



HAL
open science

Estimation et modélisation paramétrique de l'onde glottique en vue de la caractérisation objective de la voix pathologique

Mohammed Faouzi Smiej

► **To cite this version:**

Mohammed Faouzi Smiej. Estimation et modélisation paramétrique de l'onde glottique en vue de la caractérisation objective de la voix pathologique. Sciences de l'ingénieur [physics]. Université Henri Poincaré - Nancy 1, 1991. Français. NNT : 1991NAN10020 . tel-01747462

HAL Id: tel-01747462

<https://hal.univ-lorraine.fr/tel-01747462>

Submitted on 29 Mar 2018

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.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Centre de Recherche en Automatique de Nancy (URA 821)

Laboratoire d'Electricité et d'Automatique

THESE

Présentée et soutenue publiquement le 22 Janvier 1991

à l'Université de Nancy I

pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE DE NANCY I

en Automatique

par

Mohammed Faouzi SMIEJ

**ESTIMATION ET MODELISATION PARAMETRIQUE DE L'ONDE
GLOTTIQUE EN VUE DE LA CARACTERISATION
OBJECTIVE DE LA VOIX PATHOLOGIQUE**

Composition du Jury :

Président :	M. AUBRUN
Rapporteurs :	B. DUBUISSON C. HUMBERT
Examineurs :	M. LAMOTTE J. BREMONT J. RAGOT

AVANT-PROPOS

Cette étude a été menée au Laboratoire d'Automatique et d'Electricité (LEA), qui est une équipe du Centre de Recherche en Automatique de Nancy (CRAN-CNRS UA 821-UNIVERSITE DE NANCY I), sous la direction de Monsieur M. LAMOTTE. Elle entre dans un cadre d'une convention passée entre le Ministère de la Recherche et de la Technologie et le CRAN-LEA, dont l'objectif est la conception d'un système de caractérisation objective de la voix humaine.

Je tiens à exprimer ma gratitude à Monsieur le Professeur AUBRUN, Professeur à l'Université de Nancy I et Directeur du Laboratoire d'Automatique et de Recherche Appliquée, de l'honneur qu'il me fait en acceptant de présider mon jury.

Que Monsieur Le professeur M. LAMOTTE accepte ici l'expression de ma considération, ainsi que mes remerciements pour l'intérêt et le concours qu'il a portés à ce travail.

Je tiens à exprimer ici ma reconnaissance à Monsieur le Professeur J. BREMONT qui m'a accueilli dans son Laboratoire et qui a su bien m'intégrer au sein de son équipe de recherche.

Que Monsieur le Professeur B. DUBUISSON, Professeur à l'Université Technologique de Compiègne, et Directeur de l'HEUDIASYC, trouve ici mes remerciements de m'avoir consacré une partie de son précieux temps pour examiner ce travail et de m'avoir apporté ses conseils les plus pertinents.

Je remercie également Monsieur le Professeur C. HUMBERT, Professeur à l'Université de Nancy I et Directeur du Laboratoire d'Automatique (LARA-ESSTIN), de m'avoir accordé un intérêt particulier lors de la discussion sur les points épineux de ce travail par ses remarques constructives.

J'exprime mes remerciements à Monsieur le Professeur J. RAGOT, Professeur à l'Institut Polytechnique de Lorraine, qui me fait l'honneur et l'amitié de participer à mon jury.

J'adresse également tous mes remerciements et mes amitiés à Mesdames M.J VIGNERON, Maître de Conférence à l'Université de Nancy I, S. BENZEDIA-MIRZAQ, J. BOURDACHE et J. SCHWARTZ pour leur soutien moral qu'elles m'ont accordé. Je remercie tout particulièrement tous mes amis : A. BOURDACHE, M.S. RIFFI TEMSAMANI, B. HEIT, E. LEVRAT, M. KARMOUCHE , P. DE LEPPINE, M. ELJADAoui, T. SIMON et P.Y. KALIKY, pour leur fantastique enthousiasme et concours lors de la réalisation de ce travail.

Enfin, c'est avec une attention particulière que je remercie mes parents, mes amis et tout le reste de la famille pour leur présence et leurs encouragements.

SOMMAIRE

AVANT-PROPOS

INTRODUCTION 3

PARTIE I

1- Bases anatomo-physiologiques de la production de la parole	7
1-1 Eléments constitutifs de l'appareil phonatoire	7
1-1-1 La soufflerie pulmonaire	7
1-1-1-1 Les poumons	7
1-1-1-2 La cavité thoracique	8
1-1-1-3 Le système canalaire (la trachée)	9
1-1-2 Le vibreur laryngé	9
1-1-2-1 Les cartilages	9
1-1-2-2 Le système ligamentaire et musculaire	10
1-1-2-3 L'innervation	13
1-1-3 Le résonateur pharyngo-naso-buccal	13
1-1-3-1 Le pharynx	13
1-1-3-2 Le palais et le voile	14
1-1-3-3 La cavité orale	14
1-1-3-4 Les cavités naso-sinusiennes	16
1-1-4 La commande nerveuse	16
1-2 Physiologie du système de production de la parole	18
1-2-1 Physiologie de l'expiration phonatoire	18
1-2-1-1 Débit et pression sous-glottiques	18
1-2-1-2 Mécanique respiratoire en phonation	18
1-2-2 Physiologie du vibreur laryngé	19

1-2-2-1 Historique	19
1-2-2-2 Les attaques vocaliques	20
1-2-2-3 Les registres	21
1-2-3 Physiologie de la résonance	21
2- Pathologies de la voix humaine	27
2-1 Les dysphonies dysorganiques	27
2-1-2 Les troubles laryngés	27
2-1-2-1 Phénomènes morphologiques	27
2-1-2-2 Les altérations neurogènes	29
2-1-2-3 Perturbations endocriniennes et pharmacologiques	31
2-1-2 Les troubles respiratoires	31
2-1-3 Les troubles des cavités de résonance	31
2-2 Les dysphonies dysfonctionnelles	33
2-2-1 L'habitude vocale défectueuse	33
2-2-2 Les troubles relationnels et émotionnels	33
2-2-3 Le cercle vicieux de l'effort vocal	34
2-2-4 Quelques exemples particuliers de dysphonies dysfonctionnelles	35
2-2-4-1 Les troubles vocaux de la mue	35
2-2-4-2 L'aphonie psychogène	35
2-3 Conclusions	36
2-3-1 Ambivalence nosologique des troubles vocaux	36
2-3-2 Vers un système automatique d'aide au diagnostic	36

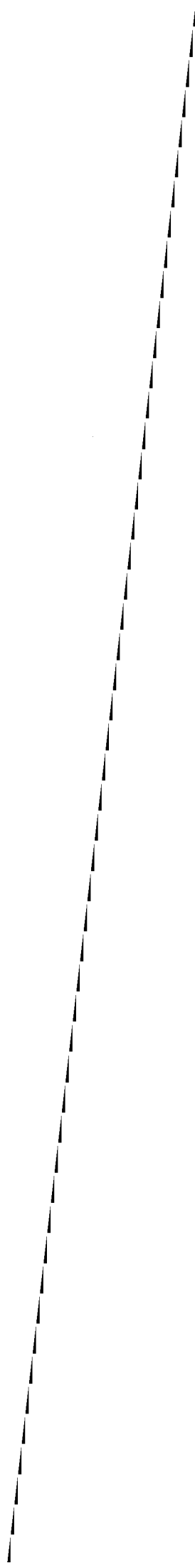
PARTIE II

3- Théorie acoustique de la production de la parole	41
3-1 Eléments de la théorie de la propagation de l'onde sonore dans le système vocal	41
3-1-1 Introduction	41
3-1-2 Hypothèses	42
3-1-3 Equations de propagation	42
3-2 Les modèles physiques de la production sonore basés sur la théorie acoustique	43

4-1-3-3	Ordre du modèle	88
4-1-4	Techniques numériques de modélisation	88
4-1-4-1	Prédictions directe et rétrograde	88
4-1-4-2	Calcul récurrent des coefficients du filtre d'analyse	90
4-1-4-3	Structure du filtre en treillis	90
4-1-4-4	Quelques algorithmes	92
4-1-4-4-1	Algorithme de Levinson-Durbin	92
4-1-4-4-2	Algorithme de Leroux-Gueguen	92
4-1-4-4-3	Algorithme de Burg	93
4-1-4-5	Comparatif des algorithmes	95
4-2	Modélisation autorégressive du signal de parole	96
4-2-1	Introduction	96
4-2-2	Quelques justificatifs de l'utilisation de la modélisation autorégressive du signal de parole	97
4-2-3	Lien entre la prédiction linéaire et le modèle des tubes acoustiques	98
4-2-4	Quelques applications de la prédiction linéaire	101
4-3	Autres modèles	102
4-3-1	Le modèle ARMA	102
4-3-2	Le modèle GAR ou GARMA	102
4-4	Conclusion	103
PARTIE III		107
5-	Description qualitative des paramètres de la voix humaine	107
5-1	Les paramètres phonémiques	107
5-1-1	L'intensité	107
5-1-2	Le voisement	108
5-1-3	Les formants	108
5-1-4	La hauteur	108
5-1-5	La nasalisation	109
5-1-6	La labialité	109
5-2	Les paramètres prosodiques	110
5-2-1	Le contour formantique	110
5-2-2	Le contour mélodique	110
5-2-3	L'intonation	111

6-2-1-2-2 Procédure d'analyse et résultats	147
6-2-1-2-3 Discussion	151
6-2-1-3 Détection des événements ouverture/fermeture par analyse adaptative du signal	152
6-2-1-3-1 Principe de la méthode	153
6-2-1-3-2 Procédure de détection des événements	154
6-2-1-3-3 Exemples d'application de la procédure	155
6-2-2 Méthodes d'extraction des paramètres de l'activité glottique	158
6-2-2-1 Paramètres mesurés	158
6-2-2-2 Méthode basée sur le résultat du filtrage inverse	159
6-2-2-3 Extraction simultanée des paramètres et de l'onde glottique	163
6-2-2-3-1 Principe de la méthode	163
6-2-2-3-2 Mise en oeuvre de la méthode	166
6-3 Programme général de mesure des caractéristiques de la voix humaine	167
CONCLUSION	173
BIBLIOGRAPHIE	175
ANNEXE	185

INTRODUCTION



INTRODUCTION

La communication parlée est le propre de l'Homme. Conscient de sa faculté de produire divers sons, il a su créer un système de représentation des concepts par un ensemble d'associations des différents sons émis : c'est la naissance d'un langage parlé, moyen de communication entre les êtres humains.

Une atteinte fonctionnelle d'un organe de la phonation peut, dans certains cas, être un handicap majeur dans le processus de la communication du sujet et avoir des répercussions sur son équilibre psychique. Notre travail a pour but de concevoir un système d'évaluation objective de la qualité fonctionnelle de la production de la parole. Précisons de prime abord, qu'en aucun cas notre objectif est de remplacer le praticien des pathologies de la voix, mais tout simplement, de lui fournir un outil automatique de mesure, pouvant l'aider dans son diagnostic. Il est clair que dans ce domaine le praticien dispose déjà d'un ensemble d'appareils de mesure, plus ou moins sophistiqués (électroglottographe, radiologie par rayons X, sonographe, cinématographie ultra-rapide, résonance magnétique...); cependant, les mesures fournies par ces appareils ne donnent qu'un aspect particulier de la phonation (un sonogramme par exemple, fournit une représentation temps-fréquence du signal de parole). Nous nous sommes fixés l'objectif de concevoir un système de mesure, pouvant fournir des renseignements sur les différents aspects du système de production. Ce système doit avoir les performances suivantes :

- il ne doit pas perturber le mécanisme de la phonation,
- les mesures doivent être très diverses, de manière à couvrir les différents aspects de la production,
- la convivialité du système et sa facilité d'utilisation sont des qualités appréciables que nous rechercherons. De plus, les traitements doivent s'effectuer dans la mesure du possible "en temps réel". En fait, tout au long de ce travail, nous nous sommes plus intéressés aux traitements permettant de caractériser la voix humaine plutôt qu'à la réalisation en temps réel.

Devant un tel problème, nous nous sommes interrogés sur la démarche à suivre pour atteindre notre objectif. Automaticien de formation, nous imaginions l'étude comme un problème d'identification d'un modèle du système de production de la parole (boîte noire). Une phase de modélisation du système était donc primordiale. Cependant, les méthodes d'identification nécessitent la connaissance à la fois du modèle et de la commande. Une telle attitude ne peut être adoptée, puisque nous ne pouvons pas intervenir sur la commande, qui dans notre cas est la source du système de production. Il a donc fallu changer de stratégie et considérer le problème comme une modélisation du signal de sortie, et identifier par la suite le

système de production (conduit vocal, éventuellement conduit vocal+conduit nasal) et la source (débit glottique).

La première partie de notre rapport, composée de deux chapitres, permet de mettre en valeur d'une part la complexité de l'anatomie du système de production et d'autre part l'origine des pathologies de la voix. Nous pensons utile d'inclure la description de l'anatomie et de la physiologie de la production de la parole, pour bien introduire ensuite les modèles physiques et mathématiques du système. Nous insistons en particulier, dans le deuxième chapitre, sur l'étiologie des pathologies laryngiennes, puisqu'il nous semble intéressant d'étudier l'activité glottique et les répercussions des atteintes du larynx sur la phonation.

La deuxième partie a pour objet principal l'établissement d'un modèle numérique du processus de la production de la parole. La théorie acoustique présentée au troisième chapitre, permet de mettre en place un modèle physique décrit par un ensemble d'équations différentielles. En revanche, le quatrième chapitre établit le passage du modèle physique à un modèle mathématique qui donne le moyen d'extraire diverses informations sur le conduit vocal à partir du signal de parole (fonction d'aire, coefficients de réflexion...). Après un rappel sur la prédiction linéaire et les processus autorégressifs, nous montrons, de même, que le signal de parole peut être modélisé par un tel processus. Nous y mentionnons d'autres modélisations souvent utilisées en traitement de la parole (ARMA, ARMAX).

Le système de mesure est décrit en détail à la troisième partie. Dans un premier chapitre (chapitre 5), nous décrivons l'ensemble des paramètres caractérisant la voix humaine. Nous les classons en deux groupes : paramètres globaux et paramètres locaux. Les premiers sont des mesures quasi-instantanées (effectuées sur des fenêtres de petite taille), les seconds sont des mesures de variation des premiers en fonction du temps (contours formantiques ...). Le dernier chapitre, constitue l'étape finale de notre travail. En premier lieu, nous présentons les moyens techniques d'extraction des signaux glottiques, ainsi qu'une synthèse des différents modèles paramétriques du débit glottique. Nous étudions ensuite, le modèle cinétique de Titze, qui permet de caractériser en détail l'activité glottique. Nous y apportons une modification en considérant séparément le comportement des deux cordes vocales. Il s'avère utile pour notre application, de considérer des paramètres supplémentaires de ce modèle, puisque les atteintes au niveau du larynx peuvent être unilatérales (deuxième chapitre). A la fin du chapitre, nous décrivons les méthodes de filtrage inverse ainsi que celles d'extraction des paramètres.

Une conclusion générale comportant notre appréciation sur notre travail et les perspectives dans ce domaine, achève le présent rapport.

PARTIE I

1- BASES ANATOMO-PHYSIOLOGIQUES DE LA PRODUCTION DE LA PAROLE

BASES ANATOMO-PHYSIOLOGIQUES DE LA PRODUCTION DE LA PAROLE

Dans ce chapitre, nous décrivons les principaux organes mis en jeu pendant la phonation, en insistant plus particulièrement sur l'aspect physiologique de la genèse de la phonation.

1-1 Eléments constitutifs de l'appareil phonatoire

Schématiquement, l'appareil phonatoire peut être subdivisé en trois parties (Fig 1-1) :

- la soufflerie pulmonaire et son appareil canalaire,
- le vibreur laryngé,
- le pavillon pharyngo-bucco-nasal, faisant office de résonateur.

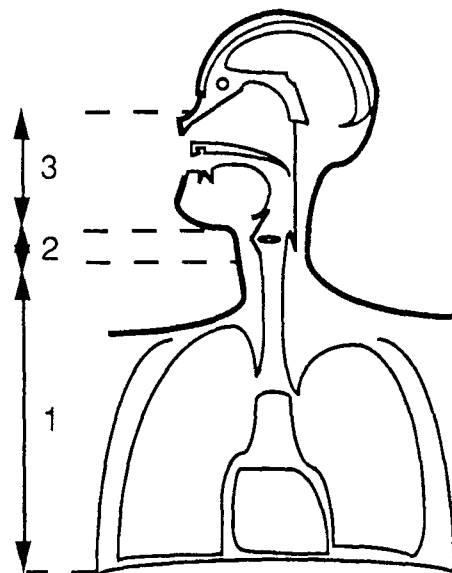


Figure 1-1. Schéma de l'appareil phonatoire :

1. Soufflerie pulmonaire et appareil canalaire
2. Vibreur laryngé
3. Pavillon pharyngo-bucco-nasal

Ce partage est désormais devenu classique, mais il est intéressant d'examiner l'unité et la synergie anatomo-fonctionnelle de ces trois éléments, ainsi que leur coordination commandée par le système nerveux.

1-1-1 La soufflerie pulmonaire

1-1-1-1 Les poumons

La fonction première des poumons est la respiration (l'hématose : oxygénation sanguine) qui est une fonction involontaire. Pour la phonation, fonction volontaire commandée par le système nerveux, les poumons jouent le rôle d'un réservoir d'air qui y est mis sous pression. Ce réservoir élastique a des parois extensibles dont la motilité est assurée par un système musculaire extrinsèque. L'élasticité des poumons est due à la structure anatomique de cet organe. En effet, chaque poumon est divisé en lobes, se partageant eux-mêmes en lobules; chaque lobule est délimité par des cloisons fibreuses doublées d'un réseau élastique. De plus, la bronche souche, qui pénètre dans chacun des poumons au niveau du hile, se ramifie en bronches extralobulaires de calibre progressivement décroissant aboutissant finalement à un lobule. Ainsi dans chaque lobule pénètre une bronchiole intralobulaire se divisant elle-même en bronchioles alvéolaires.

Dans les conditions normales, les mouvements des poumons sont étroitement liés aux modifications du volume de la cage thoracique, car chacun d'eux est solidaire de celle-ci par l'intermédiaire d'une membrane séreuse. Cette membrane dite plèvre viscérale, se présente sous la forme d'un sac sans ouverture favorisant le glissement du poumon sur la paroi de la cage thoracique.

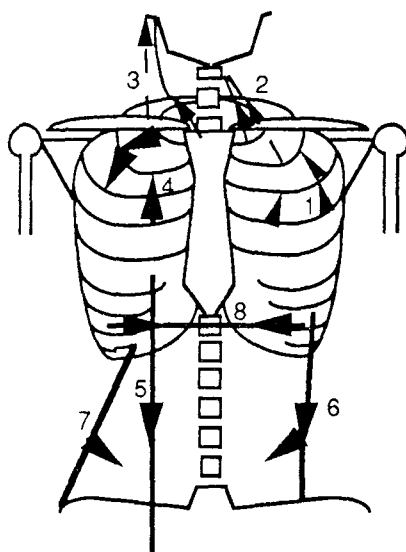
1-1-1-2 La cavité thoracique

C'est un ensemble de douze paires de côtes, maintenues en place par un tissu conjonctif et élastique et actionnées par des muscles ayant pour effet, en se contractant, de diminuer ou d'augmenter le volume de la cavité thoracique et des poumons.

L'expansion de la cage thoracique se fait dans les trois plans de l'espace :

- latéral transverse,
- antéro-postérieur,
- vertical.

En phase phonatoire, les muscles intervenant dans le mouvement de la cavité thoracique sont indiqués sur la figure 1-2.



1. Intercostaux externes
2. Scalènes
3. Sterno-cléido-mastoïdien
4. Intercostaux internes
5. Droit de l'abdomen
6. Oblique externe de l'abdomen
7. Oblique interne de l'abdomen
8. Transverse de l'abdomen

Figure 1-2. Muscles respiratoires

1-1-1-3 Le système canalaire (la trachée)

Ce système permet le transport du flux aérien jusqu'au niveau du vibreur laryngé. Il est constitué de conduits rigides dotés d'une armature cartilagineuse. La trachée est constituée de 18 anneaux cartilagineux incomplets à leur partie postérieure et en forme de fer à cheval. Elle se bifurque en deux bronches souches, une par poumon.

La rigidité anatomique du système canalaire fait de celui-ci un résonateur secondaire dans la phonation.

1-1-2 Le vibreur laryngé

Situé à l'extrémité supérieure de la trachée, le vibreur laryngé assure deux fonctions :

- fonction sphinctérienne, assurant d'une part la protection des voies respiratoires pendant la déglutition et d'autre part une consolidation de l'armature thoracique par un raidissement musculaire lors d'un soulèvement de charge (blocage glottique d'effort).
- fonction vibratoire permettant la phonation.

Nous décrivons les principales composantes du vibreur laryngé à savoir les cartilages, les systèmes ligamentaires et musculaires et enfin son innervation.

1-1-2-1 Les cartilages

Le squelette laryngé est constitué par des cartilages hyalins :

- le cartilage thyroïde impair et médian,
- le cartilage cricoïde impair et médian,
- les cartilages aryténoïdes pairs et latéraux avec leurs petits cartilages : les corniculés
- le cartilage épiglottique.

Mentionnons de plus l'os hyoïde qui joue un rôle important dans l'appareil suspenseur du larynx. La figure 1-3 présente l'aspect anatomique de ces cartilages ainsi que l'os hyoïde. Ces cartilages sont articulés entre eux et suspendus à l'hyoïde, au crâne et la mandibule, par un système ligamentaire et musculaire.

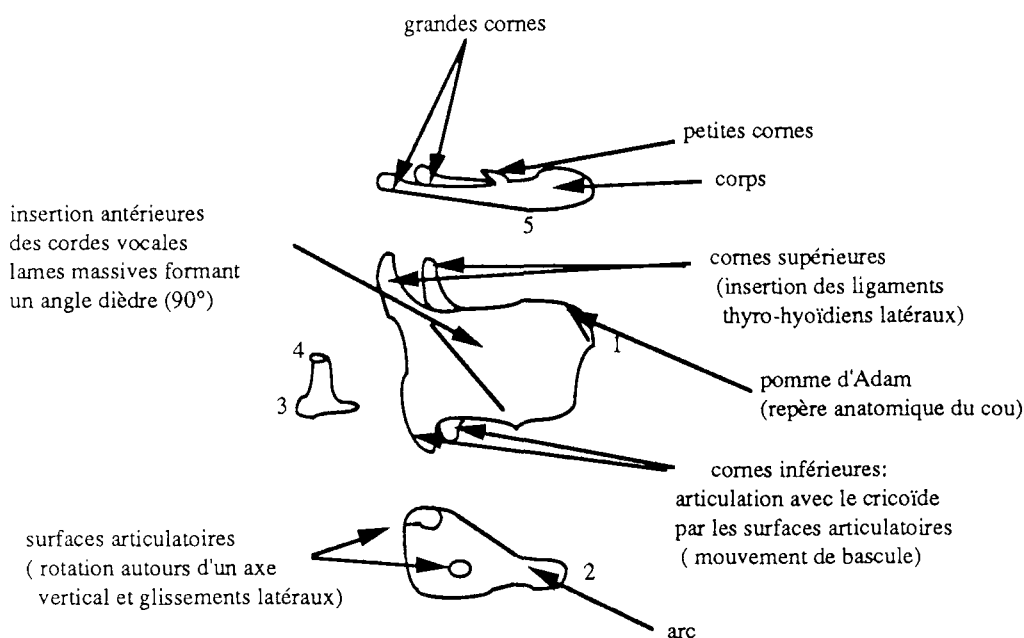


Figure 1-3. Os hyoïde et carilages laryngés :

1. Cartilage thyroïde
2. Cartilage cricoïde
3. Cartilage aryténoïde
4. Cartilage corniculé
5. Os hyoïde

1-1-2-2 Le système ligamentaire et musculaire

a- Le système ligamentaire

On distingue deux sortes de ligaments :

- extrinsèques rattachant le vibrateur laryngé à son environnement :
 - . ligaments thyro-hyoïdiens,
 - . ligament crico-trachéal,
 - . ligaments pharyngo-épiglottiques et glosso-épiglottiques,
- intrinsèques :
 - . ligament thyro-épiglottique,
 - . ligament crico-thyroïdien,
 - . ligaments ary-corniculés,
 - . membrane fibro-élastique du larynx.

b-Le système musculaire

Nous présentons dans la figure 1-4, les principaux muscles extrinsèques.

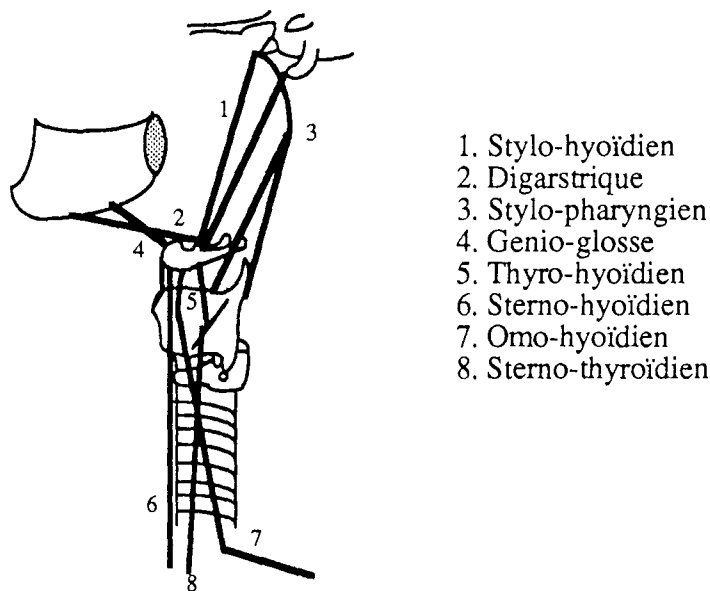


Figure 1-4. Muscles extrinsèques du vibrateur laryngé :

de 1 à 4 : muscles éleveurs

de 5 à 8 : muscles abaisseurs

Les muscles intrinsèques sont les muscles intervenant dans les fonctions d'adduction (rapprochement), d'abduction (écartement) et de tension des cordes vocales et sont présentés dans les schémas des figures 1-5 à 1-7.

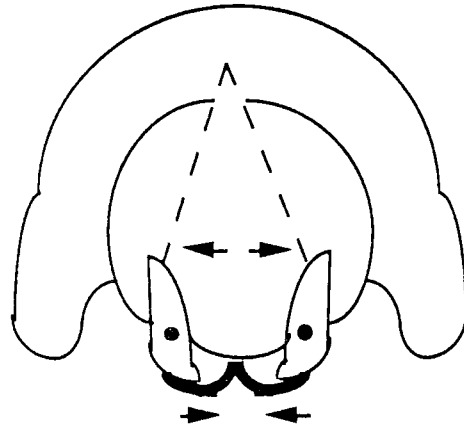


Figure 1-5. Abduction des cordes vocales par l'action des muscles crico-arythénoïdiens postérieurs

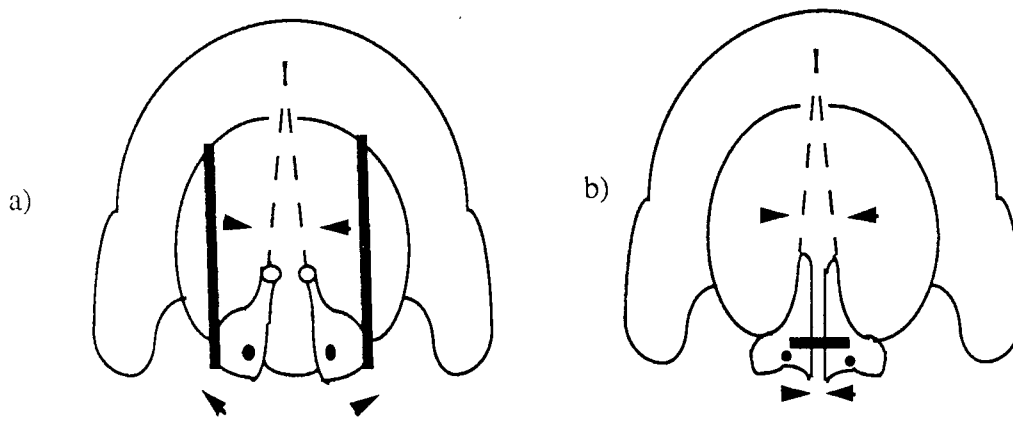


Figure 1-6. Adduction des cordes vocales par l'action des muscles :
a) crico-arythénoïdiens latéraux
b) arythénoïdiens transverse et oblique

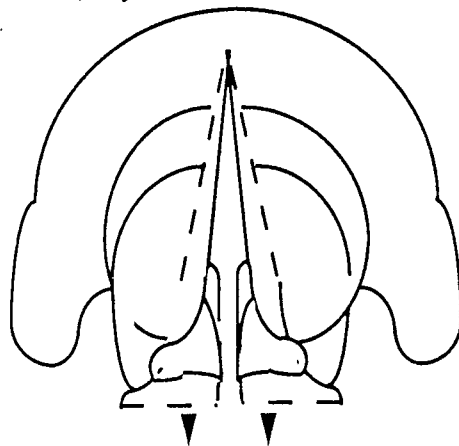


Figure 1-7. Mécanisme de tension intrinsèque de la corde vocale par l'action du muscle crico-thyroïdien (vue supérieure)

1-1-2-3 L'innervation

Le système d'innervation des muscles du vibrateur laryngé est constitué par la dixième paire et par le nerf récurrent. La figure 1-8 schématise l'innervation des muscles intrinsèques du larynx

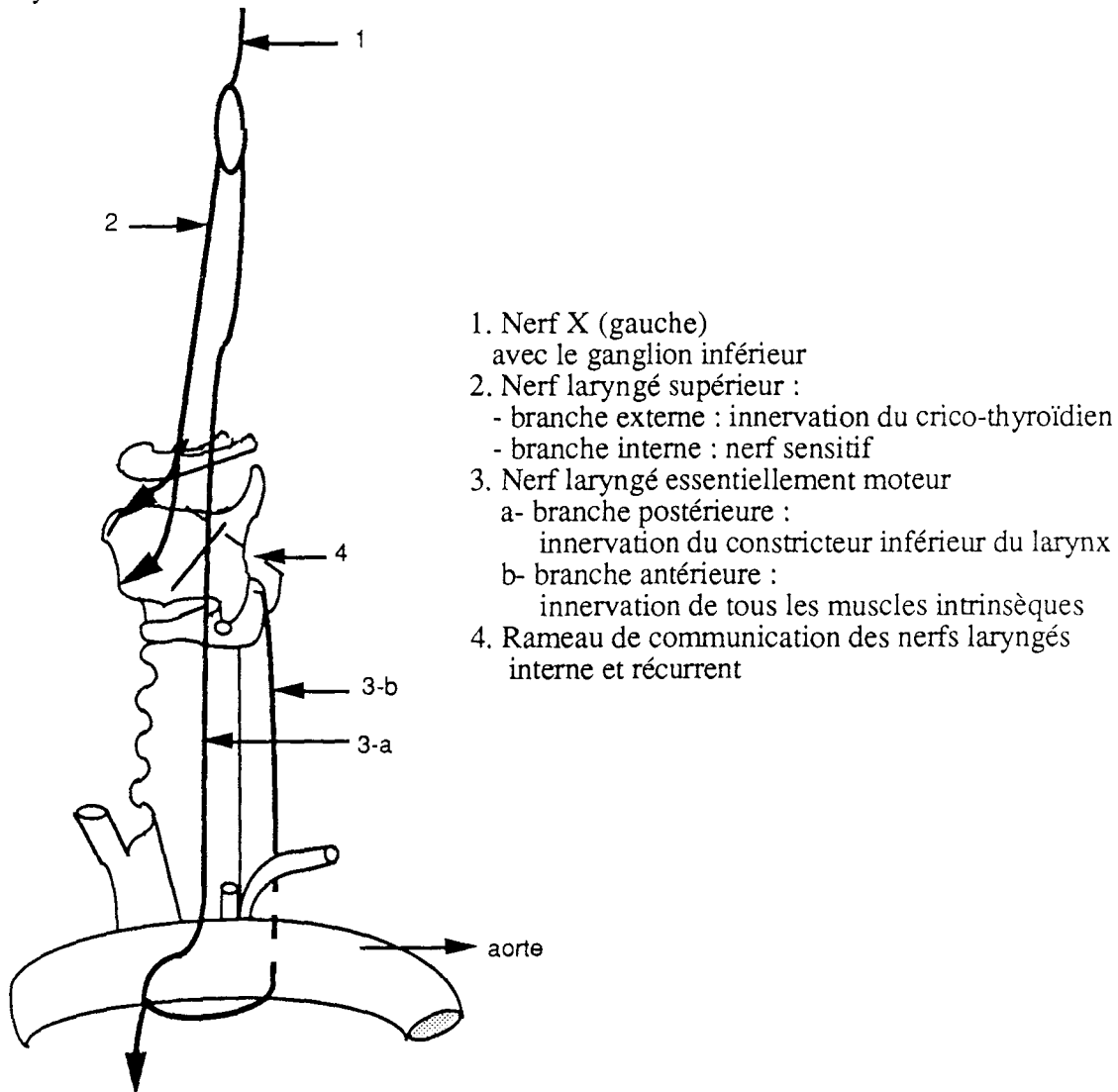


Figure 1-8. Innervation du larynx

1-1-3 Le résonateur pharyngo-naso-buccal

1-1-3-1 Le pharynx

Carrefour aéro-digestif, le pharynx constitue un pavillon musculaire s'étendant de la base du crâne jusqu'au niveau du cricoïde, où commence l'œsophage. Il est constitué de deux parois

latérales très mobiles, d'une paroi postérieure adhérent au tissu prévertébral et d'une paroi antérieure incomplète au niveau des communications pharyngo-buccales et pharyngo-nasales. Le pharynx est doté de deux systèmes vélaire, l'épiglotte et la valve vélo-pharyngée. Cette dernière joue un rôle capital dans la phonation (séparation des sons nasalisés et non nasalisés).

Le système musculaire pharyngé peut être divisé en deux groupes fonctionnels :

- les muscles constricteurs formés par des fibres circulaires; parmi ces muscles, notons le constricteur supérieur (au niveau de l'apophyse ptérygoïde), le constricteur moyen (au niveau de l'hyoïde) et le constricteur inférieur (le thyroïde et le cricoïde).

- les muscles élévateurs à fibres longitudinales essentiellement les palato-pharyngiens et les stylo-pharyngiens.

1-1-3-2 Le palais et le voile

Le palais est de structure osseuse sur sa partie antérieure et se termine en arrière à sa partie médiane par la luette. Le voile du palais contient plusieurs muscles assurant ses mouvements phonatoires et de déglutition. Le schéma de la figure 1-9 résume l'action musculaire sur le voile du palais.

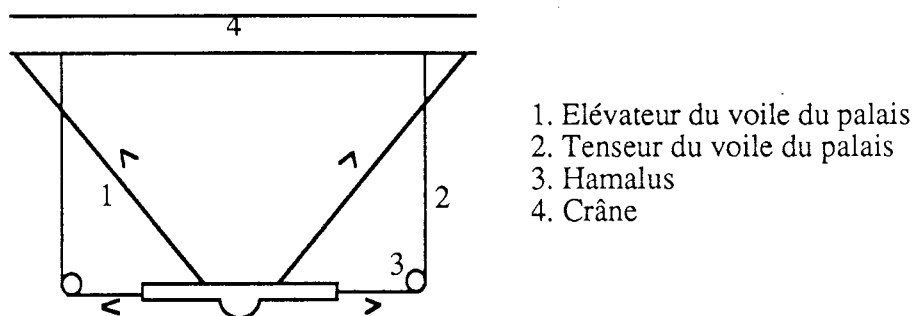


Figure 1-9. Mécanismes d'action musculaire sur le voile du palais

1-1-3-3 La cavité orale

Les arcades dentaires inférieures et supérieures séparent la cavité orale proprement dite du vestibule oral, se terminant par les lèvres. Le plancher de la cavité buccale comprend le muscle mylo-hyoïdien qui supporte la langue.

Parmi les organes de la cavité orale on note :

- la langue : organe extrêmement mobile et riche en muscles, elle joue un rôle essentiel dans la prononciation. Elle est rattachée au plancher par le frein lingual et à l'hyoïde par la membrane glosso-hyodienne. Les muscles intrinsèques et extrinsèques de la langue sont

représentés dans le schéma de la figure 1-9.

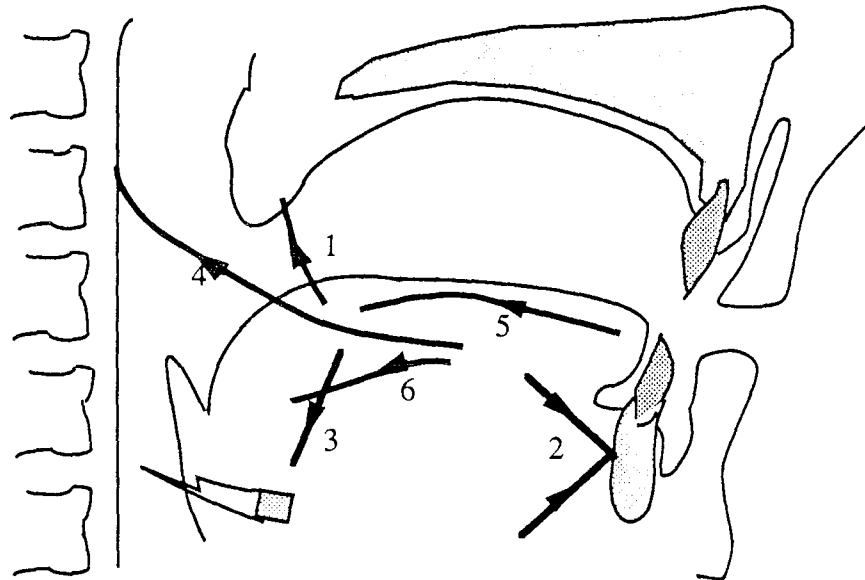


Figure 1-9. Action schématique des principaux muscles de la langue

- 1- Muscles palato-glosses
- 2- Muscles génio-glosses
- 3- Muscles hyo-glosses
- 4- Muscles stylo-glosses
- 5- Muscle longitudinal supérieur
- 6- Muscle longitudinal inférieur

- les joues sont formées essentiellement par les muscles buccinateurs.
- les lèvres : ayant une importance capitale articulaire, les lèvres ont une structure musculaire développée, représentée à la figure 1-10.
- la mâchoire inférieure : en rapport très étroit avec l'hyoïde et la langue, ses mouvements ont une influence sur la morphologie bucco-pharyngée. De par son rôle dans la physiologie masticatoire, elle possède également un appareil musculaire complexe.

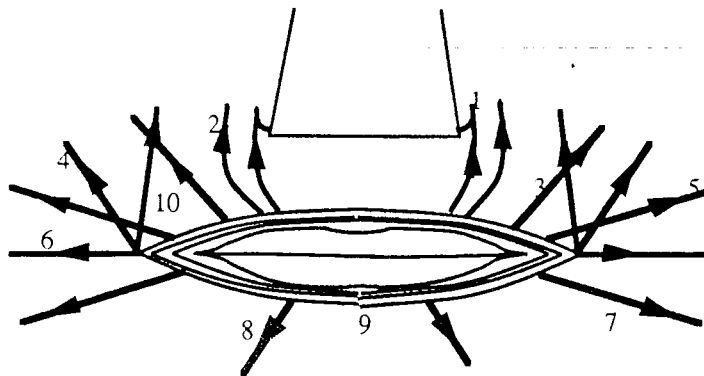


Figure 1-10. Muscles de la motricité labiale

- | | |
|--|--------------------------------------|
| 1- Relévateur naso-labial | 2- Relévateur de la lèvre supérieure |
| 3- Petit zygomatique | 4- Grand zygomatique |
| 5- Buccinateur | 6- Risorius |
| 7- Dépresseur de l'angle de la bouche | |
| 8- Dépresseur de la lèvre inférieure | |
| 9- Orbiculaire de la bouche | |
| 10- Relévateur de l'angle de la bouche | |

1-1-3-4 Les cavités naso-sinusiennes

La cavité nasale est délimitée d'un côté par les narines et de l'autre par les orifices pharyngés ou choannes. Elle est divisée plus ou moins symétriquement par le septum nasal et elle contient de chaque côté trois cornets recouverts d'une muqueuse. La cavité nasale communique avec les sinus paranasaux. Aucun muscle n'agit sur la morphologie de ces cavités pneumatiques, ce qui fait que la cavité nasale intervient dans la phonation comme résonateur statique.

1-1-4 La commande nerveuse

L'émission sonore et l'articulation verbale sont le résultat d'une synergie des actions musculaires de toutes les composantes des voies respiratoires. Elles sont commandées par des voies motrices primaires partant du cortex moteur et prémoteur (l'aire de Broca située au dessous de l'intersection des scissures de Rolando et de Silvius). Cette aire est en communication avec des voies motrices secondaires appartenant au système extra-pyramidal. Notons de plus que le cervelet joue un rôle de régulation et de coordination dans la phonation. Le tableau 1-1 résume les voies motrices nerveuses de la phonation.

NOYAUX CENTRAUX	NERFS PERIPHERIQUES	MUSCLES EFFECTEURS	ACTION MUSCULAIRE
Niveau protubérantiel			
<i>Noyau moteur du Trijumeau</i>	Nerf maxillaire inférieur (V3)	Muscles des mâchoires Tenseur du voile du palais Mylo-hyoïdien Digastrique	Fermeture et ouverture des mâchoires Tension du voile du palais Elévation du larynx
<i>Noyau moteur du facial</i>	N. facial (VII)	Muscles des joues et des lèvres Stylo-hyoïdien Digastrique	Mouvements articulaires faciaux Elévation du larynx
Niveau bulbaire			
<i>Noyau ambigu</i>	N. pneumogastrique (X) N. glosso-pharyngien (IX)	Elévateur du voile du palais Muscles du larynx Constricteurs du pharynx	Occlusion du nasopharynx Mouvements vélaire Mouvements pharyngés et laryngés
<i>Noyau de l'hypoglosse</i>	N. spinal (XI) N. hypoglosse (XII)	Muscles linguaux	Mouvements articulaires bucco-pharyngés
Niveau médullaire cervical			
1	C1 (hypoglosse)	Génio-hyoïdien Thyro-hyoïdien	Elévation du larynx Abaissement de l'hyoïde
1 - 3	C1 - C3 anse de l'hypoglosse	Sterno-hyoïdien	Abaissement et stabilisation du larynx
2 - 4	C2 - C4 racine médullaire du spinal	Omo-hyoïdien Sterno-thyroïdien	Inspiration forcée
3 - 5	C3 - C5 nerf phrénique	Diaphragme	Inspiration
1 - 4	C1 - C4 plexus cervical profond	Scalènes	Inspiration
Niveau médullaire dorsal			
1 - 11	D1 - D11 nerfs intercostaux	Intercostaux internes et externes	Mouvements respiratoires
7 - 12	D7 - D12 nerfs intercostaux	Muscles antérieurs et latéraux de la paroi abdominale	Expiration

Tableau 1-1. Voies nerveuses motrices de la phonation

1-2 Physiologie du système de production de la parole

Dans ce paragraphe, nous présenterons les aspects physiologiques les plus importants de l'appareil phonatoire, en mettant en parallèle les physiologies phonatoire et respiratoire. Nous insisterons en particulier sur le vibreur laryngé pour lequel nous présenterons brièvement un historique des différentes théories de la vibration laryngée. En fin du paragraphe, nous décrirons qualitativement la physiologie du résonateur pharyngo-bucco-nasal.

1-2-1 Physiologie de l'expiration phonatoire

1-2-1-1 Débit et pression sous-glottiques

En expiration libre, la différence de pression entre les niveaux sous-glottiques et sus-glottiques est minime car le sphincter glottique est largement ouvert. Par contre, en expiration phonatoire, équivalente à une respiration "sous contrainte", la pression sous-glottique et le débit trans-glottique sont très variables de manière à s'adapter en permanence aux contraintes du phrasé en réglant les paramètres acoustiques.

Il en résulte que :

- le temps expiratoire s'allonge considérablement,
- l'expiration devient active nécessitant la contraction musculaire expiratoire, plus particulièrement la sangle abdominale,
- le rythme respiratoire perd sa régularité pour s'adapter par de fréquentes pauses aux impératifs de la phonation (intonation, ponctuation et mélodie).

Les débits aériens au niveau glottique sont de nature différente pour une expiration phonatoire et une expiration respiratoire. Pour la première, le débit aérien est de nature turbulent à cause de la variation impédantielle glottique. Par contre, il est laminaire pour une expiration respiratoire à cause de la nature de l'écoulement de l'air et la faible pression sous-glottique.

1-2-1-2 Mécanique respiratoire en phonation

Les inspirations, qui en respiration libre ne font intervenir que la contraction du diaphragme et des muscles intercostaux externes, sont plus rapides et plus profondes au cours de la phonation. Elles sollicitent en plus des muscles inspireurs accessoires (scalènes, sterno-cléido-mastoïdiens).

L'expiration phonatoire se caractérise par une mise en tension plus ou moins ferme du

thorax par la contraction tonique antagoniste des muscles inspireurs et expirateurs, évitant ainsi un relâchement de la musculature thoraciques après les inspirations. Bien que cette posture conserve sa fermeté, elle se modifie progressivement à mesure que l'expiration phonatoire progresse.

1-2-2 Physiologie du vibrateur laryngé

1-2-2-1 Historique

Plusieurs théories tentant d'expliquer la genèse du son laryngé dans la phonation, ont vu le jour depuis plus d'un siècle.

- Théorie myo-élastique

Cette théorie due à Ewald (fin XIX ème siècle) et mise à jour par plusieurs auteurs (Van Den Berg , Vallencien, Hirato), affirme que la vibration résulte du passage du flux aérien expiratoire entre les cordes vocales plus ou moins tendues en position d'adduction. La vibration est donc passive, puisque le système nerveux n'y intervient que pour le maintien tonique de cette adduction et pour le réglage de la tension des cordes vocales.

- Théorie muco-ondulatoire

Perello (1962) explique la vibration par les propriétés ondulatoires de la muqueuse de la corde vocale. En effet, cet épithélium possède une grande liberté de mouvement par rapport au plan tissulaire, en raison de l'existence sous l'épithélium d'un espace (de Reinke) qui se comporte comme une poche séreuse.

- Théorie neuro-chronaxique

Elle repose, selon Husson (1950), sur un postulat selon lequel la vibration des cordes vocales est d'origine neuro-musculaire, donc active et induite par des influx récurrentiels à la fréquence du son émis. Bien que cette théorie comble certaines des lacunes précédentes telles que la différenciation hauteur-intensité, son postulat a été infirmé par de l'expérimentation neuro-physiologique.

- Théorie oscillo-impédantielle

La corde vocale est considérée comme un oscillateur entretenu, dont la fréquence est déterminée par l'inertie et l'élasticité de la portion vibrante et dont l'apport périodique d'énergie est constitué par l'onde de pression sous-glottique.

Le mouvement vibratoire normal de la corde vocale est tri-dimensionnel; en effet, chaque point de la corde a une trajectoire ellipsoïdale dont la principale composante est horizontale. Le bord libre présente dans le plan frontal une dynamique vibratoire différenciée, illustrée dans le schéma de la figure 1-11.

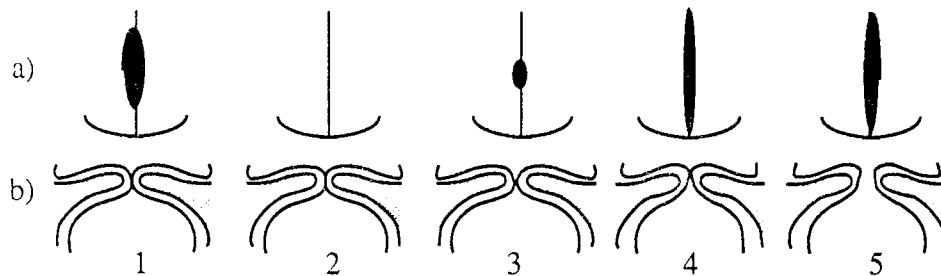


Figure 1-11. Dynamique vibratoire des bords libres des cordes vocales

a) vue supérieure

b) coupe longitudinale

Les propriétés impédantielles (masse, raideur, viscosité) sont étroitement conditionnées par le muscle vocal. Sa structure fibrale étant un enchevêtrement de petits groupes torsadés de fibres musculaires, le muscle vocal permet une régulation très fine des facteurs masse effective et raideur. Cette structure explique aussi l'incurvation macroscopique du bord libre de la corde vocale que l'on peut observer immédiatement avant la vibration.

La contraction des fibres musculaires vocales est contre-balancée par celle des crico-thyroïdiens, qui agissent par un mécanisme de bras de levier constitué par le cricoïde et les arythénoïdes. La solidarité de ces cartilages est alors assurée par la contraction des autres muscles s'insérant sur eux. Ce mouvement de bascule a pour effet un allongement appréciable des cordes vocales (20-30%), et a pour conséquence, un accroissement de la tension et une diminution de l'épaisseur. Il permet donc l'accession aux notes aiguës du registre de tête.

L'intensité sonore ne dépend, selon cette théorie que de l'amplitude de la pression sous-glottique.

1-2-2-2 Les attaques vocaliques

L'attaque vocalique permet de décrire "l'état initial" de la glotte, précédant l'émission sonore. On en distingue deux modalités physiologiques aussi bien par l'analyse acoustique que par le comportement glottique : l'attaque douce et l'attaque dure.

Pour la première, les deux arythénoïdes sont en abduction, et les bords libres des cordes vocales présentent une légère concavité, donnant à l'aire glottique une configuration fusiforme.

En phase pré-phonatoire, l'air passe à travers ce fuseau, amorçant une vibration de plus en plus ample qui a donc commencé glotte ouverte.

Une exagération de cette attaque est appelée attaque soufflée. Dans ce cas, on note un important asynchronisme entre le début de la vibration des cordes vocales et le passage d'air; un bruit aérien parasite est alors nettement audible.

Pour l'attaque dure, les cordes vocales commencent à se placer en position d'adduction obturant ainsi la glotte. Les plis vestibulaires s'écartent légèrement. La pression sous-glottique monte progressivement jusqu'à "forcer" le barrage glottique, et à déclencher le premier cycle vibratoire. La phonation a commencé glotte fermée.

Le cas extrême de cette attaque correspond au "coup de glotte". Il est caractérisé d'une part par une montée soudaine de la pression sous-glottique, et par un manque d'écartement des plis vestibulaires. Dans ce cas la sonorité présente un caractère explosif et manque de netteté et de tonalité.

1-2-2-3 Les registres

Ce sont, au sein de la tessiture, des ensembles de tons contigus, correspondant à des notes musicales, qui possèdent en commun une consonnance et un timbre propres caractéristiques, qui les différencient d'autres tons ayant également un autre timbre particulier.

Indépendamment de la fréquence fondamentale et de l'intensité, les registres permettent de caractériser le mode vibratoire du larynx. Nous distinguons comme Hollien (1972) trois registres pour la parole, dont nous résumons les caractéristiques dans le tableau 1-2.

1-2-3 Physiologie de la résonance

Un aspect important dans la morphologie du système de production de la parole est le phénomène de résonance de toutes ses cavités.

Les cavités trachéo-broncho-pulmonaires ont des fréquences propres de l'ordre de 50 - 100 à 300 Hz. Leur caractère statique fait que ces cavités n'interviennent pas dans le codage acoustico-phonémique.

Les ventricules du larynx, d'une taille très réduite de 0,5 à 1 cm³, ont un rôle accessoire de "chambre de compression".

D'autre part, les cavités pharyngo-naso-buccale, dont la morphologie est très variable, donnent naissance à des résonances ou "formants", quand elles sont couplées au vibreur laryngé.

Au départ, la voix contient les harmoniques du fondamental. Certains de ces harmoniques

vont être renforcés, d'autres étouffés, par l'agencement des pavillons articulatoires. Ainsi l'énergie acoustique va être concentrée dans le spectre sonore, autour de ces zones formantiques.

Registre	Etendue		Mode vibratoire
	Masculine	Féminine	
Registre grave ou de poitrine	Bas de la tessiture jusqu'à plus ou moins ré3-fa3	Bas de la tessiture jusqu'à ré4-fa4	<ul style="list-style-type: none"> - Registre de la phonation parlée - Cordes vocales plus ou moins tendues - Vibration des cordes vocales sur toute la longueur avec une composante appréciable d'affrontement des bords libres - Le résonateur thoracique largement sollicité : perception subjective d'une résonance dans la poitrine
Registre aigu ou de tête	de ré3-fa3 jusqu'au haut de la tessiture (exceptées celles du registre de flagelot)	de ré4-fa4 jusqu'au haut de la tessiture (exceptées celles du registre de fausset)	<ul style="list-style-type: none"> - Correspond aux sons couverts - Les cordes vocales s'allongent avec une tension accrue des bords libres et une diminution relative de la masse vibrante - Agrandissement de la cavité pharyngée dans toutes ses dimensions : légère descente du larynx, projection en avant de la langue, relèvement du voile du palais - La résonance est perçue dans la cavité crânienne
Registre de fausset (homme) de flagelot (femme)	Etendue du registre de tête à quelques degrés plus haut		<ul style="list-style-type: none"> - Voix acoustiquement grêle se rapprochant de la voix infantile - Pour les sons les plus graves : <ul style="list-style-type: none"> * tension passive des cordes vocales par l'action des crico-thyroïdiens; * relâchement des muscles vocaux; * toute la partie intermembranacée participe à la vibration. - Pour les sons les plus aigus : <ul style="list-style-type: none"> * les muscles vocaux prennent une tension active; * vibration réduite n'intéressant qu'une partie de la région intermembranacée de la glotte, le restant demeurant figé en accollement ferme; * tension maximale de tous les muscles intrinsèques

Tableau 1-2. Les registres de la voix humaine

Dans chaque spectre vocalique, on peut relever deux formants principaux pouvant correspondre d'une part à la cavité pharyngée, d'autre part à la cavité buccale.

Il est à noter les variations inter-locuteurs des formants pour un même phonème et les variations contextuelles, pour un même locuteur, dépendant de son environnement émotionnel.

2- PATHOLOGIES DE LA VOIX HUMAINE

PATHOLOGIES DE LA VOIX HUMAINE

L'élaboration d'un système automatique de mesures objectives sur la voix humaine, en vue d'une aide au diagnostic des troubles vocaux, nécessite une bonne connaissance de l'origine de ces troubles, leur localisation et dans la mesure du possible leur incidence sur la phonation.

Dans ce chapitre, nous présenterons de façon non exhaustive les dysphonies dysorganiques, les dysphonies dysfonctionnelles et celles à caractère ambivalent. Nous attacherons une importance particulière aux répercussions qualitatives sur la production de la parole.

2-1 Les dysphonies dysorganiques

Nous qualifions une dysphonie de dysorganique, celle dont l'origine est extrinsèquement mécanique; c'est-à-dire tout phénomène histologique lésionnel qui perturbe mécaniquement le fonctionnement de l'appareil phonatoire. Nous classerons ces dysphonies, selon la localisation du dysfonctionnement, en trois catégories.

2-1-1. Les troubles laryngés

Le vibreur laryngé, jouant un rôle primordial dans la phonation, tout trouble l'affectant se répercute d'une façon ou d'une autre sur la qualité de la voix.

Ces troubles peuvent avoir différentes origines et différentes formes :

- troubles affectant la morphologie du larynx,
- troubles affectant le système nerveux (périphérique ou central),
- troubles provenant d'autres phénomènes.

2-1-1-1 Phénomènes morphologiques

Ces phénomènes sont de plusieurs types et de différentes origines :

- Traumatiques extrinsèques et intrinsèques : ce sont toutes les fractures et luxations aryénoïdiennes, les blessures des muqueuses provenant de corps étrangers (intubations par exemple) ou le coup de fouet laryngien (utilisation abusive professionnelle de la voix) etc...

- Inflammatoires et infectieux aigus ou chroniques : parmi les agents des infections chroniques, notons les attaques microbiennes, virales ou mycotiques (tuberculose par

exemple), les irritants locaux et les cancérigènes de la fumée de tabac, l'alcool favorisant une congestion chronique qui facilite les autres effets irritatifs, les gaz et les poussières, etc...

Nous illustrons un exemple des effets histologiques du tabagisme dans la figure 2-1

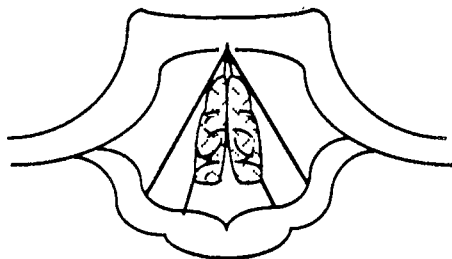


Figure 2-1. Exemple des effets du tabagisme sur les cordes vocales : Œdème de RINKE

- Effets tumoraux : l'incidence des tumeurs bénignes ou malignes sur la voix dépend plus de la localisation de la lésion que de sa taille. Ces lésions gênent la phonation normale en déformant les bords libres des cordes vocales, en empêchant une adduction complète et en inéquant les masses des deux cordes vocales. La figure 2-2 illustre quelques tumeurs bénignes et la figure 2-3 présente un exemple de tumeur maligne affectant les cordes vocales.

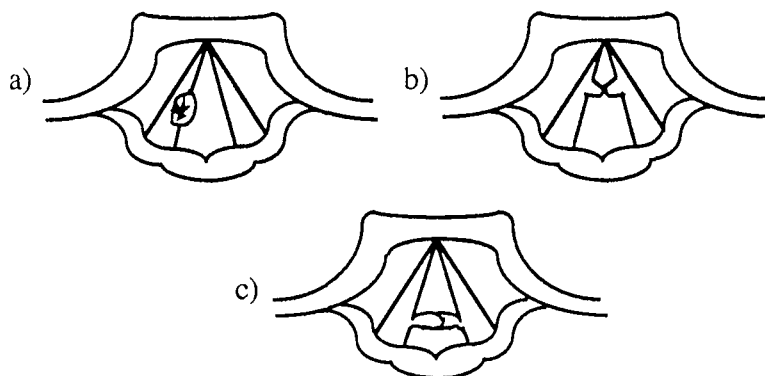


Figure 2-2. Exemples de tumeurs bénignes des cordes vocales

- a) kyste de la corde vocale droite
- b) polype bilatéral
- c) granulome d'intubation bilatéral

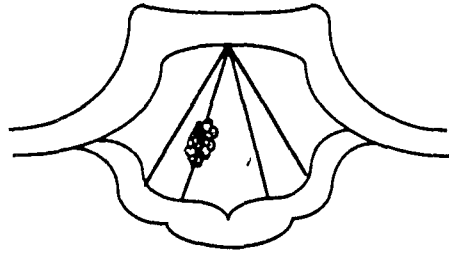


Figure 2-3. Epithélioma débutant d'une corde vocale

2-1-1-2 Les altérations neurogènes

Elles sont de deux types selon leur localisation : les altérations périphériques et centrales.

Les altérations neurogènes périphériques se manifestent au niveau du larynx par des paralysies ou parésies des cordes vocales dont la grande majorité (90%) sont unilatérales. Selon la position de la paralysie d'une corde vocale, l'incidence sur la phonation varie :

- en position d'abduction, la voix est peu altérée et parfois bitonale,
- en position paramédiane, au début de l'altération, on note un timbre modérément altéré avec parfois des cas de bitonalité,
- en position intermédiaire, la voix est souvent râpeuse, partiellement soufflée au timbre très impur, surtout lorsque le nerf laryngé supérieur est atteint,
- en adduction, position de paralysie très rare, on observe une aphonie.

La figure 2-4 montre les différentes positions de paralysie en phase respiratoire et phonatoire.

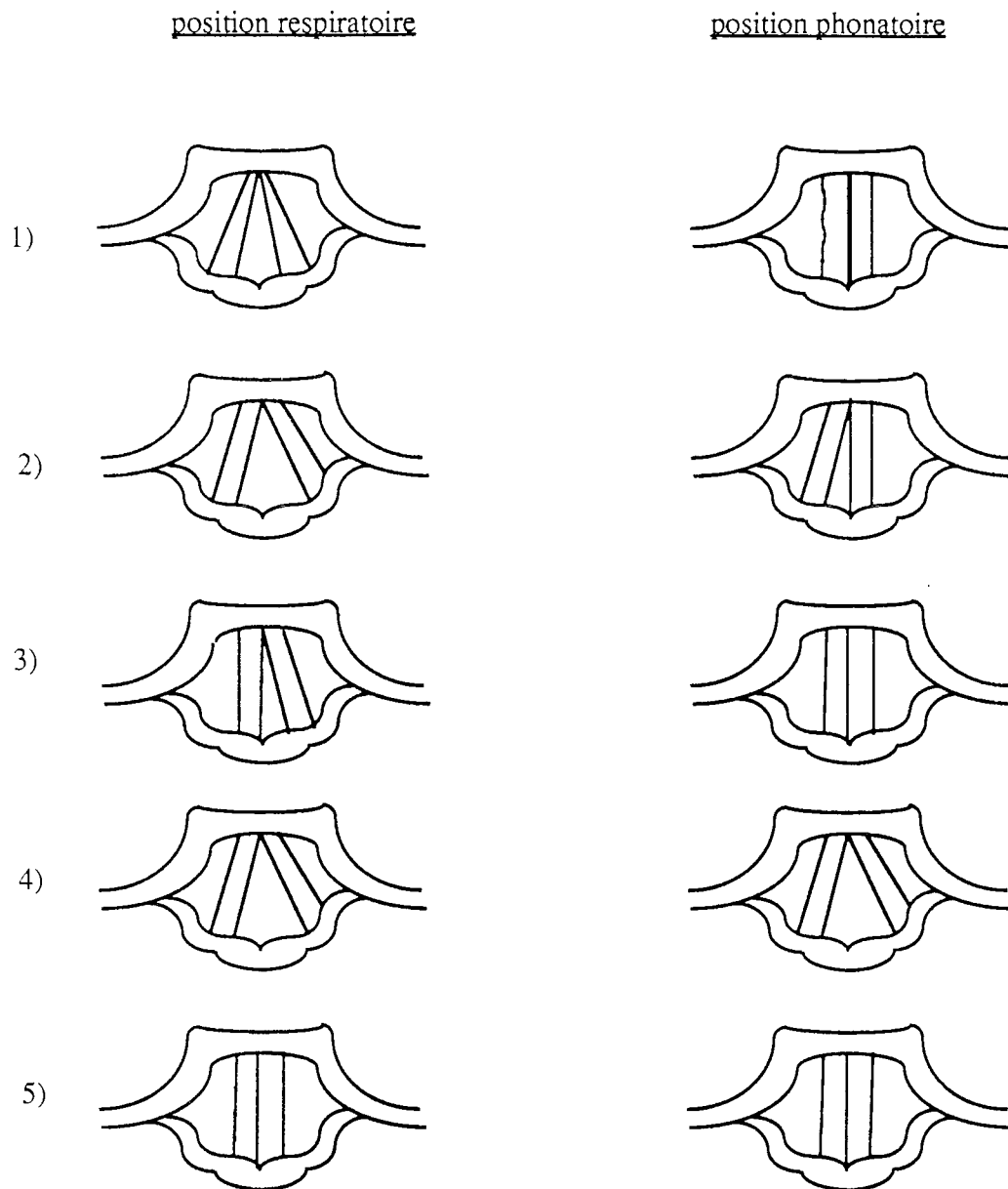


Figure 2.4. Schéma laryngoscopique de quelques paralysies laryngées classiques.

- 1) paralysie du crico-thyroïdien droit
- 2) paralysie en abduction de la corde vocale droite
- 3) paralysie en adduction de la corde vocale droite
- 4) paralysie bilatérale en abduction
- 5) paralysie bilatérale en adduction

Les atteintes neurogènes périphériques peuvent avoir comme effet, dans de rare cas, des diplégies (parésies bilatérales) soit en fermeture soit en ouverture. Dans le cas d'une diplégie en fermeture (fente glottique de 1 à 2 mm), la voix est altérée sans aphonie. Par contre pour une

paralysie en ouverture on constate une aphonie avec parfois des fausses routes alimentaires mais pas de dyspnée (gêne de respiration).

Les altérations neurogènes centrales sont des atteintes des voies nerveuses centrales de la phonation au-dessus de l'étage nucléaire bulbaire. Elles prennent origine à partir de troubles vasculaires, de tumeurs ou d'affections dégénératives. Essentiellement, ces altérations réalisent des dysarthries et se répercutent sur la voix pour donner naissance au syndrome de dysphonie dysarthrique.

Parmi ces syndromes notons le syndrome moteur pseudo-bulbaire caractérisé par une voix faible et très fatigable ou par une voix explosive saccadée et un nasonnement au timbre légèrement soufflé.

La maladie de Parkinson est un autre exemple de ces syndromes; elle se manifeste par une faible intensité, une tonalité aiguë, un léger soufflement du timbre et des attaques vocaliques en "coup de glotte".

Quant aux affections cérébelleuses, on observe une profonde perturbation rythmique et mélodique de la voix avec un timbre à composante soufflée.

Pour le syndrome de la sclérose en plaques, en plus du soufflement, de la fatigabilité et de la faiblesse de l'intensité, il se manifeste par un fondamental inconstant mais monotone accompagné parfois par des émissions bitonales.

2-1-1-3 Perturbations endocriniennes et pharmacologiques

Certaines hormones et produits pharmacologiques ont une répercussion indirecte sur la phonation. L'exemple le plus significatif est la mue observée pendant l'âge de la puberté. Notons aussi une aggravation du fondamental, une diminution de l'agilité vocale et une uniformisation de la mélodie, qui apparaissent à la période ménopausale.

2-1-2 Les troubles respiratoires

Toute trachéo-broncho-pneumopathie importante est susceptible de se répercuter sur la phonation en constituant un empêchement mécanique de la respiration phonatoire ou en entraînant des inflammations de contiguïté au niveau du larynx.

Des altérations du geste respiratoire peuvent provenir aussi bien de paralysies que d'atteintes traumatiques des muscles respiratoires constituant le "moteur" de la soufflerie.

2-1-3 Les troubles des cavités de résonance

Les pathologies infectieuses et inflammatoires des cavités de résonance influencent par proximité le larynx en favorisant dans celui-ci des sécrétions irritantes qui peuvent être à l'origine d'infection chronique de la muqueuse des cordes vocales.

Les obstructions nasales totales ou subtotaales provoquent des hyporhinophonies organiques. Elles affectent le timbre des sons nasalisés (/ẽ/, /ã/, /õ/, /m/, /n/,...).

Parmi les troubles des cavités de résonance, l'insuffisance vélo-pharyngée gêne beaucoup la phonation. Elle se manifeste par une séparation inadéquate des cavités oro-pharyngée et naso-pharyngée, avec une déperdition anormale d'air par le nez au cours de la parole (rhinolalie ouverte). Dans un certain nombre de cas, se développe un mécanisme partiellement compensateur sous la forme d'une saillie semi-circulaire horizontale dans la paroi postérieure du pharynx : le bourrelet de Passavant (fig. 2.5).

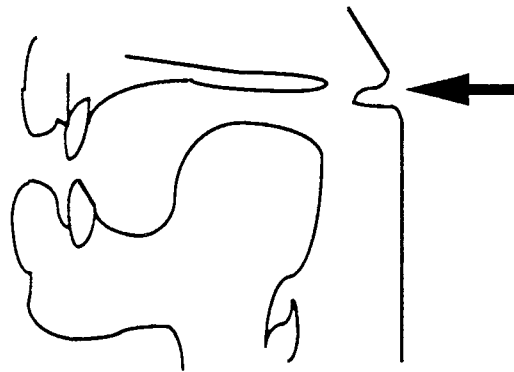


Figure 2-5. Le bourrelet de Passavant

Les symptômes de l'insuffisance vélo-pharyngée sont les suivants :

- le nasonnement caractérisé par une impression d'aggravation de la voix s'accompagnant d'une atténuation de l'énergie acoustique pour les fréquences aiguës en faveur de la zone au-dessous de 1000 Hz,
- le nasillement caractérisé au contraire par une impression de voix plus aiguë pour laquelle l'énergie acoustique est concentrée au-dessus de 2000 Hz,
- distorsions consonantiques par fuite d'air parasite audible surtout pour les occlusives sourdes,
- le souffle et le ronflement nasaux,
- les bruits phonétiques de substitution où le patient tente de remplacer les occlusives par un coup de glotte et les constrictives par un souffle rauque (rétrécissement au niveau des plis vestibulaires).

2-2- Les dysphonies dysfonctionnelles

Quand la voix est altérée avec une absence de perturbation anatomo-pathologique et une présence d'un trouble de tonicité musculaire au niveau de l'appareil phonatoire, on parle de dysphonie dysfonctionnelle. Ces dysphonies se caractérisent, au niveau de l'appareil phonatoire, par une dystonie (trouble de la tonicité musculaire) pouvant osciller entre deux pôles antagonistes : l'hypertonie et l'hypotonie.

L'origine de telles dysfonctions peut être l'adoption d'une habitude vocale défectueuse, ou un trouble d'ordre relationnel ou émotionnel. Dans la majorité des cas, la dystonie s'aggrave et s'entretient par un mécanisme dit du "cercle vicieux de l'effort" [Le Huche].

2-2-1 L'habitude vocale défectueuse

Après une déficience de la production de la parole, le phonateur peut ne pas s'affranchir d'une habitude vocale défectueuse acquise pendant la période du "trouble".

Elle comporte deux éléments :

- un élément de rendement énergétique défavorable où le mouvement fonctionnellement déficient nécessite une dépense énergétique plus importante;
- un élément de déficience quantitative ou qualitative de la production vocale.

Elle peut provenir soit d'un dérèglement mécanique ou fonctionnel du geste phonatoire.

Le dérèglement mécanique est tout phénomène d'origine extrinsèque ou intrinsèque perturbant, de manière temporaire ou prolongée, la motricité phonatoire.

Un concept fondamental est que le mécanisme peut régresser entièrement et ne laisser comme séquelle que l'habitude vocale défectueuse à laquelle il a donné naissance.

Citons quelques exemples de ce phénomène :

- un laryngite aiguë provoquant un œdème des muqueuses;
- une polypose allergique du nez perturbant la perméabilité et la résonance nasale;
- une otite séreuse qui déforme les perceptions auditives du contrôle audio-phonatoire...

Le dérèglement fonctionnel peut avoir comme origine un "abus" vocal suffisamment prolongé. Il s'agit d'un dérèglement d'un système ou d'un organe qui est appelé à fonctionner à son régime maximum trop longtemps ou trop fréquemment. On peut citer comme exemple, la jeune institutrice confrontée à la turbulence d'une classe de trente élèves, dans un local défavorable sur le plan de l'acoustique.

2-2-2 Les troubles relationnels et émotionnels

La voix est une expression corporelle qui s'insère dans un comportement d'ensemble. Sa fonction première est une fonction de relation. Chez certains patients, la dysphonie dysfonctionnelle peut se concevoir comme un trouble du comportement vocal du sujet dans la relation avec soi-même ou avec le monde dans sa réalité spatiale, temporelle ou humaine. Toutes les pulsions internes conscientes ou subconscientes et les influences extérieures agissent sur la qualité de la voix.

Le concept de dysphonie de refuge est un exemple de tels troubles relationnels, où le sujet se réfugie plus ou moins inconsciemment dans la dysphonie pour échapper à certaines contraintes qu'il subirait avec une voix tout à fait normale (refus de certaines responsabilités...).

Une dysphonie peut aussi prendre naissance à partir d'un trouble d'ordre émotionnel. En effet, un trouble émotionnel, pouvant signifier une impossibilité d'extériorisation d'un sentiment, peut distordre ou limiter la mobilité du corps par des modifications de la raideur et du tonus musculaires. Ainsi, l'hésitation, l'angoisse et le stress se traduisent par un désordre anarchique du rythme respiratoire, un accroissement de la rigidité musculaire et par des hyper-sécrétions externes au niveau des voies aériennes supérieures. Par contre, les tendances dépressives réduisent l'activité respiratoire et diminuent le tonus musculaire.

Pour le système de production, les dysphonies relationnelles ou émotionnelles ont pour effet une inadaptation des impédances glottique et sus-glottique à la pression sus-glottique et au débit trans-glottique d'une part, et une modification des caractéristiques physiques des muqueuses des cordes vocales et des cavités de résonance, d'autre part.

Sur le plan de la phonation, elles se concrétisent dans le schéma des quatre anneaux de tension:

- 1° anneau: la bouche et les lèvres (lèvres pincées);
- 2° anneau: le cou (collier de serrage);
- 3° anneau: la partie haute du thorax et la ceinture scapulaire;
- 4° anneau: la sangle abdominale.

2-2-3 Le cercle vicieux de l'effort vocal [Le Huche]

La prise de conscience d'un fléchissement au niveau de l'expression vocale, dû à une déficience qualitative quelconque, entraîne un processus de compensation. Celui-ci, dans un certain nombre de cas, s'avère inadéquat et provoque un accroissement d'énergie pour une efficacité vocale moindre. Cet accroissement résulte de l'entretien de contractions musculaires extra-phonatoires parasites, qui réduisent l'efficacité du geste phonatoire. Globalement l'effort vocal manque son but de compensation et entraîne la fatigue.

Les portes d'entrée dans le cercle vicieux de l'effort sont multiples :

- fatigue générale ou vocale, phénomène organique ou fonctionnel entraînant une diminution de la puissance, l'apparition d'une composante soufflée dans le timbre, dérèglement de l'intonation...

- manque de "portée" de la voix sur le plan signifiant qui provoque une modification du geste phonatoire quand le locuteur prend conscience du manque de "prise" sur son auditeur.

2-2-4 Quelques exemples particuliers de dysphonies dysfonctionnelles

2-2-4-1 Les troubles vocaux de la mue

La mue normale consiste en une augmentation de taille importante du larynx, avec un accroissement des dimensions des cordes vocales. Elle apparaît à l'âge de 10-12 ans chez la fille et 11-13 ans chez le garçon et dure en moyenne de 6 à 12 mois. Elle s'accompagne d'une baisse du fondamental d'environ 1 octave chez le garçon et d'une tierce chez la fille.

La période transitoire de la mue se traduit par une dysphonie caractérisée par :

- une raucité du timbre, plus ou moins permanente;
- des sauts brusques d'une octave du fondamental;
- des bitonalités par intermittence.

Parfois on observe en laryngoscopie, une béance glottique postérieure pendant la phonation (fig 2-6).

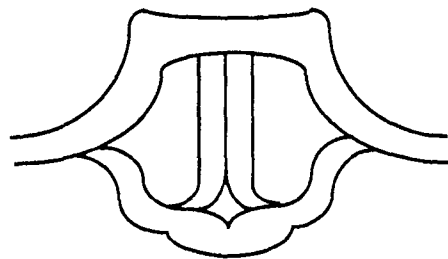


Figure 2-6. Béance glottique postérieure

La mue peut persister au delà de l'âge normal de son apparition et le sujet continue à parler dans le registre de fausset. L'origine d'une telle prolongation est exclusivement psychogène d'ordre caractériel et relationnel.

2-2-4-2 L'aphonie psychogène

D'origine soudaine, en rapport avec un contexte émotionnel ou relationnel, elle se caractérise cliniquement par une voix réduite au chuchotement. Au niveau glottique, elle se manifeste par une dyskinésie majeure qui se présente soit sous la forme d'une absence totale ou partielle d'adduction des cordes vocales, soit sous la forme d'une occlusion de la glotte juste avant la tentative d'émission.

2-3 Conclusions

2-3-1 Ambivalence nosologique des troubles vocaux

Nous avons vu que la plus grande partie des maladies de la voix font intervenir pathogéniquement les deux pôles organique et fonctionnel. Le rapport entre ces deux composantes peut, de manière schématique, revêtir deux aspects :

- le geste phonatoire défectueux d'origine fonctionnelle, une fois passé à la chronicité, entraîne à la longue des modifications anatomiques du vibrateur laryngé;
- une gêne organique, de quelque ordre qu'elle soit, qui vient perturber le geste de l'émission vocale, déclenche de la part de l'appareil phonatoire une tentative de compensation, qui manquant totalement ou partiellement son but de correction, fait entrer le sujet dans le cercle vicieux de l'effort vocal. Nous avons vu aussi que le geste fonctionnellement défectueux résultant, peut persister bien après la régression du trouble vocal.

2-3-2 Vers un système automatique d'aide au diagnostic

A l'issu de ce chapitre, nous pouvons d'ores et déjà émettre quelques réflexions concernant le système que nous nous proposons d'élaborer.

Tout d'abord, vu la complexité et l'ambiguïté des origines des troubles vocaux, l'objectif du système se limite à fournir des paramètres "objectifs" pouvant consolider les sensations de l'oreille bien entraînée du praticien, sans prétendre le remplacer (le diagnostic est du seul ressort du praticien). Par paramètre objectif, nous entendons une mesure quantitative déduite directement ou indirectement du signal de parole. Nous chercherons donc à établir une correspondance entre le système physiologique et un "modèle mathématique", pouvant caractériser de façon univoque toute modification physiologique due à un trouble organique ou fonctionnel.

Enfin, vu l'incidence directe ou indirecte de la majorité des pathologies sur le mode vibratoire du larynx, nous concentrerons une grande partie de notre effort sur la détection non

invasive de l'activité glottique en vue d'en extraire des indices décrivant au mieux le mode vibratoire.

PARTIE II

3- THEORIE ACOUSTIQUE DE LA PRODUCTION DE LA PAROLE

THEORIE ACOUSTIQUE DE LA PRODUCTION DE LA PAROLE

Le premier moyen de la communication humaine est la parole; profitant de sa vie dans l'atmosphère, l'homme a appris à communiquer en utilisant les possibilités oscillatoires de l'air. Au niveau acoustique, les signaux de parole consistent donc en une rapide fluctuation de la pression de l'air, générée et émise par l'appareil phonatoire.

Dans ce chapitre, nous commençons par présenter les lois physiques régissant d'une part une telle génération au niveau de l'appareil vocal, et d'autre part le rayonnement du son vers l'extérieur. Nous décrivons ensuite un modèle simplifié dont les équations permettent l'utilisation d'un modèle "mathématique" numérique en vue de l'analyse et de la synthèse de la parole.

3-1 Eléments de la théorie de la propagation de l'onde sonore dans le système vocal

3-1-1 Introduction

Pour une description quantitative complète de la génération et de la propagation de l'onde sonore dans le système vocal, diverses lois physiques interviennent, telles que :

- les lois fondamentales de conservation de masse, de moment et d'énergie,
- les lois de la thermodynamique,
- les lois de la mécanique des fluides compressibles à faible viscosité (l'air).

L'ensemble de ces lois peut conduire à un système d'équations différentielles décrivant le mouvement de l'air dans le système vocal dont la formulation et la résolution s'avèrent difficiles et ne conduisent pas à des solutions analytiques simples.

En outre, une théorie acoustique détaillée doit tenir compte des facteurs suivants :

- la variation dans le temps de la forme du système vocal (articulation),
- les pertes d'énergie causées par la conduction thermique et par les frottements visqueux,
- l'élasticité des bords latéraux du conduit vocal,
- le couplage nasal,
- le rayonnement de l'onde sonore par les lèvres et les narines,
- le type de l'excitation du conduit vocal.

Une telle étude aussi détaillée n'est pas dans l'optique de ce chapitre et nous renvoyons le lecteur intéressé aux références [Fant, 1970] [Flanagan, 1972]. Nous nous contentons de poser les hypothèses de départ permettant de simplifier l'étude et de décrire certains paramètres acoustiques qu'on utilisera dans la suite.

3-1-2 Hypothèses

- On considère que le système vocal est modélisé par un tube acoustique dont l'aire de la section est non uniforme et variable dans le temps.

- Pour la bande de fréquences inférieures à 4 kHz, jugée suffisante pour la perception (cf. bande téléphonique : 300-3400 Hz), les longueurs d'ondes étant grandes relativement aux dimensions du conduit vocal, on suppose que l'onde sonore est plane et se propage le long de l'axe du tube.

- Les pertes de conduction thermique et des frottements de viscosité sont négligeables.

3-1-3 Equations de propagation

Sous les conditions des hypothèses ci-dessus et partant des lois de conservation de masse, du moment et de l'énergie, Portnoff a montré que l'onde sonore dans le tube satisfait à la paire d'équations

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial \left(\frac{u}{A} \right)}{\partial t} \quad (3-1-a)$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial (pA)}{\partial t} + \frac{\partial A}{\partial t} \quad (3-1-b)$$

où $p = p(x,t)$ est la variation de la pression à un instant t et à la position x ,

$u = u(x,t)$ la vitesse de déplacement d'un volume élémentaire d'air ou vitesse volumique (débit), ρ la densité de l'air, c la vitesse du son dans l'air et $A = A(x,t)$ la surface de la section normale à l'axe du tube.

Nous remarquons la non linéarité du système (3-1) qui empêche toute tentative de résolution de type analytique dans des conditions générales; par contre un calcul numérique des solutions peut être mené à condition d'avoir des connaissances sur :

- les conditions aux limites au niveau des lèvres et de la glotte,
- la nature de l'excitation,
- la fonction d'aire.

Pour cette dernière, on peut supposer, pour des sons stables (voyelles soutenues par exemple), qu'elle ne varie pratiquement pas ou très peu en fonction du temps; cependant la difficulté de mesure de cette fonction subsiste. Les rayons X ont été utilisés comme moyen direct de mesure [Fant, 1970] et [Perkell, 1969], cependant, une méthode directe d'extraction de la fonction d'aire à partir du signal de parole a été proposée par Atal [Atal, 1974].

Dans la suite, nous décrirons des modèles plus simples qui certes ne tiennent pas compte de tous les facteurs intervenant, mais qui possèdent toutefois l'avantage de donner un bon aperçu sur la nature du signal de parole.

3-2 Les modèles physiques de la production sonore basés sur la théorie acoustique

3-2-1 Modèles du tube uniforme

3-2-1-1 Tube uniforme sans perte

Un premier aperçu sur la nature du signal vocal repose sur l'hypothèse simplificatrice qui consiste à supposer que le conduit vocal est un tube à section uniforme pour lequel on néglige toutes les formes de pertes. La figure 3-1-a schématise un tube uniforme excité par une source idéale représentée par un piston causant des perturbations de courant d'air indépendamment de la variation de pression dans le tube (hypothèse de non couplage entre la source et le système).

Dans ces conditions les équations (3-1) deviennent

$$-\frac{\partial p}{\partial x} = \frac{\rho}{A} \frac{\partial u}{\partial t} \quad (3-2-a)$$

$$-\frac{\partial u}{\partial x} = \frac{A}{\rho c^2} \frac{\partial p}{\partial t} \quad (3-2-b)$$

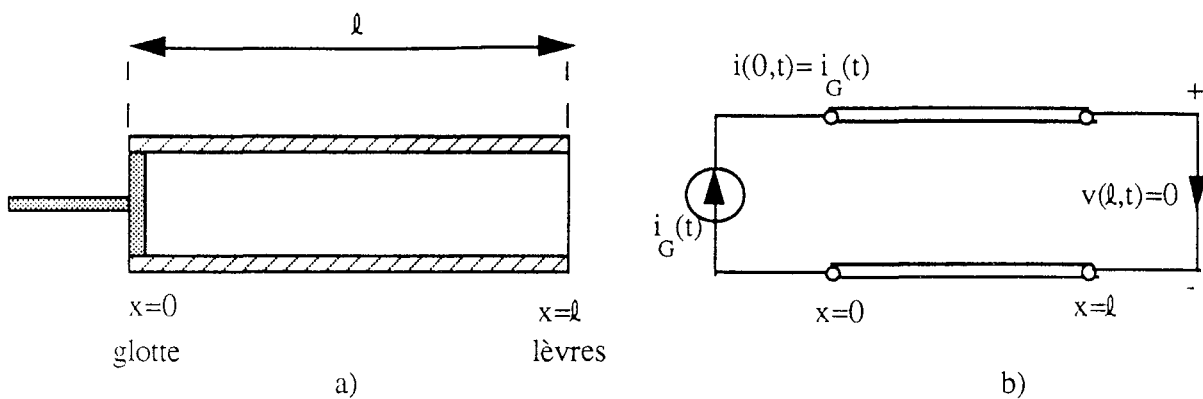


Figure 3-1. a) tube acoustique uniforme idéal
 b) ligne électrique équivalente

Nous constatons que les équations (3-2) ont la même forme que celle des équations de transmission d'une onde électrique dans une ligne sans perte

$$-\frac{\partial v}{\partial x} = L \frac{\partial i}{\partial t} \quad (3-3-a)$$

$$-\frac{\partial i}{\partial x} = C \frac{\partial v}{\partial t} \quad (3-3-b)$$

où L et C sont respectivement l'inductance et la capacitance par unité de longueur de la ligne, v et i sont respectivement la tension et l'intensité le long de la ligne.

Cette constatation nous permet d'établir une analogie acoustico-électrique (tableau 4-1), et de traiter par conséquent le tube acoustique uniforme sans perte comme une ligne électrique idéale.

Les solutions du système (3-2) sont donc une combinaison linéaire de deux ondes, l'une progressive et l'autre régressive. Par analogie avec une ligne de transmission, il est aisé de montrer qu'elles sont de la forme

$$u(x,t) = u^+(t - \frac{x}{c}) - u^-(t + \frac{x}{c}) \quad (3-4-a)$$

$$p(x,t) = \frac{\rho c}{A} [u^+(t - \frac{x}{c}) + u^-(t + \frac{x}{c})] \quad (3-4-b)$$

où $u^+(t - \frac{x}{c})$ et $u^-(t + \frac{x}{c})$ sont les fonctions d'ondes progressives et réfléchies.

Grandeurs acoustiques	Grandeurs électriques
p- pression	v- tension
u- débit	i- courant
ρ/A - inductance acoustique	L- inductance
$A/\rho c^2$ - capacitance acoustique	C- capacité

Tableau 4-1 Analogie acoustico-électrique

Comme pour les lignes de transmission, la représentation fréquentielle du tube acoustique est intéressante puisqu'elle permet d'introduire la notion de **formant** ou de fréquence de résonance du tube. Pour une telle représentation, on suppose que le système est excité par une onde sinusoïdale ou encore que la condition à la limite $x = 0$ est

$$u(0,t) = u_G(t) = U_G(\Omega) e^{j\Omega t} \quad (3-5)$$

Pour la condition à la limite $x = l$, nous utilisons l'hypothèse non réaliste au niveau de la production selon laquelle il n'y a pas de variation de pression au niveau des lèvres; ce qui est analogue à une ligne électrique en court-circuit (figure 3-1-b).

Le système étant linéaire, en considérant la condition limite $p(l,t) = 0$, on montre facilement que les solutions des équations (3-2) sont

$$p(x,t) = j Z_0 \frac{\sin\left(\Omega \frac{l-x}{c}\right)}{\cos\left(\Omega \frac{l}{c}\right)} U_G(\Omega) e^{j\Omega t} \quad (3-6-a)$$

$$u(x,t) = \frac{\cos\left(\Omega \frac{l-x}{c}\right)}{\cos\left(\Omega \frac{l}{c}\right)} U_G(\Omega) e^{j\Omega t} \quad (3-6-b)$$

où l est la longueur du tube et $Z_0 = \frac{\rho C}{A}$ est l'impédance acoustique caractéristique du tube.

Nous remarquons dans les relations (3-6) la séparabilité des fonctions temporelles et spatiales.

En posant $P(x,\Omega) = j Z_0 \frac{\sin\left(\Omega \frac{l-x}{c}\right)}{\cos\left(\Omega \frac{l}{c}\right)} U_G(\Omega)$ et $U(x,\Omega) = \frac{\cos\left(\Omega \frac{l-x}{c}\right)}{\cos\left(\Omega \frac{l}{c}\right)} U_G(\Omega)$, les

relations (3-6) peuvent s'écrire:

$$p(x,t) = P(x,\Omega) e^{j\Omega t} \quad (3-7-a)$$

$$u(x,t) = U(x,\Omega) e^{j\Omega t} \quad (3-7-b)$$

Ainsi, en considérant le débit volumique au niveau de la glotte comme l'entrée du tube, sous forme d'une "sinusoïde pure" de pulsation Ω et d'amplitude $U_G(\Omega)$, et le débit volumique au niveau des lèvres $U(l,\Omega)$ comme sortie, le rapport entrée/sortie est la réponse fréquentielle du tube donnée par

$$V_a(j\Omega) = \frac{U(l,\Omega)}{U_G(\Omega)} = \frac{1}{\cos\left(\Omega \frac{l}{c}\right)} \quad (3-8)$$

La fonction de transfert $V_a(s)$ du tube acoustique uniforme, première approximation du conduit vocal, s'obtient alors en remplaçant Ω par s/j

$$V_a(s) = \frac{2 e^{-\frac{s\ell}{c}}}{1 + e^{-\frac{2s\ell}{c}}} \quad (3-9)$$

Les pôles de $V_a(s)$ sont de nombre infini et se situent sur l'axe imaginaire

$$s_n = \pm j \left(\frac{(2n+1)\pi c}{2\ell} \right) \quad \text{où } n \text{ est un entier} \quad (3-10)$$

où s_n est le $n^{\text{ième}}$ pôle du système.

Ainsi le système est caractérisé par les fréquences de résonance qui ne dépendent que de la longueur ℓ du tube. La figure 3-2 montre la réponse fréquentielle d'un tube acoustique de longueur 17,5 cm, longueur moyenne du conduit vocal humain. Notons la répartition uniforme des formants de 500 Hz à 4500 Hz avec une distance inter-formant constante de 1000 Hz.

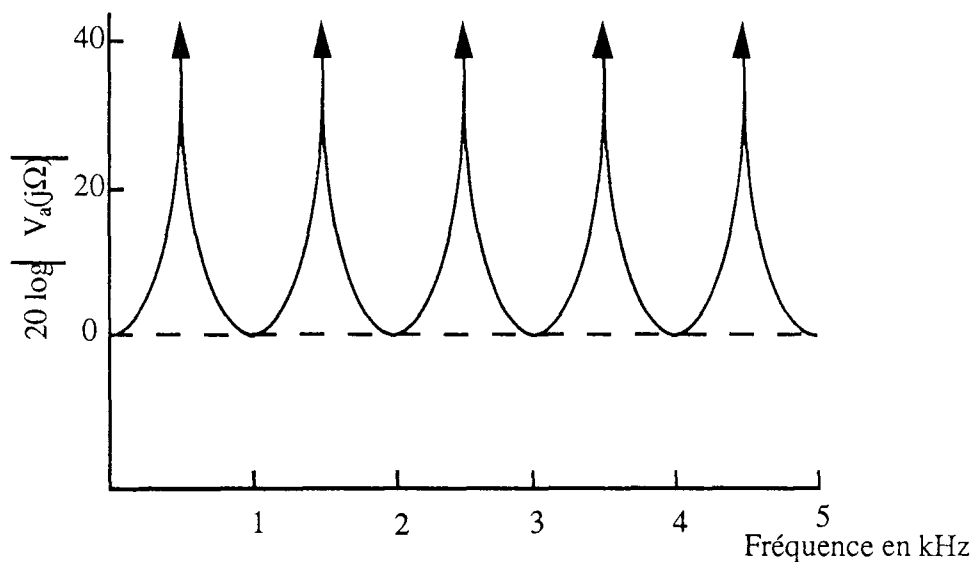


Figure 3-2. Réponse fréquentielle d'un tube acoustique uniforme sans perte, en court-circuit.

En complément de l'analogie du tube acoustique uniforme avec une ligne électrique idéale, nous déterminons l'impédance et l'admittance acoustique par unité de longueur. En substituant les expressions (3-7) dans les équations (3-1) on obtient les équations de variation des modules des ondes sinusoïdales dans le tube

$$-\frac{dP}{dx} = j \Omega \frac{\rho}{A} U \quad (3-11-a)$$

$$-\frac{dU}{dx} = j \Omega \frac{A}{\rho c^2} P \quad (3-11-b)$$

On peut définir l'impédance Z et l'admittance Y par unité de longueur par

$$Z = j \Omega \frac{\rho}{A} \quad (3-12-a)$$

$$Y = j \Omega \frac{A}{\rho c^2} \quad (3-12-b)$$

En conclusion, bien que le modèle du tube acoustique uniforme sans perte soit très schématique, il conduit à la notion des formants et à l'analogie entre un conduit vocal et une ligne de transmission électrique. Nous examinerons dans la suite de ce paragraphe, l'effet des pertes sur la réponse fréquentielle du tube acoustique uniforme, dans le but de mieux l'adapter aux réalités physiques du système de production.

3-2-1-2 Effet des pertes sur un tube acoustique uniforme

Pour l'étude des pertes dans le tube acoustique, nous nous contentons de mentionner leur influence sur la réponse fréquentielle d'un tube acoustique uniforme, en décrivant les modifications des équations (3-11) et (3-12) dues aux différentes pertes.

a) Pertes de vibration du bord interne du tube

En tenant compte de l'équation de mouvement du bord libre du tube, Flanagan [Flanagan, 1972] a montré que les équations (3-11) et (3-12) deviennent

$$-\frac{dP}{dx} = Z U \quad (3-13-a)$$

$$-\frac{dU}{dx} = (Y + Y_w) P \quad (3-13-b)$$

avec

$$Y_w = \frac{1}{j\Omega m_w + b_w + \frac{k_w}{j\Omega}} \quad (3-14)$$

où m_w est la masse unitaire du tube, b_w et k_w sont les coefficients respectivement d'amortissement et de raideur du bord du tube. Notons que les vibrations du bord libre du tube ont pour effet d'introduire une admittance supplémentaire Y_w .

La figure (3-3) montre la réponse fréquentielle du tube acoustique uniforme, tenant compte des pertes de vibrations du bord libre du tube, où les différents coefficients (masse, raideur et amortissement) sont déterminés expérimentalement et avec la condition $p(l,t) = 0$ à la limite des lèvres.

En comparant les figures (3-2) et (3-3), nous pouvons mentionner quelques effets des pertes de vibration :

- les résonances ne sont plus d'amplitude infinie, les pôles associés n'étant plus situés sur l'axe imaginaire,
- les largeurs de bande ne sont plus nulles,
- les formants sont légèrement décalés vers les hautes fréquences par rapport à ceux du modèle du tube sans perte,
- à cause de l'inertie massive de la membrane du tube, l'amplitude de ses vibrations est plus grande pour les basses fréquences que pour les hautes fréquences. De ce fait, l'effet de perte qui en résulte, est plus prononcé au voisinage des premiers formants, ce qui se traduit au niveau spectral par une amplitude plus faible des premières résonances.

b) Pertes de viscosité et de conduction thermique

Flanagan [Flanagan, 1972] a montré que l'effet de friction visqueuse se traduit par l'introduction d'un terme réel à l'impédance acoustique Z , qui devient

$$Z = \frac{S}{A^2} \sqrt{\frac{\Omega \rho \mu}{2}} + j\Omega \frac{\rho}{A_0} \quad (3-15)$$

où S est la circonférence du tube, μ le coefficient de viscosité et ρ la densité de l'air dans le tube.

Pour les pertes dues à la dissipation thermique, d'après Flanagan [Flanagan, 1972], l'admittance acoustique Y contient un terme réel supplémentaire facteur des caractéristiques thermodynamiques de l'air dans le tube

$$Y = \frac{S (\eta - 1)}{\rho c^2} \sqrt{\frac{\lambda \Omega}{2\rho c_p}} + j\Omega \frac{A_0}{\rho c^2} \quad (3-16)$$

où c_p est la chaleur spécifique à pression constante, η le rapport des chaleurs spécifiques à pression et volume constants et λ le coefficient de propagation de chaleur dans le tube.

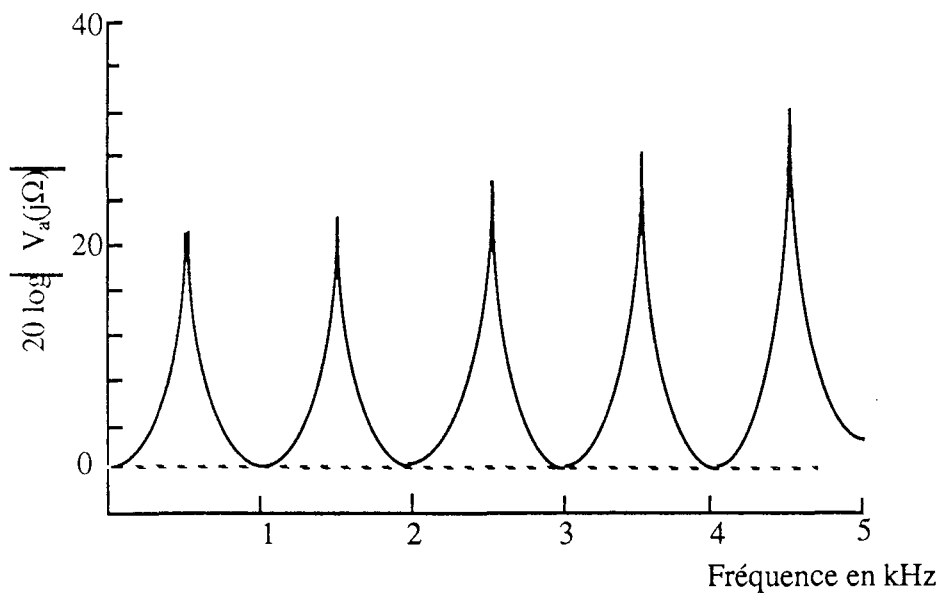


Figure 3-3 Réponse fréquentielle d'un tube uniforme, en court circuit, tenant compte des pertes de vibration du bord libre du tube [Portnoff, 1973]

Nous pouvons remarquer en première approximation les faits suivants :

- les pertes de viscosité sont proportionnelles à la partie réelle de l'impédance et à la racine carrée de la fréquence,
- les pertes thermiques sont proportionnelles à la partie réelle de l'admittance Y et à la racine carrée de la fréquence.

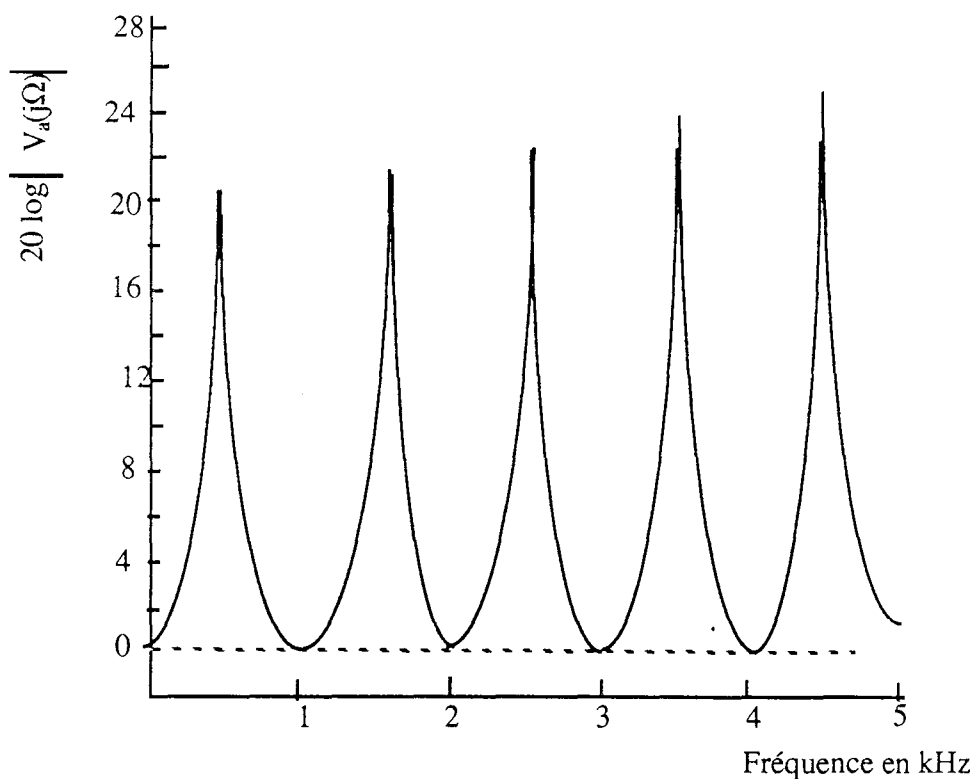


Figure 3-4. Réponse fréquentielle d'un tube uniforme, en court-circuit, tenant compte des pertes de viscosité et de conduction thermique. [Portnoff, 1973]

La figure 3-4 représente la réponse fréquentielle d'un tube uniforme de longueur 17,5 cm, tenant compte de toutes les pertes. En comparant les figures 3-3 et 3-4, nous observons que les pertes de viscosité et thermiques diminuent légèrement les fréquences centrales et augmentent plus les largeurs de bande des derniers formants.

En résumé, les pertes thermiques et de viscosité augmentent avec la fréquence et ont plus d'influence dans les hautes fréquences; tandis que les pertes dues aux vibrations du bord libre sont plus accentuées aux basses fréquences. Ces vibrations tendent à augmenter légèrement les fréquences de résonance, tandis que les autres pertes tendent à les faire diminuer. Dans la bande au-dessous de 4 kHz, l'effet des pertes thermiques et de viscosité est négligeable devant celui des vibrations; ainsi un modèle relativement correct du conduit vocal est un tube uniforme pour

lequel les pertes considérées sont les pertes de vibration du bord interne du tube, les pertes thermiques et de viscosité étant négligées.

Dans tout ce qui précède, le tube acoustique représentant le conduit vocal est considéré comme en court-circuit, c'est-à-dire qu'au niveau des lèvres il n'y a pas de variation de la pression à l'émission sonore. Dans le paragraphe suivant nous présentons l'effet de rayonnement de l'onde sonore à partir des lèvres.

3-2-1-3 Le rayonnement de l'onde sonore par les lèvres

Nous avons examiné les différentes sources de pertes internes dans le tube acoustique, en supposant la terminaison de celui-ci en court-circuit. En fait, une telle hypothèse provient plus de l'analogie acoustico-électrique, et est difficilement réalisable puisqu'elle correspond à une variation du débit volumique au niveau des lèvres ne provoquant pas une variation de la pression. En réalité, le conduit vocal se termine par une ouverture à travers les lèvres (ou les narines pour les sons nasalisés), constituant ainsi une charge acoustique. Nous exposons maintenant l'effet d'une telle charge sur les modes du tube acoustique.

Le modèle le plus fréquemment utilisé pour l'étude du rayonnement à travers les lèvres est celui représenté à la figure 3-5-a, où le conduit vocal est considéré comme un tube se terminant par une petite ouverture (lèvres) dans une enceinte (baffle) sphérique (tête). Pour des basses fréquences, l'ouverture est considérée comme une surface de rayonnement de l'onde sonore diffractée par l'enceinte sphérique. L'étude d'une telle diffraction est relativement compliquée et a fait l'objet de plusieurs travaux [Morse, 1968] [Mrayati, 1976]. Nous mentionnons la charge de rayonnement dans le cadre de l'hypothèse simplificatrice où le rayonnement se fait à travers une ouverture dans une surface plane infinie (figure 3-5-b).

Nous pouvons définir l'impédance acoustique du rayonnement comme le rapport des modules des ondes de pression et de débit volumique; ou encore :

$$P(\ell, \Omega) = Z_L(\ell, \Omega) U(\ell, \Omega) \quad (3-17)$$

Flanagan [Flanagan, 1972] a montré que l'impédance acoustique $Z_L(\ell, \Omega)$ dans le cas d'une ouverture de rayon a dans un baffle plan infini est de la forme

$$Z_L(\ell, \Omega) = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r} \quad (3-18)$$

où la résistance acoustique R_r et l'inductance acoustique L_r sont données par

$$R_r = \frac{128}{9\pi^2} \quad (3-19)$$

$$L_r = \frac{8a}{3\pi c} \quad (3-20)$$

Cette impédance intervient donc dans la nature de propagation de l'onde sonore comme condition aux limites du tube acoustique. Notons qu'au voisinage des très basses fréquences, $Z(l, \Omega)$ est voisin de zéro ($Z(l, \Omega) \approx 0$), le tube acoustique est donc en court-circuit pour les basses fréquences; pour des bandes moyennes de fréquences (telles que $\Omega L_r \ll R_r$), la charge de rayonnement se comporte comme un dérivateur ($Z(l, \Omega) \approx j\Omega L_r$) et pour des hautes fréquences l'impédance est voisine de R_r .

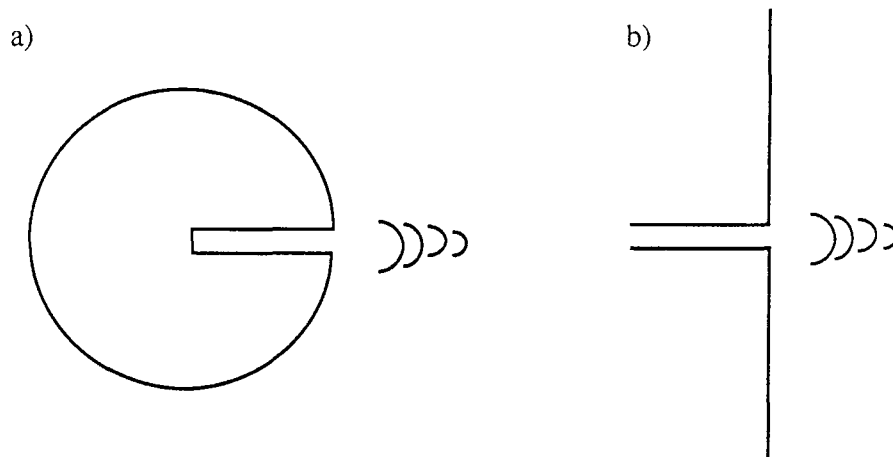


Figure 3-5. Modèles de rayonnement sonore :

a) baffie sphérique b) plan infini

En tenant compte des différentes pertes dans un tube acoustique uniforme, les équations (3-13) à (3-19) peuvent être traitées numériquement pour déterminer la réponse fréquentielle du tube acoustique uniforme. La figure 3-6 montre les réponses fréquentielles

$$20 \log \left(\frac{U(l, \Omega)}{U_G(\Omega)} \right) \text{ (fig. 3-6-a) et } 20 \log \left(\frac{P(l, \Omega)}{U_G(\Omega)} \right) \text{ (fig. 3-6-b).}$$

En comparant les figures 3-4 à 3-6, nous pouvons déduire que l'effet majeur dû à l'introduction des pertes de rayonnement est d'une part l'élargissement des bandes des résonances et la diminution de leur fréquence centrale. Nous observons en plus que l'affectation des résonances par les pertes de rayonnement est liée à la fréquence (l'impédance acoustique de charge étant dépendante de la fréquence); en effet, les plus hauts formants sont atténués et possèdent une bande plus large, tandis que les premiers formants sont plus affectés par les pertes de vibration du bord libre du tube.

En considérant la réponse fréquentielle du tube acoustique comme étant le rapport de la pression acoustique au niveau des lèvres par le débit volumique glottique (fig 3-6-b), nous observons que l'effet le plus remarquable est que les pertes de rayonnement introduisent un "zéro" à la fréquence nulle, et ont un effet similaire à celui d'un filtre de pré-emphase pour les hautes fréquences (accentuation des formants de haute fréquence).

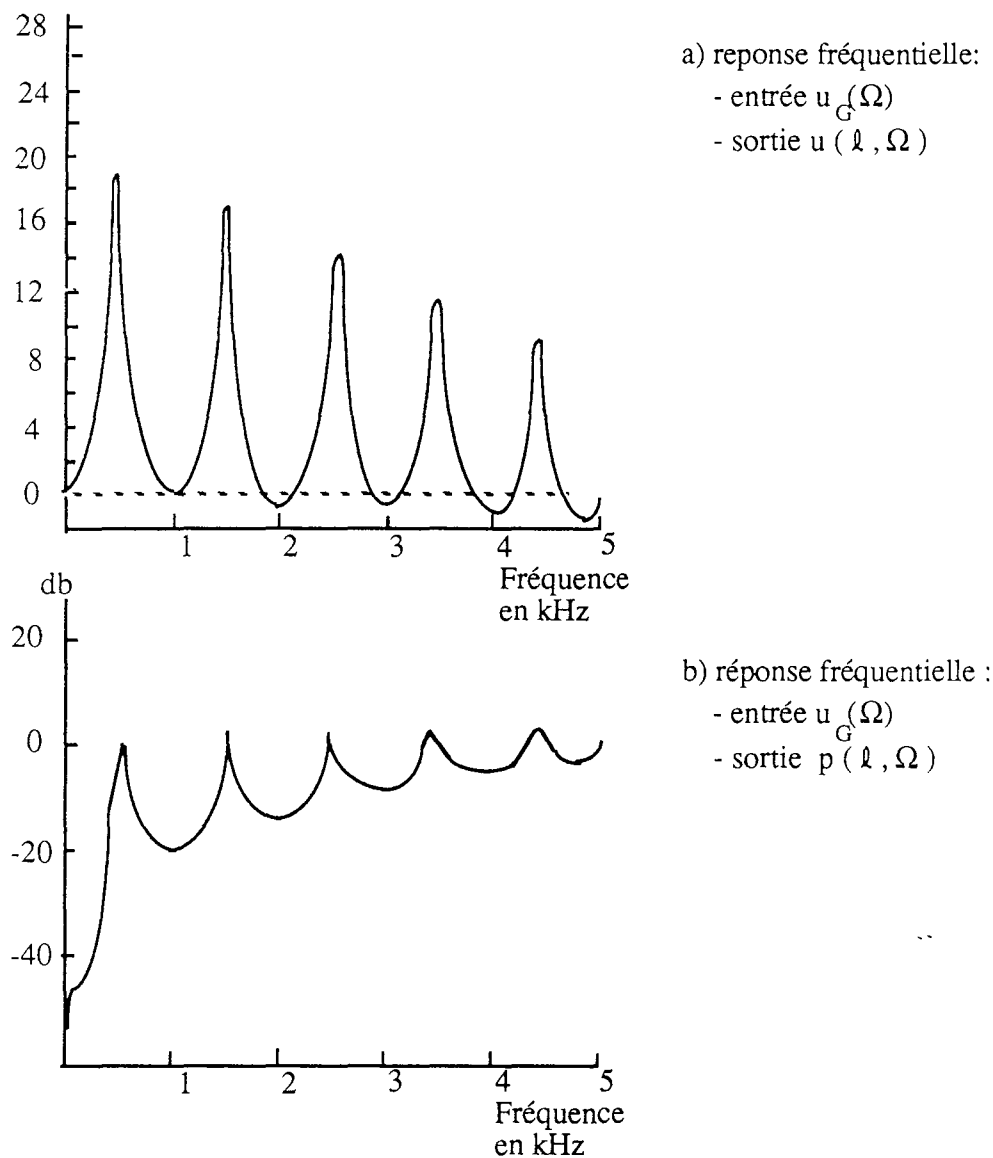


Figure 3-6. Réponses fréquentielles d'un tube acoustique uniforme en tenant compte des pertes de vibration de viscosité, pertes thermiques et pertes de rayonnement [Portnoff, 1973]

3-2-1-4 La production des sons nasalisés. Couplage naso-buccal

Dans la production des sons nasalisés, le conduit nasal est couplé avec le larynx, par l'abaissement du voile du palais. Pour les consonnes nasalisées telles que /m/, le conduit buccal peut se mettre en position de fermeture complète et le son est émis par les narines, par contre pour les voyelles nasalisées (/ẽ/, /ã/...) le rayonnement se fait principalement par la

bouche malgré l'abaissement du voile du palais. Dans un souci de simplicité, nous décrivons d'une manière approximative l'effet du couplage des deux cavités orale et nasale, en prenant les modèles des tubes acoustiques uniformes dont le schéma équivalent est donné à la figure 3-7.

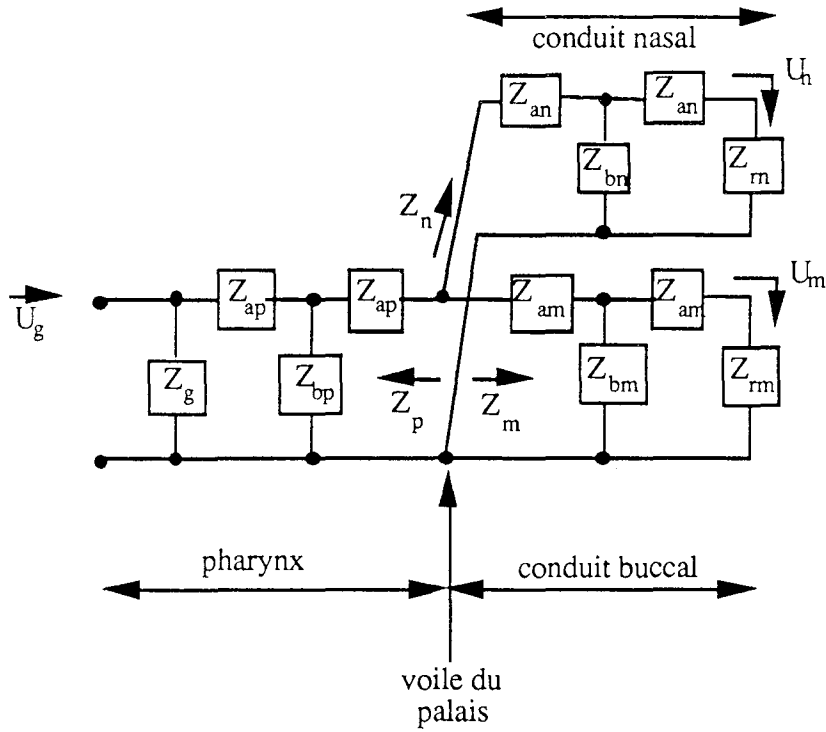


Figure 3-7. Ligne de transmission équivalente pour la production nasalisée.

Dans chaque branche de longueur l , les impédances acoustiques z_a et z_b sont données par

$$z_a = Z_0 \operatorname{th}\left(\frac{\gamma l}{2}\right) \tag{3-21}$$

$$z_b = Z_0 \frac{1}{\operatorname{sh}(\gamma l)} \tag{3-22}$$

où
$$\gamma = \alpha + j\beta = \sqrt{(R_a + j\omega L_a) + (G_a + j\omega C_a)} \tag{3-23}$$

R_a , G_a , L_a et C_a sont les caractéristiques acoustiques d'un tube uniforme

et
$$Z_0 = \sqrt{\frac{R_a + j\omega L_a}{G_a + j\omega C_a}}$$
 son impédance caractéristique.

Dans le cas d'un tube uniforme sans perte, γ et Z_0 deviennent

$$\gamma = j\beta = j\omega\sqrt{L_a C_a} = j\frac{\omega}{c} \tag{3-24}$$

et

$$Z_0 = \sqrt{\frac{L_a}{C_a}} = \frac{\rho c}{A} \tag{3-25}$$

Du fait de la mise en parallèle des conduits oral et nasal, dans la réponse fréquentielle du système apparaissent aussi bien des pôles que des zéros. Les zéros s'expliquent par le fait qu'à certaines fréquences l'impédance d'entrée d'une des deux branches est nulle, de telle manière que le flux aérien y passe complètement empêchant ainsi toute émission au niveau de l'autre branche (figure 3-8).

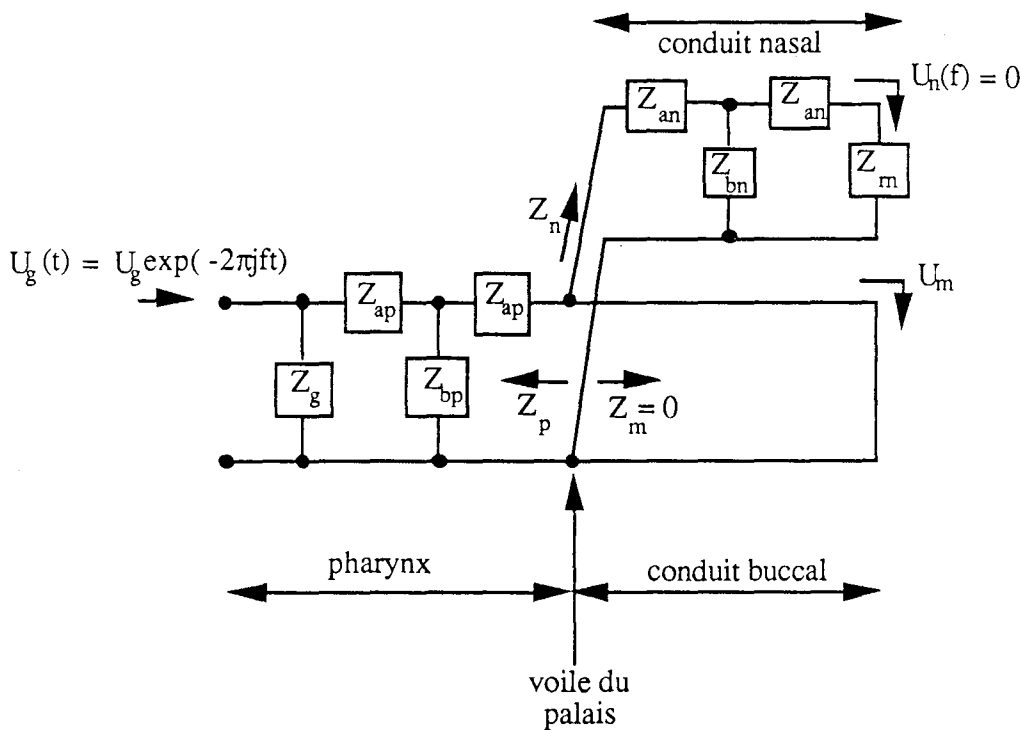


Figure 3-8. Schéma équivalent à la fréquence f correspondant à un zéro dans le cas d'une émission au niveau des narines.

Pour donner un aperçu sur la localisation des fréquences des pôles et des zéros, nous étudions le système donné par Flanagan [Flanagan, 1972], correspondant à une première approximation du système de production de la consonne /m/ (figure 3-9). Nous utilisons les hypothèses suivantes :

- les tubes correspondants à chaque partie du système de production sont uniformes et sans perte,

- l'impédance glottique est très grande ($Z_g \approx \infty$),
- les lèvres sont complètement fermées ($Z_{lm} \approx \infty$),
- l'impédance de rayonnement au niveau des narines Z_{rn} est nulle.

Rappelons tout d'abord que l'impédance d'entrée Z_t d'un tube acoustique uniforme, sans perte, de longueur ℓ , rayonnant dans une charge Z_r est donnée par

$$Z_t = Z_0 \frac{Z_r \cos(\beta\ell) + Z_0 j \sin(\beta\ell)}{Z_0 \cos(\beta\ell) + Z_r j \sin(\beta\ell)} \quad (3-26)$$

Ainsi les impédances d'entrée de chaque tube, indiquées à la figure 3-9-b sont

$$Z_p = -Z_{0p} \cotg(\beta\ell_p) = -\frac{\rho c}{A_p} j \cotg(\beta\ell_p) \quad (3-27)$$

$$Z_m = -Z_{0m} \cotg(\beta\ell_m) = -\frac{\rho c}{A_m} j \cotg(\beta\ell_m) \quad (3-28)$$

$$Z_n = Z_{0n} \tg(\beta\ell_n) = \frac{\rho c}{A_n} j \tg(\beta\ell_n) \quad (3-29)$$

L'admittance d'entrée équivalente au niveau du point de branchement (voile du palais) est donc

$$\begin{aligned} Y_e &= \frac{1}{Z_p} + \frac{1}{Z_m} + \frac{1}{Z_n} \\ &= \frac{j}{\rho c} \left(A_p \tg(\beta\ell_p) + A_m \tg(\beta\ell_m) - A_n \cotg(\beta\ell_n) \right) \end{aligned} \quad (3-30)$$

Ainsi, nous pouvons déterminer une fréquence de résonance f_r (correspondant à un pôle) pour laquelle l'admittance Y_e est nulle; elle satisfait donc à l'équation

$$A_p \tg\left(\frac{2\pi f_r}{c} \ell_p\right) + A_m \tg\left(\frac{2\pi f_r}{c} \ell_m\right) - A_n \cotg\left(\frac{2\pi f_r}{c} \ell_n\right) = 0 \quad (3-31)$$

Les fréquences de résonance pour la configuration de la figure (3-9-a) sont illustrées à la figure (3-10).

Les fréquences d'anti-résonance f_z (ou zéro spectral) sont telles que l'impédance d'entrée du système buccal est nulle; elles satisfont à l'équation suivante :

$$Z_m = - \frac{\rho c}{A_m} j \cotg\left(\frac{2\pi f_z}{c} \ell_m\right) = 0 \tag{3-32}$$

soit

$$f_z = (2n+1) \frac{c}{4\ell_m} \quad \text{avec } n \in \mathbb{N} \tag{3-33}$$

La figure 3-10 illustre la localisation des fréquences de résonance pour la configuration de la figure 3-9-a.

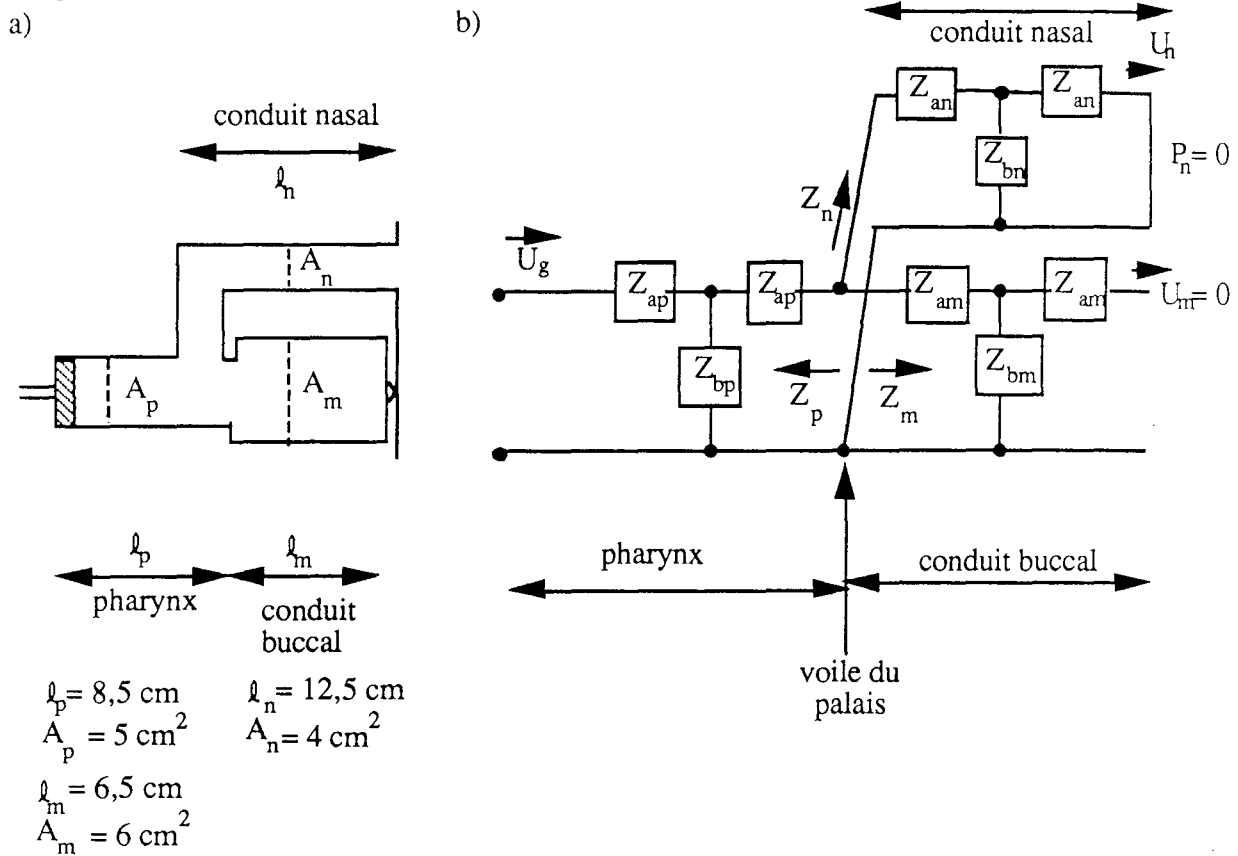


Figure 3-9. a) Approximation du système de production du son nasalisé /m/ par des tubes acoustiques uniformes [Flanagan, 1972]

b) Ligne de transmission équivalente

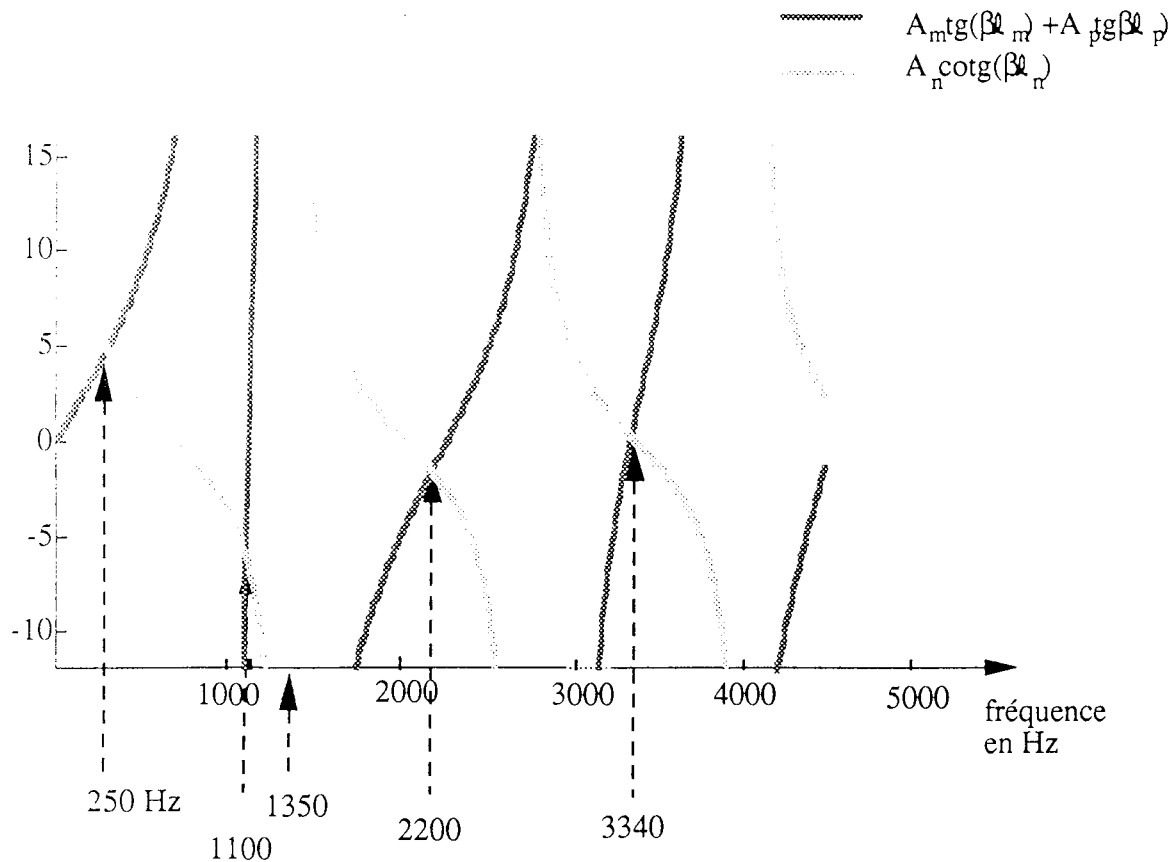


Figure 3-10. Localisation des formants de la consonne nasalisée /m/ avec la configuration de la figure 3-9-a [Flanagan, 1972].

Nous voyons à travers cet exemple que, bien que l'approximation du système vocal par des tubes uniformes soit simple, l'extraction directe des formants pour les sons nasalisés nécessite la résolution de l'équation non linéaire (3-31) pour laquelle il est nécessaire d'avoir des mesures de la fonction d'aire. Néanmoins, la résolution graphique de la figure 3-10 permet de donner une idée sur les formants de la consonne nasalisée /m/ confirmée par des analyses spectrales d'un segment stable d'un signal réel de parole correspondant au phonème /m/ [Fant, 1970].

La localisation des zéros pour un phonème nasalisé s'obtient directement par la relation (3-33) ou son équivalente pour la cavité nasale quand elle joue le rôle de circuit "bouchon". Cependant, la relation (3-33) est obtenue en supposant que l'impédance de charge Z_{rm} (ou Z_{rn} pour la cavité nasale) est nulle. En dehors de cette hypothèse, d'après l'expression (3-26) nous pouvons déterminer les fréquences f_z des zéros en remplaçant la relation (3-33) par

$$f_z = n \frac{c}{2l_m} + \delta \frac{c}{4l_m} \quad \text{avec } n \in \mathbb{N} \quad (3-34-a)$$

$$f_z = n \frac{c}{2l_n} + \delta \frac{c}{4l_n} \quad \text{où } \delta = 0 \text{ ou } 1 \quad (3-34-b)$$

Ainsi, pour un phonème nasalisé, nous pouvons, grâce à la relation (3-34), déterminer le nombre de zéros contenus dans le spectre d'un segment stable du signal, à condition de disposer des mesures des longueurs des cavités buccale et nasale.

Les largeurs de bande de chaque formant, relatif à chaque zéro, n'ont pas été mentionnées ci-dessus, puisque les équations (3-32) et (3-33 ou 4-34) sont valables sous l'hypothèse de l'approximation par des tubes sans perte. Nous avons étudié plus haut les modifications spectrales dues aux différentes sources de perte. Vu que l'aire d'une section de la cavité nasale est relativement faible, on doit s'attendre à des pertes additionnelles de viscosité et de conduction thermique. Elles ont pour effet de diminuer l'amplitude et d'augmenter les largeurs de bande des résonances de haute fréquence d'un phonème nasalisé.

En conclusion, la production nasalisée est caractérisée d'une part, par l'apparition d'anti-résonances du fait du couplage des cavités nasale et buccale, et par la diminution des amplitudes des résonances de haute fréquence due aux pertes additionnelles introduites par la forme géométrique de la cavité nasale.

3-2-1-5 Nature de l'excitation et effet de couplage source-conduit

Pour compléter le modèle acoustique, une étude des mécanismes de génération de l'onde sonore dans le conduit vocal s'impose. Dans ce paragraphe, nous nous contentons d'une présentation qualitative de ces mécanismes, que nous allons détailler plus loin, dans le but de modéliser le système glottique.

Dans les paragraphes précédents, nous avons supposé que l'excitation est modélisée par un générateur idéal de débit volumique non couplé avec le système de production. Nous nous proposons ici de revoir une telle hypothèse et d'examiner le couplage entre la source et le conduit vocal. Rappelons d'abord que des études acoustico-phonétiques ont permis de classer la production sonore en fonction de la nature et la localisation du signal d'excitation [Markel, 1976] :

- les sons voisés pour lesquels l'excitation est quasi-périodique provenant de la modulation du flux d'air par la vibration des cordes vocales,

- les sons fricatifs résultant d'une excitation de nature aléatoire proche d'un bruit blanc; elle s'explique par la naissance d'une turbulence du flux d'air, lorsqu'il passe à travers une constriction localisée en un point du conduit vocal,

- les explosifs pour lesquels l'excitation est de nature transitoire due à un changement brusque de la pression dans le conduit vocal; ce changement est dû à une libération instantanée du flux d'air après le relâchement d'une fermeture créée par un organe articulateur.

Des études sur le processus de génération des sons voisés et non voisés, basées sur la mécanique classique et la mécanique des fluides [Flanagan-Landgraf, 1968] [Flanagan, 1972] [Flanagan-Ishizaka, 1975] ont permis de fournir un modèle mathématique de la source. Sans entrer dans les détails, nous nous intéressons particulièrement à l'excitation dans le cas de la production d'un son voisé (non fricatif et non explosif). Le mécanisme de vibration des cordes vocales est illustré à la figure 3-11 dans laquelle la glotte est considérée comme une petite ouverture s'insérant entre le système sous-glottique et le conduit vocal. Le processus de vibration passe par deux phases principales. En première phase, la pression des poumons augmente, un flux d'air a tendance à passer des poumons vers le conduit vocal à travers la petite ouverture. D'après la loi de Bernoulli, la pression dans la glotte est nettement inférieure à celle des pressions de part et d'autre de l'orifice. Pour une certaine tension des muscles vocaux, la pression à l'intérieur de l'orifice étant faible, les cordes vocales sont amenées l'une contre l'autre créant une fermeture complète de la glotte. Il en résulte une augmentation de la pression sous-glottique, amorçant la deuxième phase, qui, atteignant une certaine limite, force le passage vers le conduit vocal et entraîne un éloignement des cordes vocales. Le processus recommence par la baisse de la pression glottique. Nous pouvons donc considérer un premier modèle de la source comme un oscillateur entretenu par la pression sous-glottique dont la charge est le conduit vocal (figure 3-12-a) et dont les principaux facteurs sont les pressions en amont et en aval de la glotte, l'aire au repos de la glotte, l'aire de contact des cordes vocales. Il s'avère que la fonction de l'onde glottique ($u(0,t)$) est solution d'un système d'équations non linéaires [Ishizaka-Flanagan, 1972]. Le couplage des équations différentielles non linéaires d'oscillation des cordes vocales avec celles de la transmission dans le conduit, peut être représenté par une impédance acoustique variable Z_G fonction de l'inverse de la fonction d'aire glottique $A_G(t)$ [Ishizaka-Flanagan, 1972]; ce qui explique la nature impulsionnelle du débit volumique se propageant dans le conduit vocal.

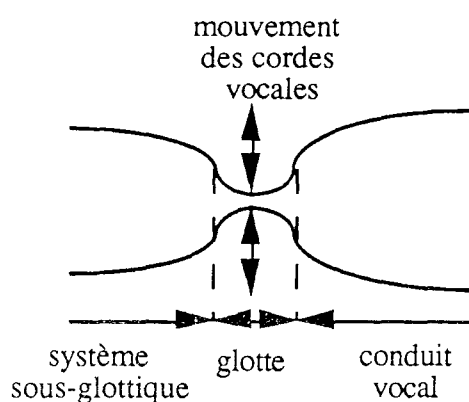
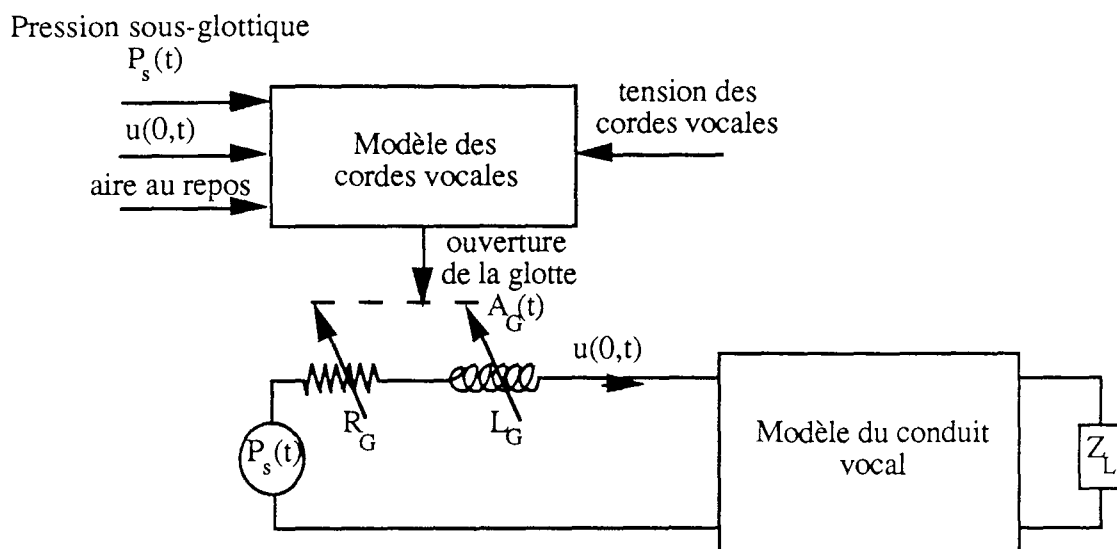


Figure 3-11. Représentation schématique de la glotte dans le système de production de la parole voisée

a)



b)

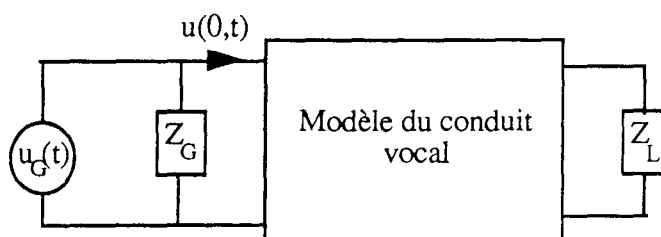


Figure 3-12. a- Diagramme du modèle des cordes vocales
b- Modèle de la source non couplée avec le conduit vocal

Pour beaucoup d'applications, telle que la synthèse de la parole, le couplage entre la source et le conduit vocal est négligé et le système glottique est modélisé par une source de débit volumique ayant une impédance interne $Z_G = R_G + j\omega L_G$ avec R_G et L_G constants (figure 3-12-b).

L'impédance de la source a un effet d'élargissement des bandes de résonance du conduit vocal. Elle élargit plus la première résonance, puisque pour les hautes fréquences, la source se rapproche plus d'une source idéale, l'impédance étant voisine d'un circuit ouvert.

3-2-2 Modèles à plusieurs tubes acoustiques

Dans le paragraphe précédent, nous avons pris le modèle simple d'un tube uniforme pour lequel nous avons examiné l'effet des différentes formes de pertes. Dans cette section, nous allons exposer un modèle couramment utilisé en traitement de la parole qui consiste à considérer le conduit vocal comme une succession de tubes uniformes dont les sections approuchent au mieux la fonction d'aire pendant la phonation.

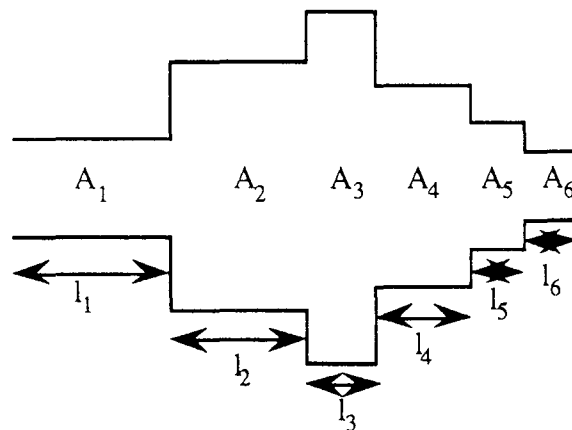


Figure 3-13. Modèle de tubes acoustiques concaténés approchant la fonction d'aire du conduit vocal

3-2-2-1 Equations de propagation d'une onde dans deux tubes sans perte

Chaque tube étant supposé sans perte, la propagation de l'onde acoustique est décrite par les équations (3-4) avec les valeurs appropriées des sections. Ainsi, si on considère le $k^{\text{ième}}$ tube de section A_k , la pression et le débit volumique vérifient les équations

$$p_k(x,t) = \frac{\rho c}{A_k} \left(u_k^+(t - \frac{x}{c}) + u_k^-(t + \frac{x}{c}) \right) \quad (3-35-a)$$

$$u_k(x,t) = u_k^+(t - \frac{x}{c}) - u_k^-(t + \frac{x}{c}) \quad (3-35-b)$$

où x est la distance au bord gauche du tube de longueur l_k , u_k^+ et u_k^- sont respectivement les ondes progressives et régressives.

A la jonction du $k^{\text{ième}}$ tube et du $(k+1)^{\text{ième}}$ tube, nous pouvons écrire les relations de continuité

$$p_k(l_k, t) = p_{k+1}(0, t) \quad (3-36-a)$$

$$u_k(l_k, t) = u_{k+1}(0, t) \quad (3-36-b)$$

qui donnent par substitution dans (3-35)

$$\frac{A_{k+1}}{A_k} \left(u_k^+(t - \tau_k) + u_k^-(t + \tau_k) \right) = u_{k+1}^+(t) + u_{k+1}^-(t) \quad (3-37-a)$$

$$u_k^+(t - \tau_k) - u_k^-(t + \tau_k) = u_{k+1}^+(t) - u_{k+1}^-(t) \quad (3-37-b)$$

avec $\tau_k = \frac{l_k}{c}$ le temps de propagation de l'onde dans le $k^{\text{ième}}$ tube.

Les équations (3-37) font apparaître une réflexion au niveau de la jonction entre les deux tubes. En résolvant l'équation (3-37) par rapport aux ondes incidentes traversant le tube ($u_{k+1}^+(t)$ et $u_{k+1}^-(t + \tau_k)$) nous pouvons obtenir les relations suivantes :

$$u_{k+1}^+(t) = (1 + r_k) u_k^+(t - \tau_k) + r_k u_k^-(t) \quad (3-38-a)$$

$$u_{k+1}^-(t + \tau_k) = -r_k u_k^+(t - \tau_k) + (1 - r_k) u_k^-(t) \quad (3-38-b)$$

avec

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \quad (3-39)$$

le coefficient de réflexion de la jonction des deux tubes. Une représentation en diagramme de fluence [Kelly-Lochbaum, 1962] [Oppenheim-Schafer, 1975] traduisant les équations (3-38), est donnée à la figure (3-14).

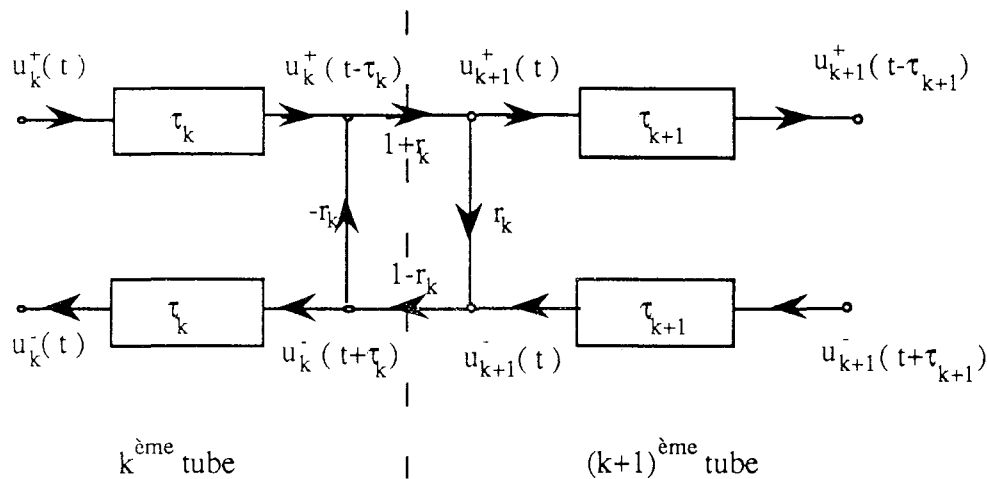


Figure 3-14. Diagramme de fluence de la jonction entre deux tubes acoustiques

Ainsi pour le modèle de tubes acoustiques uniformes, nous pouvons écrire de proche en proche l'équation 3-38 traduisant alors la propagation des ondes progressives et régressives dans les tubes, si on dispose des conditions aux limites de part et d'autre du conduit vocal.

3-2-2-2 Conditions aux limites

Supposons que le conduit vocal soit modélisé par la concaténation de N tubes uniformes sans perte. Pour compléter le schéma de la figure 3-13, nous devons tenir compte de l'effet de la source pour le premier tube et de l'effet de radiation pour le N ème tube.

En supposant que le couplage entre la source et le conduit vocal soit linéaire, nous pouvons écrire la relation suivante entre la pression et le débit volumique dans le domaine fréquentiel

$$U_1(0, \Omega) = U_G(\Omega) - \frac{1}{Z_G} P_1(0, \Omega) \quad (3-40)$$

où Z_G est l'impédance de la glotte.

Dans le cas où Z_G est supposée réelle, l'équation (3-40) se traduit dans le domaine temporel par

$$u_1^+(t) - u_1^-(t) = u_G(t) - \frac{\rho c}{A_1} \left(\frac{u_1^+(t) + u_1^-(t)}{Z_G} \right) \quad (3-41)$$

où encore si on exprime l'onde incidente $u_1^+(t)$ en fonction de l'onde réfléchie et de l'onde glottique nous obtenons facilement la relation

$$u_1^+(t) = \frac{1+r_G}{2} u_G(t) + r_G u_1^-(t) \quad (3-42)$$

où r_G est le coefficient de réflexion au niveau de la glotte, défini par

$$r_G = \frac{\left(Z_G - \frac{\rho c}{A_1} \right)}{\left(Z_G + \frac{\rho c}{A_1} \right)} \quad (3-43)$$

A partir des relations (3-42) et (3-43), nous pouvons déterminer l'onde incidente à l'entrée du conduit vocal en fonction du débit volumique glottique à condition de connaître l'impédance Z_G de la source. Le schéma de la figure (3-15-a) traduit la relation (3-42).

Pour l'effet de radiation au niveau des lèvres, si la charge Z_L est supposée réelle nous pouvons de même exprimer le débit volumique au niveau des lèvres $u_L(t)$, sortie $u_N(l_N, t)$ du $N^{\text{ième}}$ tube, en fonction des ondes incidentes et réfléchies du bord droit du $k^{\text{ième}}$ tube

$$u_{N+1}^-(t + \tau_N) = -r_L u_{N+1}^+(t - \tau_N) \quad (3-44)$$

$$u_N(l_N, t) = (1+r_L) u_{N+1}^+(t - \tau_N) \quad (3-45)$$

avec r_L coefficient de réflexion aux lèvres défini par

$$r_L = \frac{\left(\frac{\rho c}{A_N} - Z_L \right)}{\left(\frac{\rho c}{A_N} + Z_L \right)} \quad (3-46)$$

Le schéma de la figure (3-15-b) traduit les relations (3-44) et (3-45).

Ainsi, le modèle à N tubes acoustiques uniformes sans perte, représenté par le schéma de la figure 3-13, peut être construit si on dispose des mesures de la fonction d'aire pour une articulation donnée, des impédances de charge et de la source, et du signal débit volumique au niveau de la glotte. Pour ce dernier, de nombreux auteurs se contentent d'une série d'impulsions périodiques pour une production d'un son voisé ou d'un bruit blanc pour une production non voisée. Notons que ce modèle est utilisé surtout en synthèse de son isolé ou soutenu. Pour la synthèse de la parole continue, il est nécessaire d'établir d'une part une correspondance entre tous les phonèmes et la fonction d'aire correspondante, et de générer le signal d'excitation adéquat.

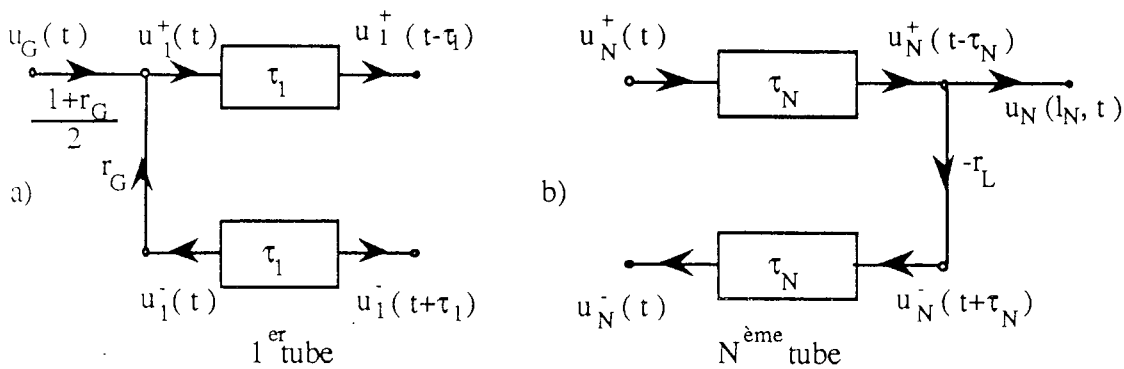


Figure 3-15. a) limite au niveau de la glotte du modèle de tubes uniformes sans perte
 b) limite au niveau des lèvres du modèle de tubes uniformes sans perte

3-2-3 Modèle de tubes acoustiques et notion du filtre numérique du conduit vocal

3-2-3-1 Introduction

Le modèle des tubes acoustiques uniformes sans perte est attractif par le fait qu'on peut le lier avec un filtre numérique. Il a donc comme avantage de faciliter l'analyse ou la synthèse du signal de parole. Cependant une telle liaison repose sur la modification du modèle précédent qui consiste à choisir les N tubes de même longueur $\Delta x = l/N$, l étant la longueur du conduit vocal.

3-2-3-2 Construction du filtre numérique modèle du conduit vocal

Les tubes étant de même longueur, le retard τ_k introduit dans les relations (3-37) à (3-45) est constant et égal à $\Delta x/c$. Il s'ensuit que la réponse impulsionnelle du conduit vocal est de la forme

$$v_a(t) = \alpha_0 \delta(t - N\tau) + \sum_{i=1}^{\infty} \alpha_i \delta(t - N\tau - 2i\tau) \tag{3-47}$$

La réponse fréquentielle du modèle est donc

$$\left\{ \begin{aligned} V_a(\Omega) &= e^{-2j\Omega N\tau} \sum_{k=0}^{\infty} \alpha_k e^{-2j\Omega k\tau} \\ &= e^{-2j\Omega N\tau} \hat{V}_a(\Omega) \end{aligned} \right. \quad (3-48)$$

Le terme $e^{-2j\Omega N\tau}$ est un déphasage constant de la réponse fréquentielle $\hat{V}_a(\Omega)$ (de déphasage nul), qui se traduit par un retard temporel constant de $N\tau$, correspondant au temps de propagation de l'onde dans le tube.

Il est aisé de montrer que la réponse fréquentielle $\hat{V}_a(\Omega)$ est périodique de période $2\pi/2\tau$; ce qui correspond à la sortie d'un filtre numérique dont la réponse impulsionnelle est

$$\left\{ \begin{aligned} \hat{v}_a(nT) &= \alpha_n \quad \text{pour } n \geq 0 \\ \hat{v}_a(nT) &= 0 \quad \text{pour } n < 0 \end{aligned} \right. \quad (3-49)$$

avec $T=2\tau$ la période d'échantillonnage.

Ainsi, si l'entrée du conduit vocal est un signal dont l'énergie est concentrée dans la bande au-dessous de la fréquence $\pi/2\tau$, le signal de parole peut être interprété comme la sortie d'un filtre numérique dont l'entrée est le signal glottique échantillonné à la fréquence $T=2\tau$. Notons que pour tenir compte du retard $N\tau$, la sortie du filtre doit être décalée de $N/2$ échantillons. En pratique, ce décalage est négligé puisque il n'a d'effet que sur le spectre de phase de la réponse fréquentielle du filtre. La figure 3-16 illustre l'analogie entre le modèle des tubes acoustiques uniformes de même longueur avec le filtre numérique.

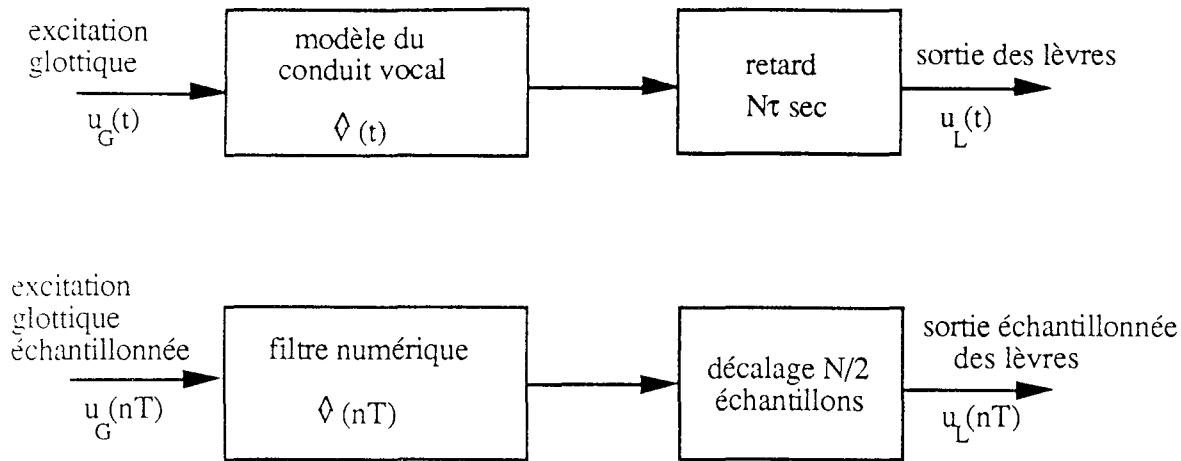


Figure 3-16. Analogie du modèle de tubes acoustiques uniformes de même longueur avec un filtre numérique (la période d'échantillonnage est $T = 2\tau$)

3-2-3-3 Fonction de transfert du modèle numérique du conduit vocal

Considérons l'entrée et la sortie du filtre numérique, comme étant les débits volumiques respectivement avant la glotte et après les lèvres et exprimons la fonction de transfert $V_a(z)$ du conduit vocal en fonction des coefficients de réflexion des tubes acoustiques. Ecrivons la transformée en z des relations (3-38), décrivant la jonction entre le $(k+1)^{ième}$ et le $k^{ième}$ tube

$$U_{k+1}^+(z) = (1 + r_k) z^{-1/2} U_{k+1}^+(z) + r_k z^{-1/2} U_{k+1}^-(z) \quad (3-50-a)$$

$$U_k^-(z) = -r_k z^{-1/2} U_k^+(z) + (1 - r_k) z^{-1/2} U_{k+1}^-(z) \quad (3-50-b)$$

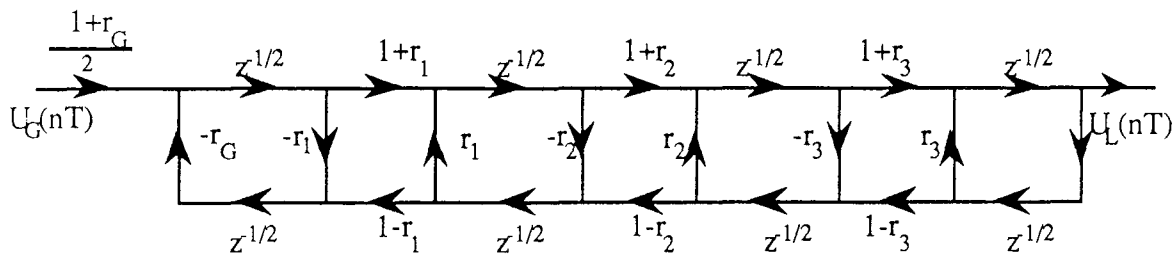


Figure 3-17. Diagramme de fluence du modèle numérique du conduit vocal correspondant au modèle de 3 tubes acoustiques ($N=3$)

La figure 3-17 représente le diagramme de fluence des équations de toutes les jonctions dans le cas d'un modèle de 3 tubes acoustiques ($N=3$).

Nous pouvons écrire les équations (3-50) sous la forme matricielle

$$U_k = M_k U_{k+1} \quad (3-51)$$

avec

$$U_k = \begin{bmatrix} U_k^+(z) \\ U_k^-(z) \end{bmatrix} \quad (3-52)$$

$$M_k = \begin{bmatrix} \frac{z^{1/2}}{1+r_k} & \frac{-r_k z^{1/2}}{1+r_k} \\ \frac{-r_k z^{-1/2}}{1+r_k} & \frac{z^{-1/2}}{1+r_k} \end{bmatrix} \quad (3-53)$$

Ainsi on peut écrire la transformée en z du vecteur d'onde du premier tube en fonction de celui de la sortie du $N^{\text{ième}}$ tube

$$U_1 = \left(\prod_{k=1}^{k=N} M_k \right) U_{N+1} \quad (3-54)$$

La transformée en z du signal glottique s'exprime en fonction du vecteur U_1 en tenant compte de la relation (3-42) sous la forme

$$U_G(z) = \begin{bmatrix} 2 \\ 1+r_G & -2 \\ & 1+r_G \end{bmatrix} U_1 \quad (3-55)$$

Puisque la sortie au niveau des lèvres est égale à l'entrée du $(N+1)^{\text{ième}}$ tube fictif, nous pouvons écrire la relation suivante

$$U_{N+1} = \begin{bmatrix} U_L(z) \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_L(z) \quad (3-56)$$

Ainsi l'inverse de la fonction de transfert du conduit vocal $V_a(z)$ est

$$\frac{1}{V_a(z)} = z^{N/2} \begin{bmatrix} 2 & -2 \\ 1+r_G & 1+r_G \end{bmatrix} \left(\prod_{k=1}^{k=N} \hat{M}_k \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (3-57)$$

où $\hat{M} = z^{-1/2} M$.

Ainsi, nous pouvons montrer que la fonction de transfert $V_a(z)$ peut s'exprimer sous la forme

$$V_a(z) = \frac{0,5(1+r_G) z^{-N/2} \prod_{k=1}^{k=N} (1+r_G)}{D(z)} \quad (3-58)$$

avec $D(z)$ un polynôme en z^{-1} défini par la relation matricielle suivante

$$D(z) = \begin{bmatrix} 1 & -r_G \end{bmatrix} \left(\prod_{k=1}^{k=N} \begin{bmatrix} 1 & -r_k \\ -r_k z^{-1} & z^{-1} \end{bmatrix} \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (3-59)$$

Si on suppose l'impédance de la source infinie ($r_G = 1$), le produit matriciel de la relation (3-59) peut être remplacé par la relation de récurrence suivante [Rabiner, 1978] [Atal-Hanauer, 1971] [Markel, 1976]

$$\begin{cases} D_0(z) = 1 \\ D_k(z) = D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}) & k=1 \dots N \\ D(z) = D_N(z) \end{cases} \quad (3-60)$$

Il est aisé à partir des relations (3-59) et (3-60) de montrer que le dénominateur de la fonction de transfert est un polynôme en z^{-1} . Il s'ensuit la forme générale de la fonction de transfert du conduit vocal

$$V_a(z) = \frac{K z^{-N/2}}{1 - \sum_{k=1}^{k=N} a_k z^{-k}} \quad (3-61)$$

où les N coefficients a_k sont fonction des coefficients de réflexion des tubes. Le choix du nombre N de tubes acoustiques dépend de la période d'échantillonnage T_e . Comme il est communément admis [Rabiner, 1978] que la densité spectrale du signal de parole est concentrée dans la bande 0-5000 Hz, la fréquence d'échantillonnage F_e suffisante pour extraire les

caractéristiques fréquentielles est de 10 kHz. Sachant que la longueur moyenne l du conduit vocal est de 17,5 cm, le modèle numérique avec une fréquence d'échantillonnage F_e correspond à un modèle de $N=10$ tubes acoustiques de même longueur (car $\frac{F_e}{2} = \frac{1}{4\tau} = \frac{Nc}{4l}$).

Le modèle numérique ainsi obtenu caractérise les résonances du conduit vocal par ses pôles. En utilisant ce modèle avec la fréquence d'échantillonnage F_e , le nombre de résonances est limité à 5 dans la bande 0-5 kHz. Nous justifierons plus loin l'utilisation d'un modèle d'ordre plus grand que 10. Nous illustrons ce paragraphe par une simulation du spectre d'une voyelle /a/, partant de l'approximation de la fonction d'aire discrète donnée par Fant [Fant, 1970] (figure 3-18).

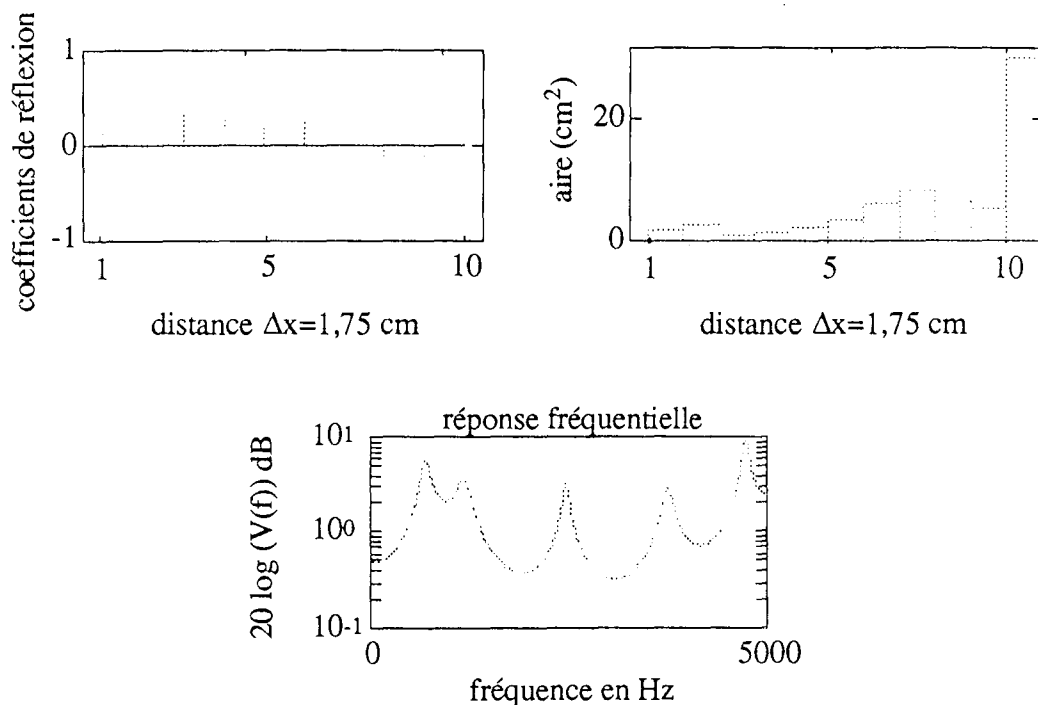


Figure 3-18. Réponse fréquentielle du modèle numérique du conduit vocal pendant la production de la voyelle /a/ [Fant, 1970]

3-3 Modèle numérique du système de production de la parole

Ce paragraphe est la charnière entre la théorie acoustique et les modèles mathématiques d'analyse et de synthèse de la parole. Le modèle numérique, donné à la figure 3-19, proposé et validé par plusieurs équipes de recherche, englobe aussi bien la génération du signal glottique,

le filtrage de ce signal par le conduit vocal (éventuellement par le couplage des conduits oral et nasal) et l'effet de radiation des lèvres.

3-3-1 Modèle d'excitation

Nous avons vu que le signal d'excitation peut être assimilé à une suite périodique d'impulsions dans le cas de la production de sons voisés ou à un bruit blanc pour les autres sons. Dans le premier cas le modèle $G(z)$ correspond à un générateur d'une forme impulsive. Rosenberg [Rosenberg, 1971] a montré que les impulsions glottiques peuvent être remplacées en synthèse sans perte de qualité, par le signal glottique $g(n)$ suivant

$$\begin{cases} g(n) = \frac{1}{2} \left(1 - \cos\left(\frac{\pi n}{N_1}\right) \right) & 0 \leq n \leq N_1 \\ g(n) = \cos\left(\pi \frac{n - N_1}{2N_2}\right) & N_1 \leq n \leq N_1 + N_2 \\ g(n) = 0 & \text{ailleurs} \end{cases} \quad (3-62)$$

Ainsi le modèle glottique est un filtre numérique de réponse impulsionnelle $g(n)$, donnée par (3-62), excité par un train d'impulsions de Dirac de période égale à l'inverse de la fréquence fondamentale.

3-3-2 Le conduit vocal

Nous avons vu que le conduit vocal peut être modélisé par un filtre numérique linéaire tous pôles. Ce modèle est relativement correct pour la production de sons non nasalisés ou non fricatifs, puisqu'il suppose que le conduit vocal n'est caractérisé que par des résonances (pôles). Cependant, il peut inclure des anti-résonances en augmentant suffisamment son ordre [Rabiner, 1978] [Markel, 1975].

3-3-3 Filtre de rayonnement

Nous avons vu que le rayonnement par les lèvres peut être modélisé par une charge Z_L . En étudiant la réponse fréquentielle d'une telle charge, nous voyons qu'elle agit comme un

passé-haut. En première approximation, l'effet de radiation peut être modélisé par un dérivateur dont la fonction de transfert est

$$R(z) = R_0 (1 - z) \tag{3-63}$$

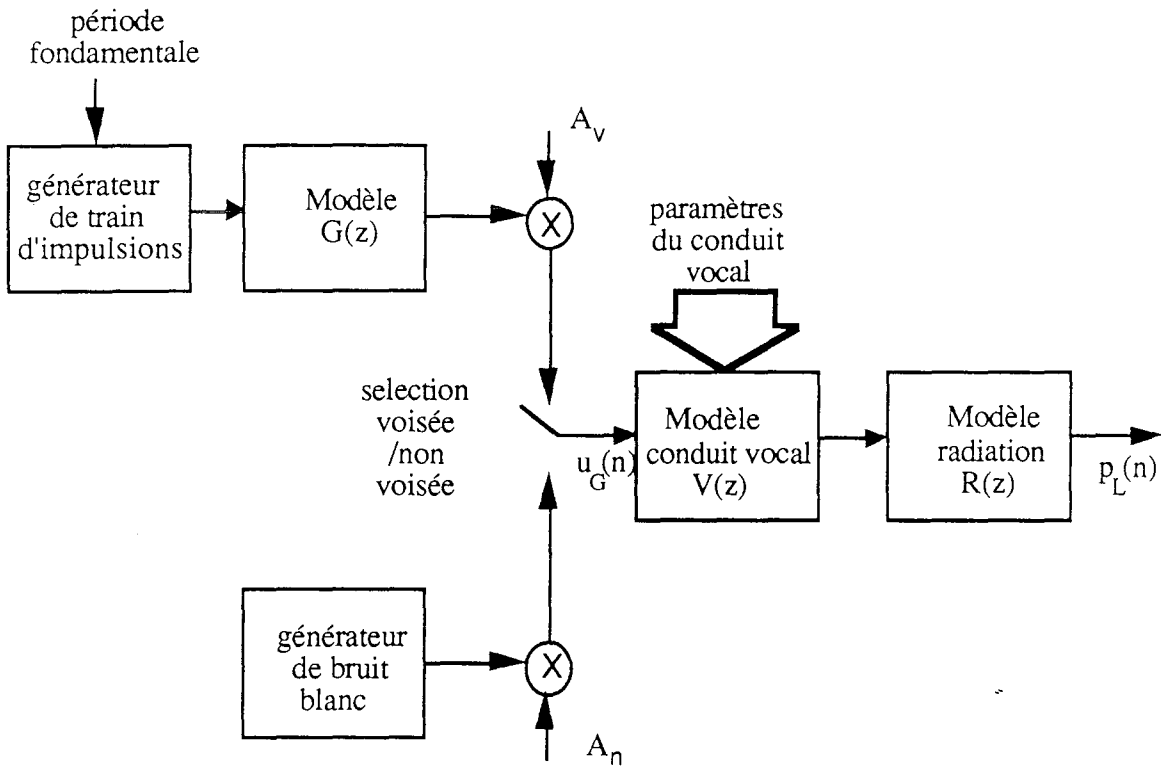


Figure 3-19. Schéma bloc du modèle numérique du système de production de la parole [Rabiner, 1978]

4- MODELISATION PARAMETRIQUE DU SIGNAL VOCAL

MODELISATION PARAMETRIQUE DU SIGNAL VOCAL

4-0 INTRODUCTION

Nous avons exposé dans le chapitre précédent des modèles de production basés sur la théorie acoustique. Ces modèles peuvent être appelés modèles "de connaissance", car ils sont construits à partir de la connaissance du geste phonatoire (configuration articulaire, nature de l'excitation...). Cependant, une telle connaissance n'est pas toujours accessible et peut elle-même faire l'objet de l'étude du signal vocal (extraction de l'onde glottique, détermination du pitch, calcul de la fonction d'aire...). Dans la littérature de nombreux modèles paramétriques d'analyse du signal vocal ont été proposés et principalement le modèle autorégressif AR et le modèle autorégressif à moyenne ajustée ARMA. A travers ce chapitre, nous nous limiterons à rappeler les principes de ces méthodes de modélisation en essayant surtout de justifier leur application dans le traitement du signal de parole. Nous commençons par exposer des généralités sur le modèle autorégressif avec une brève description des algorithmes associés, ensuite, nous présentons comment ces méthodes sont appliquées au signal de parole, et à la fin du chapitre, d'autres modèles seront décrits.

4-1 Généralités sur la modélisation autorégressive

4-1-1 Définitions et notations

Dans tout ce qui suit, nous prenons les notations suivantes :

- signal numérique ou processus $\{x(n)\}$ où $n \in \mathbb{N}$ et plus précisément $n \in \{0, 1, 2, \dots, N\}$
- échantillon du signal à l'instant nT_e : $x(n)$ ou x_n avec T_e la période d'échantillonnage du signal.

Nous donnons la définition suivante [BOITE, 1987]

- Définition d'un processus stationnaire autorégressif :

un signal stationnaire $\{x(n)\}$ est autorégressif d'ordre p , s'il existe des réels a_1, a_2, \dots, a_p , tels que, à tout instant n , on ait

$$x(n) = - \sum_{i=1}^p a_i x(n-i) + e(n) \quad (4-1)$$

où $e(n)$ est un bruit blanc de moyenne nulle et de variance constante σ^2 .

- Prédiction linéaire :

Soit un processus $\{x(n)\}$ autorégressif d'ordre p . Etant donnés les échantillons $x(n-p), x(n-p+1), \dots, x(n-1)$, nous pouvons faire une prédiction de l'échantillon à l'instant n

$$\hat{x}(n) = - \sum_{i=1}^p a_i x(n-i) \quad (4-2)$$

Ainsi, à tout processus autorégressif peut être associé un codage prédictif ou LPC où le bruit blanc peut être interprété comme une erreur de prédiction

$$e(n) = x(n) - \hat{x}(n) = x(n) + \sum_{i=1}^p a_i x(n-i) \quad (4-3)$$

- Filtre de la prédiction linéaire ou filtre inverse :

Dans le domaine fréquentiel, la relation (4-3) s'écrit en utilisant la transformée en z sous la forme

$$E(z) = X(z) \left[1 + \sum_{i=1}^p a_i z^{-i} \right] \quad (4-4-a)$$

ou encore

$$E(z) = X(z) A(z) \quad (4-4-b)$$

où

$$A(z) = \left[1 + \sum_{i=1}^p a_i z^{-i} \right] \quad (4-5)$$

est le filtre inverse de la prédiction linéaire ou filtre d'analyse et $1/A(z)$ est le filtre de la prédiction linéaire ou filtre de synthèse.

Remarquons que le filtre de synthèse est un filtre tous pôles, autrement dit le processus autorégressif peut être considéré comme la sortie d'un système n'admettant que des résonances,

et dont l'entrée est l'erreur de prédiction ou encore le bruit blanc. Cette remarque justifiera plus loin l'utilisation de la prédiction linéaire ou le modèle autorégressif pour l'analyse du signal vocal. Nous illustrons ces notions dans la figure 4-1.

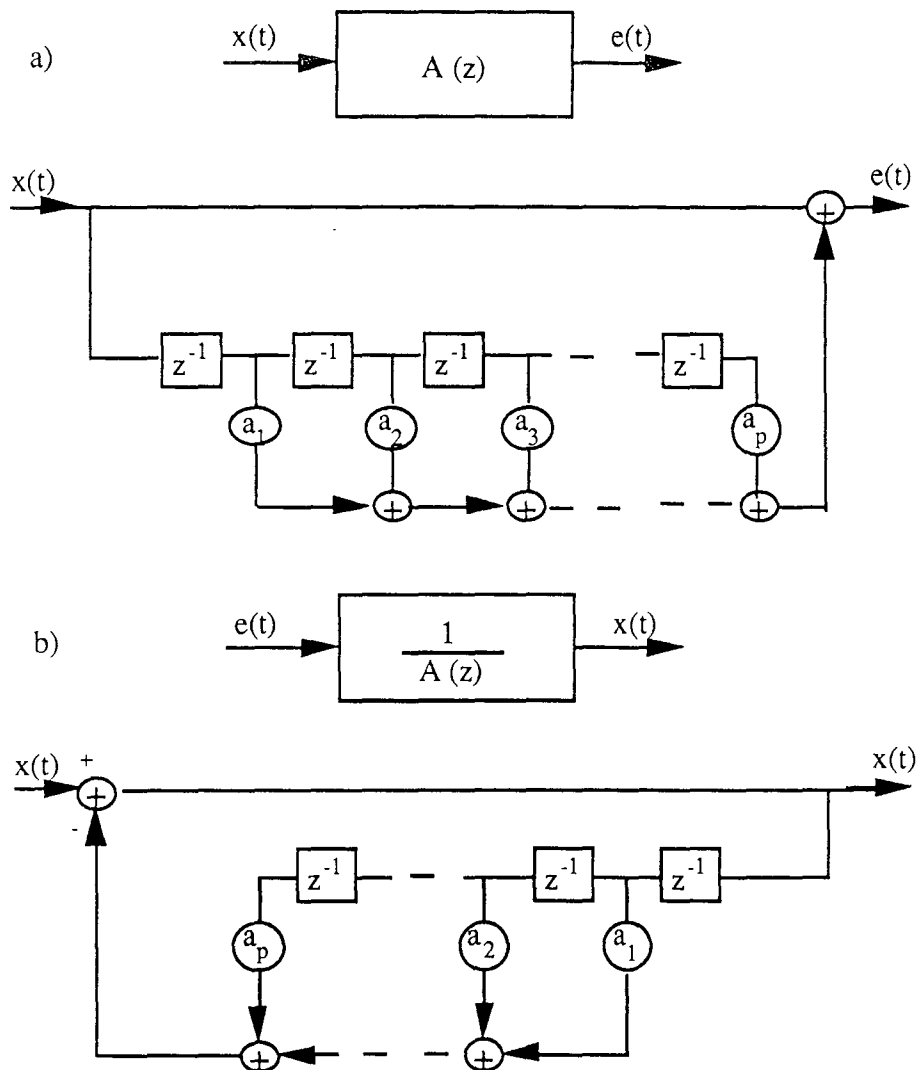


Figure 4-1. Principe de la prédiction linéaire

a) Filtre d'analyse du signal $x(t)$

b) Filtre de synthèse du signal $x(t)$

4-1-2 Modélisation autorégressive d'un signal stationnaire

Etant donné un processus stationnaire $\{x(n)\}$, supposé autorégressif d'ordre p , la modélisation de ce processus revient généralement à l'optimisation d'un critère par rapport aux coefficients du modèle. Plusieurs critères ont été proposés par les auteurs selon l'application recherchée et la nature du signal étudié, et nous les avons regroupés en deux catégories.

4-1-2-1 Critères dans le domaine temporel, dont l'objet est l'erreur de prédiction

Le critère le plus immédiat, reposant sur l'idée simple selon laquelle le modèle est satisfaisant lorsque la prédiction à tout instant n est très "proche" du signal, est la minimisation des carrés de l'erreur de prédiction. Ce critère est appelé dans la littérature critère des moindres carrés et sa formulation mathématique est

$$E_1 = \sum_n e(n)^2 \quad (4-6)$$

Un deuxième critère est basé sur la minimisation de la variance de la séquence des erreurs de prédiction $\{e(n)\}$

$$E_2 = E[e(n)^2] = \sum_{i=0}^p \sum_{j=0}^p a_i a_j E[x(n-i)x(n-j)] \quad (4-7)$$

Pour un signal stationnaire, l'espérance mathématique du produit des séquences $\{x(n-i)\}$ et $\{x(n-j)\}$ ne dépend que des décalages i et j et est par définition le coefficient de corrélation de la séquence $\{x(n)\}$ pour un retard de $|i-j|$.

D'autres auteurs [Itakura-Saito, 1968] [Itakura-Saito, 1971] sont partis d'une autre hypothèse pour établir un troisième critère appelé critère du "maximum de vraisemblance". Pour ces auteurs le processus autorégressif est considéré comme un processus aléatoire dont une séquence $\{x(n)\}_{n \in \{0, \dots, N-1\}}$ est une séquence d'observations. Le modèle recherché est celui dont les paramètres représentent la séquence d'observations de la façon la plus probable, c'est-à-dire qui maximisent la densité de probabilité conjointe des N variables aléatoires $x(0), \dots,$

$x(N-1)$. Tenant compte du théorème de la limite centrale, il est légitime de supposer que le bruit d'entrée est un bruit blanc gaussien de moyenne nulle et de variance σ_e^2 . D'après la relation (4-3), le processus $\{x(n)\}$ est aussi un processus gaussien de moyenne nulle. La densité de probabilité conjointe des N variables d'observation est donc une gaussienne fonction à la fois des paramètres a_i du modèle et de la variance σ_e^2 ; cependant le problème d'optimisation de cette densité est non linéaire par rapport aux paramètres du modèle. Pour éviter cette non linéarité Itakura et Saito ont proposé une approximation de la densité de probabilité sous la forme suivante :

$$p[x(0),x(1),\dots,x(N-1)] = (2\pi\sigma_e^2)^{-\frac{N}{2}} \exp\left(-\frac{\alpha}{2\sigma_e^2}\right) \quad (4-8)$$

où α est l'erreur quadratique définie par

$$\alpha = \sum_{n=-\infty}^{\infty} e(n)^2 \quad (4-9-a)$$

ou encore

$$\alpha = \sum_{n=-\infty}^{\infty} \left[\sum_{i=0}^p a_i x(n-i) \right]^2 \quad (4-9-b)$$

avec $x(n) = 0$ pour $n < 0$ et $n > N-1$ et $a_0 = 1$.

Nous pouvons remarquer que la densité de probabilité de (4-8) est maximale si α est minimale, et si de plus la variance σ_e^2 vérifie

$$\sigma_e^2 = \frac{\alpha}{N} \quad (4-10)$$

Ainsi, ce critère rejoint les moindres carrés avec la condition supplémentaire (4-10).

4-1-2-2 Critère dans le domaine fréquentiel : ajustement spectral

Pour certaines applications, on s'intéresse aux caractéristiques fréquentielles du signal supposé autorégressif et il est donc normal que le critère de la modélisation tienne compte de ces caractéristiques. Le critère le plus courant est celui qui traduit une bonne coïncidence du spectre du signal avec la réponse fréquentielle du filtre prédicteur. Il repose sur le fait que si la

modélisation est "bonne", le signal analysé coïncide au mieux avec la sortie du filtre prédicteur excité par une entrée aléatoire sous forme de bruit blanc. Puisque ce dernier admet un spectre constant, le spectre de la sortie du filtre prédicteur est la réponse fréquentielle de celui-ci. Une des formulations mathématiques de ce critère est la suivante

$$E_3 = \int_{-\infty}^{+\infty} \frac{P(\omega)}{\hat{P}(\omega)} d\omega \quad (4-11)$$

où $P(\omega)$ et $\hat{P}(\omega)$ sont respectivement les spectres de puissance du signal et de la réponse fréquentielle du filtre de prédiction.

Remarquons que, d'après la relation de Wiener-Khintchine selon laquelle le spectre de puissance d'un signal est égal à la transformée de Fourier de sa fonction d'autocorrélation, le critère E_3 peut se traduire dans le domaine temporel par un ajustement des séquences d'autocorrélation du signal et de la réponse impulsionnelle du filtre de synthèse. En effet, si on considère la réponse impulsionnelle $\{h(n)\}$ du filtre causal stable de synthèse, les spectres de puissance $\hat{P}(\omega)$ et $P(\omega)$ sont égaux aux transformées de Fourier des fonctions d'autocorrélation respectives de $\{h(n)\}$ et $\{x(n)\}$. Ainsi, si les séquences d'autocorrélation $\{\rho_x(n)\}$ et $\{\rho_h(n)\}$ sont égales, les spectres de puissances $\hat{P}(\omega)$ et $P(\omega)$ coïncident.

En fait, on choisit un ajustement des fonctions d'autocorrélation sur des séquences finies à $p+1$ éléments, et le critère d'erreur devient

$$\rho_x(j) = \rho_h(j) \text{ pour } j = 0, 1, \dots, p \quad (4-12)$$

où $\{\rho_x(n)\}$ et $\{\rho_h(n)\}$ sont les séquences d'autocorrélation respectives du signal et de la réponse impulsionnelle du filtre de synthèse définies par

$$\rho_x(j) = \sum_{n=0}^{+\infty} x_n x_{n+j} \quad (4-13)$$

$$\rho_h(j) = \sum_{n=0}^{+\infty} h_n h_{n+j} \quad (4-14)$$

4-1-2-3 Lien entre les différents critères

Malgré la diversité de la formulation et l'interprétation des différents critères, le résultat fondamental est que la forme de ce problème d'optimisation revient à la résolution d'un système d'équations linéaires dites équations normales ou de Yule-Walker dont la forme générale est

$$[R] \begin{bmatrix} 1 \\ \dots \\ A \end{bmatrix} = \begin{bmatrix} \sigma_e^2 \\ \dots \\ 0 \end{bmatrix} \quad (4-15)$$

où [R] est une matrice de dimension $(p+1) \times (p+1)$ dont le terme général est

$$r_{ij} = \sum_n x(n-i)x(n-j) \quad (4-16-a)$$

pour un signal déterministe, ou

$$r_{ij} = E(x(n-i)x(n-j)) \quad (4-16-b)$$

pour un signal aléatoire.

4-1-3 Considérations pratiques

En pratique, ne disposant que d'une séquence finie du signal à modéliser, la méthodologie de la modélisation doit tenir compte de cette contrainte. Deux approches, différentes par leur façon d'interpréter la séquence et par leur champ d'application, sont proposées : la méthode dite d'autocorrélation et la méthode de covariance. Pour la première, la séquence finie est considérée comme une séquence infinie d'échantillons qui sont forcés à être nuls en dehors des échantillons disponibles du signal. Pour la seconde, seuls les échantillons disponibles sont pris en compte et aussi bien le critère d'optimisation que les termes de la matrice R sont adaptés à l'intervalle fini de disponibilité des échantillons du signal.

Nous illustrons ces deux approches par la modification apportée sur le calcul des termes de la matrice R de la relation (4-15).

4-1-3-1 Méthode d'autocorrélation

- signal déterministe :

Etant donnée une séquence finie $\{x(n)\}_{n \in \{0, \dots, N-1\}}$ d'un signal déterministe la modélisation s'effectue comme précédemment sur une séquence infinie $\{x'(n)\}$ telle que pour tout $n \in \mathbb{Z}$ on ait

$$x'(n) = w(n)x(n) \quad (4-17)$$

où $w(n)$ est une fonction de pondération ou fenêtre (rectangulaire, hamming, hanning...) telle que

$$\begin{cases} 0 < w(n) \leq 1 & \text{pour } n \in \{0, \dots, N-1\} \\ 0 & \text{ailleurs} \end{cases} \quad (4-18)$$

Les termes r_{ij} de la matrice R sont donc définis par

$$r_{ij} = \sum_{n=-\infty}^{n=+\infty} x'(n-i)x'(n-j) = \sum_{n=-\infty}^{n=+\infty} w(n-i)x(n-i)w(n-j)x(n-j) \quad (4-19)$$

Nous remarquons les propriétés suivantes :

- * la sommation dans (4-19) est en fait finie puisque $x'(n) = 0$ si $n \notin \{0, 1, \dots, N-1\}$,
- * les termes de R sont les coefficients d'autocorrélation de la séquence pondérée,
- * la matrice R est symétrique et de Toeplitz.

Cette dernière propriété est exploitée par plusieurs algorithmes de résolution de l'équation de Yule-Walker.

- signal aléatoire :

Le signal étant supposé stationnaire, les termes de la matrice R sont tels que

$$r_{ij} = r_{|i-j|} = r_k$$

où
$$r_k = E[x(n)x(n+k)] \quad (4-20)$$

Pour calculer l'espérance mathématique de (4-20), il est nécessaire de disposer de la densité de probabilité du signal. En général, cette densité est supposée discrète et uniforme sur l'intervalle d'observation. Deux estimateurs de r_k sont alors proposés

- * estimateur de Bartlett

$$r_k = \frac{1}{N} \sum_{n=0}^{N-k-1} [x(n) - \bar{x}][x(n+k) - \bar{x}] \quad (4-21)$$

pour $0 \leq k \leq N-1$, où \bar{x} est la moyenne du signal dans l'intervalle d'observation. On montre que cet estimateur est biaisé, mais asymptotiquement sans biais [Meloni, 1985].

* estimateur de Blackmann-Tukey [Jerkinson et Watt, 1968] :

$$r_k = \frac{1}{N-k} \sum_{n=0}^{N-k-1} [x(n) - \bar{x}][x(n+k) - \bar{x}] \quad (4-22)$$

Cet estimateur est non biaisé, cependant sa variance augmente quand k tend vers N .

L'estimateur le plus couramment utilisé est celui de Bartlett modifié dans le sens où on suppose que la moyenne du processus est nulle et on ne tient pas compte de la constante $\frac{1}{N}$ (du fait de l'apparition de ce terme dans les membres de gauche et de droite du système des équations normales).

4-1-3-2 Méthode de covariance

Contrairement à la méthode d'autocorrélation, la méthode de covariance ne nécessite pas l'hypothèse de stationnarité du processus à l'ordre 2, car la fonction coût à minimiser est l'erreur quadratique α , calculée sur l'intervalle d'observation

$$\alpha = \sum_{n=p}^{N-1} e(n)^2 \quad (4-23)$$

Les termes de la matrice R deviennent alors

$$r_{ij} = c_{ij} = \sum_{n=p}^{N-1} x(n-i) x(n-i) \quad (4-24)$$

4-1-3-3 Ordre du modèle

Le problème du choix de l'ordre du modèle est important pour la modélisation autorégressive. Nous disposons de deux moyens pour la détermination de cet ordre :

- Si à partir d'une connaissance préalable nous pouvons affirmer que le signal est la sortie d'un système linéaire résonant, excité par une entrée proche d'un bruit blanc ou d'un train d'impulsions, l'ordre peut être choisi égal au double du nombre de résonances du système.

- En général, une telle connaissance du système de production n'est pas clairement établie. Le choix repose sur des critères de type statistiques tels que le critère d'Akaike [Akaike, 1974], de Rissanen ou de Hannan [Lekhal, 1980].

4-1-4 Techniques numériques de modélisation

Dans ce paragraphe, avant de passer en revue quelques techniques itératives de résolution des équations normales, nous introduisons le concept des prédictions directe et rétrograde sur lesquelles reposent les algorithmes rapides.

4-1-4-1 Prédictions directe et rétrograde

Etant donné un processus autorégressif, nous pouvons prédire l'échantillon du signal à l'instant n en fonction des p échantillons précédents et la prédiction est dite directe. Nous pouvons estimer, d'une manière rétrospective, l'échantillon à l'instant $n-p$ en fonction des p échantillons suivants; la prédiction est dite rétrograde ou inverse. Autrement dit, étant donné les échantillons $x(n-p), \dots, x(n-1)$, la prédiction directe est donnée à l'instant n par

$$x_p^d(n) = - \sum_{i=1}^p a_i x(n-i) \quad (4-25)$$

et la prédiction rétrograde à l'instant $n-p$ est

$$x_p^r(n-p) = - \sum_{i=0}^{p-1} b_i x(n-i) \quad (4-26)$$

Les erreurs de prédiction directe et rétrograde sont respectivement définies par

$$e_p^d(n) = x(n) - x_p^d(n) = \sum_{i=0}^p a_i x(n-i) \quad (4-27)$$

$$e_p^r(n) = x(n-p) - x_p^r(n-p) = \sum_{i=0}^p b_i x(n-i) \quad (4-28)$$

avec $a_0 = 1$ et $b_p = 1$.

La figure 4-2 illustre la prédiction directe et rétrograde.

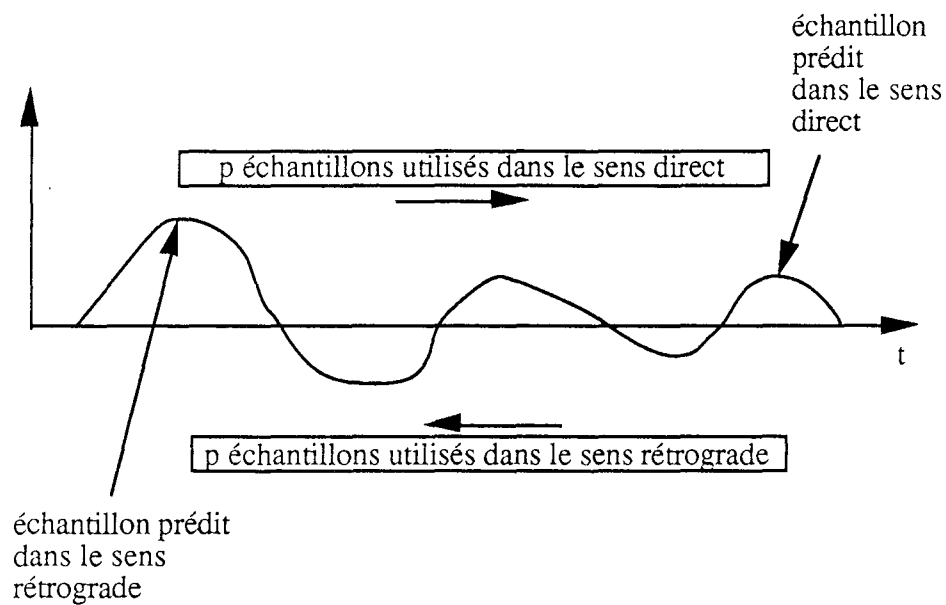


Figure 4-2. Illustration de la prédiction directe et rétrograde

Nous pouvons définir les filtres de prédiction directe et rétrograde respectivement par

$$A(z) = \sum_{i=0}^p a_i z^{-i} \quad (4-29)$$

$$B(z) = \sum_{i=0}^p b_i z^{-i} \quad (4-30)$$

Notons que pour un signal stationnaire (méthode d'autocorrélation), les coefficients des filtres rétrograde et directe sont liés par la relation

$$a_i = b_{p-i} \quad (4-31)$$

Dans la suite de ce paragraphe, nous nous limiterons pour des raisons de simplicité au cas de la méthode d'autocorrélation pour la description des algorithmes.

4-1-4-2 Calcul récurrent des coefficients du filtre d'analyse

La plupart des algorithmes que nous allons mentionner plus loin reposent sur la propriété d'orthogonalité des résidus directes et rétrogrades des filtres de prédiction de l'ordre 1 à l'ordre p [Markel, Gray, 1972]. D'après cette propriété, Markel et Gray montrent que le filtre d'analyse à l'ordre m ($1 < m \leq p$) est une combinaison linéaire des filtres prédicteurs direct et rétrograde à l'ordre $m-1$. Autrement dit, il existe un réel k_m tel que

$$A_m(z) = A_{m-1}(z) + k_m z^{-m} B_{m-1}(z) \quad (4-32)$$

En récrivant les équations normales à l'ordre m et à l'ordre $m-1$, le coefficient k_m est déterminé, par identification en utilisant (4-32), en fonction des coefficients des filtres et de l'erreur quadratique de prédiction à l'ordre $m-1$. Ces coefficients sont appelés coefficients de corrélation partielle ou coefficients de réflexion.

4-1-4-3 Structure du filtre en treillis

Pour des raisons de stabilité, et par analogie avec les techniques du filtrage numérique, la structure en treillis est souvent préférée à la structure en cascade ou en parallèle [Meloni, 1985]. Elle est basée sur les prédictions directes et rétrogrades partant de l'ordre 0 à l'ordre du modèle p . Pour la structure en treillis du filtre d'analyse nous exprimons les erreurs de prédiction directe et rétrograde partant de l'ordre 0 à l'ordre du modèle p comme le montre la figure (4-3) traduisant les relations de récurrence suivantes [Markel-Gray, 1972]

$$e_{m+1}^d(n) = e_m^d(n) + k_{m+1} e_m^r(n-1) \quad (4-33-a)$$

$$e_{m+1}^r(n) = e_m^r(n-1) + k_{m+1} e_m^d(n) \quad (4-33-b)$$

pour $m = 0 ; \dots ; m = p-1$

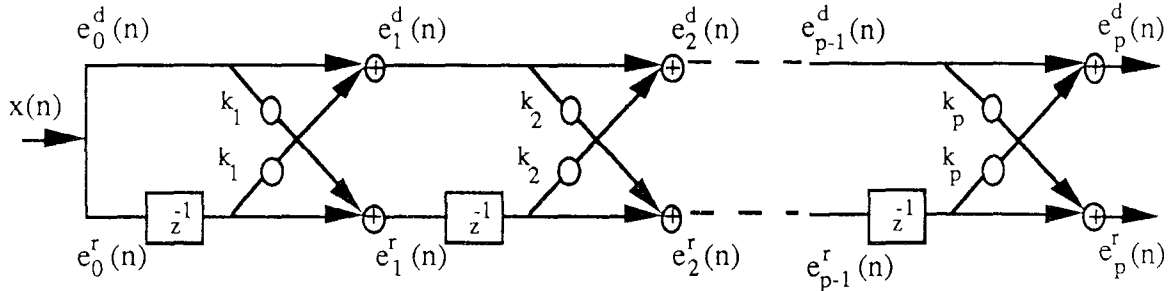


Figure 4-3. Structure en treillis du filtre d'analyse

Partant des relations (4-33), nous pouvons aussi définir la structure du filtre de synthèse de la figure (4-4) traduisant les relations rétrogrades suivantes

$$e_{m-1}^d(n) = e_m^d(n) - k_m e_m^r(n-1) \tag{4-34-a}$$

$$e_m^r(n) = e_{m-1}^r(n-1) + k_m e_{m-1}^d(n-1) \tag{4-34-b}$$

pour $m = 1 ; \dots ; m = p$

avec la condition limite suivante : $e_0^d(n) = e_0^r(n) = x(n)$

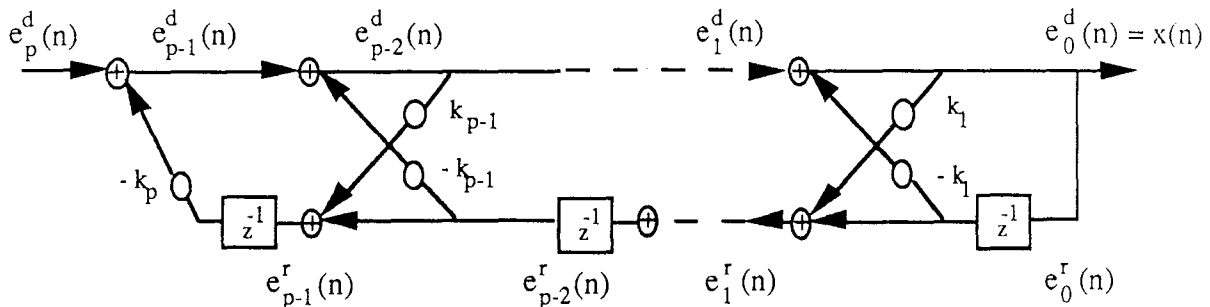


Figure 4-4. Structure en treillis du filtre de synthèse

Papoulis (1981) a montré [Meloni, 1985] que la stabilité du filtre d'analyse (toutes ses racines sont à l'intérieur du cercle unité) est assurée par la condition suivante sur les coefficients de réflexion

$$|k_m| < 1 \quad (4-35)$$

4-1-4-4 Quelques algorithmes

4-1-4-4-1 Algorithme de LEVINSON-DURBIN

L'algorithme de Levinson permet de calculer les coefficients a_i du filtre d'analyse dans le cadre de la méthode d'autocorrélation, en exploitant la structure de Tœplitz de la matrice R des équations normales (4-15).

Après avoir calculé les coefficients d'autocorrélation r_i du signal d'entrée $\{x(n)\}_{n \in \{0, \dots, N-1\}}$, par l'un des estimateurs des relations (4-21) ou (4-22), l'algorithme suit les étapes suivantes

- initialisation : (ordre 0)

$$\begin{aligned} E_0 &= r_0 \\ a_0^0 &= 1 \end{aligned}$$

- itération sur l'ordre $m = 1, 2, \dots, p$

$$\begin{aligned} \delta_{m-1} &= r_m + \sum_{i=1}^{m-1} a_i^{m-1} r_{m-i} \\ k_m &= -\frac{\delta_{m-1}}{E_{m-1}} \\ a_m^m &= k_m \\ a_i^m &= a_i^{m-1} + k_m a_{m-i}^{m-1} \quad \text{pour } 1 \leq i \leq m \\ E_m &= \left[1 - k_m^2 \right] E_{m-1} \end{aligned}$$

A l'ordre p ($m = p$), nous disposons à la fois des $p+1$ coefficients a_i du filtre d'analyse, des p coefficients de réflexion et de l'erreur quadratique totale E_p .

4-1-4-4-2 Algorithme de LEROUX-GUEGUEN

Le principe de cet algorithme est sensiblement le même que celui de Levinson. La différence réside dans la façon de mener le calcul qui utilise les coefficients d'intercorrélacion entre l'erreur de prédiction et le signal, définis par

$$C_m^p = E(e_n x_{n-p}) \quad (4-36)$$

Nous pouvons montrer aisément que ces coefficients vérifient la relation

$$C_m^p = \sum_{i=0}^p a_i^p r_{m-i} \quad (4-37)$$

Les étapes de l'algorithme sont

- initialisation :

pour $-p \leq j \leq p$ les valeurs des coefficients C_j^0 sont initialisées à r_j .

- pour $1 \leq m \leq p$ nous calculons successivement :

. k_m par

$$k_m = -\frac{C_m^{m-1}}{C_0^{m-1}} \quad (4-38)$$

. mise à jour des $2p+1$ coefficients intermédiaires C_j^m à l'ordre m par

$$C_{m-j}^m = C_{m-j}^{m-1} + k_m C_j^{m-1} \quad (4-39)$$

$$C_j^m = C_j^{m-1} + k_m C_{m-j}^{m-1}$$

pour $-p \leq j \leq p$

L'algorithme fournit en sortie les coefficients de réflexion du filtre d'analyse. Le plus grand avantage de cet algorithme est la facilité de son implantation dans des systèmes à microprocesseurs spécialisés en traitement de signal, puisque le calcul peut s'effectuer en virgule fixe [Meloni, 1985] [Leroux- Gueguen, 1977].

4-1-4-4-3 Algorithme de BURG

A partir de la structure du filtre d'analyse en treillis, la méthode de Burg a pour objet la détermination des coefficients de réflexion qui minimisent un critère fonction des erreurs de

prédiction directe et rétrograde. Ce critère est une combinaison linéaire des variances des résidus directs et rétrogrades

$$E_T^p = E\left[\left(e_p^d(n)\right)^2\right] + E\left[\left(e_p^r(n-1)\right)^2\right] \quad (4-40)$$

Les indices p, d et r indiquent respectivement l'ordre du modèle, les résidus direct et rétrograde.

On peut montrer que le p-ième coefficient de réflexion qui minimise E_T^p est la moyenne harmonique des coefficients de réflexion minimisant séparément les variances des résidus directs et rétrogrades

$$k_p = 2 \frac{K^d \cdot K^r}{K^d + K^r} \quad (4-41)$$

où

$$K^d = -\frac{E\left[e_{p-1}^d(n) \cdot e_{p-1}^r(n-1)\right]}{E\left[\left(e_{p-1}^r(n-1)\right)^2\right]} \quad (4-42-a)$$

$$K^r = -\frac{E\left[e_{p-1}^d(n) \cdot e_{p-1}^r(n-1)\right]}{E\left[\left(e_{p-1}^d(n-1)\right)^2\right]} \quad (4-42-b)$$

Pour un signal stationnaire ou localement stationnaire, les espérances mathématiques dans les relations (4-42) ne dépendent pas du temps et peuvent être approchées par

$$E\left[e_{p-1}^d \cdot e_{p-1}^r\right] = \sum_{k=0}^{N-1} e_{p-1}^d(k) \cdot e_{p-1}^r(k-1) \quad (4-43-a)$$

$$E\left[\left(e_p^d\right)^2\right] = \sum_{k=0}^{N-1} \left(e_p^d(k)\right)^2 \quad (4-43-b)$$

$$E\left[\left(e_p^r\right)^2\right] = \sum_{k=0}^{N-1} \left(e_p^r(k-1)\right)^2 \quad (4-43-c)$$

Ainsi le p-ième coefficient de réflexion est donc

$$k_p = - \frac{2 \sum_{k=0}^{N-1} e_{p-1}^d(k) \cdot e_{p-1}^r(k-1)}{\sum_{k=0}^{N-1} \left(e_p^d(k) \right)^2 + \sum_{k=0}^{N-1} \left(e_p^r(k-1) \right)^2} \quad (4-44)$$

Nous pouvons résumer les étapes de l'algorithme de Burg comme suit :

- * initialisation : $e_0^d(k) = e_0^r(k) = x(k)$ pour $0 \leq k \leq N-1$
- * calcul du coefficient de réflexion de l'étage m d'entrées $e_{m-1}^d(k)$ et $e_{m-1}^r(k)$ à partir de la relation (4-44)
- * calcul des sorties de l'étage m , $e_m^d(k)$ et $e_m^r(k)$ à partir des relations (4-33)
- * si $m = p$ alors fin, sinon revenir à l'étape 2 avec $m = m+1$

Nous pouvons émettre les remarques suivantes :

- la procédure de Burg fournit les p coefficients de réflexion du filtre en treillis,
- les p coefficients de réflexion sont de module inférieur à 1; l'algorithme de Burg détermine toujours un filtre en treillis stable.

4-1-4-5 Comparatif des algorithmes

Nous dressons le tableau comparatif (tableau 4-1) aussi bien des algorithmes récursifs présentés dans la section précédente que d'autres algorithmes directs de résolution du système d'équations normales.

	Nombres de données	complexité : nombre d'opérations proportionnel à
Inversion de GAUSS	p^2	$\frac{p^3}{3}$
Décomposition de CHOLESKY	$\frac{p^2}{2}$	$\frac{p^3}{6}$
LEVINSON-DURBIN	$2p+6$	$1,25 p^2$
LE ROUX-GUEGUEN	$2p+3$	$p^2 - p + 1$
BURG *	$3N+p$	$3Np - p^2$

Tableau 4-1. Coût des différents algorithmes de la LPC
 [Meloni, 1985], *[Marple, 1987]
 N : taille de la fenêtre d'analyse p : ordre du modèle

4-2 Modélisation autorégressive du signal de parole

4-2-1 Introduction

Le signal de parole tel qu'il apparaît à la sortie d'un microphone, présente une allure temporelle très complexe où cohabitent des zones quasi-périodiques (sons voisés) et des segments fortement bruités (sons non voisés). La première approche de modélisation paramétrique de ce signal consiste à considérer un tel processus comme un élément de l'espace fonctionnel et à le projeter sur les fonctions de base de l'espace, le représentant ainsi par ses coordonnées dans cette base. C'est dans cette approche que rentre la représentation fréquentielle par le biais de la transformée de Fourier pour laquelle la base est constituée par un continuum de sinusoides. Elle s'attache à représenter le signal dans sa globalité et doit concilier les deux impératifs antagonistes : réduction du nombre de paramètres de représentation et préservation du contenu informationnel du signal.

Une seconde approche renonce à la globalité du signal pour s'attacher localement à la façon dont ce dernier a été généré. Partant des données physiologiques (forme du conduit vocal, fonction d'aire, forme de l'onde d'excitation...) et moyennant quelques hypothèses simplificatrices, mais non irréalistes, nous avons établi au chapitre précédent un modèle linéaire

que nous pouvons qualifier de "modèle de connaissance". La modélisation paramétrique selon cette approche consiste à représenter le signal par des paramètres ayant un rapport direct avec ceux du modèle de connaissance. Cette approche bien que séduisante, nécessite des mesures directes des différentes données physiologiques qui s'avèrent souvent peu accessibles sans l'utilisation de moyens techniques qui peuvent gêner fortement la phonation (système optique par exemple).

Une troisième approche, l'objet de cette section, consiste à considérer le signal comme la sortie d'une "boite noire" caractérisée par le seul comportement externe du système. C'est dans ce cadre que le signal de parole peut être décrit par un modèle linéaire de représentation (AR ou ARMA) joint à un critère d'ajustement externe. L'idée d'appliquer à la parole des techniques de modélisation par la prédiction linéaire est principalement attribuée à Atal et Schroeder [Atal, 1971]. D'autres contributions ont affiné ces techniques par l'introduction de méthodes rapides de modélisation [Markel, 1973] [Itakura-Saito, 1971]. Une revue complète de ces techniques est fournie par [Makhoul, 1975].

Dans ce paragraphe nous commençons par exposer les principales raisons de l'utilisation du modèle autorégressif pour l'analyse et la synthèse du signal de parole. Après une présentation des avantages et des limites de cette modélisation, nous en exposerons quelques applications typiques.

4-2-2 Quelques justificatifs de l'utilisation de la modélisation autorégressive du signal de parole

Dans le chapitre précédent, nous avons vu qu'une première approximation du système physique de génération de la plupart des signaux vocaux était un système linéaire présentant un certain nombre de résonances, excité par un bruit blanc ou par un train impulsionnel. Du fait du caractère résonant (pics spectraux) du système de production, il paraît naturel de supposer que le signal de parole est modélisé par un modèle autorégressif (filtre tous pôles). L'approche de modélisation peut alors être intégrée dans une formulation générale d'identification [Gueguen, 1976]. L'information sur l'excitation n'étant pas en général disponible, on est réduit à des conjectures la concernant. Cependant, les études sur la nature de l'excitation vocale [Flanagan, 1972] montrent que le spectre de cette excitation est un spectre plat, tout au moins dans la bande fréquentielle d'analyse (0-5 kHz) et pour une brusque fermeture de la glotte, ce qui peut justifier

l'identification "brutale" de l'excitation comme étant l'erreur de prédiction. Il va de soi que le modèle obtenu par identification contient aussi bien des informations sur le système vocal que sur son excitation.

Une autre raison de l'utilisation du modèle autorégressif réside dans l'interprétation spectrale de celui-ci. En effet, certains auteurs [Rabiner, 1978] [Markel, 1972] ont montré que le spectre de puissance du filtre de synthèse constitue l'enveloppe spectrale du signal de parole; la structure formantique de ce dernier n'est donc pas fortement affectée par une telle modélisation.

Enfin l'existence d'algorithmes rapides de détermination du modèle, pouvant s'implanter dans un système à processeur spécialisé en traitement de signal, confère à une telle méthode plus d'intérêt. Cependant, la modélisation autorégressive n'est pas sans inconvénient. L'une des limites du modèle autorégressif est qu'il ne permet pas de représenter correctement certains signaux admettant des zéros spectraux (les sons nasalisés par exemple). Cet inconvénient est souvent compensé par une élévation de l'ordre du modèle [Gueguen, 1976]. Notons aussi que le signal de parole étant non stationnaire, la modélisation doit s'effectuer sur des segments de courte durée (en général 25 ms) et en toute rigueur l'ordre doit être déterminé pour chaque segment. En pratique, il est communément admis qu'un ordre fixé entre 10 et 14 est suffisant aussi bien pour l'analyse que pour la synthèse.

4-2-3 Lien entre la prédiction linéaire et le modèle des tubes acoustiques

Partant du modèle des tubes acoustiques uniformes de même longueur (chapitre précédent), nous avons établi les équations de propagation de l'onde acoustique à travers le conduit vocal. Nous montrons dans cette section que le filtre d'analyse autorégressif peut traduire ces équations de propagation et nous donnons les conditions de validité de cette traduction.

En effet, si on considère la numérotation décroissante des tubes acoustiques dans le sens glotte-lèvres, les équations (3-38) traduisant la jonction entre les tubes numéro m et $m-1$ deviennent

$$u_m^+(t - \tau) = \frac{u_{m-1}^+(t + \tau) - r_m u_{m-1}^-(t + \tau)}{1 + r_m} \quad (4-45-a)$$

$$u_m^-(t + \tau) = \frac{u_{m-1}^-(t - \tau) - r_m u_{m-1}^+(t - \tau)}{1 + r_m} \quad (4-45-b)$$

où r_m est le coefficient de réflexion à la m ème jonction défini au chapitre 3 et τ est le temps de propagation de l'onde à travers un tube ($\tau = \frac{\ell}{c}$ avec ℓ la longueur du tube).

En effectuant le changement de variables suivant

$$y_m^+(t) = c_m u_m^+(t + \tau - t_m) \quad (4-46-a)$$

$$y_m^-(t) = -c_m u_m^-(t - \tau - t_m) \quad (4-46-b)$$

où $c_m = \prod_{i=1}^m (1 + r_m)$, et

$t_m = 2(m+1)\tau$ correspondant au retard relatif à la propagation des ondes de la glotte vers les lèvres, on peut montrer [Markel, 1972]

$$y_m^+(t) = y_{m-1}^+(t) + r_m y_{m-1}^-(t) \quad (4-47-a)$$

$$y_m^-(t+T) = y_{m-1}^-(t) + r_m y_{m-1}^+(t) \quad (4-47-b)$$

avec $T = \frac{2\ell}{c}$ est le double du temps de propagation de l'onde dans l'un des M tubes.

Si les signaux $y_m^+(t)$ et $y_m^-(t)$ sont discrétisés à la période d'échantillonnage T , nous pouvons écrire les transformées en z des relations (4-47) sous la forme

$$Y_m^+(z) = Y_{m-1}^+(z) + r_m Y_{m-1}^-(z) \quad (4-48-a)$$

$$z Y_m^-(z) = Y_{m-1}^-(z) + r_m Y_{m-1}^+(z) \quad (4-48-b)$$

les conditions aux limites au niveau des lèvres et de la source se traduisent en termes des nouvelles variables par

$$Y_0^+(z) = z Y_0^-(z) = \frac{U_L}{2} \quad (4-49-a)$$

$$Y_M^+(z) = \frac{c_M U_G z^{-M/2}}{2} \quad (4-49-b)$$

Nous avons vu que le filtre inverse $A(z)$ issu de la prédiction linéaire peut être obtenu récursivement de la façon suivante :

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z) \quad (4-50-a)$$

$$z B_m(z) = B_{m-1}(z) + k_m A_{m-1}(z) \quad (4-50-b)$$

avec les conditions initiales

$$A_0(z) = 1 \text{ et } z B_0(z) = 1.$$

Nous remarquons que les relations de récurrence (4-48) et (4-50) ont la même forme. Si on suppose que les coefficients k_m et r_m sont égaux nous pouvons relier les relations de la prédiction linéaire et celles du modèle des tubes acoustiques par

$$Y_m^+(z) = Y_0^+(z) A_m(z) \quad (4-51-a)$$

$$Y_m^-(z) = Y_0^+(z) B_m(z) \quad (4-51-b)$$

pour tout $m = 0 \dots, M$

Pour $m = M$, on peut écrire d'une part

$$Y_M^+(z) = Y_0^+(z) A_M(z) = Y_0^+(z) A(z)$$

et d'autre part

$$E(z) = X(z) A(z)$$

Ainsi le filtre inverse de la prédiction linéaire est aussi le filtre de synthèse issu du modèle des tubes acoustiques

$$\frac{Y_M^+(z)}{Y_0^+(z)} = \frac{E(z)}{X(z)} = A(z) \quad (4-52)$$

Nous pouvons déduire le résultat suivant.

Si le nombre de tubes acoustiques uniformes du modèle de production est égal à l'ordre du modèle de la prédiction linéaire, les coefficients de réflexion qui définissent les rapports d'aires des tubes acoustiques sont égaux aux coefficients de réflexion de la structure en treillis du filtre de synthèse de la prédiction linéaire. Notons que la fréquence d'échantillonnage du signal de parole doit être égale au double du temps de propagation de l'onde acoustique dans un tube. Ainsi le nombre M de sections (ou l'ordre du modèle de la prédiction linéaire) est lié à la fréquence d'échantillonnage F_e et à la longueur l du conduit vocal par la relation suivante

$$F_e = \frac{Mc}{2l} \quad (4-53)$$

Par exemple si la fréquence d'échantillonnage est de 10 kHz, la longueur moyenne du conduit vocal étant de 17,5 cm alors on peut utiliser un modèle à 10 tubes acoustiques uniformes de même longueur.

Notons de plus qu'après l'analyse par la prédiction linéaire, nous pouvons obtenir la fonction d'aire relative du conduit vocal correspondant au segment analysé en utilisant la relation suivante

$$A_{m-1} = \frac{1 + k_m}{1 - k_m} A_m \quad (4-54)$$

pour $m = M, M-1, \dots, 1$ avec A_M généralement prise égale à 1.

En fait dans le signal de parole $x(n)$ (capté par le microphone) intervient l'effet de radiation par les lèvres. L'approximation du filtre du conduit vocal par le filtre d'analyse par la prédiction linéaire est valable si le signal de parole est prétraité par une préemphasis tendant à compenser l'effet de radiation. Le filtre de préemphasis $P(z)$ est en général défini par

$$P(z) = 1 - \mu z^{-1} \quad \text{avec } \mu \simeq 1 \quad \text{et } \mu < 1 \quad (4-55).$$

4-2-4 Quelques applications de la prédiction linéaire

Depuis quelques décennies, la prédiction linéaire a été largement appliquée dans le domaine du traitement de la parole. L'application qui a intéressé un grand nombre de chercheurs est le codage du signal vocal aussi bien pour améliorer la qualité et diminuer le coût de la transmission (compression par vocoder LPC), que pour un étiquetage de phonèmes en vue d'une segmentation voire même d'une reconnaissance de forme.

Une deuxième application réalisée au laboratoire CRAN-LEA touche le domaine de l'aide aux handicapés auditifs. Il s'agit d'une amélioration de la réception de la parole par les malentendants, présentant des troubles neurosensoriels dans les hautes fréquences, par une concentration de l'information portée par le signal dans les basses fréquences. Cette compression spectrale est effectuée par la modification du modèle autorégressif du conduit vocal obtenue en réduisant, de façon non uniforme, les formants et les largeurs de bande par un déplacement des pôles dans le plan complexe [Meloni, 1985].

En ce qui nous concerne, nous utiliserons la prédiction linéaire pour l'extraction de l'onde glottique (débit volumique au niveau de la glotte) à partir du signal de parole. Nous détaillerons plus loin cette technique.

4-3 Autres modèles

4-3-1 Le modèle ARMA

Une deuxième approche de modélisation consiste à considérer le signal de parole comme étant un signal autorégressif à moyenne ajustée ARMA. Nous rappelons ici, sans aucun approfondissement, l'équation récurrente du modèle ARMA :

$$x(n) + \sum_{i=1}^p a_i x(n-i) = e(n) + \sum_{j=1}^q b_j e(n-j) \quad (4-56)$$

où $\{e(n)\}$ est un bruit blanc de variance constante.

Cette modélisation bien qu'elle semble combler les lacunes de la modélisation autorégressive (existence de zéros spectraux pour les sons nasalisés par exemple), est moins utilisée en traitement du signal de parole. La principale raison, à notre avis, du désintérêt pour ce modèle provient essentiellement de la relative lenteur des algorithmes de cette modélisation. Cependant certains auteurs ont proposé des procédures d'analyse du signal de parole à partir du modèle ARMA [Grenier, 1984], dont certaines sont des procédures adaptatives [Morikawa, 1982].

4-3-2 Modèle GAR ou GARMA

Certains auteurs ont introduit un autre type de modèle mathématique dérivé des modèles précédents. Il s'agit au niveau purement formel des mêmes équations que les modèles AR ou ARMA, la différence est plutôt conceptuelle, puisque l'erreur de prédiction (bruit blanc ou train impulsionnel) est remplacée par le signal glottique de la source (GAR : "glottal autoregressive model"). L'inconvénient majeur de ce modèle est la nécessité de la capture, en parallèle et en synchronisme, des signaux de parole (par le microphone) et glottique (par un électroglottographe par exemple). Cependant, un tel modèle peut être utilisé en deuxième passe après une première approximation de l'onde glottique obtenue par le filtrage du signal de parole par un filtre d'analyse autorégressif constituant la première passe de la procédure. Comparativement au modèle autorégressif, ces modèles permettent une synthèse de meilleure

qualité (synthèse moins métallique) puisqu'ils utilisent une meilleure approximation du signal d'excitation.

4-4 Conclusion

A l'issu de ce chapitre, nous avons montré que nous disposons de plusieurs techniques de modélisation permettant d'extraire des mesures plus au moins objectives essentiellement sur le conduit vocal pendant la phonation.

PARTIE III

5- DESCRIPTION QUALITATIVE DE LA VOIX HUMAINE

DESCRIPTION QUALITATIVE DES PARAMETRES DE LA VOIX HUMAINE

Avant d'aborder la description du système de mesure sur la voix humaine, il est bon de passer en revue quelques traits, aussi bien objectifs que subjectifs, caractérisant l'émission et/ou la perception sonore humaine. L'objet de ce chapitre se limite à une description qualitative des paramètres, étant entendu que le problème de mesure ou de détection sera traité plus loin. Nous avons classé les paramètres en deux groupes : les paramètres phonémiques et les paramètres prosodiques.

5-1 Les paramètres phonémiques

Tout d'abord commençons par préciser le qualificatif "phonémique". Les linguistes définissent le phonème comme étant la plus petite unité du langage parlé, gardant des caractéristiques semblables dans un discours et pour un locuteur donnés. Nous appelons paramètres phonémiques ceux qui caractérisent cette unité dans un discours continu ou prononcée isolément (sons soutenus). D'ores et déjà nous émettons la réserve sur la variabilité de ces paramètres selon le contexte et l'état physiologique ou émotionnel du locuteur. Notons aussi que la plupart de ces paramètres appartiennent au domaine de la production de la parole et, à ce titre, possèdent un correspondant dans le domaine acoustique. Cette remarque nous permet donc de chercher pour chaque paramètre la grandeur mesurable correspondante et les techniques de mesure.

5-1-1 L'intensité

A l'audition, un sujet a une sensation différente d'un son émis selon qu'il est fort, faible ou moyen. Ceci provient de la puissance de l'onde sonore émise. Notons que le paramètre intensité ne dépend nullement du phonème prononcé mais plutôt de l'intensité du flux d'air (ou de la pression au niveau des poumons) lui donnant naissance. Cependant, la production de certains phonèmes (tels que les plosives) nécessite un niveau de puissance sonore plus intense que pour d'autres phonèmes (comme par exemple les voyelles). La grandeur mesurable associée à l'intensité est la puissance surfacique de l'onde acoustique produite (ou perçue).

5-1-2 Le voisement

Comme nous l'avons vu précédemment, l'ensemble des phonèmes se classe en deux catégories, selon la nature de l'onde d'excitation du système de production. Le paramètre voisement peut être représenté par une valeur logique indiquant si le phonème est voisé ou non. Rappelons qu'un phonème est voisé, si la vibration des cordes vocales est à l'origine de l'onde d'excitation, et qu'il n'est pas voisé, si l'excitation est de nature turbulente.

5-1-3 Les formants

Il est clair qu'une différence schématique entre les phonèmes, aussi bien au niveau de la production que de la perception, réside essentiellement dans la manière de prononcer, autrement dit, dans la configuration de l'appareil articulatoire. Il a été établi que les paramètres les plus en liaison avec le geste articulatoire sont les fréquences propres du résonateur (conduit vocal) appelées formants (cf. chapitre 3). Le codage de l'information (phonème prononcé) peut être fourni par l'ensemble de ces formants, étant entendu que l'oreille humaine effectue le décodage par leur détection (certainement pas d'une façon précise).

5-1-4 La hauteur

Ce terme, d'origine musicale, a été utilisé en parole pour désigner la fréquence propre d'un son voisé ou fréquence fondamentale notée en général F_0 . Elle peut s'interpréter de la manière suivante : étant donné un son émis, si on avait le moyen d'éliminer les résonances dues au conduit vocal, ce serait un son périodique à la fréquence F_0 , égale à celle de la vibration des cordes vocales. Cette interprétation est à la base de l'utilisation du filtrage inverse pour obtenir l'onde glottique. Comme nous l'avons précisé plus haut, elle est responsable au niveau de la perception, de la différence de tonalité des sons perçus (grave /aiguë).

En ce qui concerne la détection de la hauteur, nous décrivons plus loin une des techniques numériques de détection de la périodicité (ou pseudo-périodicité) dans un signal de parole voisée.

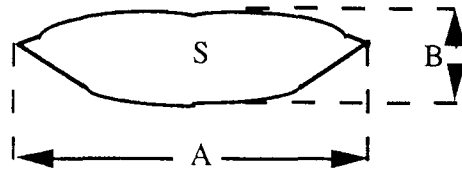
5-1-5 La nasalisation

Certains sons sont produits en faisant intervenir en parallèle les deux conduits oral et nasal. Comme le paramètre voisement, le paramètre nasalisation peut être considéré dans un premier temps comme un discriminateur entre les phonèmes nasalisés et les non nasalisés.

Cependant, le système de production des phonèmes nasalisés étant très complexe, il serait bon de disposer de plusieurs paramètres pour décrire le degré de nasalisation. Pour illustrer cette complexité, nous faisons référence à la différence entre les consonnes nasalisées (telles que le /m/) pour lesquelles le rayonnement est effectué par les narines, et les voyelles nasalisées (telles que le /ã/), pour lesquelles le rayonnement s'effectue essentiellement par la bouche avec une déperdition par les narines. Un ensemble possible des paramètres de la nasalisation peut être l'ensemble des fréquences d'anti-résonance acoustique (ou fréquences propres du système anti-résonant constitué par le conduit nasal). La détection de ces anti-formants n'est cependant pas toujours aisée dans un signal donné, et le problème de leur nombre n'est pas encore définitivement résolu, bien que certaines présomptions à ce sujet ont été déjà émises. Quant au degré de nasalisation, nous ne pouvons pas le caractériser de façon précise en l'état actuel de notre connaissance.

5-1-6 La labialité

Le rôle articulatoire et acoustique des lèvres dans l'acte phonatoire est particulièrement capital dans la mesure où leur position extrême permet une ultime intervention sur l'onde acoustique. La modification possible est double : on peut à loisir allonger ou raccourcir le tuyau résonant, ainsi qu'en ouvrir ou fermer l'embouchure. La labialité peut être caractérisée par certaines mesures géométriques de l'orifice labial définies à la figure 5-1. Ces mesures peuvent s'obtenir par des techniques radiologiques ou cinématographiques [Zerling, 1987]. Les paramètres géométriques de la figure 5-1, ont sans doute une influence sur l'impédance acoustique de la terminaison du tube acoustique. A l'heure actuelle, nous n'avons pas encore rencontré une méthode basée uniquement sur le traitement numérique du signal pour l'extraction de ces mesures.



A = écartement horizontal
 B = espace vertical
 S = aire des lèvres

Figure 5-1. Mesures géométriques de la labialité

5-2 Les paramètres prosodiques

Par opposition aux paramètres phonémiques, nous avons groupé dans ce paragraphe quelques paramètres décrivant des unités plus grandes que le phonème. Nous précisons le terme prosodique, en citant la définition de la prosodie donnée par F. Carton : "on groupe sous le nom de prosodie les faits phoniques relatifs à l'accentuation, à l'intonation, à la quantité et aux tons"[Carton, 1974]. Ces faits échappent naturellement à l'analyse "segmentale" en phonèmes, de sorte que l'école américaine les qualifie de "faits supra-segmentaux". Notons que ces faits diffèrent en général, d'un groupe humain à un autre et d'une langue à une autre. Dans ce paragraphe, nous n'avons pas la prétention de traiter tous ces faits, et nous nous limiterons à quelques traits mesurables ou représentables graphiquement.

5-2-1 Le contour formantique

C'est un graphe représentant l'évolution temporelle des formants pendant un intervalle de temps donné. Cet intervalle peut varier, selon l'application, de la durée d'une syllabe à la phrase. En segmentation de la parole, le contour formantique peut jouer un rôle non négligeable, puisqu'il peut caractériser des transitions entre les phonèmes telles que CV, VC ou la diphtongaison. En pratique, le contour formantique peut s'obtenir manuellement par analyse des tracés spectrographiques. Actuellement, par analyse des segments de parole en vue d'extraction des formants, il est possible d'avoir le contour formantique de façon automatique [Rabiner, 1978]. Il se présente sous la forme de quatre courbes décrivant chacune l'évolution d'un des quatre formants.

5-2-2 Le contour mélodique

La sensation subjective de la musicalité dans la parole provient de la variation de la fréquence fondamentale. Le contour mélodique est le moyen objectif d'étudier cette variation, nécessaire dans plusieurs domaines tels que la linguistique et la phonétique... Le contour mélodique s'obtient pratiquement par une succession d'analyses segmentales visant chacune l'extraction du fondamental.

5-2-3 L'intonation

Le terme intonation provient du mot latin "intonare" et a été traduit injustement par tonus ou ton, alors qu'il signifie en fait tonner ou faire retentir. Il a été utilisé à la fin du XIV^e siècle uniquement dans un contexte musical où il signifie "manière d'entonner la voix ou un instrument". Ce n'est qu'à partir du XVIII^e siècle qu'on a commencé à parler de l'intonation de la parole : "les tons de la parole" (Littré) et qu'au début du XX^e siècle que des auteurs ont commencé à faire la différence entre l'intonation dans le chant et la parole : "dans le chant, la mélodie procède par bonds d'une note à l'autre, tandis que dans la parole elle glisse continûment de haut en bas" [Rossi, 1981].

Après ce bref historique, nous croyons que la négligence de l'étude intonative en parole est essentiellement due à l'absence de notation graphique de l'intonation. En ce qui concerne la définition de l'intonation, nous pensons qu'elle est complexe et nécessite une définition multi-paramétrique, puisqu'elle est fonction à la fois des variations de la fréquence fondamentale au niveau de la phrase, de l'intensité et de la durée. Dans le dictionnaire DUBOIS-COLL nous rencontrons la définition de l'intonation suivante : « c'est la variation de hauteur du ton laryngien qui portent sur une suite plus longue que la syllabe ou le mot ». Martinet écrit à ce propos : « on a intérêt à réserver le terme d'intonation à ce qui reste de la courbe mélodique une fois qu'on a fait abstraction des tons et des faits accentuels » [Martinet, 1965]. Dans cette définition les tons ont une valeur distinctive au niveau du mot, alors que les faits accentuels mettent en valeur une et une seule syllabe et définissent l'unité accentuelle d'une langue donnée.

Dans le Dictionnaire linguistique [Mounin, 1968] nous pouvons lire : « l'intonation traduit la présence, à l'arrière-plan, de facteurs extra-linguistiques appartenant au domaine affectif; colère, joie, satisfaction ... ».

Pour notre part, nous caractérisons l'intonation seulement par la courbe mélodique, tracée sur des "portées vocales" contenant cinq lignes et quatre intervalles correspondant chacun à un registre (fig 5-2) [Carton, 1974]. Dans cette représentation les niveaux représentent les registres définis par Carton dans le tableau 5-1

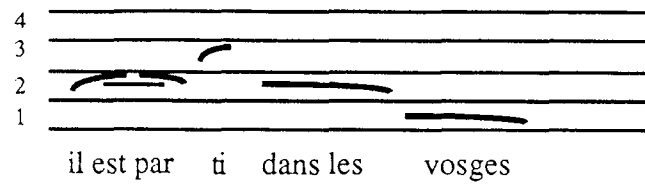


Figure 5-2. Représentation graphique de l'intonation de la phrase :
"il est parti dans les Vosges"

Registre	Niveau	Limites (Hz)	Intervalles
aigu	4	226 - 263	1,16
haut	3	191 - 226	1,18
médium	2	139 - 191	1,37
grave	1	117 - 139	1,18

Tableau 5-1. Les registres correspondant aux niveaux de la figure 5-2

Dans ce bref chapitre, nous avons énoncé de façon synthétique quelques traits de la parole humaine partant d'un niveau inférieur où l'analyse est segmentale (niveau du phonème) jusqu'à un niveau supérieur où l'analyse peut être considérée comme la juxtaposition de plusieurs analyses segmentales successives.

Nous avons vu que la détection du pitch et son évolution temporelle, ainsi que l'activité glottique, jouent un rôle primordial aussi bien au niveau inférieur qu'au niveau supérieur (syntaxique ou sémantique), ce qui explique la concentration de notre effort sur ce point.

6- MODELISATION ET EXTRACTION DES PARAMETRES DE L'ONDE GLOTTIQUE

MODELISATION ET EXTRACTION DES PARAMETRES DE L'ONDE GLOTTIQUE

L'activité glottique, comme nous l'avons déjà mentionné, joue un rôle prépondérant dans plusieurs domaines de la recherche sur la parole : identification de locuteur, synthèse de la parole... Nous avons vu en particulier, que plusieurs pathologies affectent l'activité glottique et par conséquent la phonation. Il nous a paru donc très fructueux de concentrer notre effort sur l'étude de la modélisation de l'activité glottique, en vue d'extraction de paramètres pouvant indiquer l'état de santé du larynx.

Nous commençons ce chapitre par l'examen de quelques modèles paramétriques de la glotte, en particulier le modèle paramétrique proposé par Titze [Titze, 1984] auquel nous apportons une modification faisant intervenir des paramètres supplémentaires de symétrie statique et de symétrie de vibration des deux cordes vocales. Nous décrirons ensuite les techniques d'extraction des signaux glottiques et nous présenterons en détail, les méthodes dites de filtrage inverse, permettant la détection du débit glottique à partir du signal de parole. Nous terminerons ce chapitre par une présentation de la procédure générale de mesure des différents paramètres glottiques que nous illustrerons par quelques résultats.

6-1 Modélisation de l'activité glottique

6-1-1 Modèles à masses des cordes vocales

6-1-1-1 Modèle à deux masses

Depuis déjà quelques décennies, certains auteurs [Ishizaka-Matsudaira, 1968] [Ishizaka-Flanagan, 1972] [Stevens, 1977] [Guerin-Bœe, 1977] ont étudié le modèle à deux masses des cordes vocales, proposé par Matsudaira. Ce modèle partage le muscle vocal en deux parties dont chacune est représentée par une masse soumise à une raideur (figure 6-1). Du fait que les mouvements des deux parties ne sont pas indépendants, les deux masses du modèle sont couplées par une raideur mutuelle traduisant l'existence d'une force de rappel mutuelle tendant à

ramener l'ensemble dans son état d'équilibre. Cette force est donc proportionnelle à la différence de déplacement des deux masses.

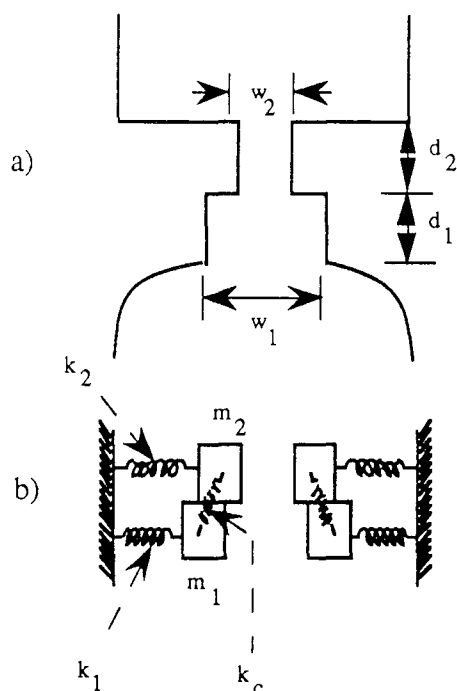


Figure 6-1. Modèle des deux masses des cordes vocales

a) représentation de la glotte

b) système physique équivalent

Ce modèle n'est pas adéquat à tous les points de vue, cependant il garde une certaine validité pour une exploration qualitative de quelques phénomènes associés à l'activité glottique.

Il nous a paru utile à titre comparatif de rapporter ici les travaux de Baer décrivant le cycle d'une corde vocale, en utilisant le modèle à deux masses et des mesures effectuées sur un larynx excisé [Baer, 1975], [Hess, 1984]. Les résultats de ces travaux sont représentés aux figures 6-2 et 6-3.

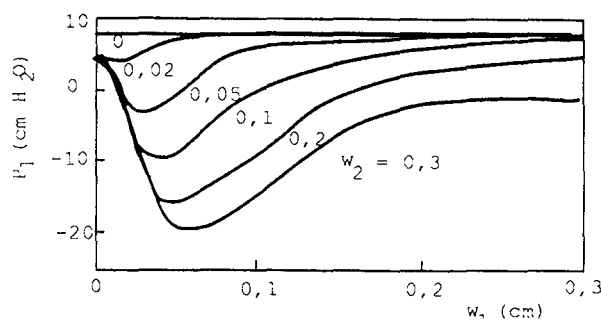


Figure 6-2. Pression dans la section basse de la glotte
en fonction des diamètres w_1 et w_2 [Baer, 1975]

Le contour représentant la séquence des configurations des deux masses pendant un cycle vibratoire est obtenu en reliant les points identifiés sur les courbes de niveaux de la figure 6-2 (pour w_1 et w_2 donnés), partant de la configuration 0 de la figure 6-3-b. Suivant le modèle des deux masses, nous pouvons décrire le cycle (état 0 à état 7) par :

- état 0 : la section supérieure (masse m_2) est complètement fermée et par conséquent aucun flux ne la traverse

- état 1 à état 2 : du fait de la pression sous-glottique, la masse m_1 subit une force externe et se déplace latéralement entraînant par effet de couplage la masse m_2 , qui reste fermée à travers une mince portion de la partie supérieure.

- état 3 à état 4 : du fait du couplage, le bord supérieur s'ouvre brusquement. Le mouvement de la masse m_2 est rapide et l'état 4 est atteint en un huitième du temps du cycle complet. Comme le montre la figure 6-3-a, la force exercée sur la masse m_1 diminue brutalement, suivant ainsi le mouvement de la masse m_2 par effet de couplage.

- état 5 : du fait du passage d'un flux d'air à travers la petite section supérieure, la masse m_1 subit une force tendant à ramener la section supérieure vers la position de fermeture

- état 6 à état 7 : la masse m_2 entraînée par la masse m_1 , continue son mouvement de fermeture, jusqu'à l'atteinte éventuelle de la fermeture totale de la section supérieure (état 7). Le cycle recommence alors à l'état 0.

Notons que d'après ce modèle la partie inférieure n'atteint jamais l'état de fermeture complète. Le sens des aiguilles d'une montre de la trajectoire de la figure 6-3-a indique que le système aéro-dynamique fournit de l'énergie au système mécanique, ce qui est nécessaire pour l'entretien de la vibration. Cette énergie est proportionnelle à l'aire de la trajectoire cyclique. La

position et la forme d'une telle trajectoire dépendent de l'ouverture statique du larynx et des propriétés physiques des cordes vocales [Stevens, 1977].

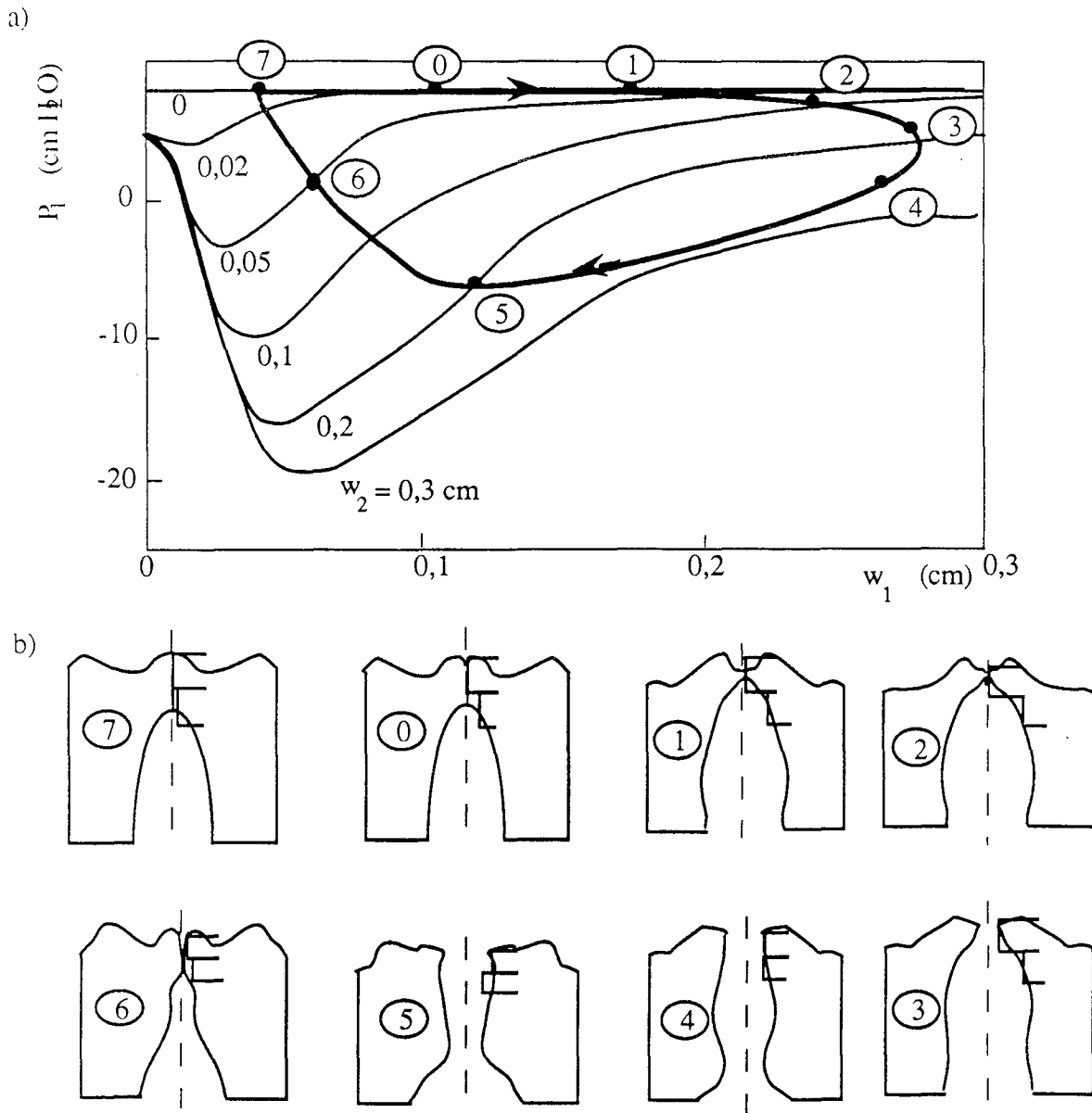


Figure 6-3. Cycle vibratoire des cordes vocales

a) variation de la pression pendant un cycle

b) configuration des deux masses pendant un cycle

Cette description qualitative des vibrations des cordes vocales suggère l'existence d'un certain nombre de paramètres agissant sur les propriétés des sons générés au larynx. Ces

paramètres de contrôle sont la pression trans-glottique, l'espacement entre les cordes vocales, les masses et les coefficients de raideur de chaque corde et le coefficient de raideur mutuelle des deux masses. C'est dans ce sens que le travail mené par Guerin et Boë [Guerin-Boë, 1977] a permis de préciser l'influence de deux paramètres sur le régime de fonctionnement du larynx : la pression sous-glottique et la tension des cordes vocales. Partant d'un ensemble de simulations du modèle à deux masses, ces auteurs ont proposé un modèle de commande physiologique utilisable en synthèse.

Bien que le modèle à deux masses soit satisfaisant au point de vue de la description qualitative du mode vibratoire des cordes vocales, il possède à notre avis les inconvénients suivants :

- le partage d'une corde vocale en deux parties simplifie considérablement l'étude du mode vibratoire en ce sens que le mouvement de la corde est décrit globalement par celui de chacune des masses, au lieu d'être décrit ponctuellement.
- les paramètres mécaniques du modèle (masses, coefficients de raideur) sont très peu liés à la physiologie du larynx et sont difficilement mesurables.
- pour l'entretien de la vibration, les deux parties inférieures des cordes vocales doivent rester en adduction pendant tout le cycle, ce qui est limitatif pour le mode vibratoire.

6-1-1-2 Modèles à plusieurs masses

Pour pallier au problème de la globalité de l'étude du mode vibratoire, certains auteurs [Titze, 1973] [Titze, 1979] [Wong, 1985] ont proposé des modèles qui considèrent une corde vocale comme la succession de plusieurs masses (de quatre à seize masses). A titre d'exemple, nous présentons ici le modèle proposé Wong, à partir duquel il a réalisé un certain nombre de simulations de l'oscillation des cordes vocales qui intègrent la possibilité d'existence d'une pathologie laryngienne. Ce modèle, à cheval entre celui de Titze [Titze, 1973] et de Ishizaki-Flanagan [Ishizaki-Flanagan, 1972], est constitué d'une distribution de deux couches de cinq masses-ressorts répartis longitudinalement (Figure 6-4).

Comme indiqué sur la figure 6-4, les masses de la couche supérieure représentent la distribution discrète de la masse des muqueuses, et celles de la couche inférieure représentent celle du muscle vocal. Dans ce modèle, Wong ne considère pour chaque masse qu'un seul degré de liberté celui de la direction latérale (dans l'axe des x sur la figure 6-4), le mouvement vertical d'ensemble n'étant pas nécessaire [Titze, 1973]. En plus des tensions exercées sur les

masses, il considère les différents amortissements dus aux pertes thermiques et de viscosité. Ces simulations tiennent compte aussi bien des forces internes exercées sur le tissu que les forces externes d'origine aérodynamique. Partant de plusieurs simulations de cas pathologiques (changement de masse locale d'une corde due par exemple à une lésion, changement de raideur ...), Wong a observé que le signal de parole simulé subit des perturbations qui peuvent être régulières ou non. Parmi les perturbations majeures observées, il cite l'introduction d'une seconde sous-harmonique (fréquence moitiée de la fréquence fondamentale) qui affecte principalement la périodicité du signal et du même coup le contour mélodique.

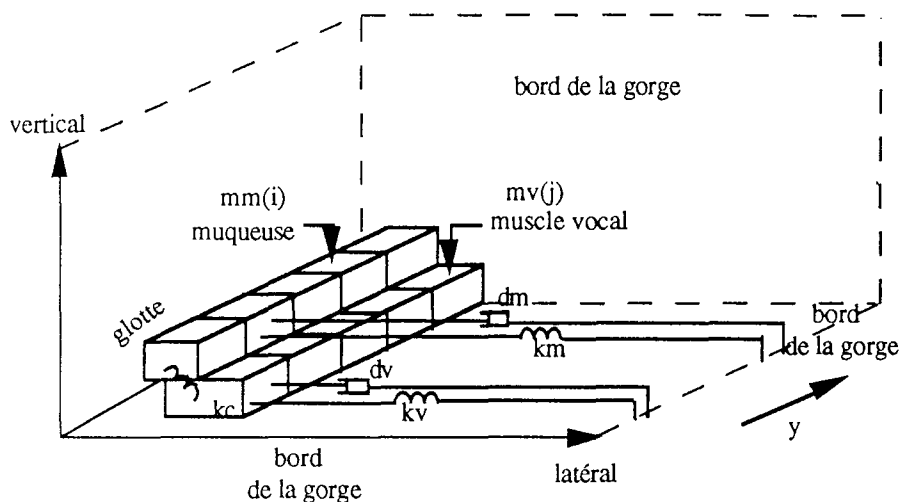


Figure 6-4. Modèle à cinq masses-ressorts d'une corde vocale [Wong, 1985]

6-1-2 Les signaux glottiques et leurs modèles

Dans ce paragraphe, nous nous intéressons aux signaux porteurs d'informations sur l'activité laryngienne. Nous décrirons dans un premier temps les signaux nécessitant des capteurs, que nous qualifions de signaux directs. Nous porterons ensuite notre attention sur les méthodes d'extraction du signal débit glottique et sur sa modélisation.

6-1-2-1 Les signaux directs

6-1-2-1-1 Le signal électroglottographique

Ce signal est très souvent utilisé par les praticiens car il mesure l'aire glottique. En fait, l'électroglottogramme mesure l'impédance électrique du larynx, qui est proportionnelle à l'aire glottique. Comme le montre la figure 6-5-a, ce capteur est constitué de deux électrodes placées de part et d'autre du cartilage thyroïdien et d'une troisième électrode, placée en arrière du cou, servant de référence pour le circuit pont. Un courant constant modulé par une tension sinusoïdale à des fréquences radio, entre 300 kHz et 3,5 Mhz, est appliqué à une des deux électrodes. Le courant capté par l'autre électrode est modulé en amplitude par la vibration des cordes. Le signal modulé est directement lié au changement de l'impédance à travers le larynx due aux vibrations des cordes. Le signal est alors soumis à une démodulation fournissant le signal électroglottographique (EGG).

Comme le montre la figure 6-5-b, le signal EGG admet deux paliers pendant un cycle vibratoire : un palier haut correspondant à la phase d'ouverture de la glotte et un palier bas correspondant à la phase de fermeture (plus l'aire glottique augmente en phase d'ouverture, plus l'impédance électrique du larynx augmente et inversement). Nous verrons plus loin que beaucoup de techniques de filtrage inverse se servent de ce dispositif. Pour notre part, l'objectif de notre travail étant l'extraction de paramètres uniquement à partir du signal de parole, nous n'utiliserons pas un tel dispositif.

6-1-2-1-2 Autres signaux

D'autres techniques ont été utilisées pour la description de l'activité glottique pendant la phonation. Parmi ces signaux nous citons :

- La photoglottographie (PGG), qui consiste à utiliser une source lumineuse et une cellule photosensible de part et d'autre de la glotte [Sonnesson, 1960]. Le signal recueilli par la cellule photosensible est proportionnel à l'aire glottique.

- Les méthodes cinématographiques ultra-rapidess (4000 images par seconde) [Farnsworth, 1940] ou aux rayons X, qui consistent à filmer directement l'activité glottique pendant la phonation. Ces méthodes sont difficiles à exploiter à cause du nombre élevé d'images. La cinématographie ultra-rapide couplée à la méthode EGG, permet de fournir plus de détails pendant les phases d'ouverture et de fermeture [Alaska, 1987].

- L'observation stroboscopique est une méthode directe d'observation grâce à une illumination stroboscopique par la bouche et un miroir placé à 45 degrés placé au fond de la bouche.

- L'image par résonance magnétique utilisant des moyens techniques sophistiqués, est basée sur les techniques de traitement d'images. Elle permet d'obtenir les dimensions du larynx [Baer, 1987].

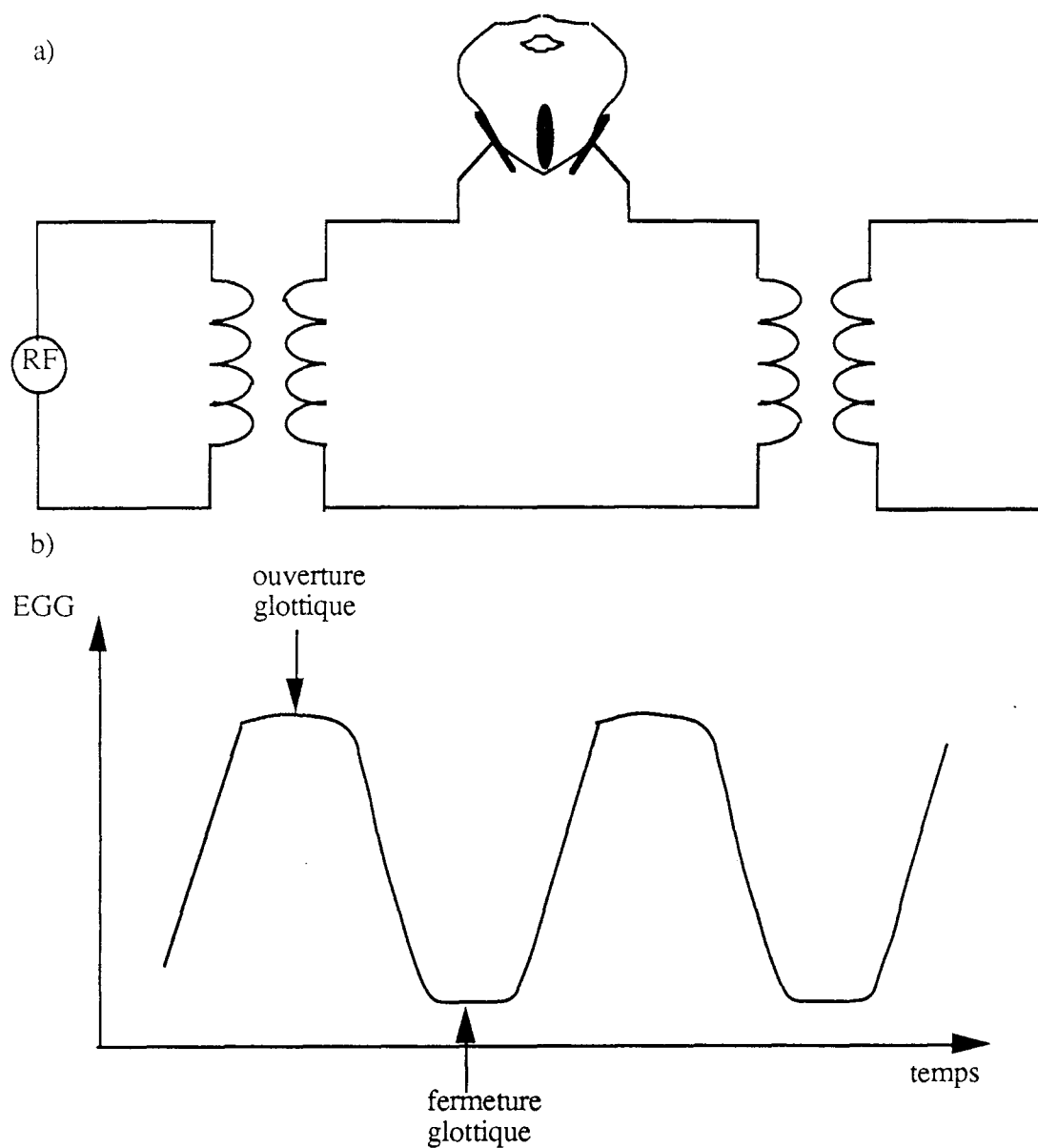


Figure 6-5. a) Schéma de principe de l'électroglottogramme

b) Exemple de signal EGG

6-1-2-2 L'onde débit glottique

L'onde débit glottique (ou simplement l'onde glottique) est d'une importance capitale dans l'étude du système de production de la parole, puisqu'elle en constitue la source. Une mesure directe d'un tel signal est une tâche extrêmement complexe et difficile et les méthodes citées plus haut ne donnent qu'une image erronée de celui-ci utilisant quelques conjectures le concernant. Citons par exemple le signal aire glottique comme image de l'onde glottique. Notons de plus que ces méthodes ont l'inconvénient majeur d'être invasives, ce qui gêne considérablement la phonation.

Dans ce paragraphe, nous présentons quelques modèles de l'onde glottique qui reposent soit sur des observations directes, soit sur des considérations théoriques ou sur des résultats de techniques de filtrage inverse.

Rotenberg [Rotenberg, 1971] a proposé un certain nombre de modèles du débit glottique dont l'amplitude et l'obliquité de l'impulsion sont ajustables; l'un de ces modèles est composé de deux segments trigonométriques avec une seule discontinuité de la pente à l'instant de fermeture (figure 6-6-a). Ce modèle a souvent servi comme référence pour des systèmes de synthèse de la parole. Hedlin [Hedlin, 1984] a utilisé un modèle similaire avec en plus une dérive de basse fréquence (figure 6-6 -b). Le modèle proposé par Fant [Fant, 1979] (figure 6-6-c) est aussi composé par des fonctions sinusoïdales mais avec la particularité supplémentaire de contrôler d'une façon indépendante la discontinuité de la fonction dérivée. Dans certains cas, le débit glottique est caractérisé par une courbe de classe C1 (fonction et dérivée continues) au voisinage de l'instant de fermeture. Ceci amène des raffinements du modèle proposé par Ananthapadmanabha [Ananthapadmanabha, 1984] (figure 6-6-d) qui considère la fonction débit glottique comme étant la somme de fonctions sinusoïdale et polynomiale ou celles proposés par Liljencrants-Fant [Liljencrants-Fant, 1985] (figure 6-6-e) qui considèrent le débit glottique comme le produit d'une exponentielle réelle et d'une sinusoïde.

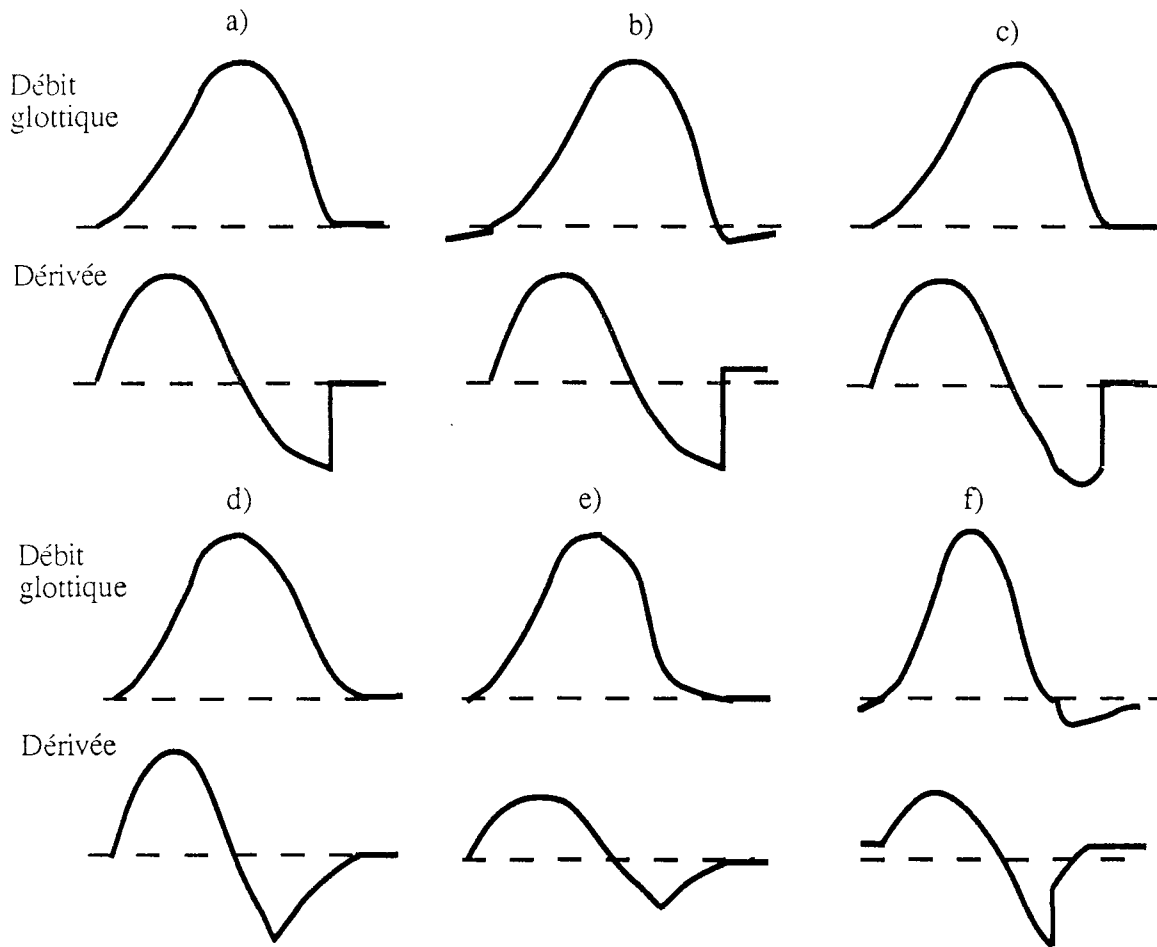


Figure 6-6. Modèles du débit glottique

[Fujisaki-Ljungqvist, 1986] [Karmouche, 1990]

a) [Rosenberg, 1971]

b) [Hedlin, 1984]

c) [Fant, 1971]

d) [Anathapadmanabha, 1984]

d) [Liljencrants-Fant, 1985]

e) [Fujisaki-Ljungqvist, 1986]

Le modèle que nous allons utiliser est celui proposé par Fujisaki-Ljungqvist [Fujisaki-Ljungqvist, 1986] qui ont proposé un modèle paramétrique défini par la fonction dérivée du débit glottique $\dot{g}(t)$ comme suit

$$\dot{g}(t) = \begin{cases} A - \frac{2A+R\alpha}{R} t + \frac{2A+R\alpha}{R} t^2 & 0 < t \leq R \\ \alpha(t - R) + \frac{3B-2F}{F^2} (t - R)^2 - \frac{2B-F\alpha}{F^3} (t - R)^3 & R < t \leq W \\ C - \frac{2(C-\beta)}{D} (t - W) + \frac{(C-\beta)}{D^2} (t - W)^2 & W < t \leq W+D \\ \beta & W+D < t \leq T \end{cases}$$

avec T la période fondamentale, $\alpha = \frac{4AR-6FB}{F^2 - R^2}$ et $\beta = \frac{CD}{D-3(T-W)}$, les autres paramètres, dont l'obliquité $S = \frac{W - R}{W}$, sont représentés à la figure 6-7.

Le choix d'un modèle polynomial a l'avantage de faciliter la variation simultanée de plusieurs paramètres et en même temps de mettre en relief des détails, en particulier l'évaluation de l'importance relative des différents paramètres temporels tels que la phase d'ouverture (W), l'obliquité de l'impulsion (S) et l'intervalle de temps séparant l'instant de fermeture à celui du maximum négatif du débit glottique (D), aussi bien que les paramètres d'amplitude (A : saut d'amplitude à l'ouverture, B et C : sauts d'amplitude respectivement avant et après la fermeture).

Fujisaki a introduit le paramètre A , qui n'est pas commun aux autres modèles, et l'attribue à l'existence d'une excitation secondaire qui provient d'un débit résiduel subsistant après l'excitation principale. L'arrondi du débit glottique à la fermeture s'explique par l'abaissement du larynx à la fermeture créant un débit glottique dans le sens inverse.

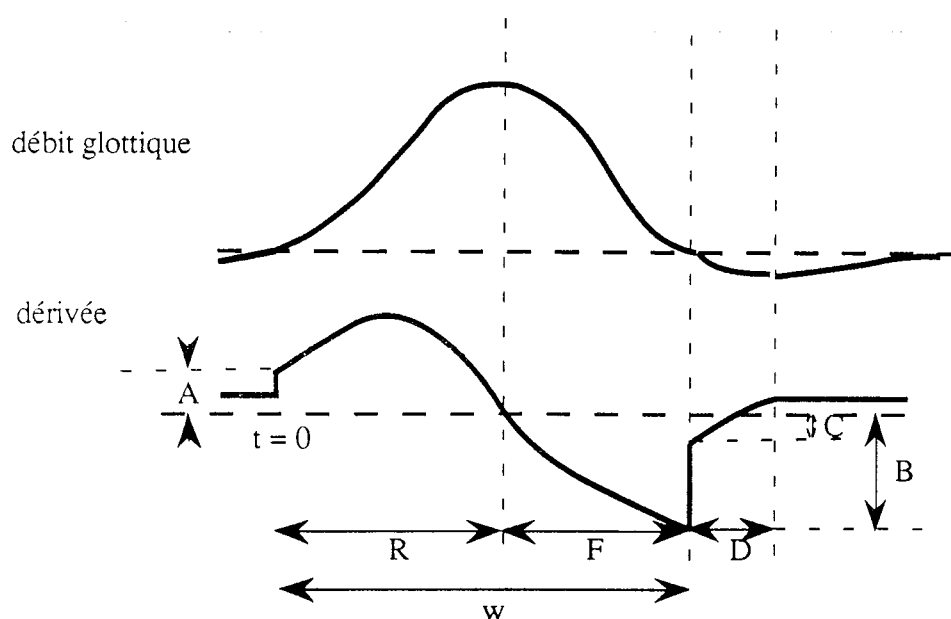


Figure 6-7. Modèle paramétrique proposé par Fujisaki

[Fujisaki-Ljungqvist, 1986]

W : largeur de l'impulsion	A : pente à l'ouverture glottique
S : obliquité de l'impulsion	B : pente avant la fermeture
D : durée de fermeture glottique	C : pente après la fermeture

6-1-3 Modèle cinétique de l'activité glottique [Titze, 1984]

Reposant sur les observations directes des signaux glottiques, Titze a proposé une interprétation des signaux glottiques en termes de cinématique des cordes vocales. Les paramétrisations ci-dessus de l'activité glottique se servent des aspects temporels et spectraux des signaux glottiques. Les dernières contributions de Rothenberg et Fant évoquent des quantités physiologiques telles que le rapport de l'inertance du conduit vocal sur la résistance de la glotte, la pression sous-glottique et l'adduction des cordes vocales. La paramétrisation de Titze repose sur les critères suivants :

- les paramètres doivent avoir une signification physiologique,
- la paramétrisation doit couvrir les différents registres de la voix humaine,
- le nombre de paramètres doit rester petit pour préserver un équilibre entre les informations articulatoires et phonatoires,
- les paramètres doivent servir à la fois pour l'analyse et la synthèse de la parole,
- les paramètres doivent être sans dimension.

Sachant que l'établissement de l'oscillation dépend d'une adduction propre de l'arythénoïde, de la forme verticale pré-phonatoire des cordes vocales (convergentes, rectangulaires, divergentes), de la différence de phase entre les mouvements des parties inférieures et supérieures des cordes vocales et des conditions de charge dans le conduit vocal, Titze a proposé pour chacune de ces conditions un paramètre spécifique que nous détaillerons plus loin.

Le modèle cinétique de Titze suppose qu'en phase pré-phonatoire la forme verticale des cordes vocales est trapézoïdale (figure 6-8). Cette configuration est représentée dans le repère (xyz) (le plan (yz) est le plan médian, l'axe des y est dans la direction postéro-antérieure, le plan (yx) étant le plan latéral) par l'équation suivante

$$\xi_0(y,z) = \left\{ \xi_{01} - \frac{\xi_{01} - \xi_{02}}{T} z \right\} \left\{ 1 - \frac{y}{L} \right\} \quad (6-1)$$

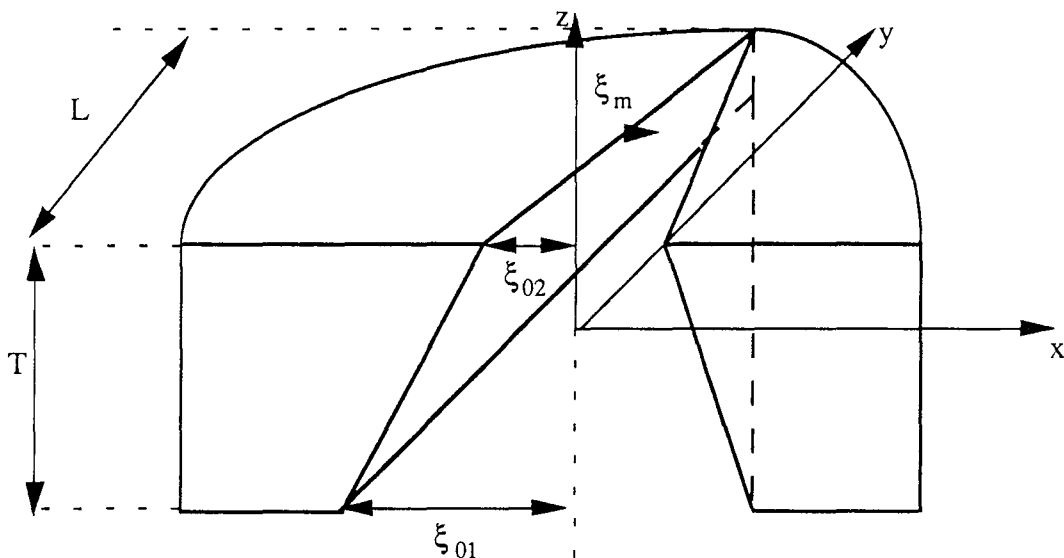


Figure 6-8. Configuration pré-phonatoire dans le cadre du modèle de Titze [Titze, 1984].

En prenant comme hypothèse supplémentaire que le mouvement latéral des cordes vocales est sinusoïdal, nul aux bords antérieur et postérieur, le déplacement dynamique à l'instant t (par rapport à la position pré-phonatoire) peut être approché par

$$\xi_1(y,z,t) = \xi_m \sin\left(\frac{\pi y}{L}\right) \sin\left(2\pi ft - \frac{\phi z}{T}\right) \quad (6-2)$$

où ξ_m l'amplitude commune des bords inférieur et supérieur ϕ le déphasage entre ces deux bords et f la fréquence fondamentale de vibration.

Pour ce modèle la symétrie des deux cordes est aussi supposée.

Partant de la configuration initiale et du mouvement relatif des cordes vocales, Titze propose les paramètres de contrôle suivants

$$\text{- quotient d'abduction : } Q_a = \frac{\xi_{02}}{\xi_m} \quad (6-3)$$

$$\text{- quotient de forme : } Q_s = \frac{\xi_{01} - \xi_{02}}{\xi_m} \quad (6-4)$$

$$\text{- quotient de phase : } Q_p = \frac{\phi}{2\pi} \quad (6-5)$$

Ainsi en combinant la position pré-phonatoire et le déplacement relatif des cordes vocales, l'équation du mouvement par rapport au plan médian en terme des paramètres ci-dessus est la suivante

$$\begin{aligned} \xi(y,z,t) &= \xi_0(y,z,t) + \xi_1(y,z,t) \quad (6-6) \\ &= \xi_m \left[\left(1 - \frac{y}{L}\right) \left(Q_a + Q_s - Q_s \frac{z}{T} \right) + \sin \left(\frac{\pi y}{L} \right) \sin 2\pi \left(ft - \frac{Q_p z}{T} \right) \right] \end{aligned}$$

L'équation du mouvement des cordes vocales de (6-6) est très idéalisée, cependant elle permet de décrire les plus importants aspects de la variation temporelle de la glotte. Par exemple, la fente postérieure peut demeurer pendant le cycle alors qu'une réduction de l'ouverture glottique peut apparaître sur la majorité de la longueur de la glotte quand le déplacement dynamique devient négatif (ceci si $Q_a > 0$ et $Q_s \geq 0$).

La normalisation des largeurs pré-phonatoires (ξ_{01} et ξ_{02}) par l'amplitude maximale de vibration est non seulement convenable mais permet aussi l'utilisation de la largeur relative de la fente glottique (par rapport à la largeur initiale) qui décrit mieux la dynamique vibratoire que la largeur absolue.

6-1-3-1 Dédution des signaux glottiques

Partant de l'équation du mouvement des cordes vocales, nous rappelons ici la démarche suivie par Titze pour déduire la fonction d'aire glottique, la fonction d'aire de contact des cordes vocales et le débit glottique.

6-1-3-1-1 L'aire glottique

Si on considère que la longueur latérale L de la glotte est divisée en séries d'éléments différentiels dy , l'aire glottique élémentaire projetée (figure 6-9) est donc :

$$da(t) = 2 \xi_{\min}(y, t) dy$$

où ξ_{\min} est la valeur positive minimale du déplacement dans le tube différentiel vertical (les valeurs négatives de $\xi(y, z)$ étant mises à zéro, car il y a alors contact).

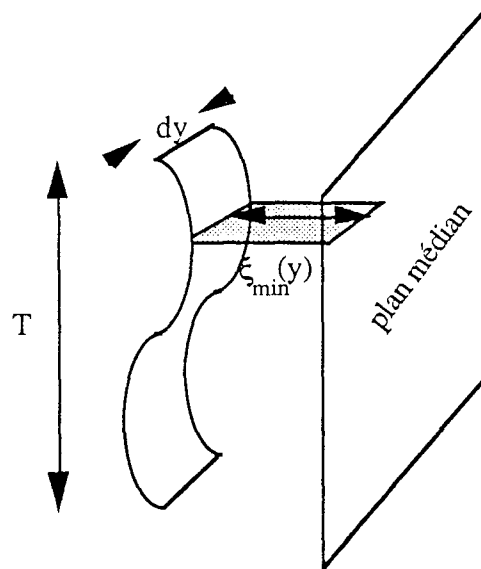


Figure 6-9. Matérialisation de l'aire élémentaire glottique

Ainsi l'aire glottique projetée aux positions minimales verticales (z_{\min}) est

$$a(t) = \int_0^L 2 \xi_{\min}(y, t) dy \quad (6-7)$$

L'évaluation numérique de (6-7) passe par une discrétisation des cordes en M tubes latéraux selon y et N tubes verticaux selon z . Pour un temps discret nT_e (T_e le pas d'échantillonnage) la fonction d'aire peut être définie par

$$a_n = \frac{L}{M} \sum_{j=1}^M 2 \xi_{\min, j, n} \quad (6-8)$$

L'évaluation de la valeur minimale $\xi_{\min, j, n}$ pour chaque $j^{\text{ème}}$ tube vertical peut être localisée par une simple comparaison de N valeurs discrètes des déplacements à l'intérieur du tube ($z = k \Delta z$ avec $k=1 \dots N$). Les positions z_{\min} des déplacements ne sont pas forcément identiques le long de la longueur; si par exemple il existe une grande différence de phase dans le

mouvement du tissu, la position z_{\min} peut être au voisinage du bord antérieur ou postérieur, et l'aire glottique projetée minimale résultante n'est plus adéquate pour le calcul du débit glottique.

Une formulation alternative de l'équation (6-8) est donnée en (6-9), en remplaçant l'aire glottique projetée par le minimum des aires glottiques, calculées pour un plan latéral situé à une hauteur (z) donnée

$$a_n = \min_k \left(\frac{L}{M} \sum_{j=1}^M 2 \xi_{j,k,n} \right) \quad (6-9)$$

Il est clair que les deux formulations de l'aire glottique ne sont pas forcément identiques, cependant nous préférons la formulation (6-9) puisqu'elle fait intervenir la globalité des tubes latéraux et évite donc l'intervention locale des bords antérieur et postérieur de la glotte.

Nous présenterons plus loin des signaux d'aire glottique simulés en fonction des paramètres décrits plus haut.

6-1-3-1-2 L'aire de contact des cordes vocales

Sachant que l'équation de mouvement est écrite par rapport au plan médian de la glotte, deux points de part et d'autre des cordes vocales sont en contact à l'instant t si leur déplacement $\xi(y,z,t)$ est négatif. En prenant la forme discrète des cordes vocales, l'aire de contact est déterminée en sommant les aires partielles ($\Delta y \Delta z = (L/M)(T/N)$) pour lesquelles le déplacement $\xi_{j,k,n}$ est négatif. Cette formulation suppose que si un point de l'aire élémentaire est en contact avec son symétrique, les deux surfaces élémentaires le sont aussi. Pour avoir une précision plus fine de l'aire de contact, une fonction d'interpolation entre les niveaux discrets est nécessaire.

La figure 6-10 représente le principe d'une telle interpolation qui est basée sur des considérations géométriques. En effet, si on considère une surface élémentaire ($\Delta y \Delta z$) en contact partiel, les centres de ces deux surfaces se trouvent sur un axe parallèle à l'axe X , mais l'orientation angulaire des deux surfaces sont arbitraires. Soit ξ la valeur du déplacement des centres (indiquant un chevauchement des deux surfaces), en ne considérant pas les contraintes bio-mécaniques pendant la collision entre les cordes vocales et sous l'hypothèse de la non compressibilité des tissus, le chevauchement est une estimation du premier ordre de l'aire de contact. Les éléments de surfaces élémentaires étant supposés planaires, le minimum de déplacement est obtenu en un des quatre coins (figures 6-10 et 6-11). Les déplacements des

quatre coins peuvent être obtenus par interpolation à partir du déplacement ξ au centre. Le minimum ξ_c peut être localisé par simple comparaison des quatre valeurs

$$\xi_c = \min \left[\xi \pm \frac{\partial \xi}{\partial y} \frac{\Delta y}{2} \pm \frac{\partial \xi}{\partial z} \frac{\Delta z}{2} \right] \quad (6-10)$$

Il y a contact si c est négatif et la surface de contact est alors soit un triangle de côtés δy et δz , soit un rectangle, soit un trapèze, soit un pentagone représenté à la figure 6-11.

Sachant que l'angle entre les directions des déplacements ξ_c et ξ est Θ , les côtés δy et δz sont alors déterminés par

$$\delta y = \min \left[0, -\xi_c / \left| \frac{\partial \xi}{\partial y} \right| \right] \quad (6-11-a)$$

$$\delta z = \min \left[0, -\xi_c / \left| \frac{\partial \xi}{\partial z} \right| \right] \quad (6-11-b)$$

Ainsi l'aire de contact élémentaire est définie par

$$\Delta c = \begin{cases} \frac{1}{2} \delta y \delta z & \text{pour } \delta y < \Delta y \text{ et } \delta z < \Delta z & \text{cas 1} \\ \frac{1}{2} \delta y \delta z - \Delta c_1 & \text{pour } \delta y > \Delta y \text{ et } \delta z < \Delta z \\ \frac{1}{2} \delta y \delta z - \Delta c_2 & \text{pour } \delta y < \Delta y \text{ et } \delta z > \Delta z \\ \frac{1}{2} \delta y \delta z - \Delta c_1 - \Delta c_2 & \text{pour } \delta y > \Delta y \text{ et } \delta z > \Delta z & \text{cas 2} \\ \Delta y \Delta z & \text{pour } \delta y \geq \frac{\Delta y \delta z}{\delta z - \Delta z} \text{ et } \delta z > \Delta z \end{cases} \quad (6-12)$$

avec $\Delta c_1 = \frac{1}{2} (\delta y - \Delta y)^2 \delta z / \delta y$ et $\Delta c_2 = \frac{1}{2} (\delta z - \Delta z)^2 \delta y / \delta z$.

A un instant donné nT_e , l'aire de contact est la somme des aires de contact élémentaires

$$c_n = \sum_{j=1}^M \sum_{k=1}^N \Delta c_{j,k,n} \quad (6-13)$$

Les dérivées partielles intervenant dans (6-11) et (6-12) peuvent être évaluées de façon analytique ou numérique.

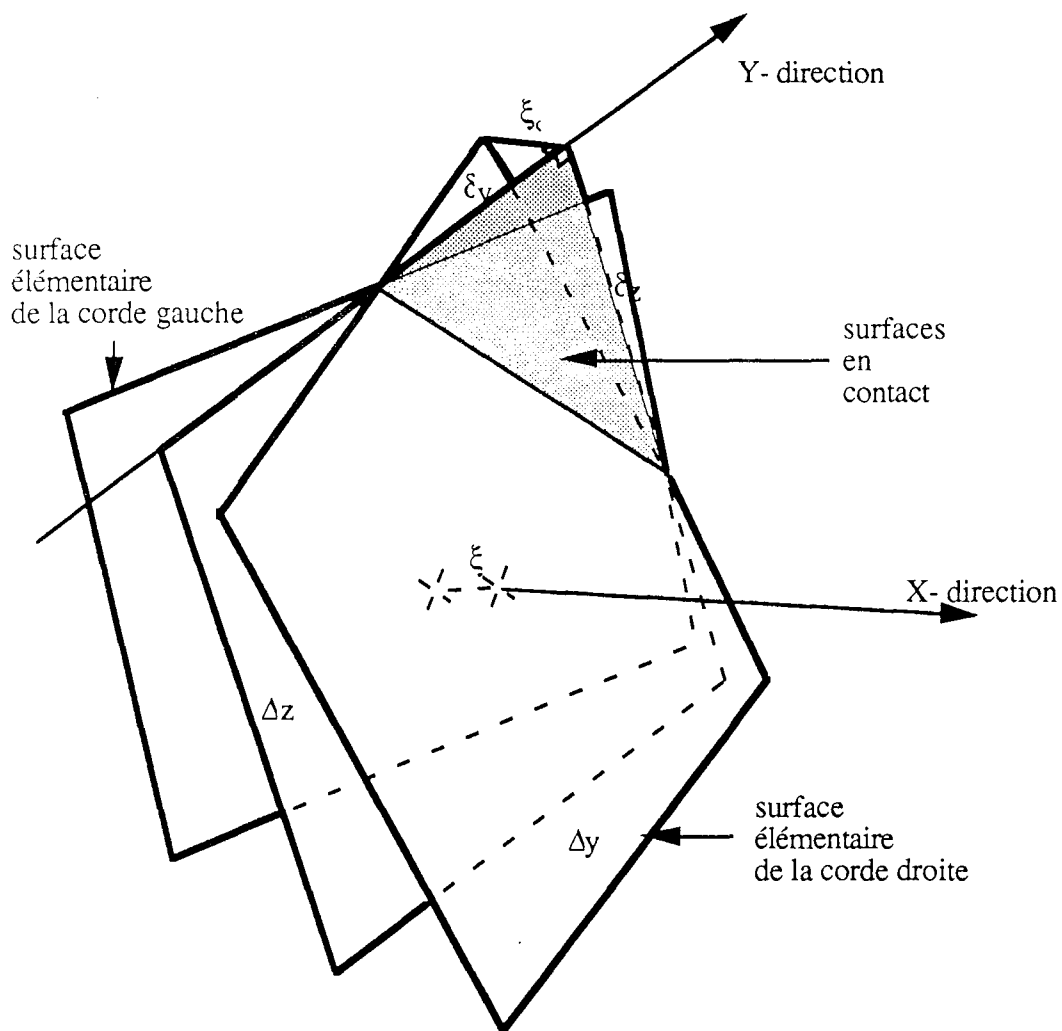


Figure 6-10. Surfaces élémentaires des cordes vocales en contact

Ainsi, à partir de l'équation de mouvement des cordes vocales, nous pouvons déduire le signal aire de contact $a_c(t)$, en négligeant les contraintes mécaniques et en utilisant l'hypothèse de non compressibilité du tissu. Nous présenterons par la suite quelques simulations de ce signal.

6-1-3-1-3 Le signal débit glottique

Considérons maintenant la déduction du débit glottique à partir de l'équation de mouvement dans le cadre du modèle de Titze. Cette déduction intègre le couplage entre la

source et les lignes de transmission correspondant respectivement au conduit vocal et au système sous-glottique (figure 6-12).

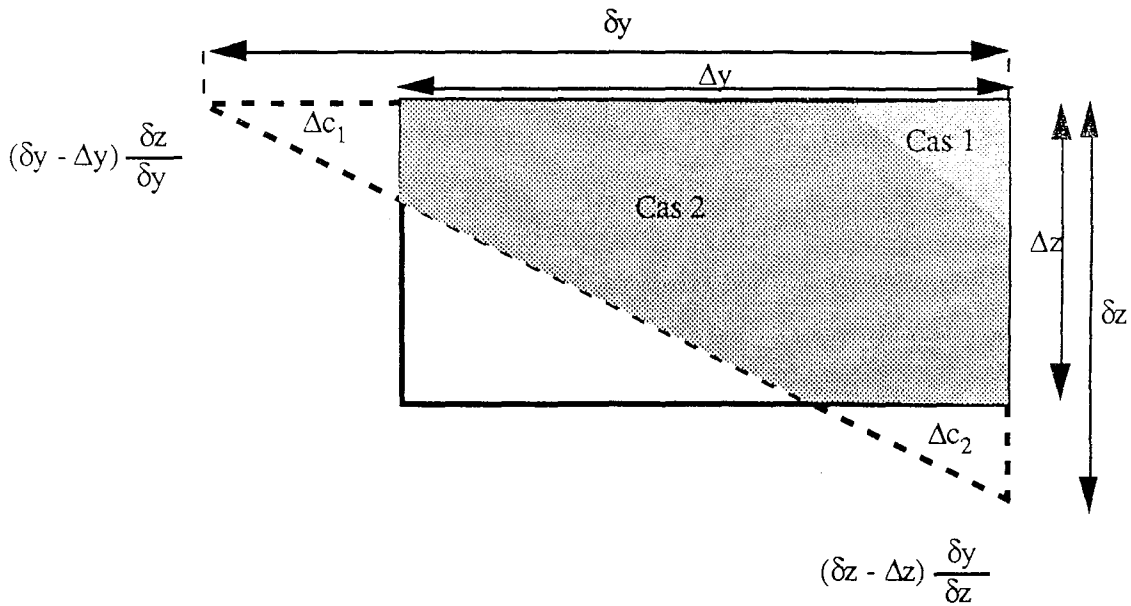


Figure 6-11. Cas d'un chevauchement de type trapézoïdal de deux surfaces élémentaires des cordes vocales

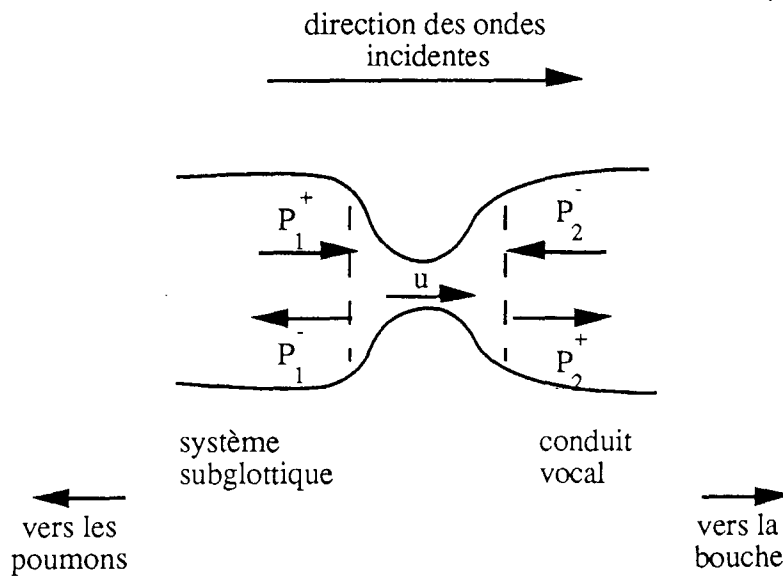


Figure 6-12. Couplage entre la source et les systèmes sous-glottique et sus-glottique

Supposons qu'à chaque instant les pressions P_1^+ et P_2^- soient connues, et prenons les hypothèses suivantes :

- l'inertie de l'air à l'intérieur de la glotte est négligeable,
- les pertes de viscosité sont incluses dans les pertes cinétiques trans-glottiques à travers un coefficient de proportionnalité k .

Dans ces conditions, nous pouvons écrire la relation suivante liant le débit volumique à la différence de pressions trans-glottiques

$$P_1 - P_2 = \left[P_1^+ + P_1^- \right] - \left[P_2^+ + P_2^- \right] = k \frac{1}{2} \rho \left(u \frac{|u|}{a^2} \right) \quad (6-14)$$

où a est l'aire glottique définie précédemment en fonction de l'équation de mouvement des cordes vocales.

En prenant comme hypothèses supplémentaires :

- l'incompressibilité de l'air à l'intérieur de la glotte,
- aucun retard n'est considéré entre les débits entrant et sortant de la glotte, puisque la hauteur des cordes vocales est considérée comme petite par rapport à la vitesse de propagation de l'onde,
- le débit introduit par le mouvement des cordes vocales est négligeable,

les relations de continuité de part et d'autre de la glotte sont alors

$$u = \frac{A_1}{\rho c} \left(P_1^+ - P_1^- \right) \quad (6-15-a)$$

$$u = \frac{A_2}{\rho c} \left(P_2^+ - P_2^- \right) \quad (6-15-b)$$

Ainsi, en éliminant dans les équations (6-13) et (6-14) les pressions inconnues (P_1^- et P_2^+), nous pouvons écrire l'équation définissant le débit glottique

$$k \frac{1}{2} \rho \left(u \frac{|u|}{a^2} \right) + \rho c \left(\frac{1}{A_1} + \frac{1}{A_2} \right) u - 2 \left(P_1^+ - P_2^- \right) = 0 \quad (6-16)$$

Le débit glottique est donc solution de l'équation du second degré (6-16)

$$u = \frac{ac}{k} \left\{ - \frac{a}{A^*} \pm \left[\left(\frac{a}{A^*} \right)^2 + \frac{4k}{c^2 \rho} \left(P_1^+ - P_2^- \right) \right]^{\frac{1}{2}} \right\} \quad (6-17)$$

où A^* est une aire effective combinant les aires sous-glottique et sus-glottique, définie par

$$\frac{1}{A^*} = \frac{1}{A_1} + \frac{1}{A_2}$$

Une fois connu le débit glottique, les pressions réfléchies sous-glottique et sus-glottique peuvent être calculées par

$$P_1^- = P_1^+ - \frac{\rho c}{A_1} u \quad (6-18-a)$$

$$P_2^+ = P_2^- - \frac{\rho c}{A_2} u \quad (6-18-b)$$

Ainsi, dans le cadre du modèle des lignes de transmission (système sous-glottique et conduit vocal) et en considérant le couplage du premier ordre entre la glotte et le conduit vocal, le débit et les pressions glottiques sont déterminés en fonction de l'aire glottique et par conséquent en fonction des paramètres de configuration géométrique pré-phonatoire précédemment définis.

L'expression du débit glottique (6-16) est similaire à l'expression quadratique du débit développée par Fant [Fant, 1982], dans laquelle il considère la forme triangulaire de l'aire glottique.

Deux comportements asymptotiques sont remarquables en fonction des conditions de charge :

- condition de charge nulle (l'aire effective A^* tend vers l'infini) : le débit glottique est proportionnel à l'aire et à la racine carrée de la différence de pressions glottiques incidentes. Notons que si l'amortissement dans le conduit vocal est fort, la pression incidente sus-glottique P_2^- est négligeable et que si la pression sous-glottique est la moitié de la pression au niveau des poumons, le débit glottique est proportionnel à l'aire glottique.

- condition de forte charge ($\frac{a}{A^*} \gg \frac{4k}{\rho c^2} (P_1^+ - P_2^-)$) : le débit glottique est proportionnel à la différence de pressions glottiques incidentes.

Entre ces deux comportements le débit est donné par l'équation (6-17).

Nous pouvons donc introduire un nouveau paramètre caractérisant la charge de la glotte comme le rapport des deux termes de la racine carrée de l'équation (6-17). Cependant, ce paramètre n'est pas constant dans le temps puisqu'il dépend à la fois de la différence de pression et de l'aire glottique. Toutefois, ce quotient de charge peut être considéré asymptotiquement constant si on considère que d'une part la moyenne temporelle de l'aire glottique est proportionnelle à l'amplitude de vibration des cordes vocales et que d'autre part celle de la différence de pressions glottiques est proportionnelle à la pression P_L au niveau des poumons. Dans ces conditions ce coefficient peut être défini par

$$Q_L = K \frac{\xi_m}{[A^*(kP_L)]} \quad (6-19)$$

6-1-3-2 Conclusion sur le modèle de Titze

Nous estimons que le modèle décrit dans ce paragraphe est très intéressant par sa paramétrisation de la configuration pré-phonatoire et des conditions de charge. Nous avons exposé comment à partir de l'équation du mouvement des cordes vocales nous pouvons déduire les principaux signaux glottiques : l'aire glottique, l'aire de contact des cordes vocales et le débit glottique. Nous pensons que les paramètres de ce modèle sont porteurs de l'information sur l'état physiologique des cordes vocales comme en témoigne les différentes simulations réalisées par Titze dans les conditions des différents registres de la voix humaine. Nous avons trouvé une lacune essentielle à ce modèle, pour une caractérisation de pathologies laryngiennes (telles que des paralysies unilatérales ou l'existence d'œdèmes, de polypes ou de nodules, cf. chapitre 2), qui considère un comportement vibratoire parfaitement symétrique des deux cordes vocales. Dans le prochain paragraphe nous allons présenter un nouveau modèle paramétrique de l'activité glottique dérivé de celui de Titze pour lequel nous introduisons de nouveaux paramètres de symétrie.

6-1-4 Nouveau modèle paramétrique de l'activité glottique

Comme pour le modèle précédent, nous prenons les mêmes hypothèses de travail concernant la forme verticale en phase pré-phonatoire, les pertes de viscosité, l'incompressibilité de l'aire et le temps de propagation nul des ondes dans la glotte. Nous allons écrire les équations de mouvement de chaque corde vocale, exposer les différents paramètres introduits et les principales modifications de déduction des signaux glottiques. Nous illustrons à la fin de ce paragraphe ce modèle par un ensemble de simulations et nous discuterons particulièrement les cas de forte asymétrie.

6-1-4-1 Equations de mouvement et différents paramètres

Nous prenons comme précédemment un repère (ox,oy,oz), où le plan (yoz) est le plan médian (du corps humain), l'axe oy a la direction postéro-antérieure et l'axe des x est parallèle à la direction latérale. Nous considérons aussi que la longueur L et la hauteur T est la même pour

les deux cordes vocales (aucune dissymétrie de dimensions des cordes vocales). Nous prenons dans la suite les indices d et g respectivement pour la corde vocale droite et gauche comme indiqué sur la figure 6-13.

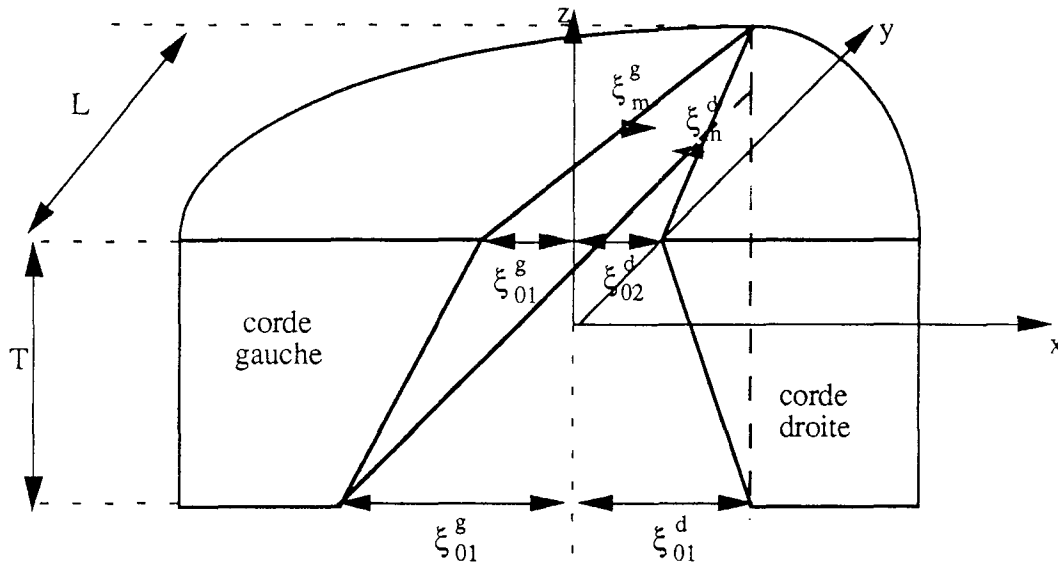


Figure 6-13. Configuration pré-phonatoire des deux cordes vocales

Sous les conditions des hypothèses ci-dessus, les positions pré-phonatoires verticales de chaque corde sont déterminées par

$$\xi_0^d(y,z) = \left\{ \xi_{01}^d - \frac{\xi_{01}^d - \xi_{02}^d}{T} z \right\} \left\{ 1 - \frac{y}{L} \right\} \quad (6-20-a)$$

$$\xi_0^g(y,z) = \left\{ \xi_{01}^g - \frac{\xi_{01}^g - \xi_{02}^g}{T} z \right\} \left\{ 1 - \frac{y}{L} \right\} \quad (6-20-b)$$

Notons que l'axe (ox) pour chacun des déplacements ou des positions, sera toujours pris dans le sens plan médian vers la corde considérée. Cette convention nous évite des déplacements de signes contraires pour chacune des cordes et nous permet la déduction des signaux glottiques, notamment l'aire glottique, comme précédemment.

Les déplacements relatifs des cordes vocales par rapport à leur position initiale sont

$$\xi_1^d(y,z,t) = \xi_m^d \sin\left(\frac{\pi y}{L}\right) \sin\left(2\pi ft - \frac{\phi^d z}{T}\right) \quad (6-21-a)$$

$$\xi_1^g(y,z,t) = \xi_m^g \sin\left(\frac{\pi y}{L}\right) \sin\left(2\pi ft - \frac{\phi^g z}{T}\right) \quad (6-21-b)$$

où ϕ^d et ϕ^g sont les déphasages entre les bords supérieurs et inférieurs de chaque corde.

Les équations de mouvement des deux cordes vocales par rapport au plan médian sont donc

$$\xi^d(y,z,t) = \xi_m^d \left\{ \left[\frac{\xi_{01}^d}{\xi_m^d} - \frac{\xi_{01}^d - \xi_{02}^d}{T\xi_m^d} z \right] \left[1 - \frac{y}{L} \right] + \sin\left(\frac{\pi y}{L}\right) \sin\left(2\pi ft - \frac{\phi^d z}{T}\right) \right\} \quad (6-22-a)$$

$$\xi^g(y,z,t) = \xi_m^g \left\{ \left[\frac{\xi_{01}^g}{\xi_m^g} - \frac{\xi_{01}^g - \xi_{02}^g}{T\xi_m^g} z \right] \left[1 - \frac{y}{L} \right] + \sin\left(\frac{\pi y}{L}\right) \sin\left(2\pi ft - \frac{\phi^g z}{T}\right) \right\} \quad (6-22-b)$$

6-1-4-2 Définitions des paramètres glottiques

Une première approche de définition des paramètres consiste à définir les mêmes paramètres que précédemment pour chacune des cordes, à savoir

$$\text{- Quotient d'abduction droit : } Q_a^d = \frac{\xi_{02}^d}{\xi_m^d} \quad (6-23-a)$$

$$\text{- Quotient d'abduction gauche : } Q_a^g = \frac{\xi_{02}^g}{\xi_m^g} \quad (6-23-b)$$

$$\text{- Quotient de forme droit : } Q_s^d = \frac{\xi_{01}^d - \xi_{02}^d}{\xi_m^d} \quad (6-24-a)$$

$$\text{- Quotient de forme gauche : } Q_s^g = \frac{\xi_{01}^g - \xi_{02}^g}{\xi_m^g} \quad (6-24-b)$$

$$\text{- Quotient de phase droit : } Q_p^d = \frac{\phi^d}{2\pi} \quad (6-25-a)$$

$$\text{- Quotient de phase gauche : } Q_p^g = \frac{\phi^g}{2\pi} \quad (6-25-b)$$

Ainsi, nous disposons de six paramètres sans dimension et de deux paramètres supplémentaires correspondant à l'amplitude de vibration de chaque corde.

Pour ne disposer que de paramètres sans dimension et dans la mesure du possible non corrélés, nous avons introduit des paramètres caractérisant la symétrie des cordes vocales du point de vue de l'amplitude de vibration, de la forme et de la phase. Nous donnons les définitions de ces paramètres

$$\text{- Quotient de symétrie d'amplitude : } Q_{\text{sya}} = \frac{\xi_m^d}{\xi_m^g} \quad (6-26)$$

$$\text{- Quotient de symétrie de la partie inférieure : } Q_{\text{sinf}} = \frac{\xi_{01}^d}{\xi_{01}^g} \quad (6-27)$$

$$\text{- Quotient de symétrie de la partie supérieure : } Q_{\text{syup}} = \frac{\xi_{02}^d}{\xi_{02}^g} \quad (6-28)$$

$$\text{- Quotient de symétrie de phase : } Q_{\text{syp}} = \frac{\phi^d}{\phi^g} \quad (6-29)$$

Ainsi, l'une des deux cordes (la corde gauche par exemple) étant prise comme référence pour les deux paramètres Q_a et Q_s , les mêmes paramètres de l'autre corde s'expriment en fonction des quotients de symétrie de la manière suivante

$$Q_a^d = Q_a^g \frac{Q_{\text{syup}}}{Q_{\text{sya}}} \quad (6-30)$$

$$Q_s^d = \frac{Q_{\text{sinf}} - Q_{\text{syup}}}{Q_{\text{sya}}} Q_a^g + \frac{Q_{\text{sinf}}}{Q_{\text{sya}}} Q_s^g \quad (6-31-a)$$

ou encore

$$Q_{\text{sinf}}^d = \frac{Q_{\text{sya}} Q_s^d + Q_{\text{syup}} Q_a^g}{Q_a^g + Q_s^g} \quad (6-31-b)$$

$$Q_p^d = Q_{\text{syp}} Q_p^g \quad (6-32)$$

Des dix paramètres définis ci-dessus, (6-23) à (6-29), nous avons choisi les sept paramètres

- le quotient d'abduction gauche,
- les deux quotients de forme droite et gauche,
- le quotient de phase gauche,
- le quotient de symétrie d'amplitude,
- le quotient de symétrie d'abduction (symétrie de la partie supérieure),
- le quotient de symétrie de phase.

Les autres se déduisent des relations (6-30) à (6-32).

Pour simplifier les écritures nous ne précisons pas l'indice gauche et droite sauf pour les deux quotients de forme.

Récrivons maintenant les équations de mouvement en fonction des paramètres choisis

$$\xi^g(y,z,t) = \xi_m \left\{ \left[Q_a + Q_s^g - Q_s^g \frac{z}{T} \right] \left[1 - \frac{y}{L} \right] + \sin\left(\frac{\pi y}{L}\right) \sin 2\pi \left(ft - \frac{Q_p}{T} z \right) \right\} \quad (6-33-a)$$

$$\xi^d(y,z,t) = Q_{sya} \xi_m \left\{ \left[Q_a \frac{Q_{syp}}{Q_{sya}} + Q_s^d - Q_s^d \frac{z}{T} \right] \left[1 - \frac{y}{L} \right] + \sin\left(\frac{\pi y}{L}\right) \sin 2\pi \left(ft - \frac{Q_{syp}}{T} \frac{Q_p}{T} z \right) \right\} \quad (6-33-b)$$

6-1-4-3 Relation entre les paramètres et le geste phonatoire pathologique

En premier lieu, précisons que nous nous sommes limité à des gestes phonatoires pathologiques dont l'étiologie est de type laryngienne. Le geste est donc perturbé par une atteinte au niveau du larynx, comme par exemple une paralysie unilatérale, un phénomène dysorganique lésionnel ou une dysfonction provenant d'une atteinte des voies nerveuses.

Nous supposons que pour une phonation normale, les paramètres de symétrie sont tous très voisins de 1, autrement dit, il existe une parfaite symétrie des mouvements des deux cordes. Dans le cas par exemple d'une paralysie unilatérale, l'amplitude de vibration de la corde paralysée est forcément très inférieure à celle de la corde saine; le paramètre quotient de symétrie sera nettement supérieur à 1 si la corde de référence est celle paralysée et très inférieur à 1 dans le cas contraire.

Le paramètre de symétrie de phase est aussi important, puisqu'il permet de caractériser la réaction d'une corde vocale par rapport à l'autre, au moment de départ de la phonation; il fournit donc une mesure comparative de la capacité de la mise en vibration des deux cordes. Dans le cadre de la théorie neuro-chronaxique [Husson, 1950], le paramètre de symétrie de phase mesure aussi, d'une manière différentielle, la réaction des muscles vocaux à l'excitation nerveuse. Ce paramètre sera probablement affecté dans le cas d'une atteinte nerveuse unilatérale (comme par exemple une hémiplégie).

Nous pensons que le quotient de symétrie d'abduction peut indiquer la présence d'une tumeur bénigne ou maligne localisée au voisinage de la partie supérieure d'une corde vocale. En effet, si les amplitudes de vibration des deux cordes sont les mêmes ($Q_{sya} \approx 1$), dans le cas d'une présence d'atteinte lésionnelle localisée sur la partie supérieure d'une corde, les positions

pré-phonatoires supérieures (ξ_{01}^d et ξ_{01}^g) ne seront pas identiques; le paramètre quotient de symétrie d'abduction sera donc très différent de 1.

6-1-4-4 Dédution des signaux glottiques du modèle paramétrique

Dans le cas d'une activité glottique "anormale", les signaux glottiques peuvent être affectés par la dissymétrie. Cependant au niveau de la déduction théorique, le développement calculatoire sous les mêmes hypothèses reste le même que pour le modèle paramétrique de Titze. Le calcul de la fonction d'aire glottique consiste à remplacer la relation (6-9) (version discrète de l'aire glottique M tubes latéraux, N tubes verticaux et à l'instant nT_e) par

$$a_n = \min_k \left(\frac{L}{M} \sum_{j=1}^M \xi_{j,k,n}^d + \xi_{j,k,n}^g \right) \quad (6-34)$$

La fonction d'aire de contact discrète et sans interpolation peut être calculée en sommant les aires de surfaces élémentaires en contact. Ces surfaces sont en contact lorsque les déplacements gauche et droite de leur centre sont tous les deux négatifs. Dans le cas où nous désirons une précision plus fine de l'aire de contact, l'interpolation linéaire basée sur les mêmes considérations géométriques peut être appliquée pour chaque corde. L'aire de contact élémentaire donnée par (6-12) reste valable à condition de remplacer δy , δz , Δc_1 et Δc_2 par les valeurs minimales de leur correspondant pour chaque corde : $\delta y = \min(\delta y^d; \delta y^g)$, $\delta z = \min(\delta z^d; \delta z^g)$,

$\Delta c_1 = \min(\Delta c_1^d; \Delta c_1^g)$ et $\Delta c_2 = \min(\Delta c_2^d; \Delta c_2^g)$ où les δy^i , δz^i , Δc_1^i et Δc_2^i sont données par les équations (6-10) à (6-12) en remplaçant ξ par ξ^i (i valant g : gauche, ou d : droite).

Le signal débit glottique, ne dépendant pas directement des différents paramètres, a la même expression (6-17) où l'aire glottique a est calculée en fonction de tous les paramètres du modèle proposé.

6-1-4-5 Simulations des signaux glottiques dans le cadre du modèle proposé

Nous avons procédé à un certain ensemble de simulations de signaux glottiques, pour estimer le bien-fondé de notre choix des paramètres de symétrie d'une part et pour étudier leurs

influences mutuelles. Nous décrivons dans ce paragraphe les différentes conditions de la simulation et nous commenterons, à la fin, les résultats obtenus.

Pour l'obtention des différents signaux glottiques, en particulier le débit glottique, nous avons besoin de simuler des conditions de charge de la glotte, et par conséquent nous sommes amenés à simuler le système de propagation de l'onde acoustique dans les tubes sous-glottiques d'une part, et le conduit vocal d'autre part. Pour le tube sous-glottique, nous avons pris quatre tubes acoustiques uniformes dont la section est connue. Pour le conduit vocal, nous nous sommes servis d'une fonction d'aire correspondant à une voyelle donnée par Mrayati [Mrayati, 1976]. Le système complet débouche sur une pression constante au niveau des lèvres P_L .

Les conditions de la simulation sont les suivantes :

- fréquence d'échantillonnage : 10 kHz,
- fréquence fondamentale : 120 Hz, correspondant à une voix masculine moyenne,
- pression au niveau des poumons : 10 cm H₂O ou 2,5 cm H₂O, correspondant à une voix d'intensité respectivement forte et faible
- amplitude de vibration des cordes vocales : 0,075 cm ou 0,3 cm.

Dans un premier temps, nous avons généré les trois signaux glottiques, le débit, l'aire d'ouverture et l'aire de contact, pour plusieurs valeurs des paramètres pour comparer d'une façon visuelle les deux configurations des cordes vocales : symétrique et asymétrique. Un exemple est donné à la figure 6-14.

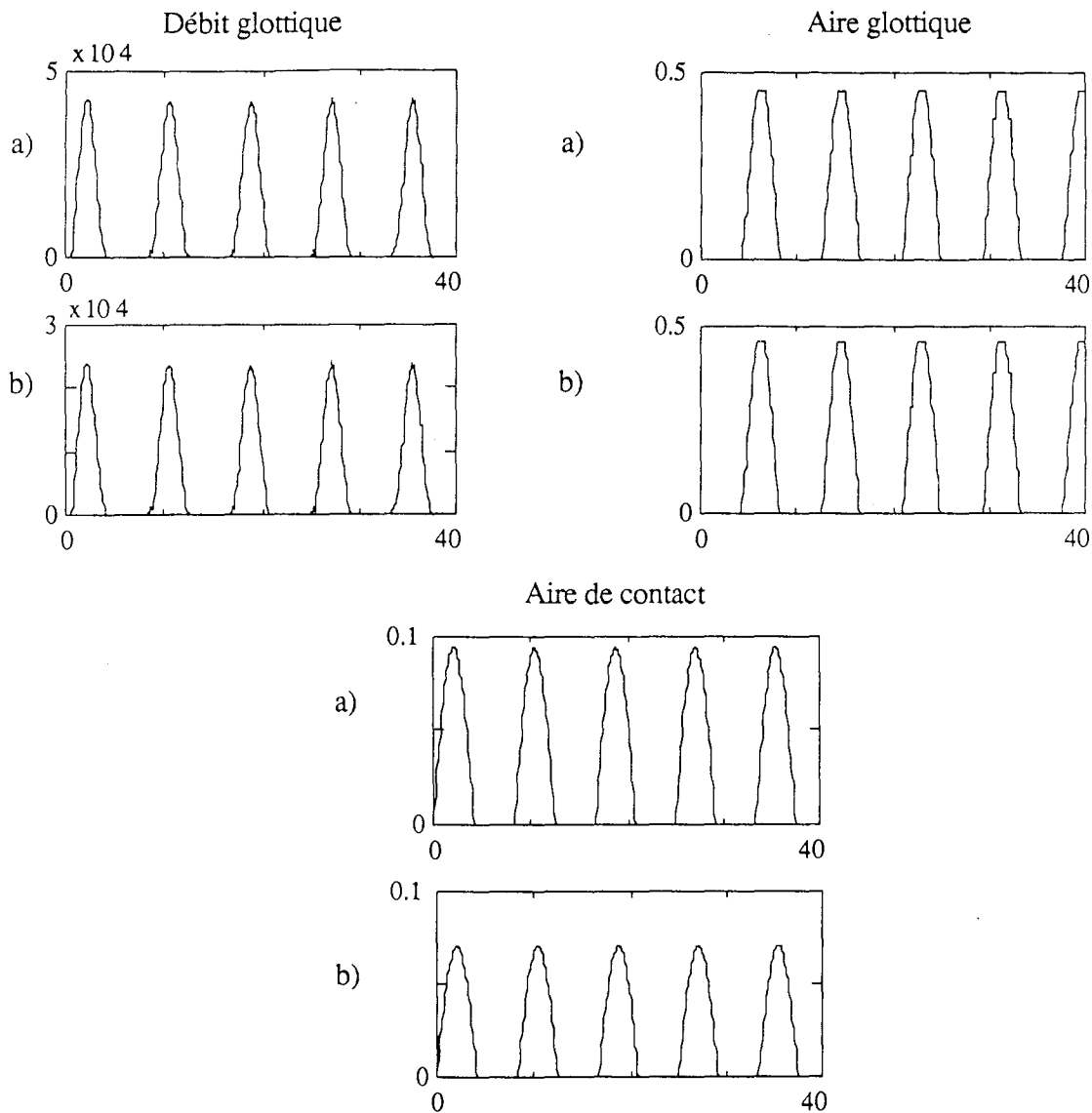


Figure 6-14. Signaux glottiques simulés

$$\xi_m = 0.075 \text{ cm} \quad PL = 10 \text{ cm H}_2\text{O}$$

$$Q_a = 0, \quad Q_s^d = 3, \quad Q_s^g = 3, \quad Q_p = 0$$

a) cas d'une configuration symétrique

b) dissymétrie d'amplitude : $Q_{sya} = 0.5$

$$(Q_{sys} = 1, Q_{syp} = 1)$$

Après examen des résultats de la simulation (voir courbes Annexe I), nous pouvons tirer les premières conclusions concernant l'introduction des paramètres de symétrie :

- pour le signal débit glottique, la forme de l'onde reste la même dans les configurations symétriques ou asymétriques; seule l'amplitude du signal diminue dans le cas asymétrique,
- pour l'aire glottique, la même constatation peut être faite,
- seule la forme de la fonction aire de contact est affectée

6-2 Extraction des paramètres de l'onde glottique

Ce paragraphe présente la procédure d'extraction des paramètres de l'activité glottique partant du signal de parole réel. Comme nous l'avons indiqué plus haut, le signal débit glottique n'est pas accessible directement par des moyens non invasifs. Nous consacrerons donc la première partie du paragraphe à la description des techniques, appelées "techniques de filtrage inverse", d'extraction de l'onde glottique partant du signal de parole.

Nous nous sommes intéressés à deux catégories de paramètres de l'activité glottique qui, à notre avis, permettent de caractériser la physiologie du larynx : les paramètres temporels et les paramètres du modèle cinétique que nous appelons les paramètres de configuration. Nous décrivons en second lieu les procédures utilisées pour mesurer ces paramètres.

6-2-1 Techniques de filtrage inverse

6-2-1-1 Position du problème

Nous avons vu (chapitres 3 et 4) que le signal de parole est la sortie d'un système physique (conduits vocal et nasal) dont l'entrée est le signal débit glottique. Nous avons présenté un ensemble de modèles mathématiques (souvent dans des conditions plus ou moins réalistes) du système de production. Le modèle numérique le plus souvent utilisé est un filtre linéaire dont les fréquences de résonance coïncident avec les formants du signal de parole. Nous avons montré aussi (chapitre 4) l'intérêt du filtre numérique tous pôles pour la production des sons non nasalisés, dont le plus grand avantage est sa liaison avec les techniques de la prédiction linéaire.

Un morceau du signal de parole supposé stationnaire, contient à la fois des informations sur l'excitation (onde glottique) et sur le filtre de production (conduit vocal). Les techniques de filtrage inverse tentent de résoudre le problème de la séparation des deux types d'informations. En termes de formalisme mathématique, si on considère le signal de parole $s(t)$, la réponse impulsionnelle du filtre modèle du conduit vocal $h(t)$ et l'excitation $g(t)$ comme des éléments de

l'espace des fonctions, dans le cadre du modèle linéaire, le problème ci-dessus revient à résoudre l'équation fonctionnelle suivante

$$g * h = s$$

où les inconnues sont à la fois les fonctions h et g .

Les méthodes de filtrage inverse que nous allons présentées sont des méthodes de déconvolution qui supposent connues à priori une des deux fonctionnelles, principalement la réponse impulsionnelle.

6-2-1-2 Méthode de filtrage inverse par analyse dans la phase de fermeture

6-2-1-2-1 Principe de la méthode

Due à Wong [Wong, 1979], cette méthode repose sur la prédiction linéaire du signal de parole, où le modèle d'analyse est recherché en synchronisme avec la période fondamentale (pitch synchronous analysis). Nous avons montré que le système de production des sons non nasalisés peut être représenté par le modèle donné à la figure 6-15

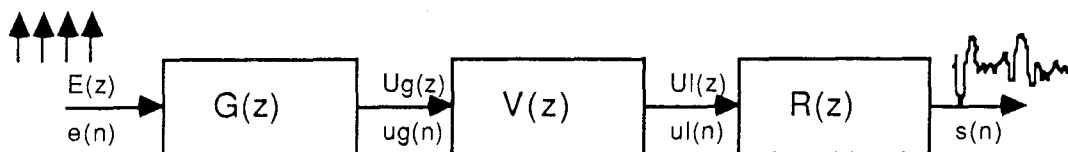


Figure 6-15. Modèle de production de la parole

où $E(z)$ ----- $e(n)$ modèle d'excitation de la glotte
 $U_g(z)$ ----- $u_g(n)$ signal débit glottique
 $U_L(z)$ ----- $u_L(n)$ signal de la vitesse volumique aux lèvres
 $S(z)$ ----- $s(n)$ signal acoustique de la parole.

$G(z)$ est le modèle du filtre de la glotte dont l'entrée $e(n)$ est un train impulsionnel pour les sons voisés.

$V(z)$ est le modèle du conduit vocal supposé être un modèle "tous-pôles" de la forme

$$V(z) = \frac{1}{1 + \sum_{i=1}^K c_i z^{-i}} \quad (6-35)$$

Quand l'ordre K est pair, $K/2$ formants sont inclus dans le modèle.

$R(z)$ est le filtre correspondant aux effets de rayonnement au niveau des lèvres. $R(z)$ est de la forme

$$R(z) = 1 - \mu z^{-1}$$

où μ est un facteur de retard dépendant de la distance du capteur aux lèvres.

Puisque $R(z)$ est le même pour les différents sons de la parole, le problème de l'estimation de l'onde glottique revient à déterminer les paramètres du filtre inverse $1/V(z)$. Le débit glottique s'obtient alors par filtrage inverse du signal de parole.

L'estimation des paramètres du filtre se fait par l'analyse du signal de parole selon la méthode de covariance pendant la phase de fermeture glottique. Le choix de la méthode de covariance se justifie par le fait que le nombre d'échantillons de la fenêtre d'analyse n'est pas suffisant pour supposer la stationarité du signal qui est nécessaire pour la méthode d'autocorrélation (partie II- chapitre 4). En effet, durant la phase de fermeture glottique, le conduit vocal n'étant plus excité, le signal de sortie se trouve dans un régime oscillatoire libre ne dépendant que des résonances du conduit vocal. Le problème le plus crucial de cette méthode réside dans la localisation des instants d'ouverture et de fermeture de la glotte. Wong utilise, comme nous le montrerons plus bas, l'erreur quadratique normalisée de la prédiction linéaire pour la localisation de ces deux instants.

Si on intègre l'effet de rayonnement par les lèvres dans le filtre glottique le système de production peut être représenté par la figure 6-16.

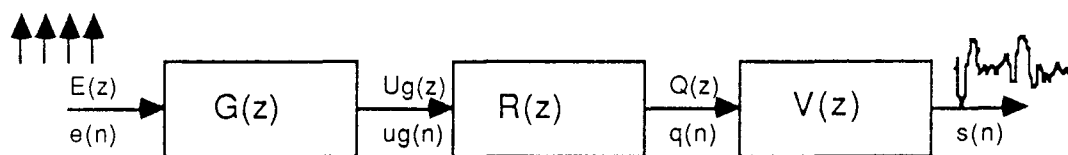


Figure 6-16. Modèle modifié de production de la parole

Le signal de parole $s(n)$ s'écrit alors en fonction de la fonction auxiliaire $q(n)$ sous la forme

$$s(n) = q(n) - \sum_{i=1}^K c_i s(n-i) \quad (6-36)$$

où $q(n)$ est le produit de convolution de $u_g(n)$ par la réponse impulsionnelle $r(n)$ du filtre $R(z)$

$$q(n) = u_g(n) * r(n)$$

ou encore

$$Q(z) = U_g(z) R(z)$$

Si n est l'instant de la phase de fermeture, le signal d'excitation $u_g(n)$ et par conséquent $q(n)$ étant nuls, l'équation (6-36) devient alors

$$s(n) = - \sum_{i=1}^K c_i s(n-i) \quad \text{pour } L_c \leq n \leq L_o \quad (6-37)$$

où L_c et L_o sont respectivement les instants de fermeture et d'ouverture de la glotte.

Soit $A(z)$ le filtre d'analyse d'ordre p ($p \geq K$) obtenu par la prédiction linéaire

$$A(z) = - \sum_{i=1}^p a_i z^{-i} \quad (a_0 = 1) \quad (6-38)$$

Le résidu $e(n)$ ou signal d'erreur de prédiction est donné par

$$e(n) = s(n) + \sum_{i=1}^p a_i s(n-i) \quad (6-39)$$

Ainsi, si les conditions suivantes

$$\begin{cases} a_i = c_i & \text{pour } i = 1, \dots, K \\ a_i = 0 & \text{pour } i = K+1, \dots, p \end{cases} \quad (6-40)$$

sont satisfaites, le résidu $e(n)$ est nul sur toute la phase de fermeture ($L_c \leq n < L_o$).

Le filtre $A(z)$ est obtenu par la méthode de covariance de la prédiction linéaire, qui minimise l'erreur quadratique sur une fenêtre d'analyse de longueur N ($[n-p, n+N-p-1]$)

$$\alpha_p(n) = \sum_{j=n-p}^{n-p+N-1} e^2(j) \quad (6-41)$$

Si les conditions (6-40) sont satisfaites (le filtre d'analyse modélise bien le conduit vocal) et si les bords de la fenêtre d'analyse se situent dans une phase de fermeture ($n \geq L_c + 1$ et $n+N-p < L_o$), l'erreur quadratique $\alpha_p(n)$ est alors nulle.

Dans ces conditions, les instants de fermeture et d'ouverture sont localisés par l'examen de l'erreur quadratique totale : le premier échantillon n_1 où $\alpha_p(n_1)$ est nul correspond à $L_c + 1$ et au premier échantillon n_2 où $\alpha_p(n_2)$ est non nul correspond à $L_o + p + 1 - N$.

En fait, les conditions d'optimalité (6-40) ne sont jamais satisfaites, puisque d'une part le modèle linéaire ne peut coïncider que partiellement avec la fonction de transfert du conduit vocal et d'autre part ce modèle dépend de l'algorithme utilisé. L'erreur quadratique n'est alors jamais nulle dans la phase de fermeture. Nous sommes donc contraint à considérer des conditions plus larges pour la détermination des instants d'ouverture et de fermeture de la glotte, et du "meilleur modèle" (noté par la suite $A_b(z)$). Pour la localisation des instants de fermeture et d'ouverture glottiques, nous comparons l'erreur quadratique normalisée $\alpha_p(n) / \alpha_0(n)$ à un seuil donné (α_0 est l'énergie du signal). Il apparaît :

- l'instant n_1 pour lequel l'erreur quadratique normalisée devient inférieure au seuil,
- l'instant n_2 pour lequel l'erreur quadratique normalisée devient supérieure au seuil.

Les instants de fermeture L_c et d'ouverture L_o sont alors (figure 6-17)

$$L_c = n_1 - 1 \quad (6-42-a)$$

$$L_o = n_2 + N - p - 1 \quad (6-42-b)$$

Le modèle du conduit vocal choisi est celui calculé sur la fenêtre d'analyse se situant à l'intérieur de la phase de fermeture et dont l'erreur quadratique normalisée est minimale.

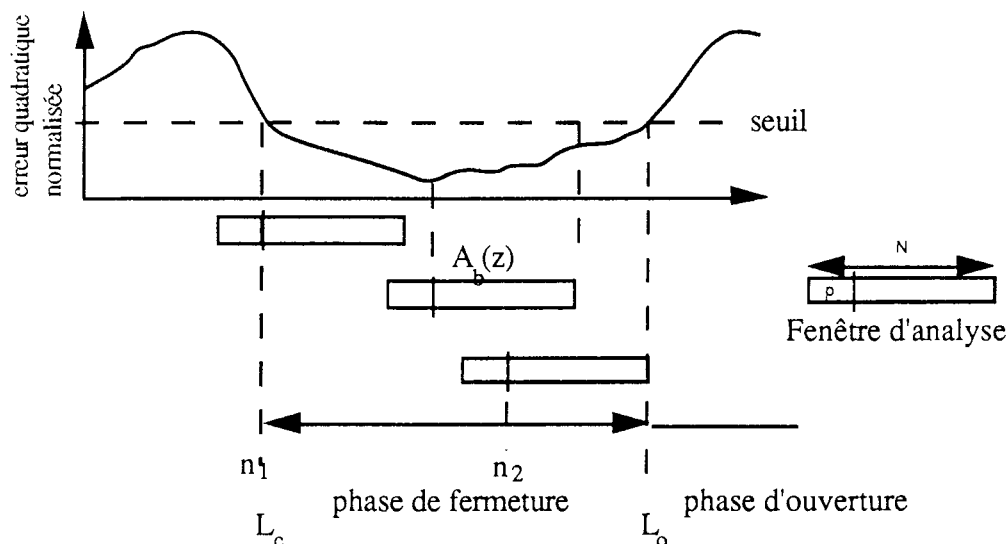


Figure 6-17. Localisation des instants de fermeture et d'ouverture glottiques

par examen de l'erreur quadratique normalisée

Le débit glottique estimé sur tout le cycle (ouverture + fermeture) est donné par les relations suivantes

$$U_g(z) = \frac{Q(z)}{1 - \mu z^{-1}} \quad (6-43-a)$$

$$Q(z) = \frac{S(z)}{A_b(z)} \quad (6-43-b)$$

6-2-1-2-2 Procédure d'analyse et résultats

La structure de la procédure de l'estimation du débit glottique basée sur l'analyse décrite ci-dessus est représentée à la figure 6-18.

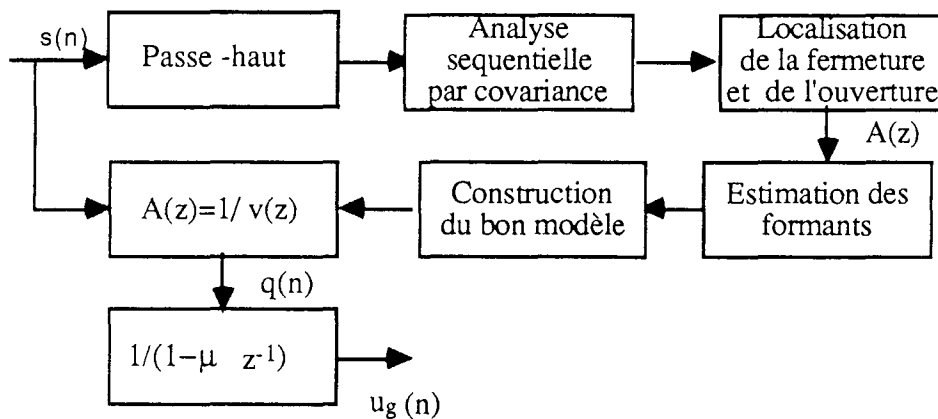


Figure 6-18. Schéma fonctionnel de l'extraction de l'onde glottique

Etant donné un segment stable d'un signal de parole voisée, la procédure d'extraction commence par un filtrage passe-haut à phase linéaire, de fréquence de coupure voisine de 50 Hz. L'utilisation de ce filtre est justifiée par la compensation des distorsions de basse fréquence introduite par le microphone ou par les conditions d'enregistrement du signal. La phase linéaire du filtre est nécessaire pour assurer le synchronisme entre le signal de parole et le débit glottique estimé, car il n'introduit qu'un décalage temporel entre les signaux d'entrée et de sortie.

Une analyse autorégressive du signal est effectuée sur une fenêtre, de longueur N , glissante sur tout le signal. La longueur de la fenêtre doit être petite (de l'ordre de 4,8 ms), ce qui justifie l'emploi de la méthode de covariance de la prédiction linéaire. Pour chaque fenêtre, l'erreur quadratique normalisée (erreur quadratique de la prédiction/énergie du signal) est

comparée au seuil "de fermeture". Ainsi, les instants de fermeture L_c et d'ouverture L_o sont déterminés comme indiqué précédemment.

Une fois que l'instant de fermeture est localisé, la recherche du modèle optimal est déclenchée. Nous prenons comme modèle du conduit vocal (y compris le filtre de rayonnement par les lèvres) celui pour lequel l'erreur quadratique normalisée est minimale dans la phase de fermeture.

Le fait d'utiliser la méthode de covariance pour le calcul du modèle, n'assure pas la stabilité du filtre prédicteur associé [Markel, 1972]. Nous testons la stabilité du modèle en examinant la position de ses pôles par rapport au cercle unité (tous les pôles à l'intérieur du cercle assure la stabilité du filtre de synthèse $1/A_b(z)$). Dans le cas où le filtre de synthèse est instable, nous avons adopté deux stratégies : rendre le filtre stable par la méthode de la factorisation spectrale qui consiste à remplacer les pôles à l'extérieur du cercle unité par leur inverse ou continuer l'analyse séquentielle jusqu'à trouver un modèle stable.

Au niveau pratique, nous déterminons les pôles du modèle par l'extraction des racines d'un polynôme d'ordre p (dénominateur du modèle). La méthode numérique utilisée est la méthode de Bairstow [Meloni, 1985]. Dans le cas de l'instabilité, les pôles externes au cercle unité sont remplacés par leur inverse et le dénominateur du modèle est le polynôme dont les racines sont les nouveaux pôles. Notons que les pôles se présentant toujours par paires de complexes conjugués, le polynôme calculé est de coefficients réels.

A la fin de la procédure, nous appliquons le filtre d'analyse au signal de parole filtré pour extraire le signal auxiliaire $q(n)$ comme indiqué par (6-43-b). Le débit glottique est alors obtenu par intégration du signal auxiliaire.

Un exemple de filtrage inverse effectué sur un signal numérique de parole (voyelle /e/ prononcée par un locuteur masculin et échantillonnée à 10 kHz 12 bits) est donné à la figure 6-19.

*

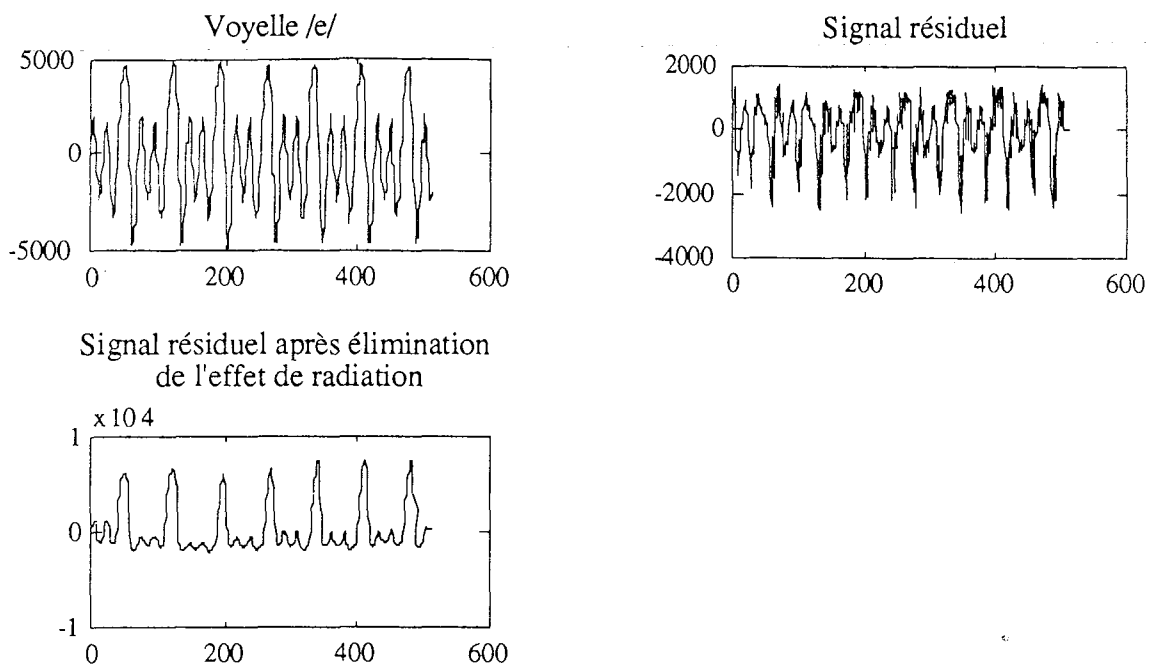


Figure 6-19. Extraction de l'onde glottique par filtrage inverse d'une voyelle soutenue /e/

6-2-1-2-3 Discussion

Malgré la simplicité relative de cette méthode, elle ne demeure point sans inconvénient. Puisque le point clé de cette méthode est la localisation de la phase de fermeture, le choix du seuil de "fermeture" est capital. Différentes méthodes peuvent être utilisées pour déterminer ce seuil :

- en parallèle à l'analyse, nous pouvons calculer le déterminant de la matrice de covariance du signal de parole sur la fenêtre considérée. D'après Strube [Strube, 1974], ce déterminant est maximal quand le début de la fenêtre coïncide avec l'instant de fermeture. Ainsi, le seuil peut être fixé à la valeur de l'erreur quadratique normalisée de la fenêtre précédente d'analyse. Ce seuil n'a pour but, dans ce cas, que la détermination de l'instant d'ouverture. Malgré des résultats satisfaisants de la localisation des instants de fermeture sur des signaux synthétiques par la méthode proposée par Strube, nous avons abandonné cette voie pour deux raisons : d'une part, le lien entre le déterminant de la matrice de covariance avec l'instant de fermeture n'a pu être justifié et d'autre part cette technique est encore plus coûteuse au niveau du temps de calcul (pour chaque fenêtre d'analyse, il faut calculer à la fois le modèle AR avec l'erreur quadratique et le déterminant de la matrice de covariance associée de dimension $(p+1) \times (p+1)$).

- Une deuxième méthode, consiste à déterminer le seuil pour chaque période en examinant la distribution des erreurs quadratiques normalisées pendant un cycle. En supposant que la durée de la phase d'ouverture est égale à celle de la phase de fermeture, le seuil sera choisi comme la valeur de l'erreur quadratique dont la densité de probabilité cumulée est voisine de 0,5. Notons que le choix de la valeur de la densité de probabilité dépend du rapport phase d'ouverture/durée du cycle (exemple : phase d'ouverture/durée du cycle = 1/4, le seuil sera choisi comme la valeur de l'erreur quadratique dont la densité de probabilité cumulée est égale à 0,75). L'inconvénient majeur de cette méthode est qu'elle part d'une hypothèse sur la durée de l'ouverture (et par conséquent sur la durée de fermeture). Un autre problème inhérent à cette méthode, reste celui d'isoler un cycle avant de commencer l'analyse. Ceci revient à détecter d'abord la fréquence fondamentale, avant de procéder à l'analyse. Nous décrirons plus loin les méthodes utilisées pour la détection du pitch, et une méthode originale de détection des instants de fermeture et d'ouverture glottique.

La procédure d'analyse proposée sous-entend que l'ordre du modèle a déjà été fixé. Nous avons choisi un ordre de 10, puisque les signaux considérés (en général des voyelles) admettent au plus cinq formants dans la bande utile du signal (0-5 kHz). Signalons qu'une fois le modèle déterminé, une approximation des fréquences et des largeurs de bande des formants pour la fenêtre considérée, peut être déduite à partir des pôles du modèle de synthèse $\frac{1}{A_b(z)}$. La relation entre les pôles z_i du filtre et les fréquences f_i et les largeurs de bande B_i des formants est

$$\text{Arg}(z_i) = 2\pi \frac{f_i}{f_e} \quad (6-44-a)$$

$$\text{modu}(z_i) = \exp\left(-\pi \frac{B_i}{f_e}\right) \quad (6-44-b)$$

où $\text{Arg}(z_i)$ et $\text{modu}(z_i)$ sont respectivement l'argument et le module du pôle z_i .

L'inconvénient majeur de cette procédure reste son temps d'exécution relativement long. En effet, pour la détermination du modèle, l'analyse par covariance est effectuée échantillon par échantillon sur une fenêtre glissant le long du signal. Cependant, l'objectif de notre travail étant essentiellement l'extraction des paramètres pertinents de l'activité glottique, nous avons privilégié l'efficacité des traitements et la robustesse des algorithmes sur le temps d'exécution.

6-2-1-3 Détection des événements ouverture/fermeture par analyse adaptative

Nous avons vu dans la section précédente que la difficulté essentielle réside dans la détection des instants d'ouverture et de fermeture, puisque le meilleur modèle du conduit vocal est obtenu par l'analyse du signal dans la région de fermeture. Nous présentons ici une méthode de localisation plus fine de ces événements en utilisant l'analyse adaptative. Ce type de méthode a déjà été effectué sur le signal de parole [Odile, 1988] [Morikawa, 1982] mais pour d'autres objectifs tels que la segmentation et la localisation des régions de transition. Notons que cette méthode est utilisée en même temps pour mesurer la fréquence fondamentale du signal de parole voisée.

6-2-1-3-1 Principe théorique et hypothèses

Comme nous l'avons souligné, le signal de parole peut être approché par un modèle autorégressif. Le principe de l'analyse adaptative est qu'à chaque instant le modèle est réajusté en fonction du modèle précédent et de l'information apportée par l'échantillon du signal. Nous prenons ce principe comme base de la détection des événements ouverture et fermeture de la façon suivante : quand le nouvel échantillon coïncide avec un instant d'ouverture ou de fermeture, il apporte une information supplémentaire de nature différente (présence ou absence de l'excitation), ce qui modifie relativement l'évolution du modèle. Ainsi nous exploitons la différence des modèles actuel et précédent pour la détection de ces instants.

Nous présentons dans ce qui suit le développement de la méthode.

Etant donné un signal de parole voisée autorégressif d'ordre p

$$y(n) = \sum_{i=1}^p a_i y(n-i) + w(n) \quad (6-45)$$

où $w(n)$ est un bruit blanc de moyenne nulle et de variance σ_w , et le vecteur $A = (a_i)_{i=1\dots p}$ des coefficients du modèle. Le critère à minimiser à chaque instant k est

$$I_k(a) = \sum_{j=1}^k \left(y_j - Y_{j-1}^T \cdot A \right)^2 + \left\| A - \hat{A}_0 \right\|_{P_0}^2 \quad (6-46)$$

où le vecteur \hat{A}_0 est une estimation a priori de A , P_0 est la valeur initiale des matrices de pondération symétriques définies positives, et la norme pondérée par P_0 d'un vecteur x est définie par $\left\| x \right\|_{P_0}^2 = x^T P_0^{-1} x$ et le vecteur $Y_{j-1} = [y_{j-1}, y_{j-2}, \dots, y_{j-p}]^T$.

Ce critère est bien conçu pour l'analyse adaptative, puisque le premier terme de (6-46) tient compte de la somme des carrés des résidus et le second terme représente une connaissance

à priori du processus. La minimisation de ce critère donne les formules itératives d'estimation à la fois du modèle, de la matrice de pondération et d'un vecteur gain K_k

$$\hat{A}_k = \hat{A}_{k-1} + K_{k-1} \left(y_k - Y_{k-1}^T \hat{A}_{k-1} \right) \quad (6-47)$$

$$K_k = P_{k-1} Y_{k-1} \left(Y_{k-1}^T P_{k-1} Y_{k-1} + 1 \right)^{-1} \quad (6-48)$$

$$P_k = P_{k-1} - P_{k-1} Y_{k-1} \left(Y_{k-1}^T P_{k-1} Y_{k-1} + 1 \right)^{-1} Y_{k-1}^T P_{k-1} \quad (6-49)$$

où le vecteur \hat{A}_0 est, soit pris égal à zéro, soit initialisé à un vecteur modèle à priori (issu par exemple de l'analyse sur une fenêtre précédente du même signal ou par une autre procédure d'analyse), et la matrice de pondération P_0 est une matrice symétrique définie positive arbitraire.

6-2-1-3-2 Procédure de détection des événements

Etant donné un signal de parole voisée, la procédure suit les étapes suivantes :

- nous fixons tout d'abord une longueur maximale d'analyse pour le cycle initial (fermeture-ouverture), correspondant à la fréquence fondamentale minimale. Nous avons choisi une longueur d'analyse de 12,8 ms (fréquence fondamentale voisine de 78 Hz);
- sur cette fenêtre nous analysons le signal par la méthode adaptative;
- à chaque instant k nous calculons le critère de proximité du modèle calculé par rapport au modèle à l'instant $k-1$:

$$J(k) = \left| \left| \hat{A}_{k-1} - \hat{A}_k \right| \right|^2 \quad (6-50)$$

- nous détectons les deux événements en localisant les deux instants où le critère est maximal;

- pour déterminer lequel des deux événements correspond à l'instant de fermeture (et par conséquent l'autre à l'instant d'ouverture), nous appliquons la règle selon laquelle le maximum du signal dans la fenêtre se trouve en phase d'ouverture glottique. Cette règle se justifie par le fait qu'en phase de fermeture, le signal peut être considéré comme la somme de sinusoïdes amorties correspondant au régime libre du conduit vocal, et qu'en phase d'ouverture, l'amplitude du signal augmente à cause de l'impulsion glottique. Notons que juste après la détection d'un événement, nous commençons l'analyse avec une matrice de pondération initiale P_0 et un vecteur modèle nul. Nous procédons de cette manière pour éviter la répercussion de l'information de l'événement sur la mémoire de l'algorithme.

- pour le prochain cycle, nous recommençons le même processus avec une longueur de fenêtre égale à la durée du cycle précédent.

Le schéma de principe de cette procédure est donné à la figure 6-20.

6-2-1-3-3 Exemples d'application de la procédure

Pour valider la méthode de localisation des instants d'ouverture et de fermeture, nous l'avons essayé sur des simulations. La figure 6-21-b représente un signal de parole synthétique correspondant à la voyelle /a/. En 6-21-f sont indiquées les fréquences des formants et leur largeur de bande du modèle tous pôles de synthèse. L'excitation du conduit vocal, représentée à la figure 6-21-a, est une onde glottique simulée suivant le modèle de Rosenberg [Rosenberg, 1972] dont la fréquence fondamentale est égale à 83 Hz. Les instants d'ouverture et de fermeture réels et ceux déterminés par la procédure sont représentés respectivement en 6-21-c et 6-21-d.

Nous avons effectué sur chacune des phases de fermeture, l'analyse séquentielle par covariance en vue d'une part de déterminer le modèle du conduit vocal, d'autre part d'effectuer le filtrage inverse. Le résultat du filtrage inverse est représenté à la figure 6-21-e. et les fréquences et les largeurs de bande sont représentés en 6-21-f.

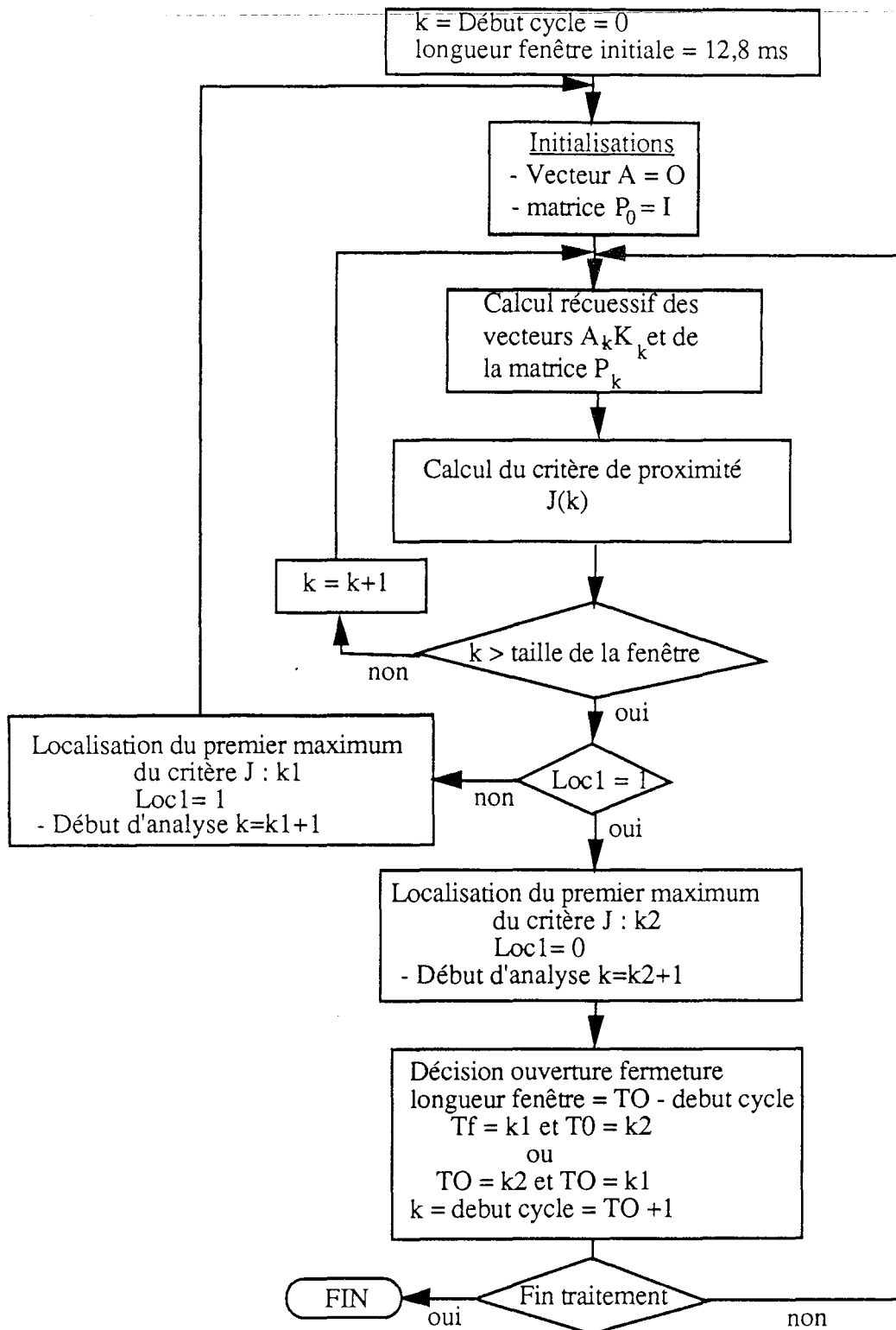


Figure 6-20. Procédure de localisation des instants de fermeture et d'ouverture pour un signal de parole voisée.

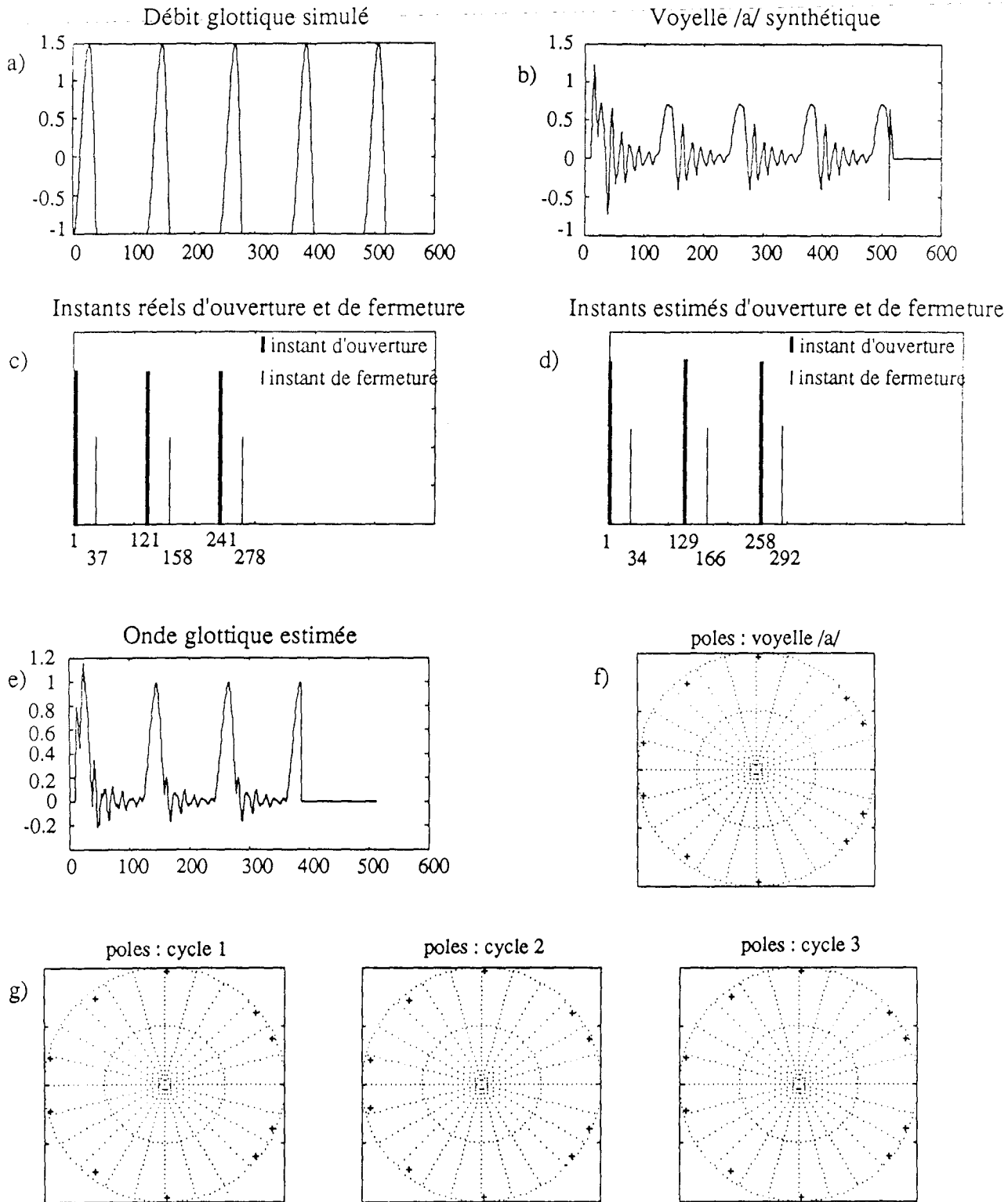


Figure 6-21. Application de la procédure de détection des instants d'ouverture et de fermeture et du filtrage inverse sur une voyelle synthétique /a/

6-2-2 Méthodes d'extraction des paramètres de l'activité glottique

Dans ce paragraphe, nous présentons deux méthodes d'extraction des paramètres de l'activité glottique.

Dans la première méthode, on calcule d'abord l'onde glottique par filtrage inverse du signal vocal. Ensuite, une procédure d'optimisation ajuste les paramètres de l'un ou l'autre des modèles, modèle de Fujisaki et celui de Titze modifié, de manière à faire coïncider l'onde glottique du modèle et l'onde glottique expérimentale.

La seconde méthode procède par une optimisation simultanée du filtre inverse et des paramètres glottiques.

6-2-2-1 Paramètres mesurés

Nous classons l'ensemble des paramètres à mesurer en deux groupes : les paramètres temporels et les paramètres de l'activité glottique.

Les premiers sont des événements directement liés à l'évolution temporelle du débit glottique pour lequel nous avons choisi le modèle de Fujisaki (1986) présenté au paragraphe I-2-2 et représenté à la figure 6-7. Nous rappelons les différents paramètres :

- R : durée de croissance d'ouverture,
- W : durée de la phase d'ouverture,
- D : durée de l'intervalle séparant l'instant de fermeture de celui du maximum négatif du débit glottique,
- S : obliquité de l'impulsion glottique $S = \frac{W-R}{W}$;
- A : pente du débit à l'instant d'ouverture,
- B : pente du débit avant l'instant de fermeture,
- C : pente du débit après l'instant de fermeture.

Les paramètres de l'activité glottique, décrits au paragraphe I-4-2, sont ceux du modèle cinétique .

Nous avons choisi de séparer l'extraction des deux groupes de paramètres, puisque nous estimons qu'il y a probablement une corrélation entre les deux groupes, que n'avons pas explicitée à l'heure actuelle de l'avancement de nos travaux.

Les deux méthodes de détermination des paramètres que nous allons exposer sont valables pour les deux groupes et sauf nécessité, nous parlerons des paramètres sans précision du groupe.

6-2-2-2 Méthode basée sur le résultat du filtrage inverse

Nous prenons ici le débit glottique comme référence pour la mesure des paramètres. La procédure de mesure consiste à déterminer les paramètres du modèle qui permettent de générer le débit glottique de référence. En d'autres termes, si on note $g_r(t)$ le signal débit glottique de référence et $g_s(t,P)$ le modèle paramétrique fonction du vecteur P des paramètres, nous essayons de résoudre l'équation suivante :

$$g_s(t,P) = g_r(t) \quad \text{pour tout } t \quad (6-51)$$

Ne disposant que des échantillons numériques de la fonction de référence, nous sommes amenés à utiliser des méthodes d'optimisation par rapport au vecteur des paramètres, selon un critère de distance entre les deux fonctions g_s et g_r . La distance que nous avons choisie est la distance euclidienne entre les deux fonctions supposées définies par N échantillons (N étant le nombre d'échantillons dans un cycle glottique : ouverture + fermeture). Le critère est alors défini par

$$C(P) = \sum_{i=1}^N [g_s(i,P) - g_r(i)]^2 \quad (6-52)$$

Nous illustrons la procédure par le schéma de la figure 6-22.

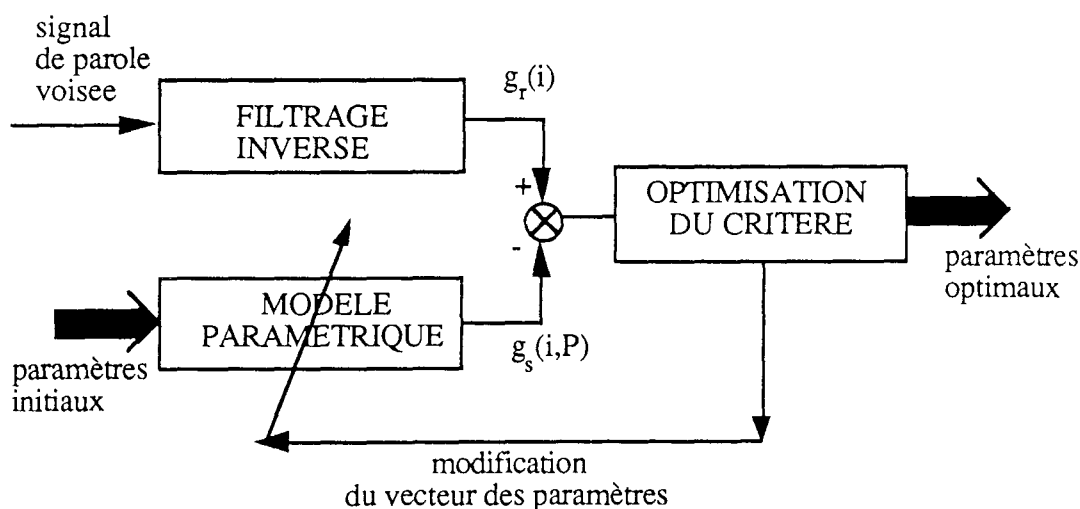


Figure 6-22. Schéma bloc de la procédure d'extraction des paramètres glottiques

Nous avons utilisé deux méthodes d'optimisation : la méthode du complexe [Box, 1965] et une méthode directe introduite par Hooke (1961) [Wilde, 1966].

La méthode du complexe est une méthode de recherche d'optimum d'une fonction $C(P)$, définie sur un domaine des contraintes D de \mathbb{R}^n , supposé convexe (nous considérons ici le cas de la minimisation de C). Les contraintes sont de deux types : des contraintes explicites de la forme

$$I) \quad c_i \leq P_i \quad \text{pour}$$

certaines valeurs de i entre 1 et n

et des contraintes implicites de la forme

$$II) \quad 0 \leq \phi_j(P) \quad \text{pour } j \text{ variant entre 1 et } q \text{ et où } \phi_j \text{ est une fonction de } \mathbb{R}^n.$$

Un complexe est un ensemble de k vecteurs de \mathbb{R}^n avec $k > n+1$ ($k=n+1$ le complexe est appelé simplexe).

La méthode du complexe consiste à construire un nouveau complexe en remplaçant le vecteur P_s du complexe précédent, pour lequel la fonction C est la plus grande, par un autre vecteur où C prend une valeur inférieure. Le procédé de changement de vecteur est dit de "sur-réflexion" où on remplace le vecteur P_s par un vecteur P_d , appartenant à la droite $(P_s P_G)$ et défini par

$$\vec{P_G P_d} = -\alpha \vec{P_G P_s} \quad \text{avec } \alpha > 1$$

où P_G est le centre de gravité des autres $(k-1)$ vecteurs du complexe.

Si au cours de l'application de la sur-réflexion le nouveau vecteur ne vérifie pas une des deux contraintes, il est ramené toujours dans la direction $(P_s P_G)$, jusqu'il soit admissible.

La procédure se termine quand la différence $C(P_s) - C(P_d)$ est inférieure à un seuil fixé.

La méthode est illustrée à la figure 6-23 pour une fonction à deux variables avec un complexe à quatre éléments.

L'autre méthode d'optimisation, à laquelle nous avons pensé, est plus simple dans le concept et plus directe pour la recherche. En effet, elle est basée sur l'hypothèse qu'il est raisonnable de considérer à nouveau tout ensemble d'ajustements des variables indépendantes qui s'est révélé fructueux au cours des expériences précédentes. Le pas de déplacement des variables, supposées indépendantes, croît au fur et à mesure que le nombre d'expériences augmente, jusqu'au voisinage de l'optimum où la recherche devient plus fine en diminuant le pas.

Nous décrivons brièvement la procédure de recherche du minimum d'une fonction C à n variables, en définissant la notion de schéma.

Le schéma peut être représenté par une trajectoire rectiligne orientée ("flèche") dont une extrémité est la base et l'autre est la pointe. La recherche commence en un point-base P_1 (P_1^j) _{$j=1\dots n$} , choisi arbitrairement. En prenant un vecteur pas Δ_1 dont la i -ième composante est δ_i , les autres étant toutes nulles, on compare les valeurs de la fonction C pour les vecteurs P_1 et $P_1 + \delta_1$. Si $C(P_1 + \delta_1)$ est inférieure à $C(P_1)$, la pointe temporaire du schéma initial est $P_1 + \delta_1$. Dans le cas contraire, on compare $C(P_1)$ et $C(P_1 - \delta_1)$, le schéma temporaire initial a pour pointe $P_1 - \delta_1$ si $C(P_1) > C(P_1 - \delta_1)$. Ainsi la pointe t_1^1 de schéma initial autour de P_1 est définie par

$$t_1^1 = \begin{cases} P_1 + \delta_1 & \text{si } C(P_1 + \delta_1) < C(P_1) \\ P_1 - \delta_1 & \text{si } C(P_1 - \delta_1) < C(P_1) < C(P_1 + \delta_1) \\ P_1 & \text{si } C(P_1) < \min [C(P_1 + \delta_1), C(P_1 - \delta_1)] \end{cases} \quad (6-53)$$

On procède de la même manière pour toutes les coordonnées, pour déterminer le schéma initial en partant du schéma temporaire de la coordonnée précédente; autrement la j -ième pointe ($j=2\dots n$) t_1^j temporaire du schéma initial est obtenue en fonction de la pointe précédente t_1^{j-1} par

$$t_1^j = \begin{cases} t_1^{j-1} + \delta_j & \text{si } C(t_1^{j-1} + \delta_j) < C(t_1^{j-1}) \\ t_1^{j-1} - \delta_j & \text{si } C(t_1^{j-1} - \delta_j) < C(t_1^{j-1}) < C(t_1^{j-1} + \delta_j) \\ t_1^{j-1} & \text{si } C(t_1^{j-1}) < \min [C(t_1^{j-1} + \delta_j), C(t_1^{j-1} - \delta_j)] \end{cases} \quad (6-54)$$

Pour $j=n$, la nouvelle base du schéma suivant est $P_2 = t_1^n$. Au lieu de procéder à une recherche autour du point P_2 , on prolonge la base à $t_2^0 = P_1 + 2(P_2 - P_1)$, et on recommence le processus. Quand le schéma autour d'une base est nul, nous pouvons soit décider que cette base est l'optimum à la précision donnée, soit procéder à une recherche plus fine autour de ce point en réduisant le vecteur pas.

Nous illustrons cette méthode par la figure 6-24.

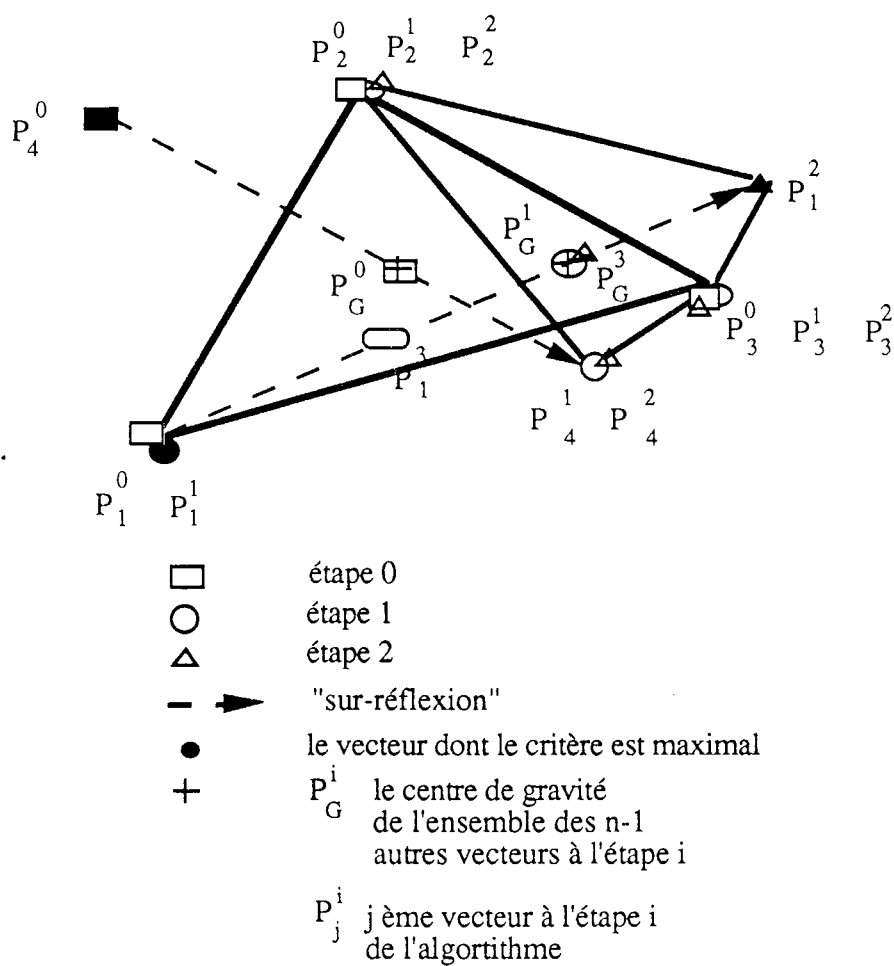


Figure 6-23. Trois étapes successives de la méthode du complexe de minimisation d'une fonction à deux variables. (le coefficient de sur-réflexion $\alpha = 1,5$)

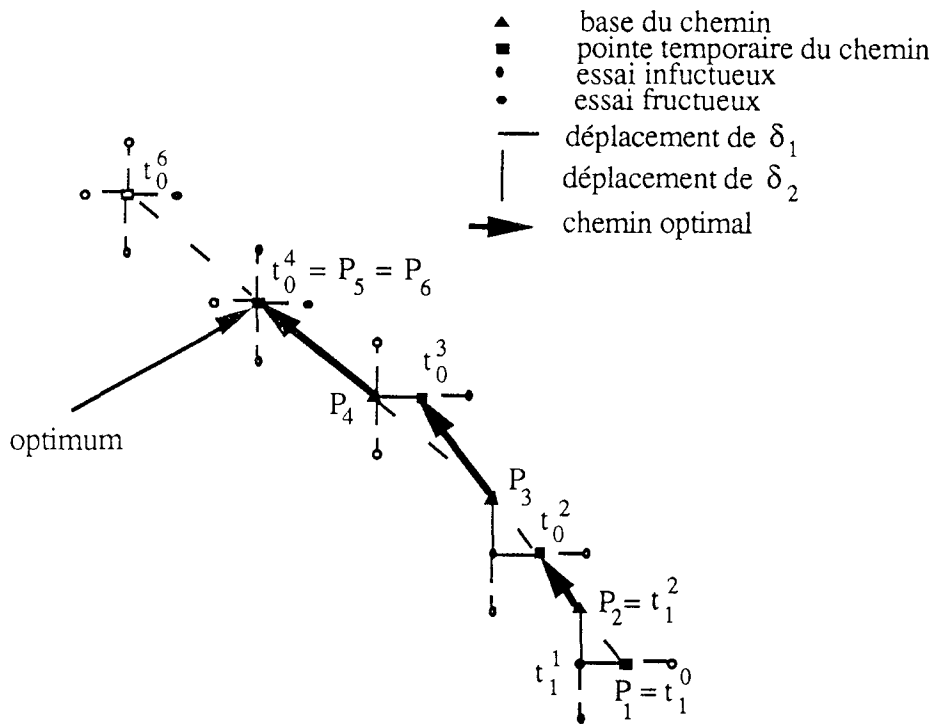


Figure 6-24. Recherche directe de l'optimum d'une fonction à deux variables.

6-2-2-3 Extraction simultanée des paramètres et de l'onde glottique

6-2-2-3-1 Principe de la méthode

Dans ce paragraphe, nous décrivons une méthode qui sert à la fois à effectuer le filtrage inverse et à extraire les paramètres de l'activité glottique à partir du signal de parole. La différence essentielle entre la présente méthode et celle développée plus haut est que la modélisation du signal de parole concerne aussi bien celle du conduit vocal que celle de l'excitation glottique. Le modèle de production de la parole reste le même sauf que l'excitation est issue du modèle paramétrique. Pour le modèle paramétrique, nous ne précisons pas ici sa nature, il peut être soit le modèle de Fujisaki ou le modèle cinétique de l'activité glottique.

Le principe de la méthode est représenté à la figure 6-25, où $s(k)$ est le signal de parole, $u_g(k)$ la sortie du modèle paramétrique, $1/A(z)$ le modèle tous pôles du conduit vocal et $L(z)$ la fonction de transfert du rayonnement par les lèvres.

Le critère à minimiser est celui des moindres carrés pondérés

$$V_N(A,P) = \sum_{k=ki}^{ki+N} \lambda_k e_w^2(k) \tag{6-55}$$

où $e_w(k)$ est l'erreur de prédiction définie par

$$E_w(z) = \left[S(z) - \frac{L(z) U_g(z)}{A(z)} \right] W(z) \quad (6-56)$$

avec $W(z)$ un filtre d'ajustement spectral. Le critère est non linéaire par rapport au vecteur A de $A(z)$. Un choix judicieux de la fonction de pondération spectrale $W(z) = P(z)A(z)$ rend le critère linéaire par rapport au vecteur A , où $P(z)$ est un filtre de pré-emphase du signal de parole $S(z)$

$$E_w(z) = A(z)S(z)P(z) - L(z)U_g(z)P(z) \quad (6-57)$$

Pour simplifier les écritures, nous posons

$$S_f(z) = P(z)S(z) \quad (6-58-a)$$

$$U_f(z) = L(z)P(z)U_g(z) \quad (6-58-b)$$

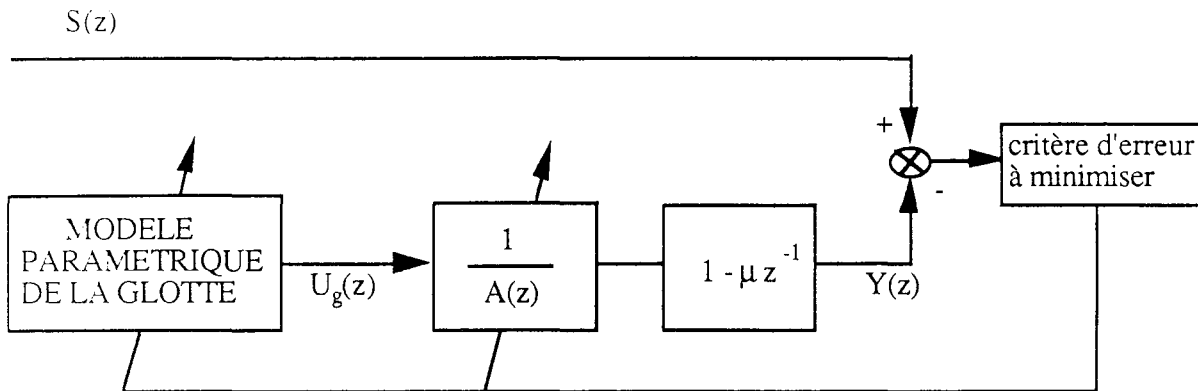


Figure 6-25. Schéma-bloc de la détermination simultanée des paramètres de l'activité glottique et du filtre d'analyse $A(z)$

L'expression de l'erreur de prédiction dans le domaine temporel est

$$e_w(k) = s_f(k) + \sum_{i=1}^p a_i s_f(k-i) - u_f(k) \quad (6-59)$$

En termes vectoriels, si on pose

$$S_N = (s_f(1), \dots, s_f(N))^T \text{ et } U_N(P) = (u_f(1,P), \dots, u_f(N,P))^T$$

où P est le vecteur paramètres de l'activité glottique, et la matrice ϕ_N définie par

$$\phi_N = \begin{pmatrix} -s_f(0) & \dots & -s_f(1-p) \\ \vdots & \ddots & \vdots \\ -s_f(N-1) & \dots & -s_f(N-p) \end{pmatrix}$$

le critère défini en (6-55) peut s'écrire

$$V_N(A,P) = (S_N(P) - \phi_N A)^T \Lambda (S_N(P) - \phi_N A) \quad (6-60)$$

où

$$\mathfrak{S}_N(P) = \mathfrak{S}_N - U_N(P) \quad (6-61)$$

et Λ une matrice diagonale contenant les facteurs de pondération.

Après avoir écrit le critère sous la forme matricielle, nous décrivons la méthode d'extraction simultanée des vecteurs optimaux \hat{P} du modèle de la glotte et \hat{A} du filtre tous pôles $1/A(z)$.

Pour un vecteur P fixé, la méthode des moindres carrés donne comme vecteur optimal \hat{A}

$$\hat{A} = \phi_N^+ \mathfrak{S}_N(P) \quad (6-62)$$

où $\phi_N^+ = (\phi_N \phi_N^T)^{-1}$ est la matrice pseudo-inverse de ϕ_N .

Un moyen efficace de calcul numérique de \hat{A} est la factorisation QR de la matrice ϕ_N , où R est une matrice triangulaire supérieure et Q une matrice orthogonale vérifiant $Q^{-1} = Q^T$. En utilisant cette factorisation le critère à minimiser (en prenant $P(z)=1$ et $\Lambda = I$) devient

$$\begin{aligned} V_N(A,P) &= \|\mathfrak{S}_N(P) - QRA\|^2 \\ &= \|Q^T \mathfrak{S}_N(P) - RA\|^2 \end{aligned} \quad (6-63)$$

La matrice R étant obtenue par triangularisation de ϕ_N , elle est de la forme

$$R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \quad (6-64)$$

où R_1 est une matrice carrée (dimension $p \times p$) triangulaire supérieure.

En posant $h(P) = Q^T \mathfrak{S}_N(P)$

$$= \begin{bmatrix} h_1(P) \\ h_2(P) \end{bmatrix} \quad (6-65)$$

où $h_1(P)$ et $h_2(P)$ sont deux vecteurs colonnes de dimension respective p et $N-p$, le critère à minimiser devient alors

$$V_N(A,P) = \|h_1(P) - R_1 A\|^2 + \|h_2(P)\|^2 \quad (6-66)$$

Le vecteur A minimisant $V_N(A,P)$, pour un vecteur paramètre P donné, est donc solution de l'équation

$$R_1 A = h_1(P) \quad (6-67)$$

De même pour le vecteur optimal \hat{A} (donc solution de (6-67), le vecteur paramètre optimal \hat{P} est celui qui minimise

$$V_N(\hat{A}, P) = \left\| \left\| h_2(P) \right\| \right\|^2 \quad (6-68)$$

Ainsi, en utilisant la factorisation QR nous avons séparé le problème d'optimisation du critère initial en deux optimisations, chacune par rapport à l'un des deux vecteurs A et P.

6-2-2-3-2 Mise en œuvre de la méthode

La procédure d'extraction des paramètres et du filtrage inverse est réalisée selon les étapes suivantes :

- Choix de la fenêtre d'analyse : la fenêtre d'analyse correspond à un cycle glottique. Il est donc nécessaire de commencer par une détection des événements instants d'ouverture et de fermeture glottique, selon l'algorithme décrit dans le paragraphe 2-2-3. Notons N la taille de cette fenêtre.

- Pré-emphase du signal sur la fenêtre d'analyse (optionnel).
- Calcul de la matrice ϕ_N (indépendante du vecteur P des paramètres).
- Factorisation QR de la matrice ϕ_N .
- Choix du vecteur initial P_0 des paramètres.
- Génération sur la fenêtre d'analyse du débit glottique $u_g(n, P_0)$ et $u_f(n, P_0)$ partant du modèle paramétrique initial.

- Calcul des vecteurs $h_1(P)$ et $h_2(P)$.
- Recherche du vecteur optimal \hat{P} des paramètres : recherche du minimum de la fonction $\left\| \left\| h_2(P) \right\| \right\|^2$ par l'une des deux méthodes d'optimisation exposées plus haut.
- Génération sur la fenêtre d'analyse du débit glottique $u_g(n, \hat{P})$ partant du modèle paramétrique optimal.

- Calcul des vecteurs $h_1(\hat{P})$ et $h_2(\hat{P})$.
- Détermination du modèle AR optimal : résolution de l'équation matricielle: $R_1 \hat{A} = h_1(\hat{P})$
- Nous pouvons recommencer la procédure pour le prochain cycle.

Remarques :

- * la procédure est séduisante puisqu'elle permet d'effectuer le filtrage inverse en même temps que l'extraction des paramètres de l'activité glottique;

- * les deux points clés de cette méthode sont la factorisation QR de la matrice ϕ_N d'une part, et l'optimisation du critère en fonction des paramètres d'autre part. Quant au temps d'exécution, la procédure est lente et dépend surtout de la convergence de la méthode

d'optimisation choisie. Nous avons utilisé la méthode de recherche directe de l'optimum. Concernant la décomposition QR, elle ne s'exécute qu'une seule fois par cycle et par conséquent elle a moins d'influence sur le temps d'exécution que la recherche de l'optimum;

* outre les paramètres de l'activité glottique, nous pouvons extraire d'autres paramètres, à partir du filtre de synthèse optimal caractérisant le conduit vocal (donc la prononciation); par exemple, la fréquence et la largeur de bande des formants, la fonction d'aire du conduit vocal.

6-3 Programme général de mesure des caractéristiques de la voix humaine

Nous avons réalisé un certain nombre de modules que l'utilisateur peut choisir et éventuellement chaîner par l'utilisation de menus. Nous présentons le schéma fonctionnel du programme général à la figure 6-26.

Nous avons groupé les modules en deux niveaux, inférieur (niveau I) et supérieur (niveau II). Le premier groupe réalise des fonctions usuelles d'acquisition du signal, de sa visualisation, son enregistrement et un "petit" module de gestion de fichiers (liste des enregistrements, date des enregistrements,...).

Le niveau supérieur est lui-même partagé en deux groupes : les traitements usuels interactifs, où l'opérateur dispose d'un certain nombre de fonctions usuelles de traitement de la parole, comme le montre le schéma fonctionnel de la figure 6-27. Le deuxième groupe est proprement dit le système de mesure des paramètres (figure 6-28). Certaines mesures, telles que la fréquence fondamentale, les fréquences et les largeurs de bande des formants, utilisent des fonctions usuelles de l'autre groupe. L'opérateur peut choisir entre différentes méthodes de mesure, comme par exemple, pour les fréquences et largeurs de bande des formants, nous disposons soit de la méthode de résolution des racines du filtre tous pôles, soit la détection des pics des spectres d'amplitude de la réponse impulsionnelle du filtre d'analyse $A(z)$. L'extraction des paramètres de l'activité glottique peut se faire comme nous l'avons explicité plus haut par deux méthodes.

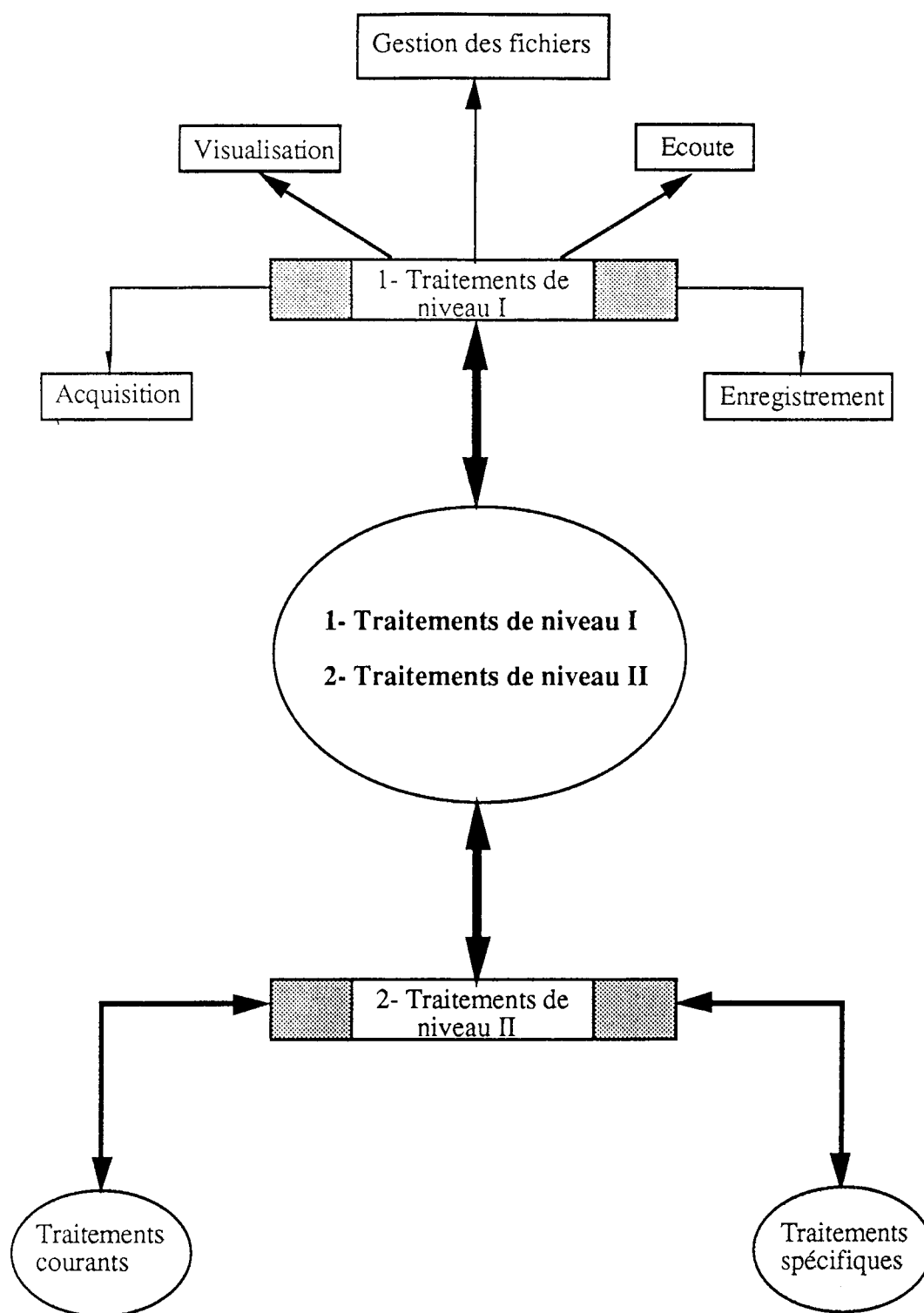


Figure 6-26. Structure fonctionnelle du programme général

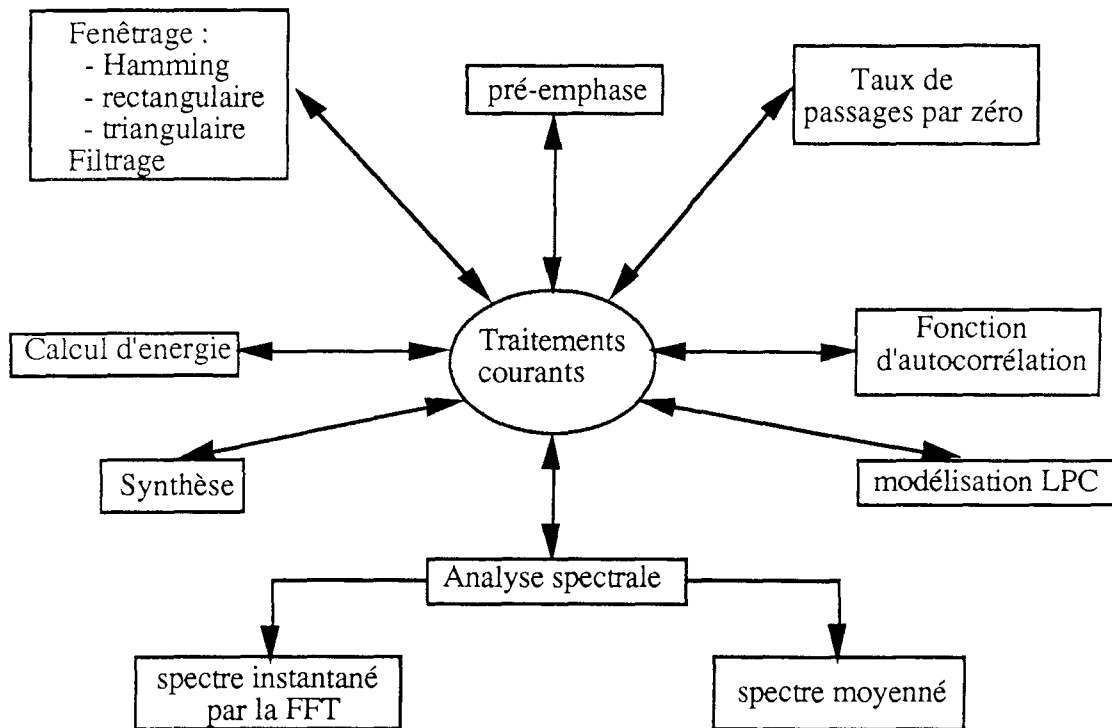


Figure 6-27. Traitements courants du niveau II du programme

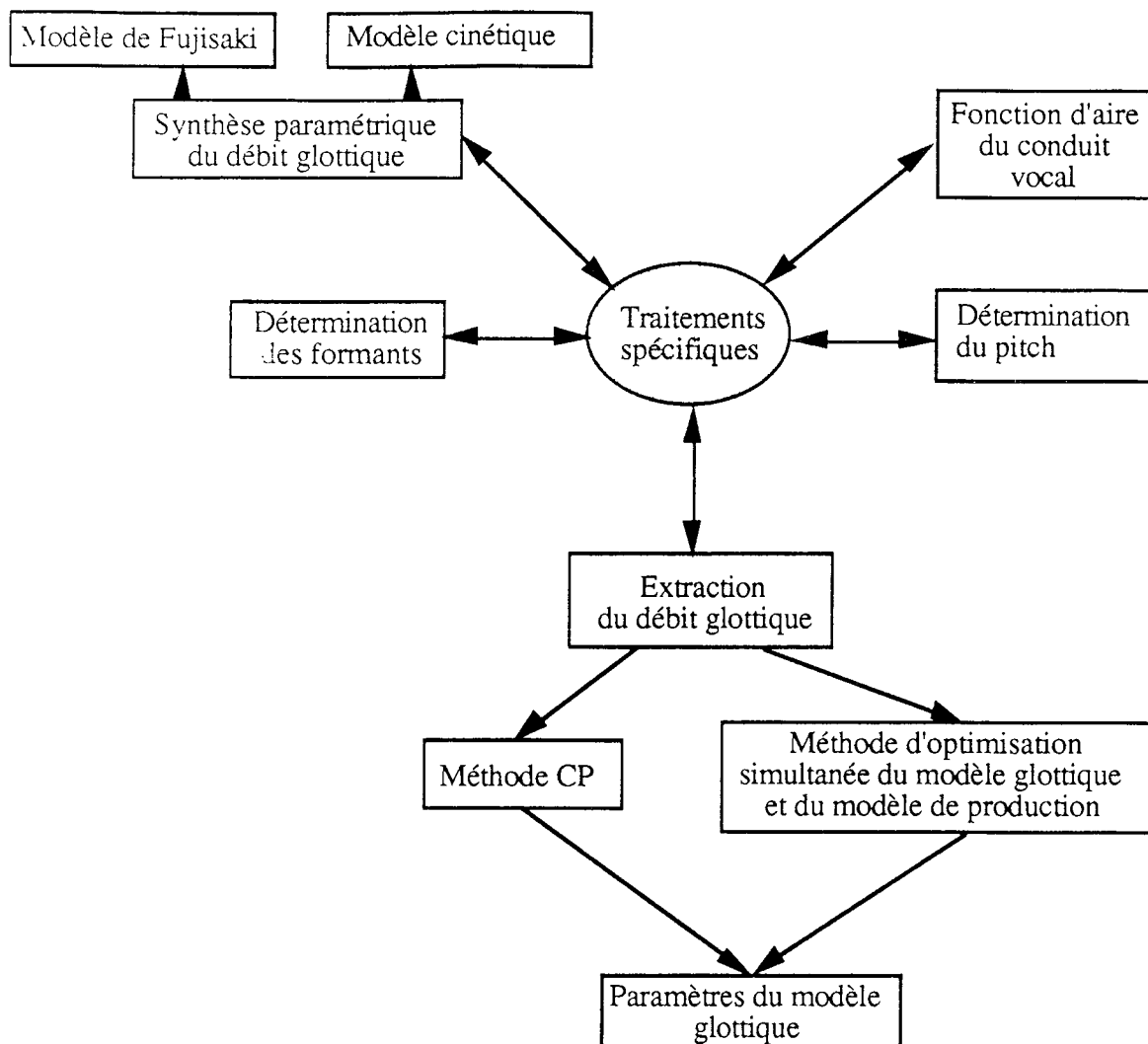


Figure 6-28. Traitements spécifiques: système de mesure

CONCLUSION

CONCLUSION

L'exposé de notre travail sur la caractérisation objective de la voix humaine prend fin ici. Ce travail entre dans un cadre général des thèmes de recherche développés par notre équipe, concernant le développement de nouveaux outils de traitement de signal et la modélisation des signaux non stationnaires.

Nous avons mis au point un certain nombre de techniques qui, outre notre application spécifique, peuvent servir éventuellement pour d'autres finalités. Citons comme exemple la méthode de détection des événements particuliers apparaissant dans un signal : la détection des instants de sauts dans un signal modulé en phase ...

Nous pensons que ce travail apporte une contribution appréciable dans le domaine de la modélisation paramétrique de l'onde glottique, puisque nous avons défini de nouveaux paramètres en étroite liaison avec le comportement différencié des deux cordes vocales. Ce sont les paramètres de symétrie, introduits au dernier chapitre. Cependant, nous estimons qu'une telle contribution mérite d'être approfondie et comptons, dans un proche avenir, étudier en détail l'influence de ces paramètres sur les différents signaux glottiques, générés par ce modèle particulier.

Quant à l'avenir du système que nous avons mis au point, nous pensons d'ores et déjà qu'il peut servir pour effectuer des mesures sur des voix pathologiques dans le but d'étudier la pertinence des paramètres. Nous considérons à ce sujet, que deux études récentes, menées dans notre laboratoire sur la classification [Levrat,1989], [Mirzak,1990], peuvent servir pour regrouper, en fonction d'un lot de paramètres, les voix en différentes classes suivant le type de pathologie. Cette suggestion concerne les équipes de recherche sur la reconnaissance de formes et la classification floue.

BIBLIOGRAPHIE

- Akaike H.** A new look at the statistical model identification, December 1974, **IEEE Trans. Aut. Control**, Vol. 19, No 6, pp 716-723.
- Alaska Y.A.** Electroglottography in speech processing, 1987, **Conf. Proc. IEEE SOUTHEASTCON'87**, Vol 1, Tampa-USA, pp 286-290
- Ananthapadmanabha T.V.** Acoustic Analysis of Voice Source Dynamics, 1984, **STL-QPSR, KTH**, No 2-3, pp 1-24.
- Atal B.S., Hanauer S.L.** Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, August 1971, **J. Acoust. Soc. Am.**, Vol-50, No. 2 (Part 2) pp 637-655.
- Atal B.S.** Towards Determining Articulator Positions from the Speech Signal, 1974, **Proc. Speech Comm. Seminar**, Stockholm, Sweden, pp 1-9.
- Baer T.** Investigation of phonation using excised larynxes, 1975, Dissertation, MIT, Cambridge MA.
- Baer T.** Application of MRI to the analysis of speech production, 1987, **Magnetic Resonance Imaging**, Vol 5, G.B., pp 1-7.
- Boite R., Kunt M.** Traitement de la Parole, 1987, Presses Polytechniques Romandes
- Box M.J.** A new method of constrained optimization and a comparison with other methods, 1965, **Computer Journal**, 8, pp 42-52
- Carton F.** Introduction à la phonétique du français, 1974, Ed. Bordas, col Etudes, PARIS

- Dejonckere P. Précis de pathologie et de thérapeutique de la voix, Encyclopédie Universitaire, Ed. J.P Delarge.
- Fant G. Acoustic Theory of Speech Production, 1970, Mouton, The Hague.
- Fant G. Glottal Source and Excitation Analysis, 1979, STL-QPSR, KTH, No 1, pp 85-70
- Fant G. Preliminaries to the Analysis of the Human Voice Source, 1982, STL-QPSR 4, Speech Transmission Laboratory, KTH, Stockholm, pp 1-27.
- Fant G., Liljencrants J., Lin Q. A Four Parameter Model of Glottal Flow, Avril 1985, French-swedish Symposium, Grenoble-France.
- Farnsworth D.W. High speed motion pictures of the human vocal cords, 1940, Bell Labs Record, 18, pp 203-208.
- Faris W.B., Timothy L.K. Linear Predictive Coding with Zeros and Glottal Wave, 1974, In Proceedings of the National Electronics Conference (National Engineering Consortium, Oakbrook, IL), vol-29, pp 409-411
- Flanagan J.L. Speech Analysis, Synthesis and Perception, 1972, 2nd Ed., Springer-Verlag, New York.
- Flanagan J.L., Landgraf L.LL Self Oscillating Source for Vocal-Tract Synthesizers, March 1968, IEEE Trans. Audio and Electroacoustics, AU-16, pp 57-64
- Flanagan J.L., Ishizaka K, Shipley K.L., Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract, March 1975, Bell Sys. Tech J., Vol-54, No. 3, pp 485-506.
- Fujisaki H., Ljungqvist M. Proposal and Evaluation of Models for the Glottal Source Waveform, 1986, Proc. Int. Conf. On Acoustics, Speech, and Signal Processing, pp 31.2.1-31.2.4

- Gueguen C. Introduction à l'analyse de la parole, Mai 1976, 7-èmes Journées d'Etude sur la Parole, Nancy.
- Gueguen C., Mathieu M. Contribution des zéros à la modélisation du signal de parole, Mai 1976, 7-èmes Journées d'Etude sur la Parole, Nancy.
- Guérin B., Bøe L.J. A two-mass model of vocal cords : determination of control parameters and their respective consequences, 1977, ICASSP-77, pp 583-586.
- Guérin B., Bøe L.J. La régulation de la vibration des cordes vocales : simulation à l'aide d'un modèle à deux masses, Mai 1979, 8-ième Journées d'Etudes Sur la parole, Aix-en-Provence, pp 37-42.
- Grenier, Y. Modélisation des Signaux non Stationnaires, 1984, Thèse d'Etat Université Paris-Sud Centre d'Orsay.
- Hannen E., Quinn B. The determination of the order of an autoregression, 1969, J. Royal Statist. Soc, Vol. 56, Series B, No. 3, pp 579-593
- Hedlin P. A glottal LPC-vocoder, 1984, Proc. Int. Conf. On Acoustics, Speech, and Signal Processing, pp 1.6.1-1.6.4.
- Hess W. Pitch Determination of Speech Signals, 1983, Springer-Verlag
- Hollien H. Three major vocal registers : a proposal, (1972), "ICPhS-7", pp 320-331
- Hooke R., Jeeves T.A. "Direct Search" Solution of numerical and statistical problems, avril 1961, J. Assoc. Comp., pp 212-229.
- Husson R. Etudes des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée, 1950, Thèse de docteur es-Sciences, Faculté des sciences, Université de Paris (17 juin 1950).

- Ishizaka K., Flanagan J.L. Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords, July-August 1972, *Bell Sys. Tech J.*, Vol-50, No. 6, pp 1233-1268.
- Ishizaka K., Matsudaira M. Analysis on the vibration of the vocal cords, 1968, *J. Acoust. Soc. Jpn*, Vol 24, pp 311-312.
- Jenkins G.M., Watts D.G. Spectral analysis and its applications, 1968, Holden-Day.
- Karmouche M. Modélisation de la glotte, 1989, *Rapport Bibliographique*, CRAN-LEA, Université de Nancy I.
- Kelly J.L., Lochbaum C. Speech Synthesis, September 1962, *Proc. Speech Comm. Seminar*, R.I.T., Stockholm, Sweden.
- Lekhal S.A. Analyse des Signaux Aléatoires Non- Stationnaires par Identification Recursive, 1980, Thèse de Docteur Ingénieur, ENST
- Le Roux J., Gueguen C. A fixed point computation of partial correlation coefficients; Juin 1977, *IEEE Trans. Acoust. Speech and Signal Proc.*, Vol ASSP-25, No 3, pp 257-259.
- Makhoul J Linear Prediction : a tutorial review, Avril 1975, *IEEE Trans. Acoust. Speech and Signal Proc.*, Vol ASSP-63, No. 4, pp 561-580.
- Markel J.D, Gray A.H Linear Prediction of Speech, 1976, Springer-Verlag, New-York.
- Marple S.L. Digital Spectral Analysis with Applications, 1987, Prentice-Hall. Inc., Englewood Cliffs, New Jersey.
- Martinet A. Elément de la linguistique générale, 1965, Presses Universitaires de France, col. "Sup", série "le linguiste", Paris

- Meloni L.G.P, Compression Spectrale du Signal Vocal par Modification du Modèle Autoregressif, 1985, Thèse de l'Université de Nancy I
- Morikawa H., Fujisaki H. Adaptive Analysis of Speech Based on Pole-Zero Representation, Feb. 1982, *IEEE Trans.Acoust. Speech and Signal Proc.*, Vol ASSP-30, No. 1, pp 77-87.
- Morse P.M., Ingard K.U. Theoretical Acoustics, 1968, McGraw Hill Book Co, New-York.
- Mounin B. Clefs pour la linguistique, 1968, Seghers
- Mrayati M. Contribution aux Etudes sur la production de la parole, 1976, Thèse d'Etat és-Sciences Physiques, Inst. Nat.Polytech. de Grenoble.
- Oppenheim A.V.,Schafer R.W. Digital Signal Processing, 1975, Prentice-Hall, Inc, Englewood Cliffs, New Jersey.
- Papoulis A. Maximum Entropy and spectral Estimation : a review, December 1981, *IEEE Trans.Acoust. Speech and Signal Proc.*, Vol ASSP-29, No 6, pp 1176-1186
- Perello J. The muco-undulatory theory of phonation, "*Ann. Oto-Laryngol.*", 1962, 79, pp 722-725.
- Perkell J.S. Physiology of Speech Production : Result and Implications of a quantitative Cineradiographic Study, 1969, MIT Press, Cambridge, Mass.
- Portnoff M.R. A Quasi-One-Dimensional Digital Simulation for the Time-Varying Vocal Tract, June 1973, M.S. Thesis, Dept. of Elect. Engr., MIT, Cambridge, Mass.
- Rabiner L.R., Schafer R.W. Digital Processing of Speech Signal, 1978, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

- Rissanen J. Shortest data description and consistency of order estimates in ARMA-processes, 1978, *Colloque Int. sur l'anal. et l'estimat. des Systèmes*, INRIA, Versailles.
- Rosenberg A.E. Effect of Glottal Pulse Shape on the Quality of Natural Vowels, February 1971, *J. Acoust. Soc. Am.*, Vol-49, No. 2, pp 583-590.
- Rossi M., Di Cristo A., Hirst D., Martin P., Nishinuma Y. L'intonation : de l'acoustique à la sémantique, 1968, Ed Klincksiek, col "Etudes linguistiques", Paris.
- Sonesson B. On the anatomy and vibratory pattern of the human vocal folds, with special reference to a photo-electrical method for studying the vibratory movements, 1960, *Acta Otolaryngol. Suppl.*, # 156, Stockholm.
- Stevens K.N. Physics of laryngeal behavior and larynx modes, 1977, *Phonetica*, Vol 34, pp 239-279.
- Strube H.W. Determination of the instant of glottal closure from the speech wave, 1974, *J. Acoust. Soc. Am.*, Vol 56, pp.1625-1629.
- Titze I.R. Parametrization of the glottal area, glottal flow, and vocal fold contact area, Feb. 1984, *J. Acoust. Soc. Am.*, Vol 75 , No 2.
- Titze I.R. The human vocal chords : A mathematical model (Part 2), *Phonetica*, Vol 29, pp 1-21.
- Titze I.R. Simulation of vocal fry, 1973, *J. Acoust. Soc. Am.*, Vol 54 (A), pp.23
- Titze I.R. Physical and physiological dimensions of intrinsic voice quality, 1979, *ICPh-Miami*, pp 217-223.

Van den Berg J. Myoelastic-aerodynamic theory of voice production, 1958, " J. Speech Hear. Res.", 1, pp 227-244.

Wilde D.J. Méthodes de recherches d'un optimum, 1966, Dunod

Wong D., Ito M.R., Nicol T. Simulation of Vocal Fold Oscillation Using a Lumped Mass-Spring Approach, 1985, **Internat. Symposium : Applied Simulation and Modeling**, ASM-85, Montreal-Canada, pp 71-76

Wong D.Y., Markel J.D., Gray A.H. Least squares glottal inverse filtering from acoustic speech waveform, 1979, **IEEE Trans. Acoustics Speech and Signal Processing**, Vol ASSP-27, pp 350-355.

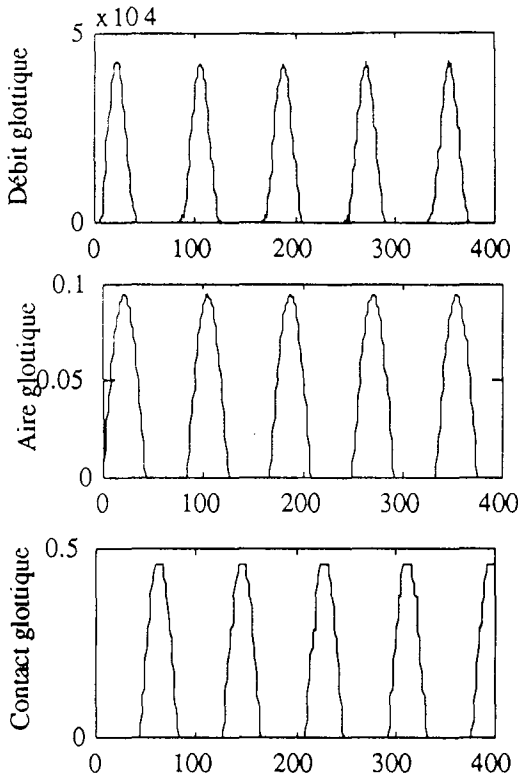
Zerling J.P. Une batterie de programmes pour l'acquisition et l'exploitation de vues frontales de l'orifice labial, 1987, Travaux de l'Institut de Phonétique de Strasbourg, No 19, pp.177-193.

ANNEXE

Dans le présent annexe, nous présentons un ensemble de simulations des signaux glottiques, en considérant le modèle de Titze modifié. Pour l'ensemble des signaux, nous avons choisi les caractéristiques suivantes :

- une fréquence fondamentale égale à 83 Hz,
- une fonction d'aire fixe, correspondante à la production d'une voyelle neutre /e/,
- la fréquence d'échantillonnage fixée à 10 kHz,
- nombre de sections élémentaires des cordes vocales 8x8 sections.

A la fin de cet annexe, nous avons étudié la variation de l'amplitude maximale de l'aire glottique en fonction des paramètres de symétrie.

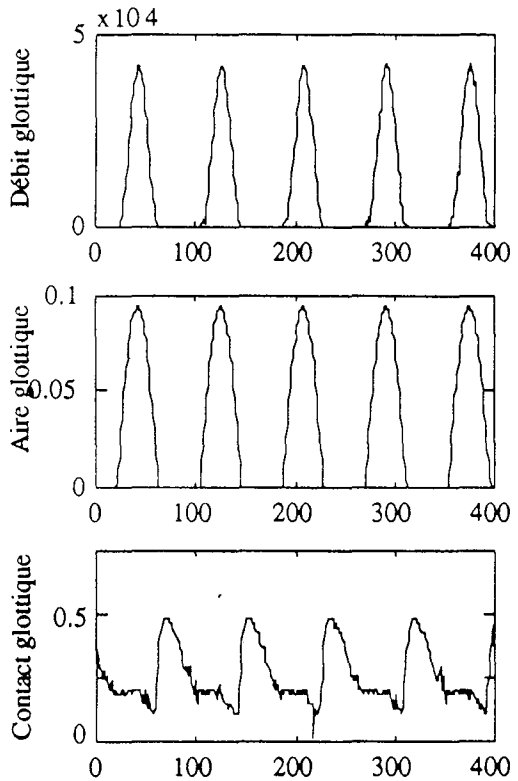


$$\xi_m = 0.075 \text{ cm}$$

$$P_L = 10 \text{ cm H}_2\text{O}$$

$$Q_a = 0 \quad Q_s^1 = 3 \quad Q_s^2 = 3 \quad Q_p = 0$$

$$Q_{sya} = 1 \quad Q_{sys} = 1 \quad Q_{syp} = 1$$

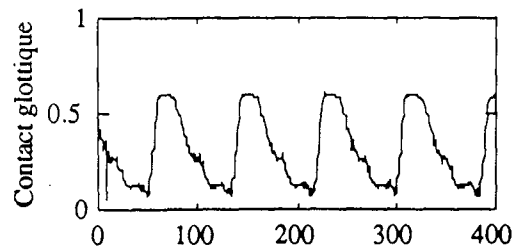
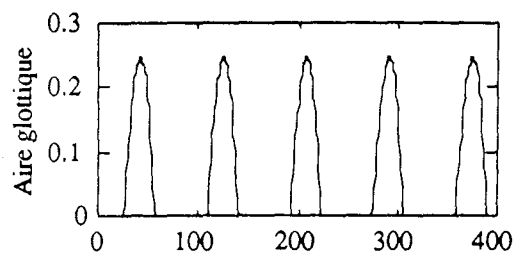
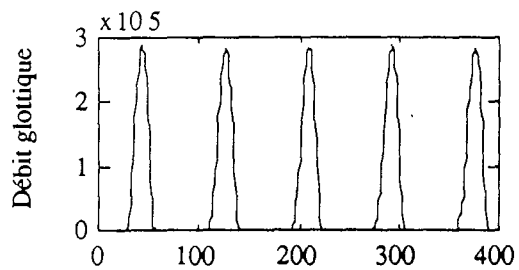
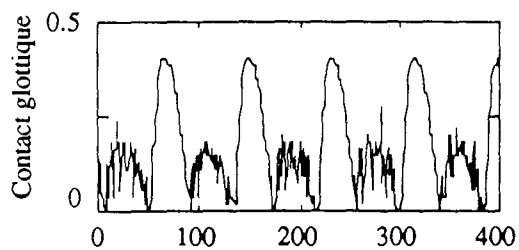
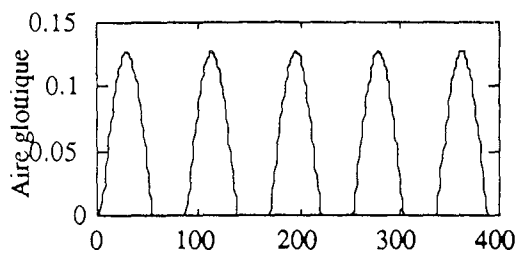
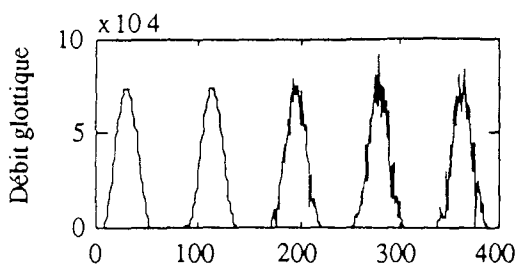


$$\xi_m = 0.075 \text{ cm}$$

$$P_L = 10 \text{ cm H}_2\text{O}$$

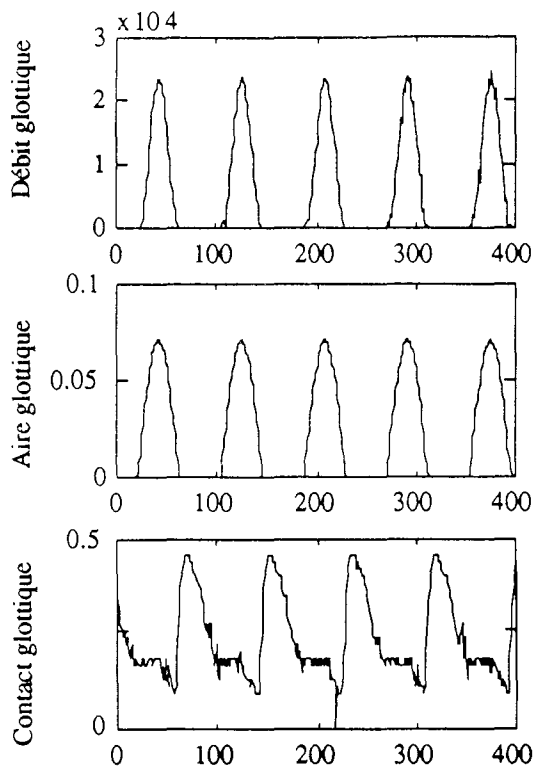
$$Q_a = 0 \quad Q_s^1 = 3 \quad Q_s^2 = 3 \quad Q_p = 0.25$$

$$Q_{sya} = 1 \quad Q_{sys} = 1 \quad Q_{syp} = 1$$

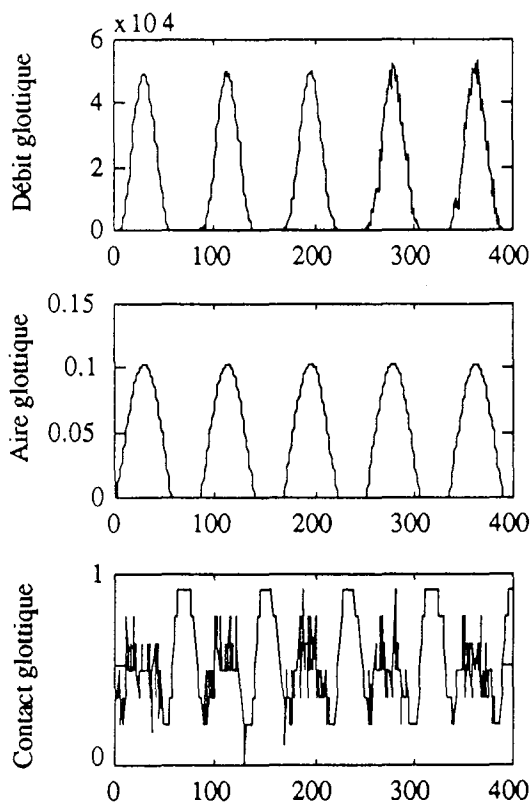


$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0.5 \quad Q_s^1 = 4 \quad Q_s^2 = 4 \quad Q_p = 0.1$
 $Q_{sya} = 1 \quad Q_{sys} = 1 \quad Q_{syp} = 1$

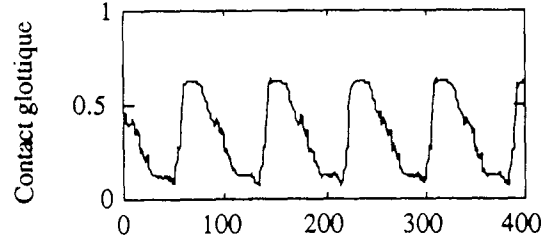
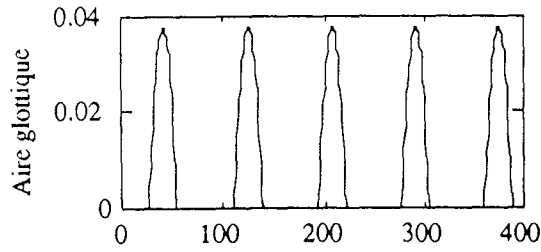
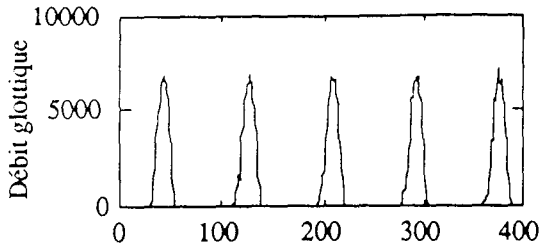
$\xi_m = 0.3 \text{ cm}$
 $P_L = 2.5 \text{ cm H}_2\text{O}$
 $Q_a = -0.5 \quad Q_s^1 = 2 \quad Q_s^2 = 2 \quad Q_p = 0.25$
 $Q_{sya} = 1 \quad Q_{sys} = 1 \quad Q_{syp} = 1$



$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0 \quad Q_s^1 = 3 \quad Q_s^2 = 3 \quad Q_p = 0.25$
 $Q_{\text{sys}} = 0.5 \quad Q_{\text{sys}} = 1 \quad Q_{\text{syp}} = 1$



$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0.5 \quad Q_s^1 = 4 \quad Q_s^2 = 4 \quad Q_p = 0.1$
 $Q_{\text{sys}} = 0.5 \quad Q_{\text{sys}} = 1 \quad Q_{\text{syp}} = 1$

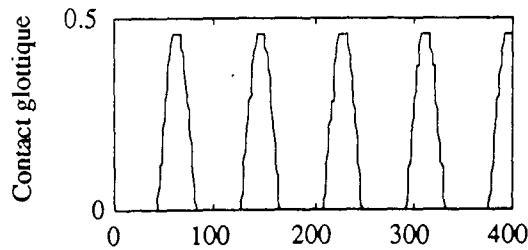
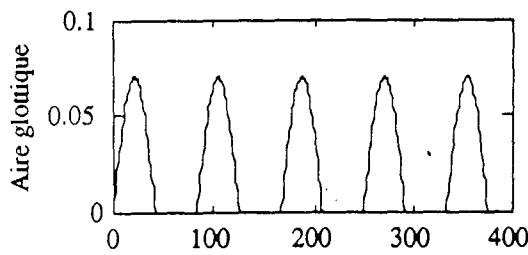
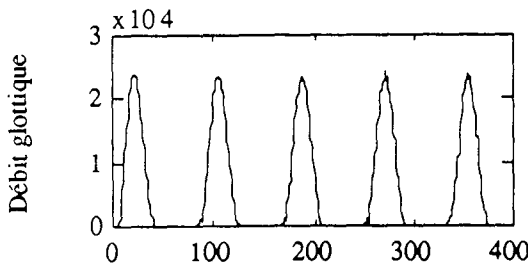


$$\xi_m = 0.075 \text{ cm}$$

$$P_L = 10 \text{ cm H}_2\text{O}$$

$$Q_a = -0.5 \quad Q_s^1 = 2 \quad Q_s^2 = 2 \quad Q_p = 0.25$$

$$Q_{\text{sys}} = 0.5 \quad Q_{\text{sys}} = 1 \quad Q_{\text{syp}} = 1$$

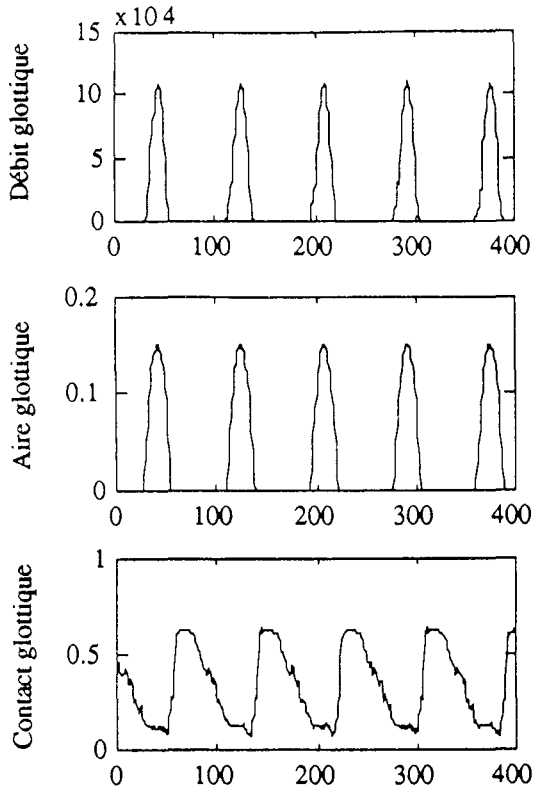


$$\xi_m = 0.075 \text{ cm}$$

$$P_L = 10 \text{ cm H}_2\text{O}$$

$$Q_a = 0 \quad Q_s^1 = 3 \quad Q_s^2 = 3 \quad Q_p = 0$$

$$Q_{\text{sys}} = 0.5 \quad Q_{\text{sys}} = 1 \quad Q_{\text{syp}} = 1$$

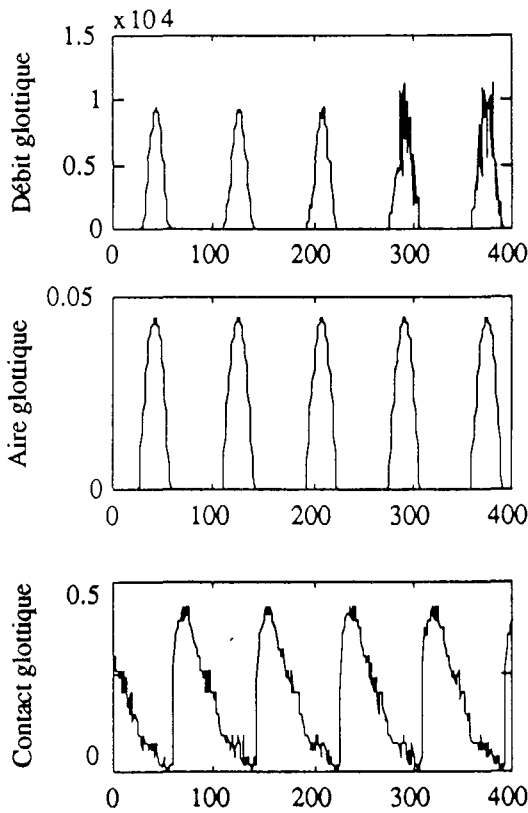


$$\xi_m = 0.3 \text{ cm}$$

$$P_L = 2.5 \text{ cm H}_2\text{O}$$

$$Q_a = -0.5 \quad Q_s^1 = 2 \quad Q_s^2 = 2 \quad Q_p = 0.25$$

$$Q_{sya} = 0.5 \quad Q_{sys} = 1 \quad Q_{syp} = 1$$

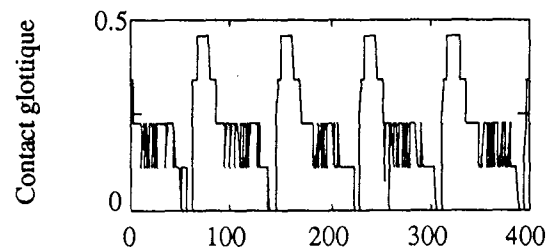
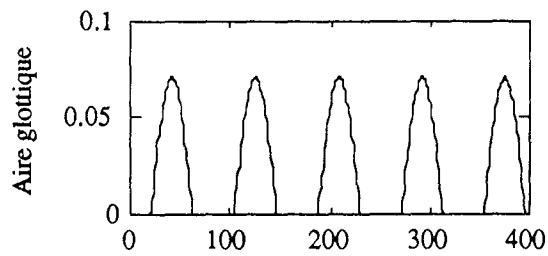
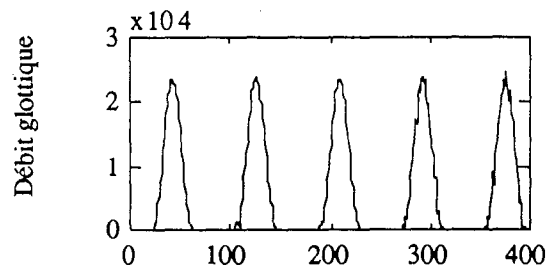
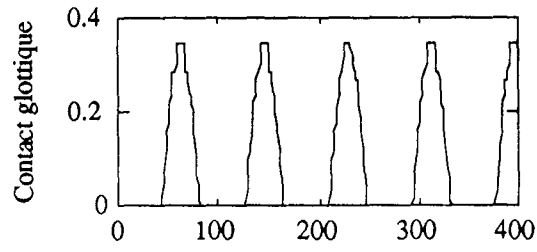
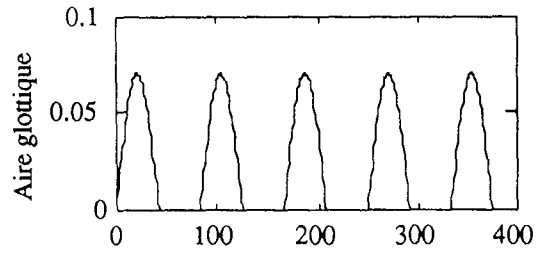
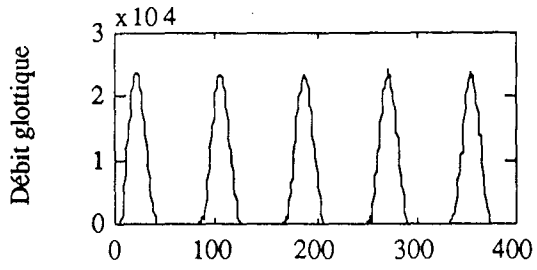


$$\xi_m = 0.075 \text{ cm}$$

$$P_L = 10 \text{ cm H}_2\text{O}$$

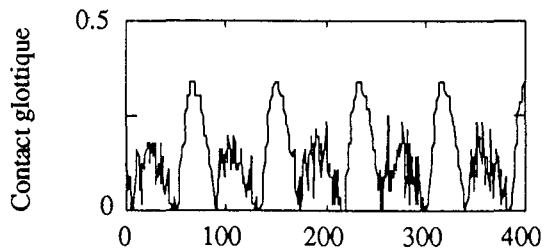
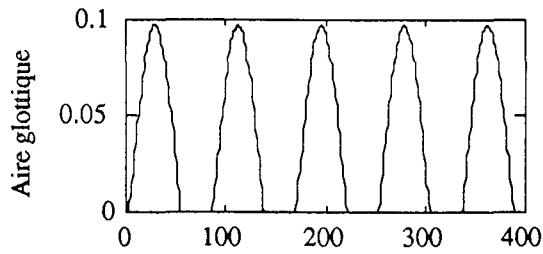
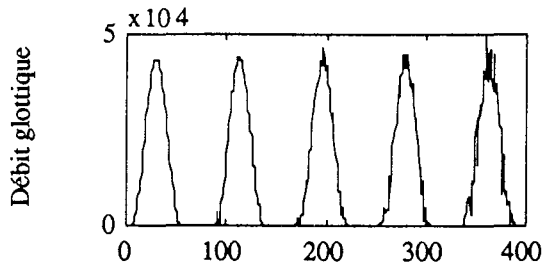
$$Q_a = -0.5 \quad Q_s^1 = 2 \quad Q_s^2 = 4 \quad Q_p = 0.25$$

$$Q_{sya} = 0.5 \quad Q_{sys} = 0.6 \quad Q_{syp} = 1$$

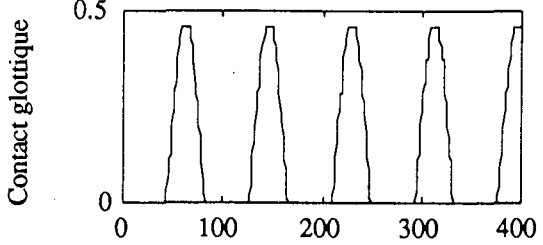
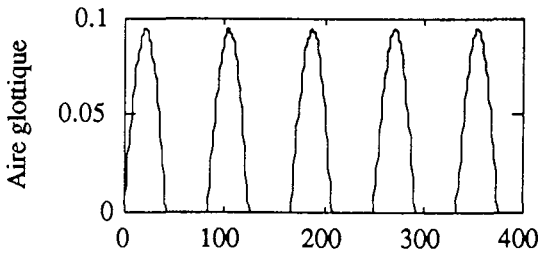
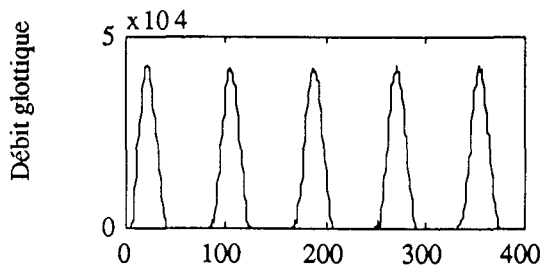


$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0 \quad Q_s^1 = 3 \quad Q_s^2 = 3.8 \quad Q_p = 0$
 $Q_{sya} = 0.5 \quad Q_{sys} = 0.6 \quad Q_{syp} = 1$

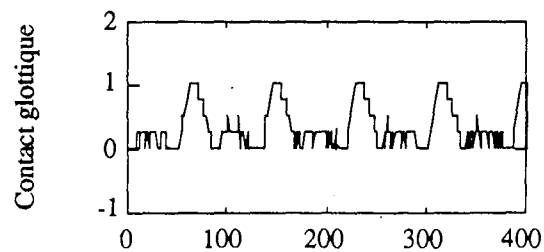
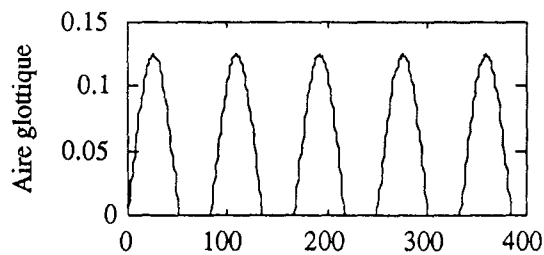
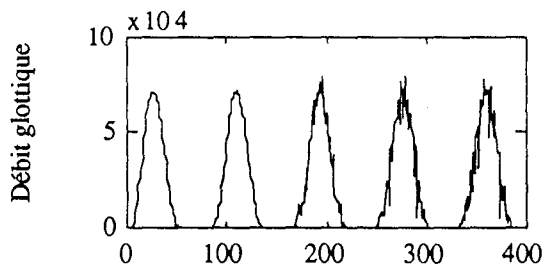
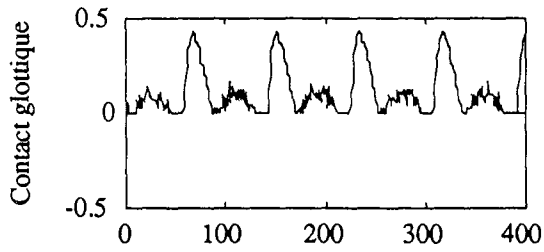
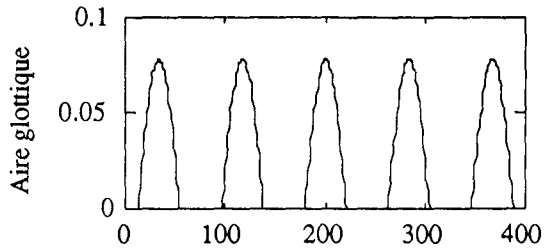
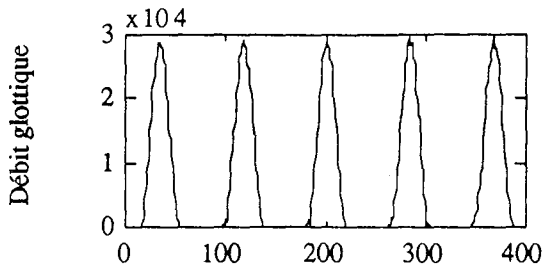
$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0 \quad Q_s^1 = 3 \quad Q_s^2 = 3.8 \quad Q_p = 0.25$
 $Q_{sya} = 0.5 \quad Q_{sys} = 0.6 \quad Q_{syp} = 1$



$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0.5 \quad Q_s^1 = 4 \quad Q_s^2 = 4.8 \quad Q_p = 0.1$
 $Q_{sya} = 0.5 \quad Q_{sys} = 0.6 \quad Q_{syp} = 1$

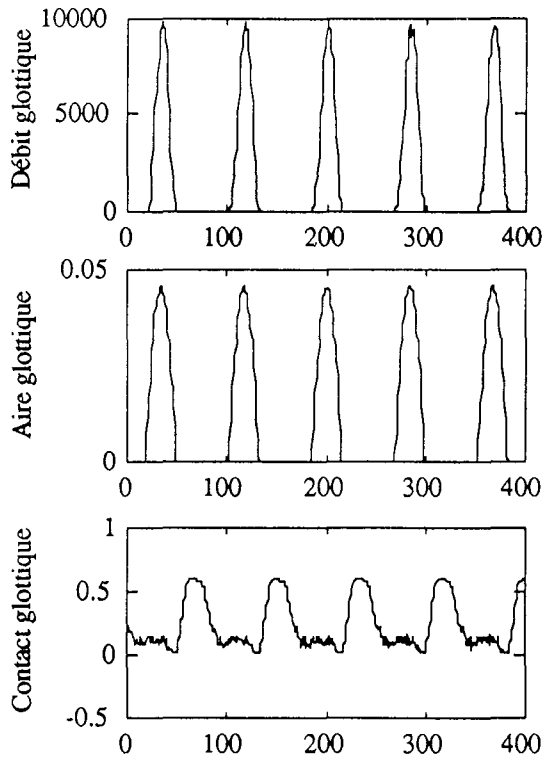


$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0 \quad Q_s^1 = 3 \quad Q_s^2 = 3 \quad Q_p = 0$
 $Q_{sya} = 1 \quad Q_{sys} = 1 \quad Q_{syp} = 0.25$

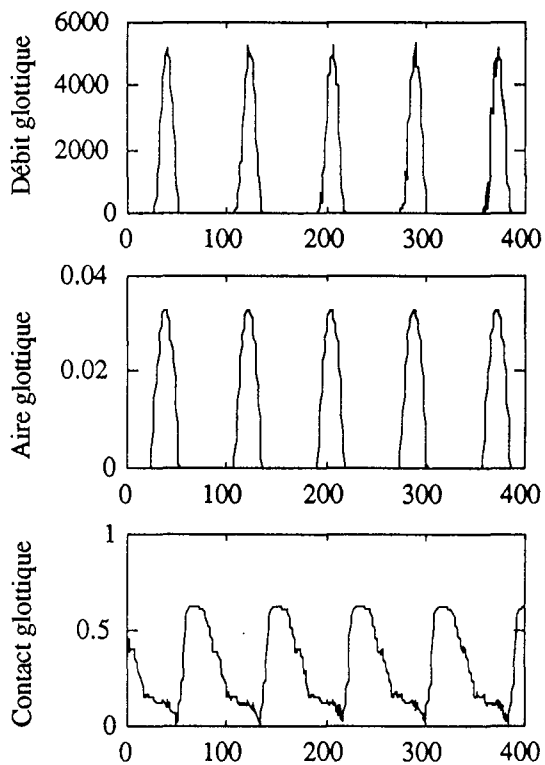


$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0 \quad Q_s^1 = 3 \quad Q_s^2 = 3 \quad Q_p = 0.25$
 $Q_{sya} = 1 \quad Q_{sys} = 1 \quad Q_{syp} = 0.25$

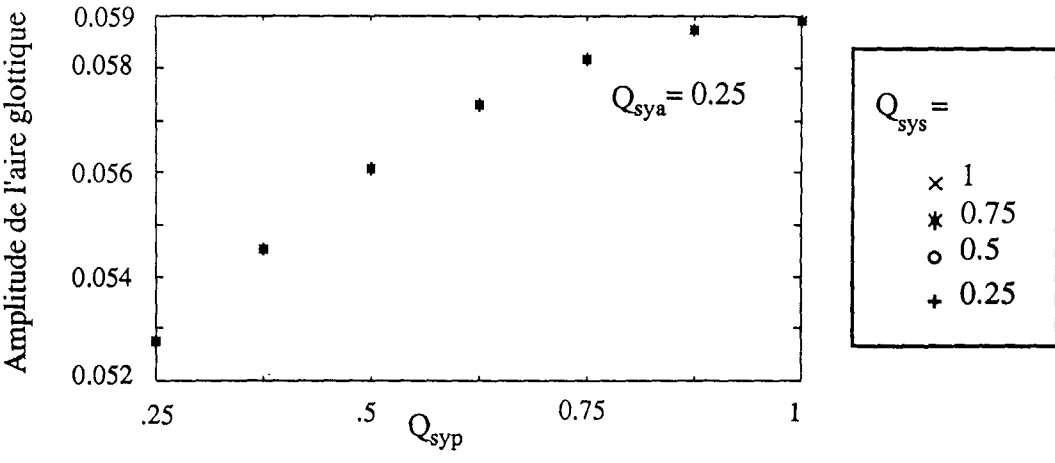
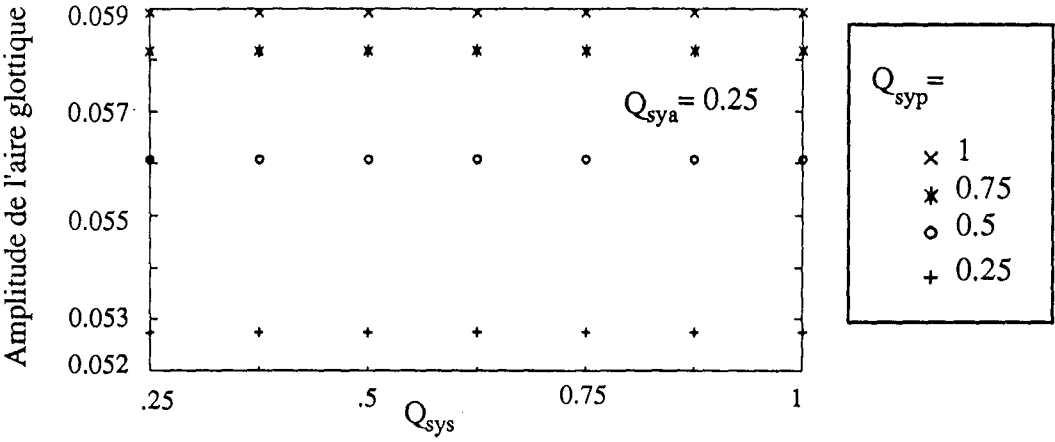
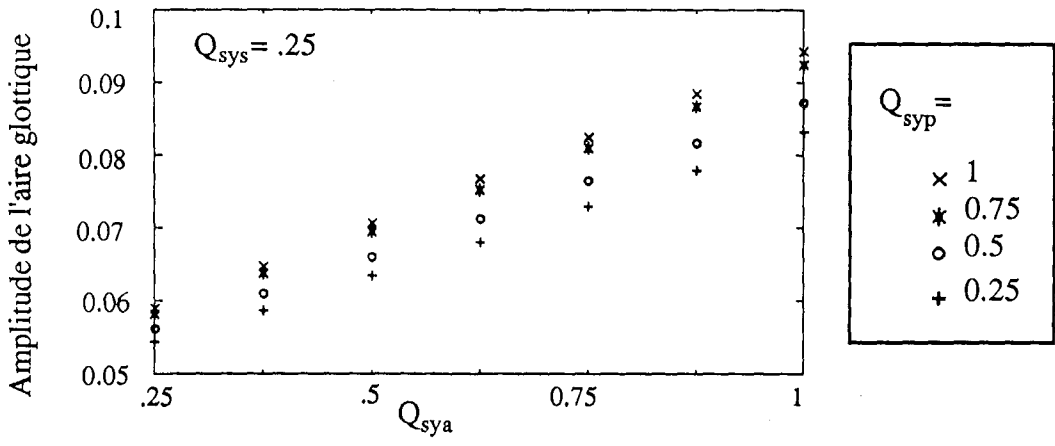
$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = 0.5 \quad Q_s^1 = 4 \quad Q_s^2 = 4 \quad Q_p = 0.1$
 $Q_{sya} = 1 \quad Q_{sys} = 1 \quad Q_{syp} = 0.25$



$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = -0.5 \quad Q_s^1 = 2 \quad Q_s^2 = 2 \quad Q_p = 0.25$
 $Q_{sya} = 1 \quad Q_{sys} = 1 \quad Q_{syp} = 0.25$



$\xi_m = 0.075 \text{ cm}$
 $P_L = 10 \text{ cm H}_2\text{O}$
 $Q_a = -0.5 \quad Q_s^1 = 2 \quad Q_s^2 = 2 \quad Q_p = 0.25$
 $Q_{sya} = 0.5 \quad Q_{sys} = 1 \quad Q_{syp} = 0.5$



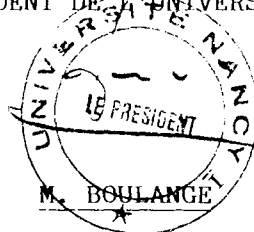
NOM DE L'ETUDIANT : SMIEJ MOHAMMED FAOUZI

NATURE DE LA THESE : DOCTORAT DE L'UNIVERSITE DE NANCY I EN AUTOMATIQUE

VU, APPROUVE ET PERMIS D'IMPRIMER

NANCY, le - 8 JAN. 1991 m^o 1

LE PRESIDENT DE L'UNIVERSITE DE NANCY I



Mots clés :

Parole, Voix humaine, Pathologies de la voix, Système de production de la parole, Acoustique, Modélisation autorégressive, Analyse adaptative, Onde glottique, Détection d'événements, Filtrage inverse.

Résumé :

La première partie de ce mémoire est une brève revue des bases anatomo-physiologiques de la production de la parole et des différentes pathologies de la voix humaine. Le premier chapitre a un double objectif : la sensibilisation sur la complexité de la physiologie phonatoire et la déduction éventuelle d'un modèle de connaissance du système de production. Le second chapitre traite des origines des pathologies notamment les pathologies laryngiennes pour mettre en valeur l'importance de la modélisation paramétrique de l'activité glottique.

La seconde partie, expose l'état de l'art sur la modélisation du système de production de la parole. Les modèles numériques, basés sur la théorie de la propagation des ondes acoustiques dans le conduit vocal, sont traités en détail, en particulier le modèle des tubes acoustiques uniformes de même longueur qui permet une analyse aisée par la prédiction linéaire. Le dernier chapitre de cette partie établit une bonne synthèse des résultats de la modélisation autorégressive où une justification de son utilisation pour l'analyse du signal vocal est longuement discutée.

Au début de la dernière partie, un certain nombre de paramètres qualitatifs pour la caractérisation de la voix humaine sont définis. Ils sont groupés en deux catégories en fonction de la "durée de mesure" : les paramètres phonémiques et les paramètres prosodiques.

Dans le dernier chapitre, après un rappel des différents modèles de l'activité glottique, il a été défini un nouveau modèle paramétrique, basé sur le modèle cinétique de Titze, où les mouvements de chaque corde vocale sont supposés indépendants. Outre les paramètres de configuration pré-phonatoire des cordes vocales définis par Titze, des paramètres de symétrie glottique sont introduits : symétrie d'amplitude, de phase et de forme. Ensuite deux méthodes d'extraction des paramètres du modèle glottique sont présentées : une méthode en deux étapes et une approche d'extraction simultanée des paramètres du modèle glottique et du modèle de production.

Pour la première, l'estimation de l'onde glottique, effectuée par la technique du filtrage inverse en phase de fermeture constitue la première étape de la méthode. Ensuite, une fois l'onde glottique estimée, l'extraction des paramètres glottiques est effectuée par un ajustement du débit glottique mesuré avec celui issu du modèle paramétrique.

La localisation de la phase de fermeture glottique, étant une étape importante du filtrage inverse par l'analyse synchrone, un nouvel algorithme, permettant cette localisation est proposé. Cet algorithme est basé sur l'analyse adaptative et les techniques de détection de "changement de modèle" dans un signal non stationnaire. L'algorithme est validé sur un exemple d'une voyelle synthétique /a/. Enfin, une présentation de l'ossature du programme général de mesure des caractéristiques de la voix humaine termine ce travail.