons voir qu'une fois encore, le rôle du corps est prépondérant dans la stabilisation et la complétude de l'évocation. Le « Je l'ai » correspond ici aux pieds qui se stabilisent et à la réintégration du corps.

Voyons ce qu'il en est dans le dernier entretien :

72. M. Et après je me dis « mais, mais la première fois que j'en ai joué » bon, je ce qui me vient c'est chez moi moi avec

73 Ma. Et là t'as quelque chose qui te vient ? (2 s.)

74. M. Là, c'est...là je pense que j'étais dans la position... c'est très...c'est pas une image... c'est moi avec la flûte.

75. Ma. D'accord.

76. M. Enfin c'est la sensation de la flûte, la sensation de quand je joue de la flûte. (geste de jouer de la flûte traversière)

252. M. Alors quand je me suis, donc, volontairement, je me suis remise, je m'suis dit « faut y aller, qu'est-ce qui te revient du moment où t'es allée chercher la flûte ? », et là c'était moi, dans mon corps devant la sonnette. Je me suis pas vu faire, c'était moi, j'avais pas de décalage.

Nous pouvons noter ici encore quelque chose de l'ordre de l'incarnation : « c'était moi, dans mon corps ».

Conclusion

En conclusion, nous pouvons noter que dans les quatre entretiens, il semble bien y avoir un lien entre la présentification et le corps. Dans chaque cas, il y a quelque chose de l'ordre de l'ancrage corporel au moment du « Je l'ai ».

Pour aller plus loin, nous pourrions nous demander si le moment du « Je l'ai » caractérise la même chose pour nous quatre. En effet, le fait qu'il semble que nous puissions corrélérer ce moment à une forme d'ancrage corporel ne garantit pas nécessairement que ce moment correspond exactement à la même expérience vécue.

Références

Maryse Maurel. (2022). Université d'été de Goutelas 2022 : Décrire le remplissage intuitif typique de l'évocation. *Expliciter*, 135.

Vermersch, P. (2014). *L'entretien d'explicitation*. ESF Éditeur.

Vermersch, P. (1996). Pour une psycho phénoménologie /2 Problèmes de validation. *Expliciter*, 14.

3.13 Psychologie / Psychanalyse

The Madness of Making Sense : Madness and Integrative Knowledge (2016)

Ioana Zamfir et Ivan Magrin-Chagnolleau. **The Madness of Making Sense : Madness and Integrative Knowledge**. Dans *Schizo : The Liberatory Potential of Madness*, édité par Irina Lyubchenko et Fiona Ann Papps. IDN Press, 2016. ISBN : 978-1-848-88460-1.

9 pages

The Madness of Making Sense: Madness and Integrative Knowledge

Ioana Zamfir and Ivan Magrin-Chagnolleau

Abstract

Madness implies a radically different view on the world than the common sense - radically different and beyond the verge of what can be called real, rational, accurate. Madness implies perceiving other sides of the given reality, sides that very often remain, for the rest of us, hidden, obscured, un-obvious. Consequent to this particular un-usual perception, madness also implies an understanding of these perceptions and an adequate expression of them. The reason why madness is kept into a closed system of queerness and meaninglessness might be the very coherence that exists between all these mind functions - perception, understanding and expression - for the mad person.

Madness can be seen as a particular kind of perceptual awareness, a sense for the paradoxical dimension of the world. Whereas a “healthy” mind lives undisturbed by frequent contradictions and loops which characterize commonsensical reasoning, madness functions as a type of knowledge adapted to dealing and going beyond inconsistency and imprecision. Therefore, studying madness complements the understanding of how the mind and consciousness work, and questions the limits of representation and communication.

Key Words: knowledge, madness, representation, experience.

1. Introduction

Madness is a wide comprising term, which can refer to subject matters varying from creativity and innovation to excess, distortion, violence and suffering. This paper will refer strictly to the connection between madness and knowledge: how can we know about madness and what does madness teach us about human nature?

Far from endorsing a sentimental or idealizing view on madness, what we claim is that, from an anthropological point of view, it is both unconstructive to try and impossible to succeed in delineating a precise and fair border between originality/authenticity and madness. The lack of a reliable authority or appropriate criteria prevents us from attaining such precision. But instead of emphasizing discrepancies, we could learn from and about madness, through appeasing the conflict between folly and normality, by focusing on what is common to them both. Eventually, this focus on their overlapping parts takes us further into

exploring what it is to be human, by accepting the variations between individuals, without seeing an inherent conflict in it.

2. What is Madness?

2.1. Madness through History of Ideas and Mentalities – Short Contextual Description of the Madness-Knowledge Link

Madness is neither a recent thing, nor a recent interest for society. Depending on time period, cultural background or disciplinary perspective, madness was what people liked, chose or happened to make of it, never defined, never exhausted. Nowadays it might get more scientific attention than ever and more social assistance, whereas the past was ascribing to it a spiritual importance.

In the Greek culture, the figure of the god Apollo reflects a mentality, which does not exclude madness from knowledge, but quite the opposite. The god of light and reason is also the god of prophecy and poetry, both of which require being “out of one’s mind”, being possessed by the god, inspired, etc. The poetic creation and the gift of prophecy were both regarded as sign of being in touch with fundamental truths, of seeing beyond past or future. An un-ordinary perspective on the world would be two-sided: mad for the society but sacred, close to the realm of the divine knowledge.

In Phaedrus, Plato praises four states of madness, which he says, permits man to access the realm of the unseen. These four are: (1) the poetic, which grants knowledge of past deeds and events; (2) the prophetic, which allows seeing the future; (3) the telestic, the mystical knowledge; and (4) the erotic, the absolute knowledge.

‘Madness is superior to a sane mind (sophrosyne) for the one is only of human, but the other of divine origin [...] But he who having no touch of the Muses’ madness in his soul, comes to the door and thinks that he will get into the temple by the help of art – he, I say, and his poetry are not admitted; the sane man disappears and is nowhere when he enters into rivalry with the madman.’¹

In a certain measure, the madman remains an outsider, but still he is surrounded with a certain aura of respect and awe, for madness remains a sine qua non condition for superior knowledge and privileged being.

Un-usual states of consciousness, such as visions, dreams, epilepsy, were considered the sign of a connection to the divine.² Madness had those days its own place and function in the society; it was integrated, accepted as human and as part of the community.

This privileged status of madness in the Antiquity is somehow obstructed in the Middle Ages, when the figure of the mad person takes the place left empty in the society by the eradication of leprosy.³ The mad becomes the persona non grata, and he is now regarded with fear, treated with cruelty and contempt.

However, in the mind frame of the Middle Ages, with an extensive taste for the fantastic worlds and legendary characters, delusional madness might have found an integrating place. In a world in which symbolic, esoteric knowledge was playing such a great role, madness wouldn't have been regarded as lack of meaning, but as hiding a deeper meaning.

This double attitude towards madness and folly oscillating between fear and fascination continued until madness became a scientific fact. With psychiatry, madness becomes the object of systematic research under the neutral eyes of science, a domain in search of results and of cures.

2.2. Us and Madness

What is madness for us nowadays? The specialized publications in psychiatry become more and more detailed, but also more and more comprising. Madness is not reserved to sick people anymore, but all of us seem to be predisposed to it, or have at least some behavioral or psychological tendencies towards one or the other psychiatric disturbance. We fell in depression, we show sometimes signs of obsessive compulsory disorder, the percentage of children with autism or ADHD grows dramatically in schools. This situation has maybe its own positive side in the effect of reducing the distance between us - the healthy and *the others* - the mad people - thus bringing us a step further towards the reintegration of the people who suffer from these disorders.

So, in virtue of all the aspects emphasized so far, madness in a certain sense is part of us, is a present issue in our society. It relates to us directly through our own experiences or those of our close friends. So we ask ourselves what would be the right way to relate to madness? What is the paradigm in which we judge and consider madness as such? What are the borders and criteria, which confer credibility and authority to mentalities and opinions about madness? In the light of the main topic of this paper - the knowledge-madness link – we will point some mad aspects of what we consider knowledge and some aspects of madness that reflect great sensibility and knowledge. We will discuss some preconceptions about commonsense and knowledge, which limit and condition contemporary attitudes and actions concerning madness.

3. What is Knowledge for Us?

3.1. Scientific Paradigm – Some Problematic Aspects

In our contemporary society, knowledge is inextricably linked with science and the academic community. Science has its foundations, its principles: rigor, logic, proof, objectivity, etc., but also its assumptions, axioms, borders and loops. In this context, madness is what falls each time outside of these borders. We'll show further on why the traditional attributes of good science remain more or less just a myth. So how reliable are they, the big and small axioms of scientific thinking, and how can we determine this?

A very good point concerning this issue is the one of Douglas Hofstadter *Metamagical Themas Questing for the Essence of Mind and Pattern* - what scientific truth relies on is commonsense. A common sense regarding what is real, about cause-effect relations happening in time, about materiality and mass, about deductive argumentation, about a binary truth-value logical system, something being either true or false, etc. However, when analyzed in detail, all these concepts are problematic.

The so-called commonsense, which is '*a subjective art more than objective science*'⁴ has in its defense that it is functional, that it gives results. But those results seem so meaningful just because they are considered inside the same commonsensical paradigm in which they had been obtained.

There are so many examples in the history of science when axioms, once taken as firm reliable foundation, are reconsidered and finally given up for a quite opposite view. Consequently, parts of what once was considered to be madness are reintegrated into good science. Axioms such as Euclid's postulate, very intuitive, which for a long time had been taken for granted by both mathematicians and lay-people, says that from a point we can never trace more than one parallel to a given line (see Fig.1).



Fig.1: Euclid Geometry.

When new hyperbolic and elliptic geometries were invented, by Riemann and Lobachevski, this obvious commonsensical foundation of geometry became nothing more than a particular case of curved metric spaces (see Fig.2).

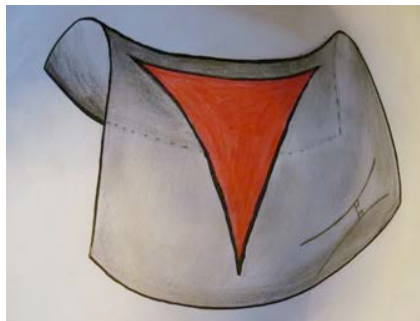


Fig.2: Riemann and Lobachevski Geometry.

Nowadays, it is acknowledged that these geometries describe better the universe's laws than the Euclidean one. Similarly, so many breakthroughs are not just counterintuitive, but even, un-elegant and very complicated, all three of these

attributes - elegance, simplicity and intuitiveness - being considered by some as marks of rational, correct proofs.

Moreover mathematics is not all about finding, but mainly about inventing, imagining mathematical objects, theorems and demonstrations. - Abstractions that end up being employed in useful applications in the concrete world. A domain, which doesn't aim at describing reality, mathematics deals with non-real objects, objects that at first appear to have no correspondence with the real world, but eventually prove to have. Octonions for example, which are the fourth algebra, after real numbers, complex numbers and quaternions, invented in 1843 by John T. Graves with no practical connection to the world at the time, have now applications in quantum mechanics.⁵

Apart from practical results in technology and elsewhere, the other main advocacy for scientific knowledge is experiment, as reliable methodology. In physics and biology and all natural sciences, we talk about experiment and falsifiability. In an experiment, there are certain rules to be respected, but also unpredictability and randomness to keep under control.

However, no matter how many measures of precaution are taken, an experiment will ultimately be a matter of interpretation.⁶ Very often, the agreement between scientists who deal with the same issue is just an appearance, when in fact that problem should still be in discussion.⁷ The successful, accepted views are sometimes the most spectacular, but not necessarily the best scientific explanations, or the correct ones. Moreover, something being functional at some point doesn't guarantee that the solution is the correct one. It can be a temporary one, a happy mistake or as it so often happens in science, a particular case of a broader problem, which requires in fact the opposite approach to be un-riddled.⁸

3.2. The Madness of Knowledge and the Sense of Madness

Without getting too much into a positivism-constructivism debate, we have to admit that most evident things, such as the existence of objects, are impossible to prove, at least existence in the classical sense.

There are so many features of reality, which go beyond our perceptual capabilities. Madness, in some of its forms, functions as an indicator of other aspects of this world, for which we have not yet developed appropriate receptors, language, and systems of representation.

But is there any commonsense at all in what is to be understood by existence, matter or time? Our commonsensical sane perceptions are contradicted by new cosmological or physical theories. Again, at our scale, classical concepts of time and matter still make sense, are still functional, but at a more essential level, they couldn't be more wrong. Time depends on speed, and can be warped by gravitation. Notions such as simultaneity or consequence change their meaning, shaking the firm and secure ground of our intuitions.

We are very familiar with distinctions such as mind-world, objectivity-subjectivity, etc. However, these distinctions do not hold anymore as absolute. It is hard to conceive nowadays a mind as a separate entity from the world. It's not just a socio-cultural context which shapes the mind, but also the genetic code we inherit, also the chemistry and electrical impulses at the level of neurons that makes us, seen at a bigger scale, indistinguishable from an outside world. The inside and outside listen to the same fundamental laws of reality – a reality in 11 dimensions, built from wave-corpucle particles and antiparticles. At this level, we don't work anymore with existent objects, but with a rate of probability, like in the famous Schrodinger Cat's paradox - there are moments in the experiment when the cat is considered both alive and dead. Energy and mass are convertible into each other. In computer science and Artificial Intelligence, we don't work anymore with traditional logic values like true or false, but with approximations. It is quite true then that this world couldn't get much madder than this.

Couldn't we then explain true madness as the translation to the existential level of these absurd counterintuitive but true universal laws that we don't quite manage to instantiate at a personal inner level of existence? Because we unconsciously maintain a fragrant contradiction between what we understand and what we perceive, between what we think and what we are.

So it seems that “sane” knowledge has some characteristics, which are not at all obvious to us, those who live inside the paradigm dictated by it. Madness functions, therefore, as an indicator of the hyper-fragmentation/“sickness” present in the contemporary knowledge systems. Madness - we all share.

In his anthropological study about madness, Jonathan Burns concludes that madness must be just another word for humanity. *‘The same “genes” that drive us mad have made us human.’*⁹

4. Art, the Translator

The postmodern world is contaminated with an excess of sense and meaning, which resembles quite a lot to the lack of any meaning - we move and live into hyper-realities.¹⁰ The appearance of one and only common sense comes from the fact that we don't have the overall picture of the universe anymore, we just deal with small portions, we work on small areas, we apply specific methods and principles. And to this level they appear to hold. What we got is small islands of sense in a sea of ever-changing meaning full reality.

There are many possible representations for a subject matter, each of it having its limits and its inadvertences, being in a certain sense incorrect and incomplete. To make sense of the world, we see things where they are not, we reconstruct, we create on top of the meaningless - the meaningful. We take David out of the marble; we take out the formed, out of the formless. We make sense out of randomness. Is it hallucination or creativity?

Art takes to the level of experience the madness of knowledge and completes its meaning. Through scientific theories, we represent the world. Through art

creation and reception, we re-become present in our representations. This can happen, because madness is itself a form of art. Madness often hurts our common sense through excess, intensity, controversy and integrative power of meaning. This is where it shares grounds with the arts and mystic experiences. What madness, art and religion have in common is a way of making a line of continuity between understanding and living.

Craziness might imply an experience of the world, which is not fragmented but continuous at the level of perception, conception and expression. For a mad person, the world can't be a simple representation but a rich experience, which is then expressed in a symbolic language we cannot completely understand. Also art and religion are lived knowledge, embodied knowledge.¹¹ An artist in a dance performance won't separate the feeling of space and time from the inner understanding of these concepts; a Tibetan monk will live accordingly to his reflections, accordingly to his philosophical but also existential experience.

Madness is regarded as a strange anomaly when confronted with statistical truth. But when confronted with absolute truth, philosophical truth, madness is part of it, is essential.

5. Conclusion

The commonsensical knowledge paradigm is built upon culturally and genetically given existential axioms, and represents just one from infinitely many other possible openings towards the world. Beyond such axioms, in the realm of madness, other paradigmatic configurations spring - with new reference points new solutions appear, and here lies the big creative power of madness.

Madness is the domain where paradox is accepted, expressed and lived, through participation, interpretation, or play of our own representations of the world.¹² It is like an artistic language, capable of making sense inside of a meaningful playful universe, transgressing the void of representation, towards the completeness of presence. What could be madder than making just one sense (a common sense) when there are so many other possible?

Notes

¹ Plato, *Phaedrus* (Project Gutenberg. Viewed on 15 June, 2014, <http://www.gutenberg.org/files/1636/1636-h/1636-h.htm>), 44a-257b.

² E. R. Dodds, *The Greeks and the Irrational* (Berkeley: University of California Press, 1951), 66-67.

³ Michel Foucault and Jean Khalfa, *History of Madness* (London: Routledge, 2006), 10-14.

⁴ Douglas R. Hofstadter, *Metamagical Themas Questing for the Essence of Mind and Pattern* (New York: Basic Books, 1985), 96.

- ⁵ Geoffrey M. Dixon, *Division Algebras: Octonions Quaternions Complex Numbers and the Algebraic Design of Physics* (Dordrecht: Kluwer Academic Publishers, 1994), 13.
- ⁶ Deborah G. Mayo, *Error and the Growth of Experimental Knowledge* (Chicago: University of Chicago Press, 1996), 214-218.
- ⁷ Michael Mulkey and G. Nigel Gilbert, "Accounting for Error: how Scientists Construct Their Social World when They Account for Correct and Incorrect Belief," *Sage Journals*, vol. 16, no. 2, (1982): 165-183. Viewed on 16 October, 2014. doi: 10.1177/0038038582016002001.
- ⁸ Gerd Gigerenzer, "I Think, therefore I Err - Errors, Consequences of Big Mistakes in the Natural and Social Sciences," *Social Research*, vol. 72, no. 1, (2005): 195-200.
- ⁹ Jonathan Burns, *The Descent of Madness Evolutionary Origins of Psychosis and the Social Brain* (London: Routledge, 2007), 181.
- ¹⁰ Jean Baudrillard, *Simulacra and Simulation* (Ann Arbor: University of Michigan Press, 1994).
- ¹¹ Francisco J. Varela, Evan Thompson and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience* (Cambridge, Mass.: MIT Press, 1991).
- ¹² Stephen Nachmanovitch, *Free Play: Improvisation in Life and Art* (Los Angeles: J.P. Tarcher, 1990).

Bibliography

- Baudrillard, Jean. *Simulacra and Simulation*. Ann Arbor: University of Michigan Press, 1994.
- Burns, Jonathan. *The Descent of Madness*. London: Routledge, 2007.
- Dixon, Geoffrey M. *Division Algebras: Octonions Quaternions Complex Numbers and the Algebraic Design of Physics*. Dordrecht: Kluwer Academic Publishers, 1994.
- Dodds, E. R.. *The Greeks and the Irrational*. Berkley: University of California Press, 1951.
- Foucault, Michel., and Khalfa, Jean. *History of Madness*. London: Routledge, 2006.
- Gigerenzer, Gerd. "I Think, therefore I Err - Errors, Consequences of Big Mistakes in the Natural and Social Sciences." *Social Research*, vol. 72, no. 1, (2005): 195-200.
- Hofstadter, Douglas R.. *Metamagical Themes Questing for the Essence of Mind and Pattern*. New York: Basic Books, 1985.
- Mayo, Deborah G.. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press, 1996.

Mulkay, Michael., and Gilbert, G. Nigel. "Accounting for Error: how Scientists Construct Their Social World when They Account for Correct and Incorrect Belief," *Sage Journals*, vol. 16, no. 2, (1982): 165-183. Viewed on 16 October, 2014. doi: 10.1177/0038038582016002001.

Nachmanovitch, Stephen. *Free Play: Improvisation in Life and Art*. Los Angeles: J.P. Tarcher, 1990.

Plato. *Phaedrus*. Project Gutenberg. Viewed on 15 June, 2014.

<http://www.gutenberg.org/files/1636/1636-h/1636-h.htm>, 44a-257b.

Varela, Francisco J., Thompson, Evan., and Rosch, Eleanor. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, Mass.: MIT Press, 1991.

Ioana Zamfir is an Art History and Religious Studies graduate. She is a Ph.D. student in the Philosophy Department of Bucharest University, with a research project in cultural anthropology. She is the Image Director of CTRRS (Center for Transdisciplinary Research in Religion and Science) at University of Bucharest.

Ivan Magrin-Chagnolleau is an artist researcher. He is an actor, director, and writer for stage and screen, as well as a musician and a photographer. He does research in Aesthetics and Philosophy of Art with the C.N.R.S., with a particular focus on the creative process, the impact of new technology on Art, and the link between Art and Spirituality. He is the founder and head of the EsPAS (Aesthetics of Performing and Spectacular Arts) Research Department inside the ACTE Institute, a joint research lab between Sorbonne Paris 1 University & C.N.R.S. He is also the founder and director of publication of the journal *p-e-r-f-o-r-m-a-n-c-e*.

**Artistic Workshops with patients of a Psychiatric
Day Hospital in France (sous presse)**

Ivan Magrin-Chagnolleau. **Artistic Workshops with patients of a Psychiatric Day Hospital in France.** *Creative Arts in Education and Therapy*, sous presse.

6 pages

Artistic Workshops with patients of a Psychiatric Day Hospital in France

Ivan MAGRIN-CHAGNOLLEAU ^{1,2}

¹ PRISM Laboratory, Aix Marseille Univ., CNRS, Marseille, France

² Chapman University, Orange, California, USA

Abstract

This paper presents a work that was led with the patients of a psychiatric day hospital in France. This work was made in the framework of institutional psychotherapy and of the therapeutic clubs, which are part of the anti-psychiatric movement of the 70's. The paper describes two activities that were led in that context, a music club and a video club. The goal of each of these clubs is to devise an activity with the patients of a psychiatric day hospital in order to abolish, on one hand, the frontier between patients and care staff, on the other hand, to implicate the patients as well as the care staff in the design and the implementation of these activities.

Keywords

Therapy, Institutional Psychotherapy, Therapeutic club, Psychiatric Day Hospital, Anti-Psychiatry, Performance, Music, Video, Enaction.

Background

The work presented in this paper was done in a Psychiatric Day Hospital in France. It is a place where patients having a mental illness can spend the day, but they are not full-time residents in the psychiatric hospital, they are just spending the days. They have some degree of autonomy allowing them to go back home in the evening and for the weekends. Most of them are stabilized through some sort of medication and they are more or less able to function in the community. The psychiatric day hospital is a place for them to get some activities, some group activities. They also get some one-on-one meetings with the psychologist in charge.

The framework for this work is called Institutional Psychotherapy. And, in that framework, the work I am presenting here is happening in what is called the Therapeutic Club. This work is in line with the works of François Tosquelles (Tosquelles 2007), Jean Oury (Oury 2003, 2016), and Pierre Delion (Delion 2018), and the whole institutional psychotherapy movement was part of the anti-psychiatric movement in the 70's (see for instance Bosseur 1974, Cooper 1978). It is also worth noting that there are strong relationships between the anti-psychiatric movement and the theories of Carl Rogers on therapy (See for instance Rogers 1980, 2003, 2004)

Aims

The aim of this kind of approaches is to get the patients out of the psychiatric hospital. In other words, the aim is to bring them to a certain degree of autonomy, where we recognize that they might not be able to function completely normally in society, but where we recognize some abilities and some competencies that allow them to do some activities and also to have a certain degree of autonomy. The idea is to welcome them during the day, mostly the weekdays, and to give them a place where they can meet other people, where they can have activities,

where they can hang around, where they can have their lunches as well, but, in the evening and during the week-ends, they go back home. While they are in the psychiatric day hospital, the goal is to help them acquire and develop competencies that will then allow them to have more autonomy, and possibly have a job in a specialized structure.

Method

The main characteristics of these types of approaches is first of all that patients and professionals are at the same level when they share activities. They design the activities together and no one is above anyone else. We assume that we can learn from each other, and also that the relationship is symmetrical.

These activities are usually conducted around some sort of artistic activities. In the day hospital where I was working, there were several of these activities. The two I am going to describe here are the ones that are related to performing arts: the music club and the video club. I am going to describe both of them, how we implemented them, and how a regular session worked.

The music club was already going on when I arrived. I was just a participant in this club, and occasionally the organizer or co-organizer. It had been implemented collectively by the psychologist in charge, some nurses, and some patients. The meetings took place once a week. For this particular club, it was on Friday mornings. The music club had a very regular structure. The first part of the session was dedicated to some music exercises, and I will describe some of them later. Then, we had a short break. Then, during the second part, we would perform songs together, well known songs. It was really two different parts with the same aim.

The video club was initiated while I was there. They had a version of it before I arrived, but we decided to do something completely different. And when I say we, again, it means the patients as well as the psychologist in charge, and the nurses, and me. It was really a collective decision.

The Music Club

Let us first go back to the music club and give a description of some of the exercises. The goal of most of these exercises is to give a space for every participant to perform some sort of music. And, when I say music, it is in a very general broad meaning of the word music. It can be hitting a drum, it can be doing some noise with hands and feet, it can be singing more or less in tune, but the important thing is to just express yourself and with a medium that is related to music and performance. The goal is to perform.

For instance, there was one exercise that was very interesting, it was called « the conductor ». And anyone who wanted to do it could do it. They would then get in the middle and conduct the rest of the group. They would move their hands and arms, and if they were high up, the volume of the other participants playing their instruments or singing had to be very high. If the arms were low, then the volume had to be low. The person in the middle is conducting the volume of the performance. It is a very interesting exercise because it gives the person who is in the middle some kind of power in a way, the power to decide how loud the music should be, and it is also very enjoyable for the person who does it. It gives a sense of accomplishing something.

There were various exercises like this one. I am not going to describe all of them, but basically the first part of the session was really about doing some exercises in order to help the participants to, first of all, have a good time, enjoy themselves, but also work on themselves one way or another.

Then, the second part was about performing a song together, or a couple of songs, depending on the days. The songs were chosen collectively, usually the session before, and then the words were printed, and we usually listened to one or two renditions of the song. We then tried to perform it together, collectively, to the best of our abilities, either following a recording of it, or by having someone who can play the keyboard accompany us.

The Video Club

The video club was a bit different. We decided together that it would be interesting to do some kind of improvisations and to film them, and then to watch them together. That was the starting point of this activity. And we were lucky enough to be able to buy a small camera and a small tripod, and then we did with what we had. Basically, we were in a room with some tables and some chairs, and we had a few costumes from a previous activity of the Psychiatric Day Hospital. We started that way, and we brainstormed about different topics that we could do the improvisations about.

What is actually surprising is that, very soon after starting the activity, we decided to do a short film together. There were some desires to do something that would have a sort of form and that could be shown. All that was decided collectively. Sometimes, when we make this kind of collective decisions, we are not necessarily always aware of all the implications, but this is also the interesting point about this way of doing things.

Once we decided we would do a short film, we needed to decide about the topic or the theme of the short film. We had to take into consideration the fact that the people who attended these activities were not always the same people from one session to the next one, because there is an established rule about these clubs: you only come when you feel like it, and when you want to come.

The first topic we decided to explore was “emotions”. It was a good topic to explore first because we could do different scenes from one session to another one, and it was not too dependent on the people in the room at a particular moment. We just wrote the names of various emotions on pieces of paper, and then everyone would choose one. I would then give a context like where they were or something like that, and then they would just improvise, and we would film it, and then we would watch it together. They could decide if they wanted to be the actor that day or not, or just watch, or if they wanted to operate the camera. Roles were assigned on a voluntary basis and at each session.

Results / Observations

I have described the two artistic clubs that I was part of, and now I want to make some observations about these types of activities and what they can do to patients with mental illness. The first thing I want to share is about the goal. Let us say the goal of these activities is really to help patients to feel better. The goal is really wellbeing. It is really mental wellbeing. It is giving them an opportunity to make progress, whatever that means. It is to have them make progress and, if nothing else, to give them at least the opportunity for a couple of hours to

enjoy themselves to the degree where they can enjoy themselves, and to give them a space where it is not so serious, it is not so anxiety triggering. It is already a lot if you can achieve that.

From what I observed, I can say that for some of them and on certain days, because of course it depends on the days and it depends on the patients, this is something that you could observe. You could observe them smiling, joking, feeling good about themselves, feeling good about each other, having fun, being creative. And that is a challenge because most of the patients that go to a psychiatric day hospital are stabilized through medications, which usually knock them out to some degree. Therefore, it is always nicely surprising when you can see them involved and wanting to be involved.

Now, I also need to put these observations into context: in order to decide if a patient is doing better, you cannot rely only on observations at a particular moment. The two persons who can really decide that are the patient, as well as the psychologist that they meet during one-on-one sessions about once a week. Besides, they take part to various activities, not just these two. Their wellbeing depends on a mixture of the activities that they do, the one-on-one sessions with the psychologist, their own process, their own medication, and so on. Therefore, it is not so easy to attribute the wellbeing of a mental patient to some particular factor or activity. But the simple fact that you can observe them having a good time is already an important result.

There is also something that is very interesting: the reactions of some professionals, the psychologist, myself, the nurses, because we were also having a good time. And this is something that is also important to notice because it is not so much that we are providing something for them, it is really more that we are sharing together an activity, and while we are sharing it, we are bringing who we are to it, and with the intent to all have a good time. What is really important is to break this barrier between patients and professionals, between patients and nurses, or patients and psychologists. It is really about bringing people together and creating a relationship with them, a real one, not just one where one is playing the role of trying to heal them and one is the one trying to be healed.

If we place all these things into perspective, I think that the main goal of these institutions is to go one step further, that is, to bring a patient to more autonomy, to the point where they can possibly have a job somewhere and leave the institution.

Conclusions

I want to close this paper with some conclusions. We give the patients a space where everyone is involved in the same activity, and by doing so, there is a certain quality to whatever is happening in the group and between people.

Another thing that I think is very important is the quality of the attention that we give them, and that they give us, and that they give each other. This quality can of course evolve during a session, but the aim is really to be all together, giving our attention to the activity at hand and sharing that moment.

It evokes Carl Rogers and the person centered therapy (Rogers 2003). I am convinced that there are strong relationships between these two theoretical approaches. If I was to continue this work, first of all, I would try to do more experimentation, maybe observe other cases. I

must say that not every psychiatric day hospital adopts this framework of the institutional psychotherapy and of the therapeutic clubs. I would even say that it is not particularly popular in France. I was therefore very lucky to find that place where they were doing it.

Pursuing this work would also require for me to dive more into the theory. I have read some books by Tosquelles, Oury and Delion, and some by Rogers, but then I would need to understand better the context for these theories, where they are coming from, and how they could be better implemented in today's context.

In the line of Rogers' work, I was also wondering if the enactive paradigm could not be applied to the relationship between a patient and a person providing care to this patient, be it a nurse, a psychologist, an intern, or even another patient. The enactive paradigm was developed by Francisco Varela (see for instance Varela 1989, 1991) and states that cognition and perception/action are looped in such a way that they are both interacting with each other. In that context, if this is mostly the relationship that is central to wellbeing, then it might be because in that relationship, there is something of an enactive nature at play. I would call that "the enactive nature of the therapeutic relationship". I intend to investigate further into that direction.

Funding

This publication was made possible through a funding from the Excellence Initiative of Aix-Marseille University, A*Midex, which is part of the French "Programme Investissements d'Avenir" - Funding #: AMX-19-IET-005.

Conflict of Interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

The author would like to thank Nader Aghakhani, the psychologist of the structure, for his supervision of the work, and for the rich discussions he provided.

References

Bosseur, C. (1974). *Clefs Pour L'anti-Psychiatrie*. Seghers.

Cooper, D. (1978). *Psychiatrie et anti-psychiatrie (Seuil)*. Seuil.

Delion, P. (2011). *Accueillir et soigner la souffrance psychique de la personne— Introduction à la psychothérapie institutionnelle*. Dunod.

Delion, P. (2018). *Qu'est-ce que la psychothérapie institutionnelle?* Éditions d'une.

Oury, J. (2003). *Psychiatrie et psychothérapie institutionnelle*. Champ Social.

Oury, J. (2016). *La psychothérapie institutionnelle de Saint-Alban à La Borde*. Editions d'une.

Rogers, C. (1980). *A Way of Being*. Houghton Mifflin.

Rogers, C. (2003). *Client Centered Therapy*. Robinson.

Rogers, C. R. (2004). *On Becoming a Person: A Therapist's View of Psychotherapy*. Constable & Robinson Ltd.

Tosquelles, F. (2007). *Education et psychothérapie institutionnelle* (1er édition). Matrice.

Varela, F. (1989). *Autonomie et Connaissance: Essai sur le Vivant*. Seuil.

Varela, F., Thomson, E., & Rosch, E. (1991). *The Embodied Mind*. The MIT Press.

3.14 Autres écrits en philosophie de l'art et en esthétique

Réflexions autour de l'expertise d'une œuvre artistique (2015)

Ivan Magrin-Chagnolleau. **Réflexions autour de l'expertise d'une œuvre artistique.** Dans *Expertise artistique*, édité par Maryvonne Saison. Ministère de la Culture, 2015.

9 pages

Réflexions autour de l'expertise d'une œuvre artistique

Ivan MAGRIN-CHAGNOLLEAU

1 Introduction

J'ai une double casquette : je fais de la recherche dans les domaines de l'esthétique et de la philosophie de l'art, et je suis également artiste. C'est donc de ce double point de vue que je m'exprime. Je vais aborder quatre thématiques. La première sera celle du transmédia. Qu'est-ce qu'une œuvre Transmédia? Comment l'appréhende-t-on ? Pourquoi les artistes se tournent-ils vers ce type d'œuvre ? Ensuite, je voudrais m'attarder un peu sur le processus créatif. J'aborderai dans une troisième partie très brève, la fonction de l'art. À quoi sert l'art ? Est-ce que l'art doit servir à quelque chose ? Et enfin je terminerai par la dimension spirituelle du travail artistique. Je vous proposerai pour finir une petite conclusion en forme d'utopie.

Je souhaite introduire mon propos par une réflexion sur l'artiste et son œuvre. Je ne pense pas qu'on puisse dissocier entièrement le projet, l'œuvre, de la personne. Il me semble que, dans le cas d'un artiste, une œuvre fait toujours partie d'un chemin de vie. Donc il me paraît essentiel de prendre en compte la personne quand on évalue une partie de son œuvre, sachant bien évidemment que cette prise en compte de la personne sera limitée à ce qu'on sait d'elle et à ce qui se passera lors de cette rencontre avec elle, lorsqu'elle est possible. Mais la raison pour laquelle ça me paraît très important, c'est en raison de l'intentionnalité. Ce qui compte dans une démarche artistique, c'est aussi l'intention, l'intentionnalité de l'artiste, comment l'artiste place sa démarche par rapport à son parcours de vie, par rapport aux échanges qu'il a pu avoir ou qu'il a avec différents interlocuteurs.

2 Le transmédia

Je voudrais aborder maintenant la notion de transmédia, de transmédialité. Et je voudrais me baser sur deux exemples pour cela. Le premier exemple est un

travail qui a été fait par Barbara Browning (*I'm Trying To Reach You*, 2012) et qui consiste en une écriture fictionnelle incluant des photos et renvoyant à des vidéos se trouvant sur internet¹. Ce travail repose donc sur l'utilisation, dans le processus de création, des nouvelles technologies, et cela représente donc un environnement de travail différent. L'artiste demande aussi au lecteur une démarche différente, celle d'aller voir sur internet des vidéos en lien avec ce qui est écrit dans le livre, démarche que le lecteur est libre d'accepter ou non. On peut se demander pourquoi l'artiste choisit d'avoir recourt à cette façon de faire, à l'utilisation des nouvelles technologies en plus de son travail d'écriture fictionnelle ? Et plus généralement, pourquoi un artiste choisit de recourir à plusieurs modalités lors de son travail de création, et de les mettre ensemble ? Je ne vais pas ici essayer nécessairement de trouver des réponses à toutes les questions que je pose, mais j'espère plutôt susciter par ces questions une réflexion qui permettra, je l'espère, d'aller un peu plus loin dans la pratique de l'évaluation.

L'autre exemple que je voudrais utiliser pour illustrer cette pratique est un projet sur lequel je travaille actuellement et qui s'intitule *Autour du lac*². Il s'agit d'un travail autour du bois de Vincennes, à côté de Paris, et plus particulièrement centré sur une zone autour du Lac des Minimes. L'utilisation du transmédia se fait plutôt dans une tentative de trouver, pour chaque chose que l'artiste essaie d'exprimer, le bon médium, ou la bonne médialité, pour le faire. Dans ce projet *Autour du lac*, l'enjeu est de pouvoir entrer au cœur du processus créatif en utilisant pour cela la métaphore de la marche. Il s'agit de pouvoir observer ce processus créatif de manière introspective, pour pouvoir notamment observer comment il se transforme au fur et à mesure du déroulé du projet. Le protocole de départ est d'écrire un texte à l'issue de chaque marche autour du lac, décrivant ce qui a été observé pendant la marche de manière introspective, et notamment en lien avec le déroulement du projet créatif. Quand le projet démarre, j'ai plutôt l'idée d'écrire des textes autour de la créativité, autour du processus créatif. Et assez vite, je me rends compte que c'est un texte qui va avoir besoin d'images, donc qui va inclure aussi des photos. Dans un deuxième temps, je prends conscience que ce projet a aussi besoin d'images animées, donc de vidéos. Je me rends compte, dans un troisième temps, qu'il y a aussi un besoin d'enregistrer des sons, notamment des sons dont les sources ne sont pas toujours filmables, par exemple tous les chants d'oiseaux. Il va donc y avoir aussi des fichiers sons dans ce projet. Ce qui pose donc assez vite la question de la présentation, de la restitution de ce travail, qui a à la fois de l'écriture, de la photo, de la vidéo, des fichiers sons. Cela contraint à un processus de présentation, de restitution qui va nécessairement être atypique. Il va notamment renvoyer de différents endroits à d'autres, de différentes médialités à d'autres. Il peut y avoir un livre avec des photos, mais avec l'impossibilité de présenter aussi

¹La chaîne youtube contenant les vidéos (mais il faut lire le livre pour comprendre comment les vidéos sont utilisées au niveau fictionnel) : <https://www.youtube.com/user/AhNethermostFun>.

²Ce travail est en cours d'élaboration et donnera lieu à des présentations sous différentes formes à partir de l'été 2015, à commencer par une exposition photo et une installation sonore au festival Les Arts ForeZtiers 2015.

les vidéos et les fichiers sons autrement que sous forme de lien internet. Il peut y avoir un document audiovisuel reprenant une partie des photos, vidéos, fichiers sons, et pourquoi pas textes lus, mais sans la dimension textuelle. Cette présentation audiovisuelle peut se faire sous forme de film, ou de vidéos d'art. Une partie de tout ce matériel peut aussi être activé (ou réactivé) par le biais d'une série de performances.

La question du transmédia est donc surtout une question du mode d'expression. Il s'agit de trouver le mode juste par rapport au questionnement en cours. Ça peut être parfois des effets de mode. Et il en existe très clairement dans l'art contemporain. Mais ça peut être aussi une démarche qui est plus de l'ordre de l'exploration, afin de trouver le médium qui va exprimer le mieux une certaine dimension du projet.

Mais cette question de la transmédiabilité est aussi liée au thème de l'exclusion et de l'inclusion. Même si je considère comme important la reconnaissance de disciplines artistiques bien établies et la connaissance de leur histoire, de comment elles se sont constituées, et l'importance pour un artiste d'avoir une formation solide dans une ou plusieurs de ces disciplines, je pense qu'il est aussi très intéressant de questionner la frontière entre ces disciplines, et de questionner aussi ces disciplines reconnues mais plutôt sur un mode inclusif. Ce qui se trouve aux frontières est souvent très intéressant, notamment parce que ça permet souvent d'inventer un nouveau langage qui, du coup, va s'appuyer sur au moins deux disciplines, mais va aussi questionner l'intermédiabilité de ces disciplines.

Et puis il y a aussi cette notion de passage d'un médium à l'autre. Passer d'un médium à l'autre permet de passer d'un paradigme, c'est-à-dire d'un terrain connu, d'un terrain avec ses codes, à un autre paradigme, c'est-à-dire à un terrain moins connu, voire inconnu, qui du coup peut s'inventer plus facilement pour l'artiste, mais tout en mettant en jeu les codes du médium de départ. Ce va-et-vient entre les médias me paraît donc être aussi une posture intéressante en terme de renouvellement de soi, de sa démarche, de son parcours.

Il y a néanmoins une difficulté par rapport à cette pratique qui est que, souvent, un artiste est identifié, au moins en premier lieu, à un médium particulier ou à une démarche particulière, et il y a parfois un phénomène d'identification voire de notoriété. Il y a donc un compromis à trouver ici entre une recherche identitaire et une exploration d'un terrain inconnu. Il y a donc ici un facteur de prise de risque qui est à la fois intéressant et représente en même temps un certain danger.

Et puis il est souvent question de l'art collaboratif, c'est-à-dire que cette transmédiabilité peut aussi se faire par le biais d'une collaboration. Ça peut être parfois entre artistes de même discipline, ce qui permet d'échanger différents regards et

différents codes au sein de cette discipline, et donc d'élargir ses horizons respectifs. Ou ça peut être une rencontre entre artistes venant de différentes disciplines, ce qui peut éventuellement conduire chacun de ces artistes à un glissement, voire un changement de paradigme. Dans tous ces cas, ce qui me paraît important est d'être guidé par le projet lui-même, par l'intuition qui était au départ de ce projet, et qui va généralement diriger le ou les artistes vers une forme d'expression ou vers une autre.

C'est par exemple le cas dans un autre de mes projets en cours dans lequel j'explore la frontière entre le féminin et le masculin³ : j'ai regardé du côté de la danse et de la chorégraphie, bien que je ne sois pas à l'origine issu de cette discipline. Il m'a semblé que le domaine de la danse était le plus approprié pour ce type d'exploration car c'était celui qui mettait sans doute le plus l'accent sur l'interaction entre les corps. Et tout en faisant ça, je suis donc privé de mes repères habituels, mais cela me donne aussi une plus grande liberté.

Ça m'amène à aborder le deuxième point qui est le processus créatif.

3 Le processus créatif

Le processus créatif, je l'aborde comme artiste, mais je l'aborde aussi comme chercheur, car c'est l'une des questions qui est au cœur de ma recherche. Il s'agit donc d'un va-et-vient entre une expérimentation de l'intérieur, des questionnements avec d'autres artistes sur leurs processus créatifs, et la confrontation de ces deux objets à des discours plus théoriques sur le processus créatif, l'esthétique et la philosophie de l'art.

Je voudrais illustrer cette dimension-là à travers un artiste qui s'appelle Robert Lepage, et qui est un acteur et metteur en scène québécois. J'ai eu la chance, en effet, d'être convié comme observateur de son travail de création pendant tout le mois d'août 2013. C'est au cours de cette période que le deuxième spectacle de la tétralogie *Jeux de Cartes*, qui s'intitule *Cœur*, a fini d'être développé⁴. C'est donc sur la base de cette observation et de toutes les discussions que j'ai pu avoir avec Robert Lepage que je vais parler maintenant de son processus créatif.

On peut dire que le processus créatif de *Cœur* s'est fait en six temps. Le premier temps correspond à l'élaboration de quelques idées en lien avec le spectacle. C'est la phase d'idéation. C'est surtout Robert Lepage qui est à l'initiative de ces

³Le projet Masculin/Féminin est en cours de développement et donnera lieu à diverses représentations à partir de 2016-2017.

⁴http://lacaserne.net/index2.php/theatre/playing_cards_hearts/

idées initiales. Mais comme pour tout artiste, ces idées sont le fruit de lectures, de rencontres, de discussions, etc. Ce premier temps est suivi assez vite par un deuxième temps, toujours d'idéation, au cours duquel Robert Lepage va inviter quelques proches collaborateurs à réfléchir avec lui au projet. À la fin de ce deuxième temps, Robert Lepage a déjà une idée assez précise au sujet des comédiens, des créateurs, et des techniciens qui vont travailler sur ce projet particulier. Le troisième temps de création consiste à inviter pour quelques jours plusieurs de ces personnes, principalement les comédiens pressentis, mais également quelques collaborateurs proches. Cette phase va permettre d'explorer plus avant les différentes possibilités du spectacle en création. Le quatrième temps de création consiste en une exploration, sous forme d'improvisations et de discussions à la table, de toutes les pistes qui ont été proposées lors de la phase précédente. Ce quatrième temps dure environ 15 jours. La plupart des comédiens y participent. Le cinquième temps de création, qui a lieu plusieurs mois après le précédent, dure environ un mois, et correspond à la mise en place à proprement parler du spectacle. Elle regroupe tous les comédiens du spectacle, tous les créateurs, et tous les techniciens. Elle comprend également deux répétitions publiques, au cours desquelles le public est invité à faire des retours. À l'issue de ce cinquième temps, un spectacle existe, et il va pouvoir commencer à tourner. Le sixième temps correspond à la confrontation au public, et il n'est pas rare qu'un spectacle de Robert Lepage évolue encore pas mal lors de ses premiers mois de tournée.

Robert Lepage a la possibilité de travailler comme ça parce qu'il a déjà acquis une certaine notoriété, et qu'il a donc plusieurs projets qui tournent déjà pendant qu'il en développe de nouveaux, ce qui permet donc d'injecter de l'argent en continu pour le développement des nouveaux projets.

D'autre part, il s'agit d'un processus de création très long. C'est un processus qui met aussi l'accent sur la participation du public comme révélateur de l'œuvre qui est en train de se construire, voire comme co-créateur. Tant qu'il n'y a pas de confrontation au public, il n'y a pas réellement d'œuvre aboutie, puisque cette œuvre va encore évoluer en fonction du public.

La première question qui se pose donc à l'artiste, lorsqu'il est en recherche de soutiens financiers pour ses créations, est celle de la projection dans le futur, et de sa capacité à parler d'un projet qui n'existe pas encore et qui, bien souvent, va beaucoup évoluer au cours de son processus créatif. L'artiste, au départ, a à sa disposition plusieurs choses : ce qu'il va écrire sur son projet, tel qu'il le connaît au début du processus créatif, qui est relié à sa capacité de produire un discours cohérent sur son projet à ce stade initial, et d'être suffisamment projectif sur son projet ; son corpus d'artiste, c'est-à-dire ce qu'il a déjà accompli, et donc comment son nouveau projet s'inscrit ou non dans une certaine continuité de son parcours. La difficulté vient donc du décalage inévitable qui existe entre la formulation initiale du projet et ce que va devenir le projet.

Cette question est encore plus cruciale pour un travail qui repose beaucoup sur de l'improvisation, de l'exploration. Pour reprendre l'exemple de Robert Lepage, lorsqu'il démarre un projet, il a une idée de la thématique du projet, et peut-être de quelques composantes de ce dernier. Mais rien de plus. Et c'est pendant tout le travail d'idéation, d'exploration, d'expérimentation et d'improvisation que le spectacle va vraiment se construire et que ses différentes composantes vont s'assembler, sans oublier sa constante évolution pendant les premiers mois de tournée.

Il y a une autre composante essentielle à prendre en compte dans un processus créatif, c'est la dimension du jeu, la dimension ludique. Et cette dimension essentielle a besoin, pour pouvoir exister, d'espace dans tous les sens du terme, et notamment d'espace de répétition, d'exploration, d'espace de laboratoire. Elle a besoin de temps. Et puis elle a besoin de non-finalité, c'est-à-dire qu'elle a besoin de ne pas être nécessairement contrainte à la production de quelque chose, tout du moins au début du processus. Ces phases-là sont parfois mises en place, par nécessité, en amont d'une demande de financement. Mais on pourrait rêver un peu en imaginant que, dans certains cas, les dispositifs de financement pourraient accompagner également ces phases de développement, ce qui est d'ailleurs déjà le cas pour certains projets.

Il y a aussi une autre question importante qui se pose, c'est la question de l'innovation. Elle pose problème parce que c'est une question cruciale, mais que le caractère d'innovation d'une œuvre entre en contradiction avec la notion de marque de fabrique, c'est-à-dire le fait qu'un artiste va être souvent reconnu par rapport à une façon de faire, par rapport à un style, à des caractéristiques de travail, etc. Et le fait d'innover, de ne pas refaire ce qui a déjà été fait, et qui est souvent ce que désirent les artistes, va les amener sur un terrain moins connu, voire inconnu, et qui va être du coup parfois plus difficile à défendre. Il y a donc là une contradiction qui n'est pas facile à dépasser.

Une dernière question qui est intéressante est celle de la cristallisation des phases d'un parcours artistique par rapport à une réalité qui, elle, est une réalité continue, c'est-à-dire que l'artiste ne se dit pas forcément qu'il va maintenant commencer un travail qui va être la première étape dans une nouvelle phase de sa carrière artistique (il ne se dit pas par exemple qu'il va commencer maintenant sa « période bleue »). Il avance dans son processus, qui est un processus continu, qui a parfois des ruptures, mais plus souvent des continuités. Mais pour être appréhendé par l'extérieur, il est parfois nécessaire de mettre en évidence des étapes dans un parcours artistique. Il y a donc ici une autre contradiction qui pose difficulté.

Enfin, il y a aussi la difficulté de prendre en compte les projets artistiques qui ne sont pas nécessairement prémédités, et qui se passent dans l'instantanéité. L'artiste n'a pas prémédité un certain projet, mais c'est le fait de se trouver dans certaines

conditions et dans un certain environnement qui va déclencher ce projet, si tant est qu'il a à ce moment-là les moyens pour le réaliser. J'illustre ce point avec un projet que j'ai réalisé récemment et qui s'intitule *Muir Woods Spirit*⁵. Je me suis trouvé à visiter, avec un ami, le parc naturel de Muir Woods à côté de San Francisco. Nous avons tous les deux nos appareils photo. Mais c'est le fait d'être au milieu de tous ces séquoias géants, dans cet environnement majestueux, qui m'a donné l'idée de prendre des photos « autrement », de « jouer » avec le lieu, et de réaliser des photos qui ne ressemblent en rien à ce que j'ai fait jusque-là, et qui semblent montrer une réalité qui est au-delà du visible. Je n'aurais probablement pas pu préméditer un tel projet. Il s'est construit dans le moment, en toute improvisation, parce que j'avais le bon outil avec moi, et que je me trouvais au bon endroit. Difficile de faire rentrer ce genre de projet dans une stratégie de recherche de financements. Ce projet illustre le paradoxe entre l'intentionnalité de l'artiste et les accidents de parcours, qui représentent souvent sa « mine d'or ».

J'en arrive maintenant à mon troisième point.

4 La fonction de l'art

Quelle est la fonction de l'art ? Je ne m'aventurerai pas à répondre définitivement à cette question. Je pense qu'on a tous des réponses différentes. En revanche, je voudrais proposer quelques pistes de réflexion parce qu'il me semble qu'il est difficile de séparer le processus créatif, la production artistique et donc son soutien institutionnel et/ou privé de la fonction de l'art. Il s'agit ici de la fonction que l'institution va attribuer à l'art.

L'exemple qui me vient est un travail du collectif Art Orienté Objet, c'est-à-dire de Marion Laval-Jantet et Benoit Mangin, quand ils ont fait cette performance autour de l'injection de sang de cheval dans le corps de Marion Laval-Jantet (*May The Horse Live In Me*, 2011)⁶. C'est un travail qui est à la fois performatif, à la fois plastique, et à la fois éthique. Outre l'injection du sang d'un cheval vivant, qui était présent lors de la performance (tout cela se faisait devant public), Marion Laval-Jantet avait également chaussé un dispositif d'échasses qui faisaient ressembler ses jambes à celles d'un cheval, et lui permettait ainsi d'être à la même hauteur que le cheval pour pouvoir avoir un échange avec lui.

⁵L'une des photos du projet a été exposée dans le cadre de l'exposition Entre-Mondes en décembre 2014. Cette exposition a eu lieu sous l'égide du Patronage Laïque Jules Vallès et regroupait des artistes chinois et français. Le commissaire de l'exposition était Michel Sicard. Une autre partie des photos de ce projet sera exposée dans le cadre du festival Les Arts ForeZtiers 2015.

⁶Voir par exemple <http://aoo.free.fr/works-2011-001.html> et https://www.youtube.com/watch?v=yx_E4DUWXbE.

Ce projet est pour moi un projet engagé, c'est-à-dire qu'il y a une dimension esthétique importante, mais aussi une dimension éminemment politique car ce travail constitue une réflexion à la fois sur l'environnement, sur la relation à l'animal, sur les traitements que l'humain peut infliger aux animaux. Cette dimension politique est dans ce cas inhérente au travail. Cela rejoint bien entendu la capacité de l'artiste à articuler un discours cohérent autour de son propre travail pour pouvoir faire prendre en compte cette dimension-là.

À partir du moment où on pose la question de la fonction de l'art, on pose aussi la question de la société que l'on souhaite. Il n'est pas anodin, dans la position d'une institution, de choisir de soutenir tel projet plutôt que tel autre. C'est la question de l'évolution de la société qui se pose dans ces choix. Se pose aussi la question de savoir qui est décideur de ces choix-là. L'institution bien sûr est en partie décisionnaire dans ces choix. Mais il y a aussi tout un pan de la création artistique qui se développe à la marge des financements et de la reconnaissance institutionnels.

Ce qui m'amène à ma quatrième et dernière partie, qui est une extension de la question de la fonction de l'art.

5 La dimension spirituelle de l'art

Je vais poser une hypothèse qui représente ma démarche d'artiste, mais je ne suis pas le seul à l'adopter. Je postule que quand il y a une démarche artistique, il y a une volonté d'exprimer quelque chose par le biais d'une expression artistique parce que ça ne peut pas être exprimé autrement, par exemple par le langage conventionnel. Et si ça ne peut pas être exprimé autrement, c'est parce que c'est indicible, et que ça ne peut être exprimé qu'à travers l'art. Certains parlent de rendre visible l'invisible. Ou d'exprimer l'indicible. Il s'agit pour moi d'une dimension de l'art qui est d'une nature spirituelle. Cette dimension pose bien sûr aussi la question de l'inspiration. Cette dimension spirituelle est l'une des dimensions de l'expérience humaine. Il ne s'agit pas exclusivement de la dimension spirituelle telle qu'elle est appréhendée par la religion, mais plutôt de la dimension spirituelle de l'humain quand il s'intéresse à son esprit, c'est-à-dire à cette expérience d'exister en dedans, d'être, et cela en deçà du langage. Je pense que c'est cette expérience de nature spirituelle que l'artiste essaie de transmettre et de partager. Il y a en chacun d'entre nous un esprit qui essaie d'exprimer quelque chose qui est de l'ordre de l'« être », de l'« être à l'intérieur ». Cette expression de l'artiste est bien sûr liée à sa vision du monde. Mais elle est aussi liée à la vision du monde des personnes qu'il a en face de lui ou dans son entourage.

Pour aller encore un peu plus loin, je dirai que dans toute démarche artistique, il y a un rêve, celui de l'artiste, et que l'artiste, pour pouvoir manifester son rêve, a

besoin de rencontrer sur son chemin des gens qui vont croire avec lui à son rêve et qui vont avoir envie de lui donner les moyens de réaliser ce rêve.

6 Conclusion en forme d'utopie

S'interroger sur les critères permettant d'expertiser des projets présentés par des artistes ou des compagnies, représente une recherche d'objectivité, qui est bien sûr louable et nécessaire, mais qui me semble un peu réductrice car je ne crois pas qu'on puisse s'affranchir de sa subjectivité. À partir du moment où on est dans la position d'évaluer un travail artistique, le premier critère, si on peut appeler ça un critère, et il est inévitable, c'est le critère de la subjectivité de l'émotion ressentie au contact de l'œuvre, de l'artiste. Ça me parle. Ça ne me parle pas. Ça me touche. Ça ne me touche pas. J'aime. Je n'aime pas. Une fois cette première réaction subjective reconnue, il est bien sûr intéressant de rechercher quelques critères objectifs complémentaires, y compris quantitatifs quand c'est demandé par l'institution. Mais je pense que la subjectivité du décideur est importante et qu'elle doit être revendiquée et assumée entièrement. C'est à travers cette subjectivité que peut se façonner une nouvelle vision de la société.

D'autre part, je ne crois pas que la responsabilité principale des décideurs soit de choisir quel travail artistique va exister et voir le jour. Cette question-là est davantage de la responsabilité de l'artiste. En revanche, il me semble que les décideurs ont plutôt la responsabilité de choisir à quelle démarche artistique, à quel projet et à quel artiste ils vont donner une voix, c'est-à-dire une certaine résonance. C'est pour moi quelque chose qui me paraît important et essentiel. Et profondément politique. Car cela va donner une certaine amplification à certaines visions du monde plutôt que d'autres.

Enfin, à travers le choix des décideurs de soutenir telle démarche artistique, tel projet et tel artiste plutôt que d'autres, il y a un choix qui est plus fondamental, et qui n'est pas anodin, c'est celui de la société que l'on souhaite construire, de l'évolution de la société que l'on souhaite soutenir, telle qu'elle est portée par le travail de tel ou tel artiste. Et donc, on pourrait dire qu'à travers ces choix-là, l'institution et les décideurs peuvent être en charge en partie de la construction d'un monde meilleur.

Futurs utopiques et dystopiques : comment la fiction nous aide à penser l'évolution (2021)

Ivan Magrin-Chagnolleau. **Futurs utopiques et dystopiques : comment la fiction nous aide à penser l'évolution.** *Arts et Sciences*, Vol. 21-5, No. 1, pp. 1-11, 2021.

11 pages

Futurs utopiques et dystopiques : Comment la fiction aide à penser l'évolution

Ivan MAGRIN-CHAGNOLLEAU, CNRS

Aix Marseille Univ, CNRS, Laboratoire PRISM (UMR 7061), Marseille

Chapman University, Orange, California

Mots clés

Art, science, fiction, science-fiction, utopie, dystopie, évolution, futur, philosophie, littérature.

Introduction

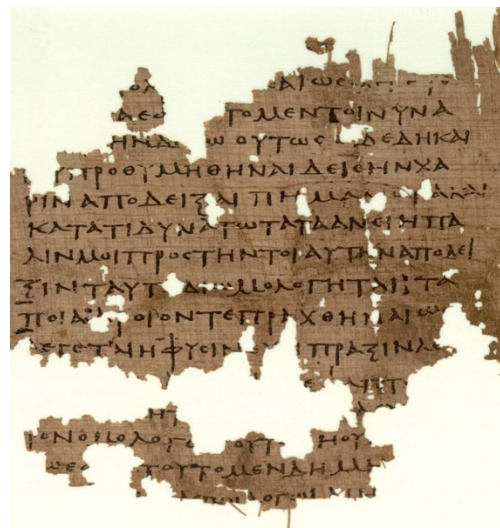
J'aborde dans cet article le thème des futurs utopiques et dystopiques tel qu'il nous est transmis à travers la philosophie et la fiction (littérature et cinéma), et montre comment ces notions nous aident à penser l'évolution. Je vais parler dans un premier temps des mots « utopie » et « dystopie », en racontant brièvement leur histoire et en proposant de les définir. Je donnerai ensuite des exemples qui appartiennent à la littérature, à la philosophie et au cinéma. Je terminerai en réfléchissant à une façon de réhabiliter l'utopie, pour pouvoir penser l'évolution aujourd'hui. J'ajoute que je parle en tant que philosophe et artiste. Mon propos n'est donc pas de faire une histoire précise des concepts d'utopie et de dystopie, ni de faire des études comparées des œuvres citées, mais plutôt de montrer comment la philosophie et la fiction peuvent nous aider à penser le futur, à penser l'évolution, et pourquoi c'est important.

Utopie et dystopie

On doit l'origine du mot « utopie » à Thomas More, un Anglais, en 1516¹. À l'époque, le terme « Utopia » est un nom propre qui désigne une île imaginaire, où tout est beau, où tout fonctionne parfaitement. Cette « Utopia » représente directement une critique de l'Angleterre de l'époque, en contrepoint de ce qui s'y passe à ce moment-là. Le mot « Utopia », qui vient du grec, signifie « un non-lieu ».



Illustration for the
1516 first edition of *Utopia*.



Papyrus d'Oxyrrhynque P. Oxy. LII 3679, IIIe siècle,
contenant des fragments de *La République* de Platon.

¹ Thomas MORE, *L'Utopie*, Paris, Folio, 2012, 384 p.

Le concept d'utopie existe bien avant Thomas More. On le trouve en particulier chez Platon dans *La République*² avec La cité idéale, qui est une façon de proposer un système politique idéal dans lequel les institutions fonctionneraient d'une façon optimale. Mais le mot en tant que tel est vraiment créé par Thomas More, et il est popularisé par Rabelais dans *Pantagruel*³ en 1532, où le mot devient alors un nom commun. À partir de ce moment-là, on peut parler d'une utopie et non pas juste de « Utopia », qui était le nom de l'île imaginaire créée par Thomas More.

Regardons maintenant comment la définition du mot « utopie » a évolué dans le temps. Au 16^e siècle, « Utopie » est un pays imaginaire où un gouvernement idéal règne sur un peuple heureux, et qui est ce qu'on pourrait dire une proposition politique utopique. Au 18^e siècle, le mot désigne le plan d'un gouvernement imaginaire, à l'exemple de la République de Platon, où l'on va planifier ce que pourrait être un gouvernement idéal, pour s'en servir de modèle pour faire évoluer les institutions politiques.

Au milieu du 19^e siècle, un nouveau glissement de la définition se produit, notamment en littérature, l'utopie devenant alors quelque chose qui ne tiendrait pas compte de la réalité. L'utopie est vue comme quelque chose de non réaliste, d'idéal, de non atteignable. Autrement dit, comme l'utopie est une version idéalisée de ce que pourrait être par exemple un gouvernement, elle n'a pas de valeur pratique, elle ne peut pas être implémentée comme telle, et donc elle n'est pas assez pragmatique pour être utilisable.

La définition évolue encore et l'utopie devient une conception ou un projet qui paraît irréalisable. L'évolution de ces définitions introduit progressivement des critiques de l'utopie, ce qui a pour conséquence de lui retirer son pouvoir de création du futur. Discréditée par la mise en avant de son caractère irréalisable, l'utopie cesse de représenter une source d'inspiration vers un avenir meilleur, et permet de mieux faire accepter la réalité et le statu quo.

Parallèlement à ces évolutions de la définition du mot « utopie », le mot « dystopie » fait son apparition au moment où la critique de l'utopie devient plus virulente, au 19^e siècle. La première utilisation du mot est attribuée à un politicien du Parlement britannique. La dystopie se définit alors comme étant une utopie qui ne suit pas la route escomptée. Une fois le projet utopique mis en place, il est fourvoyé par les personnes qui l'ont établi, et qui s'en serviront pour développer un pouvoir, notamment un pouvoir totalitaire.

Le mot dystopie n'est pas le seul au 19^e siècle à apparaître en réaction au mot utopie. Les mots « anti-utopie » et « contre-utopie » sont des alternatives fréquemment utilisées. Il y a de petites nuances entre ces différents termes. Le mot dystopie va cependant s'imposer et être davantage utilisé que les autres à partir de la fin du 19^e siècle. Depuis une vingtaine d'années, il y a un regain d'utilisation de ce mot, notamment dans les séries télé. La dystopie devient même un genre à part entière en littérature et au cinéma.

Une des critiques qui est formulée au sujet de l'utopie est qu'elle serait le discours des dirigeants, et la dystopie, la réalité vécue par le reste de la population. C'est cette conception de la paire utopie-dystopie qui va souvent se retrouver dans les fictions contemporaines.

² PLATON, *La République*, Paris, Folio, 1993, 560 p.

³ RABELAIS, *Œuvres complètes*, Paris, Gallimard : La Pléiade, 1994, 1888 p.

Il y a à mon sens une autre façon de regarder la paire utopie-dystopie qui, en réalité, n'est pas vraiment une paire de mots opposés, mais recouvre plutôt des façons très différentes de concevoir un projet du futur. L'utopie, ce serait être pour quelque chose, alors que la dystopie, ce serait être contre quelque chose. L'utopie peut aussi être vue comme une critique politique ou sociale, mais comme une critique positive. On parle de la situation courante, et on essaie de réfléchir à ce qu'on pourrait faire idéalement pour l'améliorer. Alors que dans la dystopie, il y aurait peut-être une forme de résignation : on montre des excès, on montre des dysfonctionnements, mais en étant à court de propositions constructives. Mais l'impact d'une représentation dystopique peut aussi parfois servir de déclencheur à une prise de conscience, même si la solution n'est pas encore proposée ni envisagée.

Exemples d'utopies et de dystopies en fiction

Je voudrais aborder toutes ces questions à travers des exemples en littérature, en philosophie et au cinéma. Il existe de nombreux exemples pour faire une chronologie de l'utopie et de la dystopie. J'invite le lecteur à noter au préalable ceux qui peuvent lui sembler pertinents, qui font partie de ses référents culturels. Les choix que j'ai faits reposent sur le rôle que certaines œuvres ont jouée à mon sens dans la fabrication d'un inconscient collectif autour de ces deux termes. Ils sont aussi liés à mes propres référents culturels.

En littérature et en philosophie, j'ai déjà parlé de Platon, de *La République*⁴, avec La cité idéale, et de Thomas More, qui propose son île imaginaire « Utopia »⁵. On trouve un autre exemple au 18e siècle chez Voltaire, dans *Candide*⁶, où il est question d'une visite au pays d'Eldorado, qui est un pays magique, magnifique, où tout se passe bien. Le personnage principal le quitte pour l'unique raison de retrouver sa bien-aimée Cunégonde. Dans cette œuvre de Voltaire, le pays d'Eldorado est utilisé par l'auteur comme un prétexte pour exposer ses idées politiques au sujet d'une gouvernance idéale.

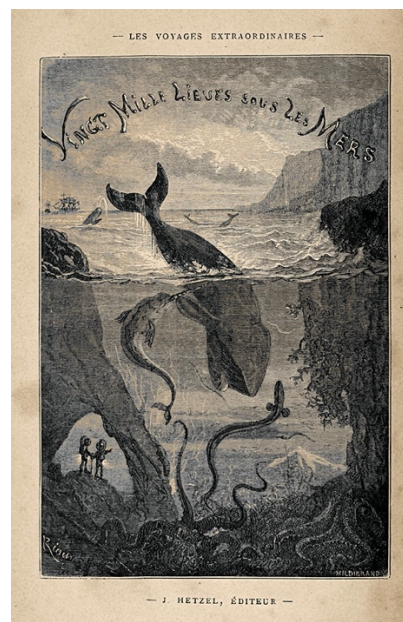
CANDIDE,
OU
L'OPTIMISME,
TRADUIT DE L'ALLEMAND
DE
MR. LE DOCTEUR RALPH.



M D C C L I X.

Édition princeps.

« M. le Docteur Ralph » est un des nombreux pseudonymes de Voltaire.



Page de garde de la première édition Hetzel.

⁴ PLATON, *La République*, Paris, Folio, 1993, 560 p.

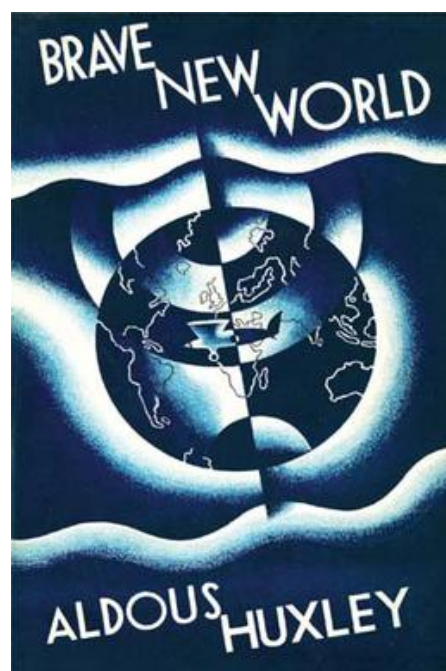
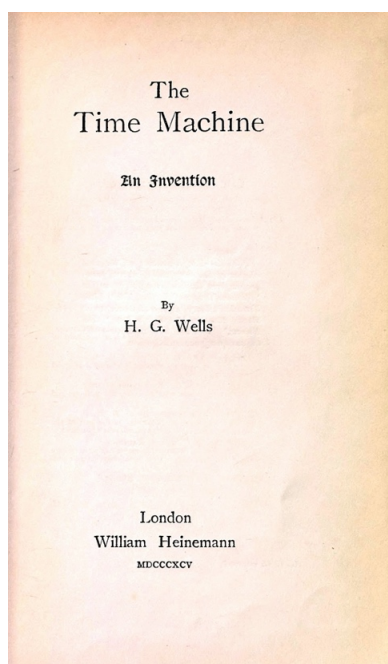
⁵ THOMAS MORE, *L'Utopie*, Paris, Folio, 2012, 384 p.

⁶ VOLTAIRE, *Candide et autres contes*, Paris, Folio, 1992, 512 p.

Au 19e siècle, on assiste à la naissance de la science-fiction qui, très vite, va s'emparer de l'utopie, puis de la dystopie, et va construire beaucoup de récits de science-fiction ou d'anticipation sur cette terre utopique ou dystopique.

La catégorie de la science-fiction utopique comporte un certain nombre d'œuvres de Jules Verne, notamment *Vingt mille lieues sous les mers*⁷ en 1870. Dans cette veine de science-fiction utopique, l'auteur imagine des futurs où l'homme arrive à conquérir une nouvelle frontière technologique (comme ici la capacité de construire un sous-marin qui permet d'explorer les fonds marins à une profondeur exceptionnelle), politique ou morale. Il est essentiellement question de thématiques liées à la technologie, à l'évolution, au progrès de la science. Et ces récits s'inscrivent dans l'ère du positivisme : la science va permettre d'élargir la connaissance et de faire de nouvelles découvertes. C'est une période où, à l'instar d'Auguste Comte⁸, une grande partie des scientifiques, et des auteurs qui en sont proches comme Jules Verne, croit que la science permettra de faire des progrès considérables, et finira un jour par conquérir les territoires de la connaissance qui sont pour l'instant inaccessibles.

À partir de la fin du 19e siècle, des exemples de dystopie en fiction apparaissent et deviennent plus fréquents. Ils illustrent une crainte que peuvent avoir certaines personnes vis-à-vis d'un progrès scientifique qui ne serait pas suffisamment encadré par l'éthique. Un exemple frappant est le roman de H. G. Wells, *La machine à explorer le temps*⁹ en 1895. J'appelle cela explorer le revers de la médaille : un explorateur du futur revient au temps présent et raconte son voyage, en commençant par une terre paradisiaque, où les gens vivent comme s'ils étaient dans le jardin d'Éden. Puis, on découvre petit à petit qu'une autre espèce, qui a également évolué à partir de l'humain, vit dans les souterrains. Ce sont les Morlocks, et ils se nourrissent des habitants qui vivent sur la surface de la planète. C'est le revers de la médaille.



⁷ Jules VERNE, *Vingt mille lieues sous les mers*, Paris, Le Livre de Poche, 2001, 606 p.

⁸ Auguste COMTE, *Philosophie des Sciences*, Paris, Gallimard, 1997, 462 p.

⁹ Herbert George WELLS, *La machine à explorer le temps*, Paris, Folio SF, 2016, 176 p.

Dans *Le meilleur des mondes*¹⁰ d'Aldous Huxley en 1931, l'auteur aborde d'autres thématiques. Il y décrit une société qui a évolué vers un système de castes, mais il existe quelques endroits qui ont échappé à cette évolution, et où vivent des « sauvages ». Ces sauvages n'ont pas conscience de la répartition en castes où chacun a sa place et est éduqué pour faire le mieux possible ce qu'il doit faire. Nous assistons alors à cette confrontation entre, d'une part, le sauvage, que l'on pourrait aussi appeler le naïf, celui qui ne sait pas que la société a été hiérarchisée, et d'autre part, cette société extérieure à la réalité de ce sauvage, qui a été complètement codifiée et programmée. Ce qui est important, c'est cette confrontation entre ceux qui n'ont pas été formatés par cette société idéale et ceux qui, ayant été formatés, ne se posent même plus la question de savoir si c'est réellement une société idéale.

En 1943, le roman *Ravage*¹¹ de René Barjavel explore l'évolution d'une autre façon. On pourrait se demander si c'est une dystopie ou une utopie. L'auteur explore ce que serait finalement un retour à un monde où la technologie aurait disparu. Il invente donc une catastrophe technologique qui serait liée à des rayonnements solaires et qui fait que toute technologie s'arrête de fonctionner du jour au lendemain. Cela occasionne une période de chaos, qui remet tout le monde au même niveau. Les gens qui sont en haut de l'échelle sociale n'ont pas plus de chances de survivre dans ce Nouveau Monde que ceux qui sont en bas de l'échelle. Et petit à petit, ceux qui survivent sont ceux qui arrivent à développer à la fois des stratégies de solidarité, d'entraide, et qui sont capables de se réadapter très vite aux forces de la nature et à de nouvelles façons d'y faire face. Le personnage qui permet la recréation d'une société basée sur l'entraide se fait finalement tuer par quelqu'un qui voudrait réintroduire la technologie. Et le lecteur comprend à demi-mot que le monde est en train de repartir dans cette espèce de spirale technologique, que le roman questionne, mais ne condamne pas nécessairement.

Il y a bien sûr le roman de George Orwell *1984*¹², en 1949, qui, lui, pose vraiment de front la question de la surveillance, dans un contexte de régime totalitaire et donc de surveillance à outrance et de privation de liberté.

Puis Robert Merle, en 1972, propose, dans un contexte d'escalade nucléaire, un roman post-apocalyptique, *Malevil*¹³. Il y a eu juste avant le début de l'histoire une catastrophe nucléaire, et le roman raconte comment des gens essaient de survivre après cette catastrophe.

Tous ces romans proposent des questionnements sur différentes thématiques de l'évolution, en partant du présent et en se disant : « Et si ça allait dans cette direction ? » ou bien : « Et si ça dérapait ? » dans le cas d'un questionnement dystopique.

Je mets un peu à part Isaac Asimov, qui a produit une œuvre monumentale. C'est un auteur qui, en fonction de ses romans, de ses cycles, de ses périodes, va osciller entre

¹⁰ Aldous HUXLEY, *Le meilleur des mondes*, Paris, Pocket, 2017, 320 p.

¹¹ René BARJAVEL, *Ravage*, Paris, Folio, 1972, 313 p.

¹² George ORWELL, *1984*, Paris, Folio, 1972, 438 p.

¹³ Robert MERLE, *Malevil*, Paris, Folio, 1983, 635 p.

utopie et dystopie. Un de ses cycles les plus connus est celui sur les robots¹⁴, dont plusieurs nouvelles ont été adaptées au cinéma, et qui pose la question de l'intelligence artificielle et des relations homme-machine. Ce cycle s'est développé à une époque où l'on croyait beaucoup à la cybernétique et à son développement. L'auteur y propose notamment les trois lois de la robotique, censées garantir la sécurité de l'humain face aux robots. Puis, ayant une formation de logicien à l'origine, il déconstruit petit à petit ces trois lois et nous dépeint comment, tout en les respectant, des robots vont être en mesure de les détourner finalement.

Dans son cycle *Fondation*¹⁵ qui est également très connu. Isaac Asimov propose, le concept de psychohistoire, c'est-à-dire l'idée qu'avec la loi des grands nombres, en contrôlant suffisamment bien les conditions initiales, on peut prédire l'avenir de l'humanité dans ses grands mouvements. De nouveau, il va introduire petit à petit des éléments perturbateurs dans ces grandes lois, afin d'expliquer comment ces prédictions pourraient dévier par rapport à ce qu'elles étaient initialement.

Aujourd'hui, la dystopie est très courante en littérature et au cinéma. On en a des exemples assez récents avec *Hunger games*¹⁶ ou *Divergent*¹⁷. Si elle s'adresse plutôt à des adolescents, cette littérature reprend le thème de la dystopie en le poussant à l'extrême, ce qui permet l'émergence d'un héros, en l'occurrence une héroïne, qui va se révolter contre cette dystopie et qui va essayer de faire changer les choses. On est vraiment dans l'idée de l'émancipation. C'est à la fois une dystopie et une proposition de comment pouvoir faire changer les choses, le plus souvent par une forme de révolte.

Le cinéma fournit aussi de nombreux exemples. Je me suis limité à quelques choix que je considère comme étant vraiment emblématiques de l'évolution de la thématique de la dystopie au cinéma. Dans le film de Fritz Lang *Metropolis*¹⁸ en 1927, il est question de lutte des classes et d'inégalités sociales. En présentant une situation exacerbée, l'auteur met le doigt sur ce qu'il considère comme étant un problème dans la société actuelle. Ce n'est pas nécessairement une proposition de résolution – on est bien dans une dystopie – mais plutôt une façon de montrer ce que serait le paroxysme de cette situation problématique.

Dans *Fahrenheit 451* – qui est d'abord un roman de Ray Bradbury¹⁹ en 1953, puis un film de François Truffaut²⁰ en 1966 – il est à la fois question du rôle de la connaissance, notamment par les livres, et d'une société où ces derniers sont interdits et brûlés – Fahrenheit 451 étant la température d'auto-combustion d'un livre. Et dans ce monde-là, qui est totalitaire puisque les gens n'ont plus accès à la connaissance, un pompier, qui est en charge de brûler les livres, se met à questionner son travail.

¹⁴ Isaac ASIMOV, *Le cycle des robots, Tome : Les robots*, Paris, J'ai Lu, 2012, 284 p.

¹⁵ Isaac ASIMOV, *Le cycle de Fondation, I : Fondation*, Paris, Folio SF, 2009, 416 p.

¹⁶ Suzanne COLLINS, *The Hunger Games, Book One*, New York, Scholastic Press, 2008, 384 p.

¹⁷ Veronica ROTH, *Divergent*, New York, Katherine Tegen Books, 2012, 496 p.

¹⁸ *Metropolis*, Dir. Fritz LANG, Perf. Alfred ABEL, Gustav FRÖHLICH, Rudolf KLEIN-ROGGE, Universum Film, 1927, Film.

¹⁹ Ray BRADBURY, *Fahrenheit 451*, Paris, Folio SF, 2000, 224 p.

²⁰ *Fahrenheit 451*, Dir. François TRUFFAUT, Perf. Julie CHRISTIE, Oskar WERNER, Cyril CUSACK, Universal Film, 1966, Film.



Theatrical release poster by John Alvin.



Theatrical release poster.

Il y a aussi le film *2001, L'odyssée de l'espace*²¹, en 1968, film de Stanley Kubrick. Il s'agit d'un film très riche, qui aborde beaucoup de thématiques, notamment celle de l'intelligence artificielle et de la communication homme-machine. Le film aborde aussi la question du temps, le temps tel qu'il est perçu, le temps qui peut être différent pour un humain par rapport à une machine, ce qui pose un certain nombre de questions philosophiques.

Un peu plus récemment, le film *Soleil vert*²² de Richard Fleischer, en 1973, aborde la question de la surpopulation, de la famine, de la limitation des ressources et de la solution trouvée par cette société pour y remédier.

*Blade Runner*²³ de Ridley Scott en 1982 parle aussi beaucoup de cette question de cybernétique, d'androïdes, dans une planète post-apocalyptique où les gens les plus riches ont fui la Terre pour aller vers d'autres planètes, où des colonies ont été installées. Ce film, qui est une adaptation du roman de Philip K. Dick *Les androïdes rêvent-ils de moutons électriques* ²⁴, pose les questions suivantes : « Quelle est la différence entre un androïde et un humain ? », « Comment faire la différence entre les deux ? », et, « Quelle est l'évolution, non plus de l'humain, mais de l'androïde ? »

Il faut noter le magnifique film *Children of Men*²⁵ d'Alfonso Cuarón, en 2006, qui traite directement de la question de la stérilité, qui imagine notre monde dans quelques décennies, à un moment où les humains sont devenus stériles, et où l'humanité est vouée à disparaître en tant qu'espèce. Le film raconte l'histoire d'une femme enceinte, ce qui

²¹ *2001, l'odyssée de l'espace*, Dir. Stanley KUBRICK, Perf. Keir DULLEA, Gary LOCKWOOD, William SYLVESTER, Metro-Goldwyn-Mayer, 1968, Film.

²² *Soleil vert*, Dir. Richard FLEISCHER, Perf. Charlton HESTON, Leigh TAYLOR-YOUNG, Chuck CONNORS, Metro-Goldwyn-Mayer, 1973, Film.

²³ *Blade Runner*, Dir. Ridley SCOTT, Perf. Harrison FORD, Rutger HAUER, Sean YOUNG, Warner Bros., 1982, Film.

²⁴ Philip K. DICK, *Les androïdes rêvent-ils de moutons électriques ?*, Paris, J'ai Lu, 2014, 283 p.

²⁵ *Children of Men*, Dir. Alfonso CUARÓN, Perf. Clive OWEN, Juan Gabriel YACUZZI, Michael CAINE, Universal Pictures, 2006, Film.

n'était pas arrivé depuis plus de vingt ans. Et il pose cette question des conséquences de nos choix.

Nous pouvons aussi souligner, plus récemment, le film *Interstellar*²⁶ de Christopher Nolan sorti en 2014, qui pose aussi cette question de la limitation des ressources et de la catastrophe climatique et écologique. Le film questionne le temps, avec une histoire de voyage dans le temps, de passage par un trou noir, et les conséquences que cela peut avoir sur le temps tel qu'on le vit quand on est en train de traverser ce trou noir, par rapport à ceux qui sont restés sur la planète.

En ce qui concerne les séries, on peut observer ces dernières années un regain d'intérêt important pour les dystopies. Un des exemples les plus connus est la série *Black Mirror*²⁷, qui s'est construite comme une sorte d'anthologie de la dystopie. L'épisode de la troisième série, « San Junipero », a reçu de nombreux prix. Il pose la question du transhumanisme et d'un monde où l'humain transcende la mort, en enregistrant la conscience, et en permettant à cette conscience enregistrée de continuer à vivre une certaine existence idéalisée après que le corps biologique est mort.

*La servante écarlate*²⁸ (« The Handmaid's Tale »), sortie en 2017, reprend un roman de Margaret Atwood²⁹ de 1985, et revient sur la question de la pollution et de la baisse de fertilité. Puis, très récemment, la série *Maniac*³⁰ questionne à la fois l'intelligence artificielle et la santé psychique.

Réhabiliter l'utopie

Je pense que la prolifération de la dystopie correspond à des évolutions de la société qui rendent plus difficile le fait de penser le futur, ce qui pourtant n'a jamais été aussi important qu'aujourd'hui. Cette évolution de la dystopie peut être liée à différents facteurs : le développement des réseaux sociaux, le flux d'information continu, l'obsolescence de plus en plus rapide, qui fait qu'il y a peut-être une transformation dans l'inconscient collectif de la notion de temps psychologique. Le temps va de plus en plus vite, les informations sont de plus en plus nombreuses, et par conséquent, penser le futur devient peut-être de plus en plus difficile. Est-ce qu'on arrive encore à penser à un futur qui dépasserait les quelques jours ou quelques semaines, alors que les évolutions sont de l'ordre de quelques semaines ou quelques mois ? Cette question me paraît essentielle.

Deux ouvrages abordent cette question du temps d'une manière qui me paraît pertinente par rapport à cette notion d'accélération du temps. Le premier est un ouvrage de François Hartog intitulé *Régimes d'historicité : Présentisme et expériences du temps*³¹. Dans cet ouvrage, l'auteur propose une réflexion sur les différentes façons de concevoir le temps de

²⁶ *Interstellar*, Dir. Christopher NOLAN, Perf. Ellen BURSTYN, Matthew MCCONAUGHEY, Mackenzie FOY, 20th Century Fox, 2014, Film.

²⁷ *Black Mirror*, Creator Charlie BROOKER, Perf. Daniel LAPAINE, Hannah JOHN-KAMEN, Jesse ARMSTRONG, Channel 4 Television Corporation & Netflix, 2011-, Série.

²⁸ *La servante écarlate*, Creator Bruce MILLER, Perf. Elisabeth MOSS, Yvonne STRAHOVSKI, Joseph FIENNES, Hulu, 2017-, Série.

²⁹ Margaret ATWOOD, *La servante écarlate*, Paris, Robert Laffont, 2017, 544 p.

³⁰ *Maniac*, Creator Cary JOJI FUKUNAGA, Perf. Jonah HILL, Emma STONE, Sonoya MIZUNO, Netflix, 2018, Série.

³¹ François HARTOG, *Régimes d'historicité : Présentisme et expériences du temps*, Paris, Seuil, 2003, 272 p.

l'histoire, et rappelle à la fois la subjectivité de cette notion de temps historique, ainsi que l'influence de la distance temporelle à l'événement observé sur cette notion. Ainsi, cette impression que le temps s'accélère serait, d'une part, subjective, d'autre part, en lien avec le fait qu'il s'agit d'un temps vécu au présent et non plus d'un temps historique observé avec le recul nécessaire.

Quant au livre d'Eugène Minkowski intitulé *Le temps vécu*³², il propose précisément une réflexion sur le temps comme expérience vécue, c'est-à-dire une réflexion sur la phénoménologie du temps.

Le roman de Frank Herbert *Dune*³³ occupe une place à part. Il propose une dystopie écologique, qui questionne à la fois le climat et la question de l'eau. Ces deux questions sont selon moi les plus importantes auxquelles nous faisons face aujourd'hui. *Dune* présente une planète complètement désertique dont on s'aperçoit, au fur et à mesure du roman, que c'est notre planète devenue aride. Mais le projet du personnage principal de cette saga est utopique, et consiste à recréer sur cette planète aride les conditions pour que le cycle de l'eau soit de nouveau possible. Le réchauffement climatique ainsi que la dégradation de l'environnement (notamment la qualité de l'air, de l'eau, etc.) occupent une place de plus en plus importante dans la dystopie contemporaine³⁴.

Je termine en proposant de réfléchir à la façon dont nous pourrions réhabiliter l'utopie pour penser l'évolution aujourd'hui. Ce qui est finalement assez rassurant, c'est que nous assistons aussi, récemment, à un retour des utopies, mais des utopies pragmatiques et pratiques. Trois livres sont sortis récemment, qui ne sont pas de la fiction, mais qui traitent précisément de cette question d'une utopie pragmatique : *Utopies réalisables*³⁵ de Yona Friedman en 1974 déjà, réédité en poche en 2015 ; *Utopies réalistes*³⁶ de Rutger Bregman en 2017 ; et *Utopies réelles*³⁷ d'Erik Olin Wright en 2017. Ces livres sont des enquêtes sociologiques sur des endroits où des gens ont proposé des solutions utopiques, mais où ces propositions sont devenues des réalités, et ont permis de résoudre localement certains problèmes, par exemple la monnaie locale, la culture locale, etc. Ces ouvrages et ces exemples montrent qu'il y a finalement une nécessité de repenser l'évolution autrement, et surtout que cela est possible.

Il est essentiel de réhabiliter l'utopie et sa vocation à proposer des alternatives politiques, sociales, écologiques, économiques, morales. L'utopie peut être une force de proposition. Pour le dire autrement, pour pouvoir créer un futur, il faut pouvoir le rêver. Ce qui est important, c'est de briser la chaîne des prédictions néfastes. Les utopies du passé et du présent construisent la réalité de demain. Si nous n'avons plus d'utopies aujourd'hui, quelle réalité allons-nous construire demain ?

³² Eugène MINKOWSKI, *Le temps vécu*, Paris, PUF, 2013, 432 p.

³³ Frank HERBERT, *Dune*, Paris, Pocket, 2012, 832 p.

³⁴ Voir par exemple Paolo BACIGALUPI, *Water knife*, Montpellier, Au Diable Vauvert, 2016, 487 p.

³⁵ Yona FRIEDMAN, *Utopies réalisables*, Paris, L'éclat poche, 2015, 240 p.

³⁶ Rutger BREGMAN, *Utopies réalistes*, Paris, Le Seuil, 2017, 256 p.

³⁷ Erik Olin WRIGHT, *Utopies réelles*, Paris, La Découverte, 2017, 624 p.

Je terminerai par un exemple d'utopie politique. Il s'agit du livre *Le bien commun*³⁸ de Riccardo Petrella, sorti en 1996. Ce livre propose à mon sens une véritable utopie, qui est la réhabilitation du concept de bien commun. Depuis la sortie du livre, on peut dire que la société a continué à suivre le chemin opposé à celui que l'auteur propose, lui qui a identifié très tôt que l'eau était notre bien le plus précieux, et qui souhaitait, par ce petit livre, remettre l'humain au cœur des politiques, et non les marchés.

Bibliographie

- ASIMOV, Isaac, *Le cycle des robots, Tome 1 : Les robots*, Paris, J'ai Lu, 2012, 284 p.
- ASIMOV, Isaac, *Le cycle de Fondation, I : Fondation*, Paris, Folio SF, 2009, 416 p.
- ATWOOD, Margaret, *La servante écarlate*, Paris, Robert Laffont, 2017, 544 p.
- BACIGALUPI, Paolo, *Water knife*, Montpellier, Au Diable Vauvert, 2016, 487 p.
- BARJAVEL, René, *Ravage*, Paris, Folio, 1972, 313 p.
- BRADBURY, Ray, *Fahrenheit 451*, Paris, Folio SF, 2000, 224 p.
- BREGMAN, Rutger, *Utopies réalistes*, Paris, Le Seuil, 2017, 256 p.
- COLLINS, Suzanne, *The Hunger Games, Book One*, New York, Scholastic Press, 2008, 384 p.
- COMTE, Auguste, *Philosophie des Sciences*, Paris, Gallimard, 1997, 462 p.
- DICK, Philip K., *Les androïdes rêvent-ils de moutons électriques ?*, Paris, J'ai Lu, 2014, 283 p.
- FRIEDMAN, Yona, *Utopies réalisables*, Paris, L'éclat poche, 2015, 240 p.
- HARTOG, François, *Régimes d'historicité : Présentisme et expériences du temps*, Paris, Seuil, 2003, 272 p.
- HERBERT, Frank, *Dune*, Paris, Pocket, 2012, 832 p.
- HUXLEY, Aldous, *Le meilleur des mondes*, Paris, Pocket, 2017, 320 p.
- MERLE, Robert, *Malevil*, Paris, Folio, 1983, 635 p.
- MINKOWSKI, Eugène, *Le temps vécu*, Paris, PUF, 2013, 432 p.
- MORE, Thomas, *L'Utopie*, Paris, Folio, 2012, 384 p.
- ORWELL, George, *1984*, Paris, Folio, 1972, 438 p.
- PETRELLA, Ricardo, *Le bien commun*, Paris, Éditions Labor, 1996, 96 p.
- PLATON, *La République*, Paris, Folio, 1993, 560 p.
- RABELAIS, *Œuvres complètes*, Paris, Gallimard : La Pléiade, 1994, 1888 p.
- ROTH, Veronica, *Divergent*, New York, Katherine Tegen Books, 2012, 496 p.
- VERNE, Jules, *Vingt mille lieues sous les mers*, Paris, Le Livre de Poche, 2001, 606 p.
- VOLTAIRE, *Candide et autres contes*, Paris, Folio, 1992, 512 p.
- WELLS, Herbert George, *La machine à explorer le temps*, Paris, Folio SF, 2016, 176 p.
- WRIGHT, Erik Olin, *Utopies réelles*, Paris, La Découverte, 2017, 624 p.

Filmographie

- 2001, l'odyssée de l'espace*, Dir. Stanley KUBRICK, Perf. Keir DULLEA, Gary LOCKWOOD, William SYLVESTER, Metro-Goldwyn-Mayer, 1968, Film.
- Black Mirror*, Creator Charlie BROOKER, Perf. Daniel LAPAINE, Hannah JOHN-KAMEN, Jesse ARMSTRONG, Channel 4 Television Corporation & Netflix, 2011—, Série.
- Blade Runner*, Dir. Ridley SCOTT, Perf. Harrison FORD, Rutger HAUER, Sean YOUNG, Warner Bros., 1982, Film.
- Children of Men*, Dir. Alfonso CUARÓN, Perf. Clive OWEN, Juan Gabriel YACUZZI, Michael CAINE, Universal Pictures, 2006, Film.
- Fahrenheit 451*, Dir. François TRUFFAUT, Perf. Julie CHRISTIE, Oskar WERNER, Cyril CUSACK, Universal Film, 1966, Film.
- Interstellar*, Dir. Christopher NOLAN, Perf. Ellen BURSTYN, Matthew MCCONAUGHEY, Mackenzie FOY, 20th Century Fox, 2014, Film.

³⁸ Ricardo PETRELLA, *Le bien commun*, Paris, Éditions Labor, 1996, 96 p.

La servante écarlate, Creator Bruce MILLER, Perf. Elisabeth MOSS, Yvonne STRAHOVSKI, Joseph FIENNES, Hulu, 2017—, Série.

Maniac, Creator Cary JOJI FUKUNAGA, Perf. Jonah HILL, Emma STONE, Sonoya MIZUNO, Netflix, 2018, Série.

Metropolis, Dir. Fritz LANG, Perf. Alfred ABEL, Gustav FRÖHLICH, Rudolf KLEIN-ROGGE, Universum Film, 1927, Film.

Soleil vert, Dir. Richard FLEISCHER, Perf. Charlton HESTON, Leigh TAYLOR-YOUNG, Chuck CONNORS, Metro-Goldwyn-Mayer, 1973, Film.

3.15 Quelques critiques de cinéma

Official Competition

by Mariano Cohn and Gastón Duprat

When Ego Takes Control

Official Competition is the new feature film by Mariano Cohn and Gastón Duprat (*The Distinguished Citizen*). It tells the story of a very rich man, Humberto Suárez (José Luis Gómez), who wants to do something important before he dies, in order to leave behind him something worthwhile. He decides to produce a feature film. He asks Lola Cuevas (Penélope Cruz), who is the famous writer-director of the moment, to be in charge of the project. She decides to hire Félix Rivero (Antonio Banderas), famous self-made film actor of the moment, and Iván Torres (Oscar Martínez), older stage actor who went through classical training, to play the two lead parts of the film she wrote. The film shows part of the rehearsal process preceding the shooting of the film, which happens in the villa of Lola Cuevas.

The film is a confrontation between three huge egos with very different work methods and work ethics. Everyone is trying to show who has the most power by imposing their views and rhythm on the rehearsal process. What starts as a comedy, playing ludicrously on the oppositions between the three characters and their approach to art, and in particular film and character interpretation, turns into a dark fable about the disaster created by the lack of love and compassion and kindness.

I see this film as a reflection on our modern world and societies, where the appearance has become more important than who we really are deep inside, and where image is everything to the detriment of authenticity. Games, in any arena of power, are never without consequences, and the ones portrayed in the film are the most extremes. The film will finally be made, in spite of all the obstacles, but at what cost? Is it worth it? Everyone can interrogate their own self about the value of fighting for who we pretend to be and want to be perceived as by others instead of defending our authenticity, our heart, our ability to love.

The mise-en-scène of the film is very well mastered, and reinforces the coldness of the environment and of the relationships between these three characters, who are brilliantly interpreted and directed.

A dark fable about our modern world.

Directors : Mariano Cohn and Gastón Duprat

Screenplay : Mariano Cohn, Andrés Duprat, Gastón Duprat

Producer : Jaume Roures

Cinematography : Arnau Valls Colomer

Editing : Alberto del Campo

Sound : Leandro de Loredo

Music : Eduardo Cruz

Cast :

Penélope Cruz (Lola Cuevas)

Antonio Banderas (Félix Rivero)

Oscar Martínez (Iván Torres)

José Luis Gómez (Humberto Suárez)

The Story of my Wife

by Ildikó Enyedi

Social Hypocrisy versus Authenticity, or How to Lose Your Soul

The Story of my Wife is the new feature film by Hungarian filmmaker Ildikó Enyedi (*On Body and Soul, 20th Century*), adapted from a book by Hungarian writer Milán Füst. This is a book that Ildikó Enyedi read a long time ago and which appealed to her greatly at the time. She had wanted to do a film adaptation of the book for a long time.

The Story of my Life tells the story of Captain Jakob Störr (Gijs Naber), who is a ship captain, and has been lonely for a long time, without a woman in his life. One day, while having a meeting in a restaurant with his long-time friend Kodor (Sergio Rubini), he makes the bet he will propose to the next woman walking through the door. It happens to be Lizzy (Léa Seydoux), a French woman. When he finally finds the courage to go to her and tell her about the bet, she accepts.

The film then tells the story of this couple, between a ship captain and a refined French young woman, an ill-assorted couple with very different needs and takes on the meaning of life.

What is really striking while watching this film is the treatment of the story by Ildikó Enyedi. The *mise-en-scène* is really amazing. Her way of filming the sea, the life on a cargo, the loneliness of the captain, who rarely mingles with the rest of the crew reinforces the contrast with the other aspect of his life, trying to live within a finely decorated apartment and in a very refined high-society. Ildikó Enyedi creates a striking opposition between this simple man with basic needs and expectations in life that his wife tries to model to her image, while at the same time having a life on her own, while he is away, and sometimes even when he is here. The formal work on the image and on the lighting by cinematographer Marcell Rév is also amazing.

The film reflects on the topic of risking losing yourself and your soul when not remaining authentic to who you really are, and how one can be corrupted by another person when blinded by love. In spite of what happens to him, Captain Jakob Störr tries to find salvation at the end of the film when he finally confronts the truth about having been fooled by his wife.

The beauty of the film resides also in the complexity of the characters, and the fact that they are not really who we expect them to be. In spite of her foolishness, Lizzy is genuinely in love with her husband, at times. And in spite of being the one playing the role of the fool, Jakob ends up being the one developing the most profound philosophy of life. These characters are carried by excellent actors, and in particular Gijs Naber playing Jacob, who is really amazing in the film.

There is something that stays with you after watching this film, a sort of sadness for these two persons who wasted their chance at true love in the name of social hypocrisy and having to pretend to be someone they are not. In this sense, this film is very modern, and reflects on our own modern world, where authenticity is not as much praised as it used to be.

Director : Ildikó Enyedi

Screenplay : Ildikó Enyedi based on the book by Milán Füst

Cinematography : Marcell Rév

Editing : Károly Szalai

Sound : Matthias Schwab

Music : Ádám Balázs

Cast :

Gijs Naber (Captain Jakob Störr)

Léa Seydoux (Lizzy)

Illusions perdues

de **Xavier Gianoli**

Où est passé le cinéma utopiste de Xavier Gianoli ?

Illusions perdues est le nouveau long métrage de Xavier Gianoli. Il s'agit d'une adaptation du roman éponyme d'Honoré de Balzac. Le roman et le film racontent l'histoire d'un jeune poète de province, Lucien Chardon, dit Lucien de Rubempré, qui monte à Paris dans l'espoir d'y faire carrière et d'y être publié.

Cette adaptation de Xavier Gianoli est très réussie. La mise en scène est brillante, très dynamique, et le film a beaucoup de rythme. La direction d'acteurs est aussi très précise, et les acteurs excellents. À noter notamment Benjamin Voisin (Lucien), qui produit un travail de création de personnage très détaillé et très crédible. On notera aussi Vincent Lacoste (Etienne Lousteau) et Salomé Dewaels (Coralie), deux rôles secondaires du film, également brillamment interprétés.

Le sujet du film est très actuel, et propose une réflexion en miroir sur notre société. En arrivant à Paris, Lucien se retrouve à écrire des critiques pour un journal populaire, et s'éloigne petit à petit de son rêve littéraire. Il travaille pour un journal dont l'objectif est de faire et défaire les réputations d'artistes, sans égard pour la valeur artistique des œuvres de ces artistes. Le ragot est devenu plus important. Le bon mot est devenu plus important. Ce que l'on pense vraiment, ce que l'on ressent n'est plus si important, au risque de se perdre, de perdre son authenticité.

Si ce nouveau long métrage de Xavier Gianoli est particulièrement réussi, et que l'on retrouve sa grande maîtrise de la mise en scène et de la direction d'acteurs, il est aussi désenchanté. Je regrette personnellement le Xavier Gianoli d'À l'origine ou de Marguerite, qui nous livrait, à travers ces deux films, un cinéma de l'utopie. Et dans ce genre, non seulement Xavier Gianoli excelle, mais il est aussi un des rares cinéastes français d'aujourd'hui à s'y frotter. On attend donc de nouveau avec impatience un film de Xavier Gianoli dans cette veine utopiste, car on en a grand besoin.

Réalisation : Xavier Gianoli

Scénario : Xavier Gianoli & Jacques Fieschi

Photographie : Christophe Beaucarne

Montage : Cyril Nakache

Son : Didier Lozahic & François Musy

Interprétation :

Benjamin Voisin (Lucien Chardon, dit Lucien de Rubempré)

Vincent Lacoste (Etienne Lousteau)

Salomé Dewaels (Coralie)

Xavier Dolan (Nathan d'Anastazio)

My Very Own Circus

by **Miryam Bouchard**

I want to be a normal girl

My Very Own Circus, by Canadian director Miryam Bouchard, is partly autobiographical, and tells the story of Laura, who grows up with her dad, who is a clown. Being his assistant in his shows, she misses a lot of school and travels around quite a lot. What was first a sort of fairy tale for a little girl soon becomes a burden and Laura aspires to a more stable life and to the possibility of going to a good school.

One of the reasons why we chose this film unanimously is because this is the film that moved us the most. We laughed, we cried, we cheered up for Laura. The film relies on a very strong storytelling, based on a solid script, a skillful mise-en-scène, a beautiful cinematography, and a poetic music. The unity of all these elements contributes to create an unforgettable experience.

The actors are amazing, particularly Patrick Huard playing the dad-clown. In the press kit, Miryam Bouchard, the director, admits that Patrick Huard was so good in his appropriation of the character that she finally forgot he was portraying her own dad and embraced the work as a work of fiction.

The film is also a wonderful tribute to the art of clowning. Through this art, the film is a reflection on conformity, and the need to fight against it at times. The film is also a beautiful plea for tolerance, and the acceptance of people who are different.

Let us hope this film will get all the credit it deserves, and will be distributed widely.

Director : Miryam Bouchard

Screenplay : Miryam Bouchard & Martin Forget

Cinematography : Ronald Plante

Editing : Valérie Héroux

Sound : Luc Boudrias

Cast :

Patrick Huard (Bill)

Jasmine Lemée (Laura)

Nevada

de Laure de Clermont-Tonnerre

Un très beau plaidoyer pour l'équithérapie

Nevada, de Laure de Clermont-Tonnerre, raconte l'histoire de Roman Coleman (Matthias Schoenaerts), un détenu ayant une lourde peine de prison et une grosse difficulté à gérer ses émotions, qui participe à un programme carcéral de dressage de chevaux sauvages dirigé par Myles (Bruce Dern).

À travers cette expérience, Roman va apprendre non seulement à dresser un cheval, à le monter, mais aussi à construire des relations avec son mustang, Marquis, avec Myles, le directeur du programme, avec certains de ses codétenus, et avec sa fille, Martha (Gideon Adlon). Il va aussi apprendre à exprimer et à mieux maîtriser ses émotions. Et commencer à former un début de projet d'avenir qui va lui donner enfin l'envie de sortir de prison.

L'interprétation de Matthias Schoenaerts (*De Rouille et D'Os*, *Frères ennemis*) est magistrale, l'image (Ruben Impens : *Alabama Monroe*) est splendide, et la mise en scène ainsi que la direction d'acteurs (Laure de Clermont-Tonnerre) parfaitement maîtrisées. À noter en particulier la façon très intimiste de filmer le lien entre Roman et son cheval.

Un très beau plaidoyer pour l'équithérapie, et un film magnifique et très émouvant à voir absolument.

Réalisatrice : Laure de Clermont-Tonnerre

Producteur : Ilan Goldman

Scénario : Laure de Clermont-Tonnerre, Mona Fastvold, Brock Norman Brock

Photo : Ruben Impens

Musique : Jed Kurzel

Montage : Géraldine Mangelot

Son : Zsolt Magyar, Gabor Erdei, Severin Favriau, Stéphane Thiebaud

Interprètes principaux :

Matthias Schoenaerts (Roman Coleman)

Jason Mitchell (Henry)

Bruce Dern (Myles)

Gideon Adlon (Martha)

Nice Girls Don't Stay for Breakfast

de Bruce Weber

Mais qui est vraiment Robert Mitchum ?

Le nouveau film de Bruce Weber, *Nice Girls Don't Stay for Breakfast*, dresse un portrait très atypique et par petites touches de l'acteur américain Robert Mitchum, acteur qui a occupé une place bien à part dans le milieu hollywoodien. Bruce Weber, à qui l'on doit le magnifique documentaire *Let's Get Lost* sur Chet Baker, a filmé pendant plusieurs mois Robert Mitchum. Il a également interviewé plusieurs personnalités qui l'ont bien connu (Johnny Depp, Clint Eastwood, etc.). Il a aussi rassemblé des images d'archives incroyables et le plus souvent inédites. Pourquoi ce film a mis plus de 20 ans à sortir reste un grand mystère, mais heureusement que ce film s'est fait et que l'on peut enfin le voir !

C'est un document d'une grande valeur. Il est construit de façon très atypique, un peu à la manière d'un portrait impressionniste en peinture. Il juxtapose des images d'archives, des extraits de films, des interviews, des images de Robert Mitchum que Bruce Weber a lui-même tournées. Cette juxtaposition semble n'obéir à aucune logique, si ce n'est celle du moment. On fait des va-et-vient dans le temps, on passe d'un extrait de film à une séance d'enregistrement de chanson, on passe d'un lieu à un autre, d'un âge à un autre. Tout cela forme un tourbillon dans lequel on se laisse emporter. Et au détour de l'un de ces tourbillons, on est saisi par

l'émotion d'une séquence, par le charme d'une autre, par un regard furtif capté par la caméra, ou par un moment de vérité où Robert Mitchum livre quelque chose de très intime devant la caméra de Bruce Weber.

Un film à voir absolument. Pour découvrir ou redécouvrir cet immense acteur. Pour pénétrer au cœur de la vie publique et privée d'un homme qui a occupé une place à part dans le cinéma américain. Un acteur plein de contradictions (qui n'en a pas?) et en même temps tellement attachant dans sa vulnérabilité. Un film qui ne laisse pas indifférent!

Réalisateur : Bruce Weber

Meltem

de Basile Doganis

Retour aux sources

Meltem (qui est le nom d'un vent violent en Grèce) est le premier long métrage de fiction de Basile Doganis, à qui l'on doit les deux documentaires *On The Edge* (2008) et *Altérations / Kô Murobushi* (2019), ainsi que les deux courts métrages de fiction *Le Gardien de son frère* (2007) et *Journée d'appel* (2014).

Meltem raconte le voyage initiatique d'Elena (Daphné Patakia), une jeune fille d'origine grecque vivant en France avec son père. Neuf mois après le décès de sa mère, qui était restée en Grèce, Elena décide d'entreprendre un voyage estival sur l'île de Lesbos, où sa mère vivait avec son nouveau compagnon, Manos (Akis Sakellariou), dans une maison dont elle a maintenant hérité. Elle est venue vendre cette maison, ou tout du moins c'est la raison officielle de son voyage. Elle se fait accompagner par deux amis de l'école d'hôtellerie, Nassim (Rabah Naït Ouffela) et Sekou (Lamine Cissokho).

Ce séjour sur l'île de Lesbos est l'occasion pour Elena de faire son deuil, de faire aussi la paix avec son passé et avec Manos, de découvrir ses sentiments amoureux, et de se confronter au problème des migrants à travers le personnage d'Elyas (Karam Al Kafri), migrant syrien qui vit clandestinement sur l'île, et qui croise régulièrement le chemin des trois amis.

Même si l'intrigue paraît à première vue un peu complexe, le film fonctionne à merveille. Cela est dû principalement à la mise en scène, qui reste toujours très près des personnages et qui fait aussi la part belle à l'environnement idyllique de l'île, et au fait que l'histoire est toujours racontée du point de vue d'Elena, ce qui donne une grande unité au film. Le film évite aussi de tomber dans les clichés de ce genre d'histoire, et présente la question des migrants d'une façon profondément humaniste. Le casting est parfait et la direction d'acteurs très bien maîtrisée. La performance de Daphné Patakia et celles de Rabah Naït Ouffela et Lamine Cissokho sont remarquables. Et même les rôles un peu plus secondaires de Karam Al Kafri (acteur non professionnel) et d'Akis Sakellariou sont magnifiquement interprétés. Le travail sur l'image de Konstantinos Koukoulios et la musique de Kyriakos Kalaitzidis sont splendides et accompagnent l'histoire en parfaite adéquation.

Une très grande réussite. À voir absolument. On attend le prochain avec impatience !

Réalisateur : Basile Doganis

Producteurs : Denis Carot, Marie Masmonteil, Fenia Cossovitsa

Scénario : Basile Doganis et Fadette Drouard

Photo : Konstantinos Koukoulios

Musique : Kyriakos Kalaitzidis

Montage : Riwanon Le Beller

Son : Simon Apostolou

Interprètes principaux :

Daphné Patakia (Elena)

Rabah Naït Ouffela (Nassim)

Lamine Cissokho (Sekou)

Karam Al Kafri (Elyas)

Akis Sakellariou (Manos)

Féodor Atkine (Edward)

La Chute de l'Empire Américain

de Denys Arcand

Le pouvoir de l'argent

La Chute de l'Empire Américain, le nouveau long métrage de Denys Arcand, est un film sur le pouvoir de l'argent. Conçu initialement comme

un film indépendant des précédents, et qui devait avoir le titre plus explicite « Le Triomphe de l'Argent », c'est au moment du montage que son réalisateur a noté les points communs importants avec *Le Déclin de l'Empire Américain* (qui explorait surtout la sexualité) et *Les Invasions Barbares* (qui explorait le thème de la mort).

Ce nouvel opus, qui du coup s'inscrit maintenant dans une série dont il constitue le troisième volet, a un ton plus léger que les précédents, et pourtant comporte une part de cynisme beaucoup plus importante. Selon Denys Arcand, l'argent l'aurait emporté sur tout, serait devenu roi. Le film raconte l'histoire de Pierre-Paul (Alexandre Landry), docteur en philosophie mais ayant choisi un boulot de livreur plutôt que d'enseignant de philosophie à la fac parce que ça paye mieux, se retrouve par hasard en possession de deux gros sacs remplis de billets. Il s'est simplement trouvé là au bon moment, a été témoin d'un hold up qui a mal tourné, et s'est retrouvé avec les deux sacs littéralement à ses pieds.

Il fait alors équipe avec une call girl, Aspasia (magnifique Maripier Morin), qu'il a aussitôt appelée pour commencer à profiter de son nouveau butin, et avec Sylvain Bigras (Rémy Girard avec les cheveux longs!), détenu récemment sorti de prison et qui a appris, pendant son incarcération, la gestion financière. Le film nous montre alors à quel point il est devenu facile de faire disparaître de l'argent douteux et de le transformer, après quelques transactions passant par des paradis fiscaux, en argent propre. Tout l'enjeu du film est alors de savoir si les protagonistes vont s'y prendre de la bonne façon, et surtout avant d'être rattrapés par la police qui enquête ou les malfrats à qui cet argent appartenait initialement.

Ce qui marque le plus, et qui est à mon avis un choix très judicieux, est le contraste entre le cynisme du thème traité et la légèreté de traitement. On est clairement dans le registre de la comédie et en même temps cette comédie est à la fois jubilatoire et fait froid dans le dos. Le film repose aussi sur les épaules de ses trois interprètes principaux, Alexandre Landry, Maripier Morin et Rémy Girard, qui campent avec brio des personnages que tout éloigne, et qui pourtant vont réussir envers et contre tout à nouer entre eux une relation forte et profonde bien qu'au départ totalement improbable. C'est l'autre point fort du film.

Une réflexion cynique et amusée (sans doute aussi un peu désabusée) sur notre société contemporaine et sur sa dérive de plus en plus amoral.

Réalisateur : Denys Arcand
Productrice : Denise Robert
Scénario : Denys Arcand
Photo : Van Royko
Musique : Mathieu Lussier, Louis Dufort
Montage : Arthur Tarnowski
Son : Martin Desmarais, Louis Gignac
Interprètes principaux :
Alexandre Landry (Pierre-Paul Daoust)
Maripier Morin (Aspasie/Camille Lafontaine)
Rémy Girard (Sylvain « the brain » Bigras)

Meerkat Moonship

by **Haneke Schutte**

The Difficult Work of Mourning

Meerkat Moonship is the second feature film by South African writer-director Haneke Schutte. It tells the story of Gideonette, a teenage girl beautifully interpreted by young actress Anchen du Plessis, who is sent to her grandparents' house in the country side after the sudden death of her dad. Her mom can't deal with the death and take care of her daughter at the same time.

When Gideonette arrives at her grandparents' place, she is filled with fear, about an old curse supposedly affecting her family, and about everything else. She stays in her bedroom and doesn't even want to go outside. She takes most of her meals in her bedroom. But after a while, she can't avoid noticing a little deaf kid, Bhubesi (Themba Ntuli), playing in front of her window.

Little by little, first out of curiosity, and then because she probably wants to change her inner state, she starts connecting with Bhubesi and he is the one who will manage to make her go outside to play with him and finally overcome her fears. Part of Bhubesi's daily schedule, which also becomes part of Gideonette's daily schedule, consists in building a ship to go to the moon, and therefore training for this trip.

What is striking the most when watching the film is the way Haneke Schutte chooses to film her story. We are with Gideonette from the very

beginning of the film. The camera is very close to her most of the time, and allows the audience to feel her fears and to follow her journey towards being freed from her interior demons as if we were ourselves following that same journey. The framing is innovative, and quite minimalist. The camera also often films through a window or through a curtain, metaphorically suggesting that Gideonette only sees the world through the walls of her protected space. And it's only at the end that she'll manage to get a first-hand experience of the world.

The work of cinematographer Willie Nel is splendid. He has a way to film interiors and exteriors that emphasizes the visual poetic quality of the spaces. The editing by Warwick Allan is very thorough and conveys to the film a slow and cyclic rhythm accompanying the slow transformation of Gideonette's subconscious mind and is perfectly right for it. And the outstanding music by Clare Vandeleur accompanies beautifully this poetic journey.

Director : Haneke Schutte

Screenplay : Hanneke Schutte

Cinematography : Willie Nel

Editing : Warwick Allan

Sound : Charlotte Buys

Music : Clare Vandeleur

Cast :

Anchen du Plessis (Gideonette)

Themba Ntuli (Bhubesi)

Pony and Birdboy

by Mari Rantasila

The Reality of my Imaginary Friend

Pony and Birdboy was one of the serious contenders. It is the fourth feature film by Finnish director Mari Rantasila. It tells the story of Pony, 7 (Jenni Lausi), who is afraid to go to a new school. She wants to stay in her childhood world where she can put on a pony muzzle, pretend she is herself a pony, and play with her imaginary friend Birdboy (Aapo Puusti). He is a fun companion, has all the answers, and allows her to have a good time even though, at times, real things are a bit disturbed or displaced by

their numerous adventures.

But she will have to confront her fear of going to her new school and learn to put childhood's fantasies at their right place, and to grow up. But she has wonderful parents who indulge in their daughter's needs while at the same time accompanying her and helping her in that difficult transition.

This is by far one of the best films I have seen for that age category. Through the wonderful directing of Mari Rantasila, the film captures the energy so particular to that age, and the story is always filmed from Pony's point-of-view. The colors are vivid, and even though Birdboy is an imaginary friend, he is totally real to Pony and to us, the audience. It's only towards the end and only on rare occasions that the director suggests Birdboy is imaginary by giving us some glimpses of the real world, but even then, we are just doubtful about his imaginary status.

Timo Heinänen, the cinematographer, has worked on all of Mari Rantasila's films. This long time collaboration allows the cinematography to be totally in sync with the storytelling and the visual choices of the director : Pony's point-of-view, colorful images, dynamic camera moves, and the candor and innocence of childhood.

Director : Mari Rantasila

Screenplay : Hannamaija Matila, Veera Salmi

Cinematography : Timo Heinänen

Editing : Hanna Kuirinlahti

Sound : Petter Fagelund

Music : Markku Kanerva

Cast :

Jenni Lausi (Pony)

Aapo Puusti (Birdboy)

Little Ticks

by **Andréa Bescond and Éric Métayer**

Pedophilia is the Ultimate Abuse of Trust

Little Ticks started as a one-woman show written and interpreted by Andréa Bescond and directed by Éric Métayer. François Kraus saw it and immediately decided he wanted to produce a film based on the show.

His producing partner, Denis Pineau-Valencienne, agreed and, three years later, this amazing film exists. It tells the story of Odette, an eight-years old innocent girl, who likes drawing and dancing, and is sexually abused by a very close friend of her parents. Once adult, Odette finally manages to face her childhood trauma through dance and therapy, and with the help of some of her close friends.

This film is amazing. First because it manages to speak about a very difficult topic with a lot of delicacy and a lot of empathy. The authors Andréa Bescond and Éric Métayer play very skillfully with a blurring between fantasy, memory, and reality and, through this device, explore masterfully the impact such a trauma can have on the psyche of a little girl. The film explores the difficulty for a little girl to go through such a trauma, but also the difficulty of the adult she became to deal with her past and get free from it. Dance plays an important role in the process, and the dancing sequences serve as a counterpoint and also a metaphor to the expression of some inner rage and anger that Odette, once adult, feels inside her.

Andréa Bescond interprets herself, like in the show, Odette adult, and delivers a great performance, full of delicacy, emotion, rage, and always truthfull. The other actors are also great. Cyrille Mairesse, who plays Odette when she is a girl, brings such an innocence to her character and at the same time a strength which already prefigure what Odette will then become as an adult. And Pierre Deladonchamps managed to play Gilbert, the perpetrator of such horrible acts, without judging him and by finding the inner logic of the character and embracing his point of view and its downward spiral.

This is by far the best film I have ever seen on this topic. It needs to be shown and discussed everywhere, to help liberate everyone who has been or is a victim of such a horrible act.

Directors : Andréa Bescond, Eric Métayer

Screenplay : Andréa Bescond, Eric Métayer

Cinematography : Pierre Aïm

Editing : Valérie Deseine

Sound : Thomas Gauder

Music : Clément Ducol

Cast :

Andréa Bescond (Odette Le Nadant)

Karin Viard (Mado Le Nadant)

Clovis Cornillac (Fabrice Le Nadant)
Pierre Deladonchamps (Gilbert Miguié)
Cyrille Mairesse (Odette enfant)

The Big Bath

by Gilles Lellouche

Gilles Lellouche where we didn't expect him

The Big Bath, directed by Gilles Lellouche, and featuring Mathieu Amalric, Benoit Poilvoerde, Guillaume Canet, Virginie Efira and a whole bunch of other well-known French actors, tells the story of several men living mediocre lives and having mediocre relationships, who decide to join a leisure practice of masculine synchronised gymnastic.

Through that decision of joining a practice that is usually associated with women, they discover and develop their ability of connecting to their feminine side and expressing their emotions and feelings, and discover the value of male bounding and friendship.

After deciding, in spite of their amateurship, to participate to the world championship in Norway, they will face a lot of mockery, find some unexpected support, and raise to the challenge together. They will all gain some self-knowledge and some pride from the experience, and finally become able to face their responsibilities and look at reality with less fear. They will find their inner strength and joy.

This film is very moving at times, very funny at other times. It is well written and the characters, particularly well developed, are endearing. The directing is very good, with some great ideas. The actors are excellent. And the film rises some very interesting questions about the society we live in.

Director : Gilles Lellouche

Screenplay : Gilles Lellouche, Julien Lambroschini, Ahmed Hamidi

Cinematography : Laurent Tangy

Editing : Simon Jaquet

Sound : Cédric Deloche, Gwénolé Leborgne, Marc Doisne

Music : John Brion

Cast :

Mathieu Amalric (Bertrand)
Guillaume Canet (Laurent)
Benoît Poelvoorde (Marcus)
Jean-Hugues Anglade (Simon)
Virginie Efira (Delphine)
Leïla Bekhti (Amanda)
Marina Foïs (Claire)
Philippe Katerine (Thierry)

Happy as Lazzaro

by Alice Rohrwacher

Embodied Goodness

With *Happy as Lazzaro*, Alice Rohrwacher delivers a fable about goodness. She portrays Lazzaro, a young man who is only capable of goodness and knows nothing else. Using the conventions of the fable, she first places Lazzaro, her anti-hero, in a place where some peasants are exploited by a woman who has hidden to them social progress and take advantage of them and their credulity. Lazzaro is the one who is always willing to help anyone who needs it, without any thought about what he could gain from it, or even what he might want to do for himself. A man without ego, who is only living from his heart.

The same young man is then propelled into the modern world, where, contrarily to the other characters from his past who have aged, he remains the same, same age, same appearance, same goodness. He meets again some of the characters from his past and, again, faithful to his nature and what he is, is trying to help them the best he can. His heart is untouched by wickedness, and he can only do what seems right given the situation he is in, without thinking about the possible consequences for himself.

This film is extremely moving. It reaches a deep place inside of us, if this place is still open, where we know what our truth is, in spite of the modern world and the media who would like to make us think this truth doesn't exist anymore. And the strength of *Happy as Lazzaro* is to bring us to that inner place, to show us that it always exists, whether we are aware of it or not.

The film is beautifully shot, using 16 mm, which is for the director a way to renounce to an absolute control over the image and accept that the film will only reveal later what it captured during shooting. This process is a beautiful metaphor of what art can do to us : it shows us something, but sometimes what we really see is only revealed to us later. It resonates inside us but we become aware of that inner resonance at a later time, often through an emotional experience, a heartfelt and soulful resonance.

The actor playing Lazzaro, Adriano Tardolo, had never acted before, and brings to the role its own innocence face to the film world. He doesn't have any preconceptions about what acting is, and under the right guidance of the director, delivers a very truthful and moving performance, even in his body attitude, which conveys something of the abandonment to each moment and its needs.

This is a film which I hope will have the place it deserves, the place of an artwork that can change our consciences. And it is very much needed.

Director : Alice Rohrwacher

Screenplay : Alice Rohrwacher

Cinematography : H el ene Louvart

Editing : Nelly Quettier

Sound : Maxence Ciekawy, Fran ois Musy

Cast :

Adriano Tardiolo (Lazzaro)

Agnese Graziani (Antonia bambina)

Alba Rohrwacher (Antonia)

Image Book

by **Jean-Luc Godard**

How to Give Back Meaning to Image and Sound

In *Image Book (Le Livre d'Images)*, Jean-Luc Godard continues the exploration he started in his precedent films, and in particular in *Goodbye to Language (Adieu au Langage)*. He adopts a similar approach, deconstructing the cinematographic language we are used to, by proposing a new language with images and sounds, that doesn't respect the usual conventions.

For instance, he discontinues images and sounds at a point that is unexpected. When you expect the moving image or the sound to continue, he stops them. Or he repeats a particularly significant bit of image or sound twice or more. The cuts in the images and sounds are usually not synchronized. One usually stops while the other continues, and it is almost never in the same way each time. Or he creates silence in the sound while the image continues, or silence in the image (black screen) while the sound continues. He also transforms images (playing with colors and saturation and ratio) and sounds (playing with transformations and spatialisation). He also plays with imperfections, at the image and sound level (using as a language things that are usually hidden).

By doing all that, he reintroduces significance in a media that sometimes lost it. When not knowing what to expect, the audience is forced to pay particular attention to what's going on. This is a poetic statement since it creates a new language, one that is peculiar to its author. This also creates a new rhythm, its own rhythm, with a density which at times saturate the mental ability to make sense and put the audience in a state of trance.

All that is not done for nothing, it is put in service of a statement, a statement which is the juxtaposition of images, sounds, music, and texts that are meaningful to Godard, and through which he is trying to tell us something, and maybe even to awaken consciences. Some of his favorite themes are present, like revolution, what it is, and if it is possible or even desirable. Or meaning : how do you achieve it and for what purpose. Or power, with its derivatives, dictatorship and oppression. And Godard also proposes a reflection on the Arabic world, visiting history, ancient and recent, as well as pictural representations associated to it. He also questions us about what it means to represent something, and the violence of this act of representing which necessarily makes choices and obliterate part of what is represented.

Like for *Goodbye to Language (Adieu au Langage)*, this virtuosity of form and content and meaning and symbols merging together at times, or creating chaos at other times, leaves us confused, full of wonder, certain not to have understood or grasped everything, but, and this is the best strength of last Godard films, stays with us for days to come, and feeds the thread of our reflection, almost in spite of us. It is an experience that is really worth living and is necessarily transforming.

Director : Jean-Luc Godard

Editing : Jean-Luc Godard, Fabrice Aragno, Jean-Paul Battaglia, Nicole Brenez

Yomeddine

by **A.B. Shawky**

Sublime !

Yomeddine, first feature film by Egyptian director A.B. Shawky, is a wonderful film, full of humanity. It tells the story of a former leper, Bashay, who in spite of his condition and harsh life, has kept a certain innocence about life. One day, after his wife passed away, he decides to begin a journey back to where he is from, in hope to reconnect with his family who abandoned him when he was a kid. He is accompanied in his journey by Obama, an orphan Bashay is taking care of, and who also would like to know about his parents.

This journey is the opportunity for Bashay to discover who he truly is, to forge new friendships, and to tell the world he is also a person, a human being, and not just a leper. The film, with all its humanity, teaches us to see difference in a different way, with an open heart.

The mise-en-scène is brilliant and the story is told in a very minimalistic way. The image by Federico Casca is beautiful. The music by Omar Fadel is amazing. And the actors are phenomenal, and wonderfully directed. Everything contributes to this lesson in humanity, but always with a lot of truthfulness and without ever moralizing. A very rare film. I hope it will get all the consideration it deserves and will be wildly distributed.

Director : A.B. Shawky

Screenplay : A.B. Shawky

Photography : Federico Cesca

Editing : Erin Greenwell

Music : Omar Fadel

Sound : Sherif Zein

Set Decorator : Laura Moss

Cast :

Rady Gamal (Beshay)

Ahmed Abdelhafiz (Obama)

Rafiki

by **Wanuri Kahiu**

Love Against Prejudices

Rafiki, first feature film by Kenyan director Wanuri Kahiu, tells the story of Kena (portrayed by Samantha Mugatsia) and Ziki (portrayed by Sheila Munyiva), who fall in love in a Kenya where homosexuality is not only very negatively judged upon but also outlawed. The strength of this film is to show us first and foremost a true love story. And from that point of view, the film is beautifully directed by Wanuri Kahiu, and the actresses portraying the two protagonists are amazing. The image by Christopher Wessels is beautiful, and the sounds (Frédéric Salles) and the music accompany perfectly the story.

The direction is very mature, always chooses beauty and poetry, even in scenes where the tension is at its paroxysm, like when the population violently beats up the two young women. But telling the story of this unconventional love in a country where it is outlawed (the film has been forbidden in Kenya) is also a pretext for Wanuri Kahiu to depict Kenya today, with its contradictions, and also its beauty and humanity, with the best and the worse. The film is very political. This is not a coincidence that the dads of these two women are politicians and opponents in a local election. On top of increasing the stakes for the love story, it is also an opportunity for the director to show us the political intricacies in Kenya. The film also shows the presence of religion and the strong influence it has on some people.

This is also in a way a story of emancipation for these two young women. Both of them will have to make a choice between their father and their heart, between being faithful to the past or being themselves no matter what the consequences will be. And this path is not easy, even with love, passion, and conviction. And some of the battles are lost along the way, and even physically and emotionally painful. But the message of the film is to keep going at it no matter what, trust that in the end, by remaining true to yourself, you will find and know who you are.

A brilliant and very sensitive film.

Director : Wanuri Kahiu

Screenplay : Wanuri Kahiu, Jenna Bass

Cinematography : Christopher Wessels

Production Design : Arya Laloo

Editing : Isabelle Dedieu

Sound : Frédéric Salles

Cast :

Samantha Mugatsia (Kena)

Sheila Munyiva (Ziki)

Neville Misati (Blacksta)

Wildlife

by Paul Dano

Slow Desintegration of a Family

The Critic's Week opened this year with the remarkable *Wildlife*, first feature film by actor Paul Dano (*Little Miss Sunshine*, *There Will Be Blood*, *Youth*, *Okja*). He wrote the screenplay with Zoe Kazan, the grand-daughter of Elia Kazan, and it is an adaptation from the book by Richard Ford.

In a director's statement collected by Charles Tesson for the Critic's Week catalogue, Paul Dano says he was empowered by what Richard Ford told him when he asked for the permission and the rights to adapt the book for the screen : "I'm grateful to you for your interest in my book ; but I should also say this - in hopes of actually encouraging you : my book is my book ; your picture - were you to make it - is your picture. Your movie-maker's fidelity to my novel is of no great concern to me. Establish your own values, means, goals ; leave the book behind so it doesn't get in the way." What a great lesson on how to adapt a book, and what a great way to start the long process of making a feature film.

I'd say that Paul Dano did an excellent job following that piece of advice. The film is very personal, this is a very personal statement. It is brilliantly directed. Paul Dano and his producer Alex Saks have chosen the right people to work on this film. The image by Diego Garcia is beautiful, and together with the *mise-en-scène*, conveys an intimate feeling, making us be part of this family, its desintegration, and its reconstruction. Also, the point of view is almost always of Joe, the pre-teen son. We sees what he sees, and we feel what he feels.

The music by David Lang reinforces the mood of something falling apart, and the choice of where to put music and where not is very accurate given the type of story and its arch.

The actors are amazing. Carey Mulligan (*An Education, Drive, Shame, The Great Gatsby, Suffragette*) portrays Jeannette Brinson, the mom, and confirms once more her talent and the cleverness of her choices as to which projects to be involved with. Jake Gyllenhaal (*Brokeback Mountain, Zodiak, Brothers*) portrays Jerry Brinson, the dad, who is unable to face the challenges that arise in his relationships with his wife and son, and chooses to run away to fight a fire rather than confront his own demons. Jake Gyllenhaal is excellent in this film, and find the right dosage between a sort of mutism when running away, and the intensity of his anger when he comes back to find that the family has desintegrated. Ed Oxenbould, who portrays Joe Brinson, the son, is amazing and carries the film on his shoulders. He is the witness of this story, we see it unraveling through his eyes and emotions, and he is always completely truthful.

Again in the director's statement collected by Charles Tesson, Paul Dano shares an entry from his journal : "I want this film to feel personal. I want it to explore feelings I have had, through Joe. I want it to ask questions about family and parents. I want to explore a loss of hope, a family unraveling, and then finally surviving. I want to explore how even when the worst thing happens, we can still survive. We can still be family. We may never be the same, but we still have love."

Not only has Paul Dano managed to do that, but he did so much more and shares with us a film that is extremely well made, is indeed personal, and very moving. I can't wait for his next film !

Director : Paul Dano

Production : Alex Saks

Screenplay : Paul Dano, Zoe Kazan

Cinematography : Diego Garcia

Editing : Matthew Hannam, Louise Ford

Sound : Jacob Ribicoff

Production Design : Akin Mckenzie

Music : David Lang

Cast :

Carey Mulligan (Jeanne Brinston)

Ed Oxenbould (Joe Brinston)

Bill Camp (Warren Miller)
Jake Gyllenhaal (Jerry Brinston)

Loveless (Faute d'amour)

by Andrey Zvyagintsev

In the Middle of a Loveless Divorce

Loveless is the fifth feature film by Andrey Zvyagintsev after *Leviathan* in 2014. It tells the story of a couple, Zhenya (Maryana Spivak) and Boris (Alexey Rozin), getting divorced while their son, Alyosha (Matvey Novikov), is going missing. Zhenya and Boris don't really love each other anymore, and are barely concerned by the disappearance of their son. They have to at least pretend they care. But in the middle of the search for their missing son, feelings and emotions are exacerbated, including the hatred they have for each other.

What strikes first in this film is the pace of it. Andrey Zvyagintsev really takes the time to explore places, the relationship between the characters and these places, and what these places tell us metaphorically about the people in them. There are these big buildings where so many people live next to each other, the nature around it with a forest and the cold winter, the big apartment where Zhenya, Boris, and Alyosha used to live as a family, and the abandoned building at the fringe of the city where Alyosha used to go play.

The acting is also amazing. I happen to know that between *Leviathan* and *Loveless*, Andrey Zvyagintsev took some acting classes in Moscow with Lola Cohen, a friend of mine and teacher of Method Acting at the Lee Strasberg Institute in New York. He wanted to be better at directing actors. Speaking of an artist dedicated to his art! And in *Loveless*, the direction of the actors is amazing indeed!

But *Loveless*, like *Leviathan*, is also a very political film. It points the inability (or indifference?) of the police force to the disappearance of people, and in the case of the film, of a kid. And it shows how the people take care of it as a last resort, through an association, « Liza Alerte », which really exists and is composed of people who look for missing people for free.

A great film that confirms Andrey Zvyagintsev as one of the major Russian film directors.

Director : Andrey Zvyagintsev

Screenplay : Oleg Negin, Andrey Zvyagintsev

Cinematography : Mikhail Krichman

Production Design : Andrey Ponkratov

Music : Evgeny Calperin

Editing : Anna Mass

Sound : Andrey Dergachev

Cast :

Maryana Spivak (Zhenya)

Alexey Rozin (Boris)

Matvey Novikov (Alyocha)

Marina Vasilyeva (Masha)

Andris Keishs (Anton)

Alexey Fateev (Coordinator)

**DIALOGUES ENTRE LES ARTS ET LES SCIENCES :
VERS UNE PHÉNOMÉNOLOGIE DE LA CRÉATION**

Ivan MAGRIN-CHAGNOLLEAU

HABILITATION À DIRIGER DES RECHERCHES

TOME 2

MÉMOIRE DE SYNTHÈSE

TABLE DES MATIÈRES

| | |
|---|----|
| Remerciements..... | 3 |
| Ouverture..... | 5 |
| Épistémologie..... | 7 |
| Point(s) de départ..... | 10 |
| La voix dans tous ses états..... | 13 |
| Rencontres..... | 15 |
| Transitions et Tournants..... | 18 |
| Multi-, Inter- et Trans-Disciplinarité..... | 20 |
| La voix, encore !..... | 23 |
| Géophysique..... | 25 |
| Langage et Cognition..... | 27 |
| La méthode de l'Actors Studio..... | 28 |
| Théâtre, performance et cinéma..... | 30 |
| Création Recherche..... | 33 |
| Phénoménologie du processus créatif..... | 36 |
| Énaction..... | 40 |
| Explicitation..... | 43 |
| Psychologie et Création Artistique..... | 48 |
| S'impliquer comme chercheur dans la transition écologique..... | 50 |
| IA+CA..... | 52 |
| Le sens de la recherche : vers une recherche qui a du sens..... | 55 |
| Bibliographie..... | 56 |

REMERCIEMENTS

Patrick Duvaut, qui le premier a su détecter mon goût pour la recherche, et l'encourager.

Frédéric Bimbot, qui a dirigé ma thèse et m'a montré un modèle de chercheur atypique et stimulant.

Aaron Rosenberg, qui a été mon mentor lors de mon post-doc à AT&T et avec lequel j'ai eu de nombreuses discussions passionnantes sur de très nombreux sujets.

Richard Baraniuk, qui m'a appris comment diriger une équipe de recherche dans la bienveillance et avec efficacité et focus.

Nathalie Vallès-Parlangeau, qui a été une partenaire privilégiée dans mon cheminement de chercheur, avec laquelle j'ai développé plusieurs projets, et qui m'a toujours encouragé sans jugement sur mes voies transversales.

Jean-François Bonastre, qui a été, avec Frédéric Bimbot, un partenaire au long cours tout au long de la première partie de ma carrière, avec fidélité et rigueur.

Jean-Marie Hombert, à qui je dois en grande partie mon intégration au CNRS, et qui a su créer un laboratoire de recherche où il faisait bon vivre et chercher.

Élisabeth Saint-Blancat, qui a éveillé en moi le désir profond d'une carrière artistique, et qui a accompagné avec bienveillance et encouragements mon désir de théâtre.

Marcia Haufrecht, qui a été mon professeur de théâtre en première année de l'Actors Studio et qui a su nous initier avec un grand talent à la méthode de Lee Strasberg.

Arthur Storch, qui a eu le don de me pousser vers des personnages qui m'obligeaient à explorer en dehors de ma zone de confort.

Gene Lasko, qui a encouragé ma passion pour Shakespeare et qui a su nous faire comprendre en profondeur cet immense artiste.

Cynthia Reynolds, qui a été mon professeur de technique Alexander et qui m'a éveillé à un autre regard sur la vie et sur moi-même.

Ron Leibman, qui m'a permis de faire éclater les derniers blocages que j'avais en tant qu'acteur, et grâce auquel j'ai enfin expérimenté le vrai plaisir de jouer.

Bob LuPone, qui m'a encouragé à poursuivre conjointement une carrière d'acteur et de metteur en scène, et qui m'a donné de précieux conseils pour cette dernière.

Costin Mireanu, à qui je dois ma réintégration au CNRS après un changement radical de direction de recherche, et qui m'a donné une grande liberté pour définir le projet d'une nouvelle équipe.

Joëlle Aden, qui m'a encouragé pendant plusieurs années sur les chemins très atypiques que je poursuis en tant qu'artiste chercheur.

Sylvie Dallet, qui est artiste chercheur, et m'a offert de nombreuses opportunités de création recherche et un soutien inconditionnel lors de traversées institutionnelles plus compliquées.

Richard Kronland, qui m'a permis de rejoindre un nouveau laboratoire plein de promesses et où la création recherche ainsi que la phénoménologie ont la part belle, un laboratoire inter- et trans-disciplinaire où les nouvelles expériences sont possibles.

Christine Esclapez, qui a accompagné une étape de ce travail d'HDR, en me rappelant notamment les exigences du milieu académique par rapport à cet exercice.

Claire Lahuerta, qui a accepté d'être la garante de ce travail et dont les suggestions et retours ont aidé à son façonnage final.

Mon jury d'HDR qui a accepté de prendre le temps de porter un regard critique et bienveillant sur mon parcours d'artiste chercheur.

OUVERTURE

Je voudrais commencer ce mémoire de synthèse en remerciant toutes les personnes qui m'ont permis d'être là où je suis aujourd'hui. Certaines seront nommées dans ce document. Mais il y en a d'autres qui ont joué des rôles importants et qui ne seront pas nommées. Que toute personne qui m'a aidé à avancer d'une façon ou d'une autre sur mon chemin personnel et professionnel soit ici remerciée du fond du cœur. On avance avec les autres, et avec soi-même.

Mon ambition, dans ce document, est de retracer comment je suis arrivé à ce point de ma vie où je prends le temps de réfléchir à mon passé, à mon cheminement, et à l'émergence des thèmes qui me préoccupent le plus actuellement. J'ai lu récemment le livre de Paul Ricœur intitulé *Réflexion faite*¹, et dont le sous-titre est *Autobiographie intellectuelle*. Je ne suis pas sûr que ce mémoire soit uniquement auto-biographique, ni qu'il se situe uniquement sur le plan intellectuel. J'ose espérer que non. Non pas que je dénigre les activités intellectuelles qui ont occupé une bonne partie de ma vie professionnelle, mais parce que je ne crois pas à la possibilité de se raconter en se cantonnant au niveau intellectuel. Parce que je ne suis pas seulement un intellect, même si je l'ai longtemps cru. Et parce que ce qui m'anime aujourd'hui emprunte non seulement au domaine de l'intellect, mais aussi au domaine du sensible, et au domaine de l'affect.

Je souhaite aussi adopter dans ce document le format du **fragment**, qui est proche aussi de celui du **propos**. Parce que ma pensée est fragmentaire, parce qu'elle se donne à moi sous forme de petits morceaux, et qu'il serait illusoire de vouloir lui donner une autre forme. Je ne crois pas aux grandes formes de pensée. En cela, je me rapproche je crois de la façon d'écrire de Roland Barthes², ou encore de celle du philosophe Alain³. J'espère que la juxtaposition de ces fragments racontera une histoire qui aura une unité, qui aura un fil rouge, un fil conducteur. Mais je ne peux que l'espérer, pas le planifier. Si j'essaie de trouver ce fil rouge, de l'explicitier, de le plaquer, comme on me demande instamment de le faire, je ne ferai que réduire, appauvrir la richesse de mes expériences.

¹ RICŒUR, P. (2021). *Réflexion faite*. Points.

² BARTHES, R. (2002). *Oeuvres complètes, tome 1 à 5*. Seuil.

³ ALAIN. (1985). *Propos sur le bonheur*. Folio.

Enfin, si j'ai intitulé ce prologue ouverture, c'est parce que la **musique** a toujours joué un grand rôle dans ma vie, à la fois ma vie de créateur, mais aussi ma vie de chercheur. C'est donc une forme d'hommage que je rends à la musique en ouvrant ce document par cette ouverture.

Ce mémoire de synthèse va faire l'objet d'une publication sous forme de livre en 2024 sous le titre « Dialogues entre les arts et les sciences : vers une phénoménologie de la création ». Le reste du document n'est donc pas reproduit ici.

BIBLIOGRAPHIE

- ADEN, J., CLARK, S., & POTAPUSHKINA-DELFOSSÉ, M. (2019). Éveiller le corps sensible pour entrer dans l'oralité des langues : une approche éactive de l'enseignement de l'oral. *Lidil. Revue de linguistique et de didactique des langues*, 59.
- ADLER, S. (1988). *The Technique of Acting*. Bantam.
- ADLER, S. (2000). *The Art of Acting*. Applause.
- ALAIN. (1985). *Propos sur le bonheur*. Folio.
- BARTHES, R. (2002). *Oeuvres complètes, tome 1 à 5*. Seuil.
- BITBOL, M. (2014). *La conscience a-t-elle une origine? Des neurosciences à la pleine conscience: Une nouvelle approche de l'esprit*. Flammarion.
- BITBOL, M., & PETITMENGIN, C. (2017). Neurophenomenology and the Microphenomenological Interview. In *The Blackwell Companion to Consciousness* (pp. 726–739). Wiley-Blackwell.
- BOLESZAVSKY, R. (1933). *Acting: The First Six Lessons*. Routledge.
- BRECHT, B. (2000). *Écrits sur le théâtre*. NRF.
- BROOK, P. (2012). *The Empty Space*. Penguin Classics.
- CANDY, L., & EDMONDS, E. (2018). Practice-Based Research in the Creative Arts. *Leonardo*, 51(1), 63–69.
- CHAPMAN, O., & SAWCHUK, K. (2015). Creation-as-Research: Critical Making in Complex Environments. *RACAR : Revue d'art Canadienne / Canadian Art Review*, 40(1), 49–52. Érudit.
- CHEKHOV, M. (1997). *Être acteur*. Pygmalion.
- CHEKHOV, M. (1997). *L'imagination créatrice de l'acteur*. Pygmalion.
- CLURMAN, H. (1983). *The Fervent Years*. Da Capo Press.
- CLURMAN, H. (1997). *On Directing*. S&S.
- CLURMAN, H. (2000). *The Collected Works of Harold Clurman* (M. Loggia & G. Young, Eds.). Applause.
- COHEN, L. (2010). *The Lee Strasberg Notes*. Routledge.
- COHEN, L. (2017). *The Method Acting Exercises Handbook*. Routledge.
- CRAIG, G. (1998). *De l'Art du Théâtre*. Circé.

- CROUSE, M. S., NOWAK, R. D., & BARANIUK, R. G. (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46(4), 886–902.
- DALLET, S., & VEITL, A. (2001). *Du sonore au musical: Cinquante années de recherches concrètes*. Editions L'Harmattan.
- DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics.
- DELION, P. (2011). *Accueillir et soigner la souffrance psychique de la personne - Introduction à la psychothérapie institutionnelle*. Dunod.
- DELION, P. (2018). *Qu'est-ce que la psychothérapie institutionnelle ?* Éditions d'une.
- DEPRAZ, N., VARELA, F. J., & VERMERSCH, P. (2011). *À l'épreuve de l'expérience*. Zeta Books.
- DUZERT, M. (2016). *Vivre ensemble son autonomie*. Editions L'Harmattan.
- EASTY, E. D. (1981). *On Method Acting*. Ivy Books.
- ELKINS, J. (2009). The Three Configurations of Studio-Art PhDs. In S. R. Riley & L. Hunter (Eds.), *Mapping Landscapes for Performance as Research: Scholarly Acts and Creative Cartographies* (pp. 107–113). Palgrave Macmillan UK.
- ELKINS, J. (Ed.). (2014). *Artists with PhDs: On the New Doctoral Degree in Studio Art*. New Academia Publishing.
- FÉRAL, J. (2002). *Mise en scène et jeu de l'acteur - Entretiens: Tome 1, L'espace du texte*. Lansman.
- FÉRAL, J. (2002). *Mise en scène et jeu de l'acteur - Entretiens: Tome 2, Le corps en scène*. Lansman.
- FLANDRIN, P. (1998). *Temps-Fréquence—2ème édition*. Hermes Science Publications.
- FOREMAN, E. (1998). *Transformative voice*. Pro Musica Press.
- GARAUDY, R. (1980). *Parole d'homme*. Seuil.
- GIACCO, G., DIDIER, J., CHATELAIN, S., & VERRY, F. (Eds.). (2020). Définir l'identité de la recherche-crédation: État des lieux et au-delà. EME éditions.
- GOSSELIN, P., & LE COGUEC, É. (Eds.). (2006). *La recherche création*. Presses de l'université du Québec.
- HEIDEGGER, M. (1992). *Être et Temps*. Gallimard.

- HUESCA, R. (2012). *Danse, art et modernité: Au mépris des usages*. PUF.
- HUGO, V. (2002). *La Légende des siècles*. Gallimard.
- HUSSERL, E. (1985). *Idées directrices pour une phénoménologie*. Gallimard.
- HUSSERL, E. (1992). *L'idée de la phénoménologie*. PUF.
- HUSSERL, E. (2000). *Méditations Cartésiennes*. Librairie Philosophique Vrin.
- KLEIN, J.-P., & MALDINEY, H. (2015). *Penser l'art-thérapie*. Presses Universitaires de France.
- LEBLANC, J., PHILIPPON, A., MAGRIN-CHAGNOLLEAU, I., LACHAUD, M.-H., & BOUTRAIS, M. (2021). Pratiquer l'entretien d'explicitation à distance : Petit ABC. *Explicititer*, 130.
- LEVINAS, E. (1992). *Le temps et l'autre*. Presses Universitaires de France - PUF.
- LEWIS, R. (1958). *Method—Or Madness?* Samuel French.
- MACLEOD, K., & HOLDRIDGE, L. (Eds.). (2006). *Thinking through art: Reflections on art as research*. Routledge.
- MAGRIN-CHAGNOLLEAU, I., & BARANIUK, R. G. (1999). Empirical mode decomposition based frequency attributes. *Proceedings of the 69th SEG Meeting, 1949–1952*.
- MAGRIN-CHAGNOLLEAU, I., CHOI, H., VAN SPAENDONCK, R., STEEGHS, P., & BARANIUK, R. G. (1999). Multiscale texture segmentation of dip-cube slices using wavelet-domain hidden Markov trees. *Proceedings of the SEG Meeting*. SEG.
- MAGRIN-CHAGNOLLEAU, I., BARKAT, M., & MEUNIER, F. (2002). Intelligibility of reverse speech in French: A perceptual study. *Proceedings of ICSLP '02*. ICSLP.
- MAGRIN-CHAGNOLLEAU, I. (2015). Du chaos à la création: Une expérience de pédagogie énaactive pour expérimenter le processus créatif en performance. *p-e-r-f-o-r-m-a-n-c-e.org*, 2.
- MAGRIN-CHAGNOLLEAU, I. (2017). L'énaaction dans la création artistique : théâtre, cinéma et performance. In *Action, Énaaction* (pp. 213–224). L'Harmattan, Collection « Ouverture Philosophique ».
- MAGRIN-CHAGNOLLEAU, I., NAHOUM, R., & WEISEN, M. (2019). Le projet Sipapu: Vers une micro-phénoménologie de la création. In X. Lambert (Ed.), *Les Enjeux Cognitifs de l'Artefact Esthétique—Tome 2* (pp. 187–194). L'Harmattan.
- MAGRIN-CHAGNOLLEAU, I. (2021). Actors Studio. In *Encyclopædia Universalis*.
- MAGRIN-CHAGNOLLEAU, I. (2021). Hasard et Création. In *Le hasard, la calcul et la vie*. ISTE Editions.

- MAGRIN-CHAGNOLLEAU, I. (2021). Chance and the Creative Process. In *Chance, Calculation and Life* (pp. 169–183). ISTE Editions.
- MAGRIN-CHAGNOLLEAU, I. (2021). Futurs utopiques et dystopiques: Comment la fiction aide à penser l'évolution. *Arts et Sciences*, 5(1).
- MAGRIN-CHAGNOLLEAU, I. (2022). A Phenomenological Approach to Color Theory and Aesthetics. *Proceedings of AIC 2022*.
- MAGRIN-CHAGNOLLEAU, I. (2022). Reflections on the impact of creation research, practice-based research, and practice as research, and how to define and maximize that impact. *Media Practice and Education*, 23(2), 105–111.
- MAGRIN-CHAGNOLLEAU, I. (2022). The Relationship between Photo Retouching and Color Grading. *Color Culture and Science Journal*, 14(1), pp. 79-86.
- MAGRIN-CHAGNOLLEAU, I. La Méthode de l'Actors Studio : Genèse, Pratique et Phénoménologie. In *Être en scène, être en jeu. Statuts, esthétiques, processus et techniques pour l'interprète et le performeur*. Sous presse.
- MAGRIN-CHAGNOLLEAU, I. Expérience phénoménologique et énaïve d'une oeuvre de poésie spatialisée visuelle et sonore. *Corporéité et construction du sens dans la poésie visuelle : Énaction, perception, expérimentation*. Sous presse.
- MAGRIN-CHAGNOLLEAU, I. *La Méthode de l'Actors Studio*. Deuxième Époque. Sous presse.
- MAGRIN-CHAGNOLLEAU, I. Création énaïve et pédagogie énaïve : expérimentations, bilan, perspective, prospective. *L'énaïve en perspective et en prospective*. Sous presse.
- MAGRIN-CHAGNOLLEAU, I. The Enactive Nature of the Relationship between Criticism and Aesthetics: From Practice to Theory to Pedagogy. *Teaching Philosophy*. Submitted.
- MAGRIN-CHAGNOLLEAU, I. Artistic Workshops with patients of a Day Psychiatric Hospital in France. *Creative Arts in Education and Therapy*. Submitted.
- MARTINEZ THOMAS, M., & NAUGRETTE, C. (2020). *Le doctorat et la recherche en création*. L'Harmattan.
- MEISNER, S. (1996). *On Acting*. Vintage.
- MERLEAU-PONTY, M. (1976). *Phénoménologie de la perception*. Gallimard.
- MEYERHOLD, V. (2001). *Écrits sur le théâtre: Tomes 1 à 4. L'âge d'homme*.
- NACHMANOVITCH, S. (1991). *Free Play*. Tarcher / Putnam.
- NOVAK, A. (2019). Affective spaces. Experiencing atmosphere in the visual arts. *Archimaera*, 2019(8).

- OULIPO. (1998). *La littérature potentielle*. Paris: Folio Essais.
- OURY, J. (2003). *Psychiatrie et psychothérapie institutionnelle*. Champ Social.
- OURY, J. (2016). *La psychothérapie institutionnelle de Saint-Alban à La Borde*. Editions d'une.
- PETITMENGIN, C. (2003). *L'expérience intuitive*. L'Harmattan.
- PETITMENGIN, C. (2006). Describing one's subjective experience in the second person: An interview method for the science of consciousness. *Phenomenology and the Cognitive Sciences*, 5(3–4), 229–269.
- PIAGET, J. (1972). *Epistémologie des sciences de l'homme*. Gallimard.
- POTROVIC, L. (2017). *What a Body can Become: Cartography between Dance and Philosophy*. Université Sorbonne Paris Cité.
- REY, A. (2013). *Le Grand Robert de la Langue Française*, version électronique.
- RICŒUR, P. (2021). *Réflexion faite*. Points.
- SCHECHNER, R. (1988). *Performance theory*. Routledge.
- SCHECHNER, R., & BRADY, S. (2013). *Performance studies: An introduction* (3rd ed.). Routledge.
- SERRES, M. (1968). *Hermès: Tome 1, La communication*. Editions de Minuit.
- SMITH, H., & DEAN, R. T. (Eds.). (2009). *Practice-led research, research-led practice in the creative arts*. Edinburgh University Press.
- STANISLAVSKI, C. (1937). *An Actor Prepares*. Methuen Drama.
- STANISLAVSKI, C. (1950). *Building A Character*. Methuen Drama.
- STANISLAVSKI, C. (1988). *Creating A Role*. Methuen.
- STANISLAVSKI, C. (1996). *My Life in Art*. Routledge.
- STÉVANCE, S., & LACASSE, S. (2013). *Les enjeux de la recherche-crédation en musique. Institution, définition, formation*. Presses de l'Université Laval.
- STEWART, J., GAPENNE, O., & DI PAOLO, E. A. (2010). *Enaction: Toward a New Paradigm for Cognitive Science*. MIT Press.
- STRASBERG, L., & HETHMON, R. H. (1965). *Strasberg at The Actors Studio*. Theater Communications Group.
- STRASBERG, L. (1987). *A Dream of Passion*. Plume.

- SULLIVAN, G. (2009). *Art Practice as Research*. Sage Publishing.
- SUMARTOJO, S., & PINK, S. (2017). *Atmospheres and the Experiential World: Theory and Methods*. Routledge.
- TOSQUELLES, F. (2007). *Éducation et psychothérapie institutionnelle*. Matrice.
- TROCMÉ-FABRE, H. (1987). *J'apprends, donc je suis*. Editions d'Organisation.
- TROCMÉ-FABRE, H. (1999). *Réinventer le métier d'apprendre*. Editions d'Organisation.
- VAKHTANGOV, E. (2000). *Écrits sur le théâtre*. L'Âge d'Homme.
- VALÉRY, P. (2002). *Variété III, IV et V*. Folio Essais.
- VARELA, F. (1989). *Autonomie et Connaissance: Essai sur le Vivant*. Seuil.
- VARELA, F., THOMSON, E., & ROSCH, E. (1991). *The Embodied Mind*. The MIT Press.
- VARELA, F., & MATURANA, H. (1994). *L'Arbre de la connaissance*. Addison Wesley France.
- VARELA, F. (1996). *Invitation aux sciences cognitives*. Seuil.
- VARELA, F., & SHEAR, J. (1999). *The View From Within: First-Person Approaches to the Study of Consciousness*. Imprint Academic.
- VARELA, F. (2004). *Quel savoir pour l'éthique ?* La Découverte.
- VARELA, F. (2017). *Le cercle créateur—Écrits (1976-2001)*. Seuil.
- VERMERSCH, P. (2007). Bases de l'auto-explicitation (1). *Expliciter*, 69.
- VERMERSCH, P. (2008). Introspection et auto-explicitation: Bases de l'auto-explicitation (2). *Expliciter*, 73.
- VERMERSCH, P. (2012). *Explicitation et phénoménologie*. PUF.
- VERMERSCH, P. (2014). *L'entretien d'explicitation*. ESF Éditeur.
- VIGNY, A. de. (1986). *Œuvres complètes, tome I*. Gallimard.

**LA MÉTHODE
DE L'ACTORS STUDIO**

Ivan MAGRIN-CHAGNOLLEAU

Habilitation à Diriger des Recherches

Tome 3 : Inédit

*À Blandine,
qui m'a encouragé depuis le début
à écrire ce livre, puis à le terminer.*

*À Élisabeth,
qui a permis à mon envie de théâtre
de s'éveiller en moi.*

Table des matières

| | |
|--|-----------|
| Introduction..... | 5 |
| Remerciements | 7 |
| Pourquoi une méthode de jeu ? | 9 |
| D'où vient la méthode de l'Actors Studio ? | 10 |
| Pourquoi est-il nécessaire d'apprendre cette méthode avec un professeur ? | 11 |
| Relaxation | 13 |
| Position pour se relaxer | 14 |
| Observer sa respiration | 15 |
| La lumière qui se déplace | 16 |
| Se relaxer avec des sons | 17 |
| Parler ses pensées : le monologue intérieur | 18 |
| Song and Dance | 20 |
| 1ère partie | 20 |
| 2ème partie | 20 |
| Mémoire sensorielle..... | 22 |
| Stanislavski et les émotions..... | 22 |
| Lee Strasberg et les émotions | 22 |
| Quelques exemples d'exercices de mémoire sensorielle | 25 |
| Boisson du matin | 25 |
| Nourriture | 28 |
| Repas..... | 28 |
| Douche ou bain..... | 29 |
| Miroir | 30 |
| Maquillage ou rasage..... | 31 |
| Sensation de soleil sur la peau | 31 |
| Exploration des déclencheurs | 32 |
| Une séquence d'exploration | 33 |
| Tous les sens ne sont pas égaux | 33 |
| Pratique de la mémoire sensorielle en cours | 34 |
| Mémoire affective..... | 35 |
| Moment Privé..... | 38 |
| L'exercice de l'incorporation animale | 40 |
| Présentation de l'exercice..... | 40 |
| Observation..... | 40 |
| Incorporation..... | 41 |
| Transformation en personnage | 43 |

| | |
|---|-----------|
| Jeu d'acteur et incorporation animale | 44 |
| Comment travailler un rôle | 48 |
| Les différentes étapes du travail sur un rôle | 48 |
| Les premières lectures | 48 |
| Les questions à poser | 50 |
| Un exemple de progression pédagogique | 58 |
| Cours No. 1 | 58 |
| Cours No. 2 | 59 |
| Cours No. 3 | 59 |
| Cours No. 4 | 59 |
| Cours No. 5 | 60 |
| Cours No. 6 | 60 |
| Cours No. 7 | 60 |
| Cours No. 8 | 61 |
| Cours No. 9 | 62 |
| Cours No. 10 | 62 |
| Cours No. 11 | 63 |
| Cours No. 12 | 63 |
| Cours No. 13 | 65 |
| Cours No. 14 | 65 |
| Cours No. 15 | 66 |
| Cours No. 16 | 67 |
| Cours No. 17 | 68 |
| Cours No. 18 | 68 |
| Cours No. 19 | 69 |
| Cours No. 20 | 69 |
| Cours No. 21 | 70 |
| Cours No. 22 | 70 |
| Cours No. 23 | 71 |
| Cours No. 24 | 71 |
| Cours No. 25 | 72 |
| Cours No. 26 | 72 |
| En conclusion | 74 |
| Bibliographie | 76 |
| Livres en français | 76 |
| Livres en anglais | 76 |

Introduction

Pourquoi écrire un livre sur la méthode de l'Actors Studio aujourd'hui en français ? D'abord parce que c'est une méthode de jeu qui a toujours été mal comprise, surtout en France, et surtout par celles et ceux qui ne l'ont jamais pratiquée, ou qui ne l'ont approchée que de très loin, par un ou deux stages. Il y a de nombreux préjugés sur cette méthode et j'espère, en écrivant ce livre, en éliminer un certain nombre.

Ensuite parce qu'en dépit de l'efficacité de cette méthode de jeu, comme en témoigne une hégémonie très nette des actrices et acteurs qui la pratiquent, notamment aux États-Unis, elle est encore peu pratiquée en France et peu utilisée comme technique de direction d'acteurs par les metteurs en scène français.

Enfin, parce qu'il n'existe aucun livre, à ma connaissance, qui à la fois présente la méthode dans son contexte, décrit les exercices en détails, et surtout explique comment les utiliser concrètement dans un travail de personnage.

Un petit mot sur mes compétences pour écrire ce livre. D'abord, j'ai été formé comme acteur et metteur en scène pendant trois années intensives à l'école de l'Actors Studio à New York, et tous les professeurs que j'ai eus ont tous été formés par Lee Strasberg lui-même, le créateur de cette méthode. J'ai également, pendant ces trois années et après, assisté à de nombreuses séances de travail par des acteurs professionnels au sein du laboratoire d'acteurs de l'Actors Studio.

Ensuite, j'ai continué à jouer tous mes personnages après l'école en

utilisant cette méthode, et j'ai interprété plus d'une vingtaine de rôles principaux au théâtre et au cinéma.

J'ai aussi dirigé au théâtre et au cinéma et également coaché de nombreuses actrices et de nombreux acteurs en utilisant cette méthode. J'ai enseigné cette méthode, sous forme de cours, de stages, de masterclasses.

J'ai également accès depuis de nombreuses années aux archives de Lee Strasberg, qui sont conservées au Lee Strasberg Institute à New York. Cet accès m'a été très gentiment permis par Anna Strasberg et Victoria Krane, grâce notamment à l'aide de Lola Cohen. Ces archives donnent un accès inestimable à des séances de cours que Lee Strasberg a données peu de temps avant sa mort, et qui représentent l'aboutissement d'un travail de toute une vie.

Enfin, étant également chercheur au CNRS, je suis notamment spécialisé sur les techniques de jeu, et encore plus spécifiquement sur cette technique de jeu.

Tous ces éléments font que j'ai une connaissance à la fois organique et viscérale de cette méthode, mais aussi une connaissance théorique très poussée.

Il m'a semblé très naturel de consigner finalement toute cette connaissance acquise par l'étude et par la pratique dans un livre. Et c'est ce livre que vous tenez dans les mains.

J'espère que vous prendrez autant de plaisir à le lire que j'ai pris de plaisir à l'écrire. Je vous souhaite une très bonne lecture.

Remerciements

J'ai tenu à écrire une page spéciale de remerciements car il y a bon nombre de personnes sans qui, d'une part, ma carrière d'acteur n'aurait pas été possible, et d'autre part, ce livre n'aurait pas pu voir le jour.

La première personne que je souhaite remercier est Elisabeth Saint-Blancat. Elle a été la directrice artistique du théâtre des Clochards Célestes à Lyon, théâtre de création qui a aidé quasiment toutes les jeunes compagnies de la région lyonnaise à démarrer. Elisabeth a été ma première professeur de théâtre et ma première fan. Je lui dois la découverte de la méthode de l'Actors Studio (elle a elle-même été l'élève à Paris d'Andreas Voutsinas, ancien assistant de Lee Strasberg). Je lui dois aussi de m'avoir encouragé à poursuivre mon rêve de devenir comédien quand d'autres professeurs d'autres écoles que je ne nommerai pas me disaient que j'étais bien trop vieux pour ça. Et de m'avoir aidé à préparer l'audition pour entrer à l'Actors Studio.

Je souhaiterais également remercier ma famille et mes amis, dont la plupart m'ont soutenu dans ce rêve qui n'allait pas de soi. Et qui ont accepté de me voir partir trois ans à New York, même si cela signifiait qu'ils allaient me voir moins souvent.

Je voudrais remercier le CNRS qui m'a permis de pouvoir partir faire l'Actors Studio pendant trois ans à New York, tout en ayant la possibilité de retrouver un poste de chercheur au retour, ce qui a fortement facilité cette prise de risque. Et encore aujourd'hui, m'étant placé sur le créneau de la création recherche, je bénéficie de cette double articulation entre création et recherche, et c'est mon poste de chercheur au CNRS qui rend cela possible.

Je souhaiterais remercier vivement mes professeurs à l'Actors Studio. J'ai eu la chance d'avoir chaque année des professeurs exceptionnels. Marcia Haufrecht en première année, qui a su non seulement nous enseigner les exercices de base, mais aussi la façon de les utiliser dans un travail de personnage. Marcia est devenue une très bonne amie et je lui suis très reconnaissant de tout ce qu'elle m'a transmis. Arthur Storch en deuxième année, décédé depuis, et qui a su nous pousser à réfléchir par nous-même sur la méthode, et aussi à prendre de la distance par rapport à tous les exercices, pour arriver à construire de manière autonome notre propre méthode de travail. Ron Leibman en troisième année, également décédé depuis, qui a su affiner notre travail et lever les derniers obstacles qui se trouvaient sur notre route d'acteurs.

Mes camarades de classe à l'Actors Studio avec lesquels j'ai partagé ces trois magnifiques années et qui ont été tour à tour mes partenaires de jeu, mes comédiens, mes metteurs en scène, mes amis, mes aiguillons. J'ai une pensée toute particulière pour certaines et certains d'entre eux avec lesquels j'ai noué des liens plus forts : Florin, Michelle, Andy, Emily, Zoe, JC, Tawny, ...

Je voudrais enfin remercier Lola Cohen, qui enseigne au Lee Strasberg Institute, et qui m'a ouvert les portes de ses classes et les portes de l'institut. C'est grâce à elle que j'ai pu rencontrer Victoria Krane, que je remercie également vivement, car elle m'a donné accès aux archives de Lee Strasberg. Lola partage comme moi la passion de la méthode, de son enseignement, de sa transmission, et de la recherche toujours plus fine à travers les archives et l'expérimentation sur le terrain.

Cet inédit est en cours de publication sous forme de livre et paraîtra en 2024 sous le titre « La méthode de l'Actors Studio ». Le reste du document n'est donc pas reproduit ici.

Bibliographie

Livres en français

- Chekhov, M. (1997). *Être acteur*. Pygmalion.
- Chekhov, M. (1997). *L'imagination créatrice de l'acteur*. Pygmalion.
- Magrin-Chagnolleau, I. (2021). *Actors Studio*. Encyclopædia Universalis.
- Meyerhold, V. (2001). *Écrits sur le théâtre: Tomes 1 à 4*. L'âge d'homme.
- Stanislavski, C. (2015). *La formation de l'acteur*. Payot.
- Stanislavski, C. (2021). *La Construction du personnage*. Pygmalion.
- Strasberg, L. (1986). *Le travail à l'Actors Studio*. Gallimard, NRF.
- Vakhtangov, E. (2000). *Écrits sur le théâtre*. L'Âge d'Homme.

Livres en anglais

- Adler, S. (1988). *The Technique of Acting*: Bantam.
- Adler, S. (1999). on Ibsen Strindberg and Chekhov: Alfred A. Knopf.
- Adler, S. (2000). *The Art of Acting*: Applause.
- Adler, S. (2013). on America's Master Playwrights: Vintage.
- Benedetti, J. (Ed.). (1991). *The Moscow Art Theatre Letters*. Routledge.
- Benedetti, J. (1998). *Stanislavski and the Actor*. Routledge.
- Benedetti, J. (2016). *Stanislavski: An Introduction*. Methuen Drama.
- Boleslavsky, R. (1933). *Acting: The First Six Lessons*: Routledge.
- Chekhov, M. (2016). *To the Actor: On the Technique of Acting*. Martino Publishing.
- Clurman, H. (1983). *The Fervent Years*: Da Capo Press.
- Clurman, H. (1997). *On Directing*: S&S.

Clurman, H. (2000). *The Collected Works of Harold Clurman* (M. Loggia & G. Young, Eds.). Applause.

Cohen, L. (2010). *The Lee Strasberg Notes*. New York: Routledge.

Cohen, L. (2017). *The Method Acting Exercises Handbook*: Routledge.

Cole, T., & Strasberg, L. (2014). *Acting: A Handbook of the Stanislavski Method* (Illustrated edition). Martino Fine Books.

Easty, E. D. (1981). *On Method Acting*: Ivy Books.

Gordon, Mel. *The Stanislavsky Technique: Russia*

Gordon, Mel. *Stanislavsky in America: An Actor's Workbook*

Hagen, U., & Frankel, H. (1973). *Respect for Acting*. Macmillan.

Hagen, U. (1991). *A Challenge For The Actor*. Charles Scribner's Sons.

Lewis, R. (1958). *Method - or Madness?* : Samuel French.

Meisner, S. (1996). *On Acting*: Vintage.

Moore, S. (1984). *The Stanislavski System: The Professional Training of an Actor*. Penguin Books.

Stanislavski, C. (1937). *An Actor Prepares*: Methuen Drama.

Stanislavski, C. (1950). *Building A Character*: Methuen Drama.

Stanislavski, C. (1988). *Creating A Role*. Methuen.

Stanislavski, C. (1996). *My Life in Art*. Routledge.

Strasberg at The Actors Studio. (1965). (R. H. Hethmon Ed.): Theater Communications Group.

Strasberg, L. (1987). *A Dream of Passion*: Plume.