HAL
open science

Recherches sur les indices perceptifs des voyelles orales et nasales : application à la structure du système vocalique français et de diverses autres langues

François Lonchamp

▶ To cite this version:

François Lonchamp. Recherches sur les indices perceptifs des voyelles orales et nasales : application à la structure du système vocalique français et de diverses autres langues. Linguistique. Université Nancy 2, 1978. Français. NNT : 1978NAN21001 . tel-01776374

HAL Id: tel-01776374
https://hal.univ-lorraine.fr/tel-01776374

Submitted on 24 Apr 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.

UNIVERSITÉ DE LORRAINE

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

UNIVERSITE de NANCY II

INSTITUT de PHONETIQUE

LN78/1

FRANÇOIS LONCHAMP

RECHERCHES SUR LES INDICES PERCEPTIFS DES VOYELLES ORALES ET NASALES

APPLICATION A LA STRUCTURE DU SYSTEME VOCALIQUE FRANÇAIS ET DE DIVERSES AUTRES LANGUES

THESE DE DOCTORAT
DE IIIe CYCLE

SOUS LA DIRECTION DE
MONSIEUR LE PROFESSEUR F. CARTON

UNIVERSITE NANCY BIBLIOTHEQUE

MAI 1978

469

A MA MÈRE

R E M E R C I E M E N T S

- Monsieur le Professeur F. CARTON a bien voulu accepter de diriger cette thèse. Qu'il trouve ici d'abord nos remerciements pour la confiance qu'il nous a toujours témoignée. Avec un enthousiasme jamais démenti, il nous a encouragé à explorer des domaines et des techniques nouvelles. L'intérêt constant qu'il a pris pour ce travail nous a permis de ne pas nous noyer dans les problèmes techniques, de ne jamais perdre de vue l'essentiel, c'est-à-dire la nécessaire liaison entre les résultats et leur utilisation linguistique. Nous n'avons pas toujours suivi ses conseils, et nous devons reconnaître que cela a nui à la qualité de ce travail. Qu'il soit enfin remercié pour la qualité des rapports humains qu'il a su développer à l'Institut de Phonétique.

- Madame le Professeur P. SIMON nous a ouvert les portes de l'excellente bibliothèque de son Institut de Phonétique de STRASBOURG. Qu'elle soit assurée de notre gratitude pour son accueil et ses encouragements.

- Les Journées d'Etude annuelles du G.A.L.F., dont Monsieur le Professeur M. ROSSI est l'un des principaux organisateurs, m'ont permis de prendre connaissance des derniers travaux et de nouer de fructueux contacts. Qu'il soit assuré de la reconnaissance d'un jeune chercheur pour cette tâche indispensable. Le chapitre XI est nourri des réflexions inspirées par son étude sur les traits acoustiques.

- Monsieur le Professeur M. LAMOTTE nous a permis d'utiliser à notre guise les magnifiques installations que son groupe de recherche possède au sein du Laboratoire d'Electricité et d'Automatique de l'Université de NANCY I. Qu'il soit remercié de l'aide inestimable qu'il a apportée au "littéraire" venu d'une université voisine.

- Monsieur l'Ingénieur G. ROGER a bien voulu nous confier une série de voyelles synthétiques réalisées par son groupe de recherche aux Laboratoires de MARCOUSSIS (C.G.E.). L'expérience ainsi acquise au début de cette recherche nous a considérablement aidés par la suite.

- J.P. ZERLING nous a fourni les documents qui nous ont permis de rédiger une partie du chapitre VIII ; le chapitre VI est fondé sur l'excellent travail qu'il a mené à bien en 1974. Qu'il soit remercié pour les heures de discussion, mais également de détente, que nous avons passées ensemble.

- Cette thèse n'aurait pas été terminée à temps sans SIMONE, ma femme, qui a accepté le pire en gardant l'espoir du meilleur, après...

- La frappe du manuscrit a été réalisée avec compétence et diligence par Mlle Danièle LOUIS.

CHAPITRE I

INTRODUCTION

Cette thèse a pour objet l'étude des indices de perception des voyelles orales et nasales en français. Il nous faut préciser immédiatement que seuls seront envisagés les problèmes posés par la perception des voyelles en dehors d'un contexte linguistique. Les études perceptives portant sur des voyelles isolées ont souvent été critiquées, car elles ne peuvent rendre compte des caractéristiques spécifiques de la perception contextuelle que nous présenterons plus loin. Il nous faut donc justifier dès maintenant notre choix.

I - LA PERCEPTION DES VOYELLES EN CONTEXTE

La notion de "contexte", appliquée à la perception des voyelles doit être précisée. Nous faisons référence principalement à l'influence du cadre consonantique de la syllabe sur la probabilité de catégorisation d'une voyelle dans une classe phonologique donnée. Nous excluons donc de cette discussion les mécanismes de suppléance (WARREN 1976) qui infèrent cette catégorisation d'informations syntactico-sémantiques contenues dans l'élément de discours où s'insère la voyelle.

Trois possibilités existent quant au rapport entre la perception des voyelles isolées et en contexte :

a) la perception des voyelles en contexte se fait par un mécanisme totalement différent de celui mis en oeuvre pour les voyelles isolées ;

b) la perception des voyelles isolées fait appel à une partie seulement du mécanisme général de la perception vocalique ;

c) il n'existe aucune différence dans les mécanismes mis en jeu.

Comme nous le verrons plus loin, il est possible, en l'état actuel des connaissances, de rejeter la troisième hypothèse qui suppose un parallélisme strict entre la perception des voyelles isolées et en contexte. Nous laisserons aux partisans de la première le soin de défendre cette thèse extrême. L'hypothèse que nous croyons la plus fructueuse est celle qui prévoit que la perception en contexte fera appel à

la fois à un mécanisme commun, que l'on peut isoler par une étude portant sur des voyelles isolées, et à un mécanisme spécialisé lui permettant d'utiliser des informations de type articulatoire, en provenance d'un centre intégrateur sensori-moteur. Un tel mode opératoire répondrait d'ailleurs à un critère d'économie. L'information supplémentaire éventuellement transmise par le contexte consonantique est utilisée, mais son absence ne constitue pas un obstacle insurmontable.

Cette thèse a donc pour objet l'étude de la partie du mécanisme de perception qui est utilisée, lors de la perception des voyelles isolées, et qui peut donc être mise en évidence par des expériences portant sur de tels stimuli. Mais pourquoi avoir choisi cette partie du mécanisme, et non l'autre qui est moins bien connue ? Notre choix a été dicté par deux considérations.

La première est qu'il paraît difficile d'étudier la prise en compte d'informations acoustiques provenant du contexte consonantique, sans posséder au préalable une idée précise de la manière dont les informations venant de la voyelle elle-même sont utilisées. Si des progrès ont été réalisés, notre travail prouve que nous ne disposons pas encore d'un modèle de perception satisfaisant pour les voyelles isolées.

Une deuxième considération, d'ordre plus pratique, a dicté ce choix. Une revue des travaux consacrés à la perception des voyelles en contexte révèle l'absence de paradigmes expérimentaux "prometteurs", et la nécessité de disposer d'un dispositif de synthèse permettant de créer des syllabes de haute qualité. Comme l'écrit FLANAGAN (1971) ".. La recherche d'un modèle général de la perception continue à ressembler à la quête du Saint-Graal... Aucune ligne d'attaque n'apparaît clairement... en raison de la difficulté à contrôler les différents paramètres expérimentaux (p. 68)". C'est cette difficulté que nous voudrions illustrer par une revue rapide des travaux récents dans ce domaine. Nous avons regroupé les aspects spécifiques de la perception des voyelles en contexte en deux catégories générales : la perception catégorielle et

la normalisation en fonction du contexte consonantique et de la fréquence fondamentale.

A/ La perception catégorielle des voyelles en contexte

A la suite de tests comparant la discrimination et l'identification de voyelles isolées et en contexte, STEVENS (1968) (cf. aussi FUJISAKI et KAWASHIMA 1970) conclut que la perception des voyelles en contexte est catégorielle. FRY et al. (1962) avaient montré que la perception des voyelles isolées était continue. Mais cette dichotomie n'est sans doute pas correcte. SACHS (1968) a constaté que des voyelles isolées d'une durée de 150 msec étaient également perçues de façon catégorielle, tout comme ces mêmes voyelles dans un contexte de mot. Ces résultats ont été confirmés par FUJISAKI et KAWASHIMA (1970) et PISONI (1971, 1975). Les auteurs japonais ont proposé un modèle de l'activité discriminante dans la méthode de test ABX. Celui-ci prend en considération le rapport variable des informations phonétiques (identification) et des informations perceptives (discrimination) pour expliquer les différents degrés de catégorisation observée. PISONI (1973, 1975) souligne que l'utilisation de la méthode ABX pose un sérieux problème méthodologique : elle favorise la perception catégorielle. De plus le rôle spécifique joué par le contexte consonantique ne peut être distingué du rôle de tout contexte, même extra-linguistique (bruit blanc par exemple), dans l'accroissement du degré de perception catégorielle. CROWDER (1973) interprète également les résultats d'expériences de mémorisation de voyelles, en et hors contexte, par des degrés de perception catégorielle. Ce degré dépend de la quantité d'information stockée dans une mémoire acoustique pré-catégorielle à court-terme (PAS). A notre connaissance, aucune technique ne permet d'étudier de façon spécifique l'effet de la perturbation acoustique que représente la présence d'une consonne à proximité immédiate d'une voyelle. Notons enfin que la perception catégorielle ne saurait expliquer le processus de normalisation que subissent les voyelles en contexte.

B/ Le processus de normalisation

La nécessité d'une "normalisation" de l'information spectrale avant le stade de catégorisation linguistique a été reconnue depuis les travaux de PETERSON et BARNEY (1952), qui ont montré un degré de recouvrement important des zones spectrales, pour des locuteurs de sexe ou d'âge différents. Cette normalisation explique également la possibilité de reconnaître des segments acoustiques relativement dissemblables comme appartenant à la même catégorie linguistique. Ces différences acoustiques proviennent de la longueur variable du conduit vocal en fonction de l'âge et du sexe du sujet. Mais des difficultés d'analyse ont conduit NORDSTRÖM (1977) à affirmer que les différences anatomiques entre les sujets ne pouvaient expliquer qu'une partie des variations spectrales observées. Ceci supposerait une articulation spécifique selon le sexe ou l'âge!

Cette normalisation peut être due en partie à la fréquence fondamentale (CARLSON et al. 1975). Ces derniers proposent une explication fondée sur une interaction spectrale entre la fréquence du premier formant et la fréquence fondamentale. Or SCOTT (1976) suggère que la perception du premier formant est fondée sur une analyse temporelle.

Il est probable que la consonne, par l'intermédiaire des transitions, joue un rôle important dans le processus de normalisation, mais également dans la compensation des variations par rapport à la "cible" vocalique ("overshooting" et "undershooting" : tir au-dessus et en-dessous de la cible). Le meilleur exemple est fourni par MILLAR et AINSWORTH (1972). Nous l'avons extrait des matrices de confusions présentées, car il n'est pas relevé par les auteurs. Une voyelle synthétique isolée, perçue par 68 % des sujets comme [ɛ] (et par 14 % comme [ʌ]), est identifiée dans 38 % des cas comme un [æ] si elle est insérée dans le contexte (h - d). Ce changement du timbre perçu ne peut être dû qu'à la présence des transitions. FOURCIN (1968), RAND (1971) et SUMMERFIELD et HAGGARD (1975) indiquent que la perception des consonnes elle-même dépend d'un

processus de normalisation. STRANGE et al. (1976) montrent que la présence d'un contexte consonantique réduit fortement l'ambiguïté de voyelles prononcées par des locuteurs de sexe et d'âge différents. Mais la poursuite et l'extension de ces travaux supposent que l'on dispose d'un modèle de perception des consonnes. Si la perception des consonnes dépend de celle des voyelles, c'est alors d'un modèle de perception syllabique dont il faut disposer. On conçoit l'immensité de la tâche et sa difficulté. LINDBLOM et STUDDERT-KENNEDY (1967) suggèrent un mécanisme de compensation des phénomènes de "tir au-dessus ou en-dessous de la cible" ("overshooting" et "undershoot ng"). Un examen détaillé des résultats indique que seuls certains sujets effectuent cette compensation. Une expérience de contrôle effectuée par MERMELSTEIN (1975) s'est soldée par un échec car les résultats suggèrent que les sujets utilisent une moyenne temporelle des évolutions formantiques. Si les travaux de FUJISAKI et SEKIMOTO (1975) révèlent que les sujets sont capables d'extrapoler la fréquence du deuxième formant au-delà de la fréquence effectivement atteinte, KLATT, dans la discussion, souligne qu'une interprétation en termes de décision volontaire et consciente est plus probable.

Enfin, les travaux sur l'influence des voyelles dans les syllabes adjacentes montrent des effets de contraste mais également d'assimilation, qui défient encore l'analyse (SHELDON 1973).

Cette revue rapide confirme le pessimisme de FLANAGAN (1971). L'absence de résultats préliminaires cohérents et de perspectives "prometteuses" nous a conduit à ne pas choisir les problèmes de perception contextuelle comme sujet de notre travail.

Il est un deuxième aspect de notre recherche que nous voudrions justifier. Il s'agit de la place importante que tient l'analyse critique des résultats publiés, et le fait que ceux-ci ne concernent pas tous les voyelles françaises.

II - PROBLEMES DE METHODOLOGIE

Notons tout d'abord qu'il n'existe aucune présentation synthétique de l'ensemble des résultats expérimentaux accumulés depuis une trentaine d'années. On peut sans doute en estimer le nombre à près d'un millier. Tout au plus dispose-t-on de quelques essais dus à STUDDERT-KENNEDY (1974, 1976). STEVENS et HOUSE (1972) et DARWIN (1976). Mais les auteurs ne peuvent, en quelques 50 pages, traîter de toutes les questions ; le problème de l'espace perceptif des voyelles n'y est pas abordé. Nous ne sommes pas loin de penser que le travail le plus urgent dans notre domaine est une revue critique des résultats obtenus. Nous serions prêts à soutenir que presque toutes les hypothèses concevables peuvent faire l'objet d'une évaluation préliminaire, qui n'utiliserait que les résultats expérimentaux déjà publiés. Il est un fait que, si les bibliographies s'allongent, on ne voit guère de discussions critiques des résultats antérieurs. Certains travaux semblent tomber dans l'oubli, bien qu'ils contiennent des résultats qui ne sont pas aisément reconciliables avec l'ensemble de nos connaissances. Nous sommes là en présence d'une mine très riche de renseignements originaux, qu'il ne faut pas négliger. Nous avons mentionné plus haut les indices que nous avons relevés dans un travail de MILLAR et AINSWORTH, et ce travail en fournira d'autres exemples.

Cette thèse, sans prétendre aucunement à l'exhaustivité, comportera l'étude critique détaillée de toutes les études antérieures qui nous ont parues significatives. Le chapitre sur l'analyse multidimensionnelle comporte, par exemple, l'étude détaillée des 15 analyses qui ont précédé la nôtre. Ce "corpus" ainsi rassemblé permet de tirer des conclusions qui ne dépendent plus aussi fortement des positions théoriques des auteurs et de la méthodologie utilisée. Cette étude critique s'accompagne d'une nouvelle analyse chaque fois que les résultats le permettent.

Est-il besoin d'ajouter que ce travail contient également de nombreux résultats originaux ?

L'utilisation de résultats portant sur d'autres langues que le français nous paraît indispensable. On ne peut en effet rejeter systématiquement tous les résultats obtenus sur l'anglais américain, ou le japonais, sous prétexte que la structure vocalique de ces langues est différente de celle du français. Mais il convient de prendre garde aux généralisations hâtives. Il est toujours nécessaire de vérifier que les caractéristiques de la langue dont on utilise les résultats ne sont pas à l'origine d'un comportement particulier : nous pensons par exemple au trait de tension qui distingue, avec le trait d'allongement et de non-stabilité temporelle, certaines voyelles américaines, au trait non-labialisé de la voyelle postérieure du japonais, ou au trait d'allongement qui distingue les voyelles du néerlandais. C'est toute la science du phonéticien que de savoir déjouer les pièges éventuels.

III - MOTIVATIONS ET PLAN DE CETTE ETUDE

Le contenu-même de ce travail a été dicté par trois facteurs. Le premier est, tout naturellement, l'enchaînement des questions que posait l'interprétation des résultats de l'analyse multidimensionnelle de l'espace perceptif des voyelles. Il y a quelques années, nous avions publié (LONCHAMP 1973) quelques réflexions sur un modèle de perception, dont on trouvera une version améliorée au chapitre IV. Le deuxième est l'heureuse intervention de notre directeur, qui a su nous convaincre de la nécessité de ne pas nous borner à l'étude des indices perceptifs pour eux-mêmes, mais d'envisager l'application des résultats aux problèmes de typologie et de description des structures phonologiques. Nous lui sommes reconnaissants des perspectives qu'il nous a ouvertes. Le troisième facteur est lié aux possibilités croissantes qui se sont offertes dans l'utilisation des moyens informatiques. Pour la mini-informatique, l'accueil que nous a réservé M. LAMOTTE au Laboratoire d'Electricité et d'Automatique de l'Université de Nancy I nous a permis de réaliser un synthétiseur à formants sur un mini-ordinateur NOVA 2. Pour des raisons de temps, nous n'avons pas,

malheureusement, été en mesure de l'utiliser pleinement dans ce travail. Nous espérons que ce n'est que partie remise. Pour la macro-informatique, l'installation d'un terminal lourd à la Faculté des Lettres nous a permis, dans un premier temps, de mettre au point l'analyse multidimensionnelle par la méthode INDSCAL (LONCHAMP 1974), puis d'utiliser certaines des techniques nouvelles qui n'ont cessé de voir le jour, comme la méthode de calcul des fonctions d'aire par prédiction linéaire (WAKITA 1973).

Cette thèse s'ouvre par un chapitre consacré aux techniques "monodimensionnelles" d'étude de l'espace perceptif vocalique. Une revue critique des expériences publiées nous conduit à rejeter ces techniques. Le troisième chapitre présente une contribution originale à l'étude des indices perceptifs, par analyse multidimensionnelle des indices de dissimilarité entre des couples de voyelles françaises. Le chapitre contient également une étude critique et de nouvelles analyses des 15 études antérieures publiées. Les problèmes posés par l'interprétation d'un indice nous ont conduit à étudier en détail, au chapitre IV, la mesure du "deuxième formant équivalent (F'2)" dégagée en synthèse de la parole (FANT 1959). Nous suggérons une nouvelle méthode d'estimations de F'2 et présentons un modèle de catégorisation des voyelles dans l'espace bi-dimensionnel du timbre. L'utilisation du principe perceptif de "différenciation maxima dans l'espace du timbre" pour déterminer la liste des structures vocaliques attestées est en échec, comme le montre une réanalyse qui infirme des conclusions de LILJENCRANTS et LINDBLOM (1972). Il paraît probable que des contraintes articulatoires jouent un rôle important dans la structuration des systèmes vocaliques. Leur étude exige que l'on soit en mesure de passer des informations acoustiques aux paramètres articulatoires et vice versa. Les chapitres VI et VII ont pour but de présenter les deux "outils" qui permettent la liaison entre ces domaines : calcul des fréquences formantiques à partir de la forme du conduit vocal, et calcul de la fonction d'aire à partir des fréquences formantiques. Le chapitre suivant est consacré à l'étude de deux critères articulatoires ; le premier explique la labialisation "naturelle"

des voyelles postérieures et le second rend compte du fait que les voyelles se répartissent, de façon privilégiée, sur des "axes" allant, par exemple, de [i] à [ɑ] pour l'un d'eux. Le chapitre IX étudie plus particulièrement l'évolution diachronique [u > y], pour laquelle nous proposons une solution originale. Le chapitre X présente quelques observations sur la structure acoustique et la perception des voyelles nasales. C'est l'existence d'un hiatus considérable entre notre connaissance de la structure des voyelles nasales, comparée à celle des voyelles orales, qui nous a conduit à regrouper en un chapitre distinct ces quelques réflexions. Enfin le chapitre XI propose quelques remarques sur la théorie des traits distinctifs. Nous suggérons que seul un système de traits, fondé à la fois sur des critères articulatoires et acoustico-perceptifs permet à la phonologie d'aller plus loin qu'une simple notation des processus phonologiques (alternances synchroniques, mutations diachroniques, prédiction des systèmes attestés) et de décrire véritablement la "substance" du phénomène. Nous présentons un système de traits permettant de décrire le sous-système des voyelles fermées.

Après ces quelques remarques justifiant le choix de l'étude des voyelles hors-contexte, et présentant rapidement l'enchaînement des chapitres, nous pouvons présenter une première analyse de l'espace perceptif des voyelles à partir de techniques "monodimensionnelles".

C H A P I T R E II

ANALYSES MONODIMENSIONNELLES DE L'ESPACE PERCEPTIF DES VOYELLES

I - INTRODUCTION

C'est devenu un lieu commun d'affirmer que les sons du langage peuvent et doivent être étudiés du point de vue articulatoire, acoustique et perceptif. Si FISCHER-JØRGENSEN (1967) attribue la paternité de ce point de vue à R. JAKOBSON (1941), il faut cependant noter que des travaux sur les dimensions perceptives des stimuli simples définis au plan acoustique par la fréquence, l'intensité et la durée, sont antérieurs à 1941 (par ex. STEVENS & DAVIS (1938), STEVENS & VOLKMANN (1940), pour la définition du Mel ; cf. UNGEHEUER (1971), pour un intéressant aperçu historique).

De plus, comme le rappelle FONAGY (1963) dans le premier chapitre de "Die Metaphern in der Phonetik", les descriptions articulatoires mais surtout perceptives des voyelles ne manquent pas depuis le XVI° siècle. Le mérite de JAKOBSON a donc surtout été d'insister sur la nécessaire intégration de la triple description des sons.

Si les problèmes de la description articulatoire et acoustique des voyelles ont fait l'objet d'un nombre considérable d'études qui ont sérieusement défriché le terrain, l'étude de l'aspect perceptif se heurte à des difficultés beaucoup plus grandes. Celles-ci apparaîtront dans les chapitres II et III de ce travail. Cette division en deux chapitres souligne en effet une dichotomie essentielle des techniques d'analyses. Le chapitre II traite des descriptions qui définissent l'espace perceptif comme une somme de dimensions dont on trouve directement la trace dans les réponses des sujets. Le chapitre III étudiera et apportera une contribution à la description de l'espace perceptif à partir de la notion de distance subjective entre stimuli. Nous avons appelé les études du chapitre II "monodimensionnelles", car elles présupposent que le sujet peut placer chaque voyelle tour à tour sur chacune des dimensions. L'espace global n'est que la somme des dimensions. Monodimensionnel n'implique pas que le sujet ait une seule dimension à sa disposition. Une étude récente de CHIPMAN & CAREY (1975), que nous analyserons dans le chapitre

suivant, fait clairement le point sur les rapports entre analyses mono- et multidimensionnelle, telles qu'elles sont définies ici.

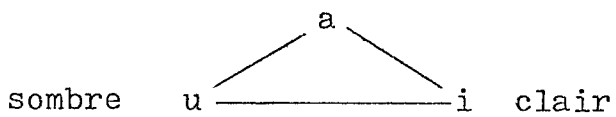
Plutôt que de poser d'abord les problèmes de l'analyse perceptive en termes théoriques, il nous a semblé préférable de décrire et commenter les travaux sur ce sujet jusqu'en 1977. C'est ce que font d'ailleurs tous les auteurs, qui ne présentent dans la plupart des cas qu'un cadre théorique rudimentaire. Il faut cependant mentionner les points suivants : l'hypothèse fondamentale est que le sujet, après une analyse acoustique du stimulus vocalique, procède (ou peut procéder) au positionnement de celui-ci sur plusieurs dimensions. Cette opération, de caractère phonétique, servira de point de départ à l'analyse phonologique (c'est-à-dire au classement dans un petit nombre de catégories spécifiques à la langue du sujet). Les rapports entre ces dimensions ne peuvent être précisés. Le problème de l'orthogonalité par exemple (c'est-à-dire de l'indépendance des dimensions) n'a guère de sens dans ce contexte théorique. Le mot "espace", d'ailleurs, est presque un abus de langage ; il faudrait parler plutôt d'une collection de dimensions. L'analyse monodimensionnelle présuppose de plus que le sujet est capable d'associer un label sémantique à chaque dimension (ou une couleur), même s'il n'est pas capable de préciser le nombre de dimensions qu'il utilise. On voit que le modèle est relativement contraignant, pour ne pas dire improbable.

II - FISCHER-JORGENSEN (1967)

E. FISCHER-JORGENSEN (1967) (FJ) décrit très brièvement une série d'expériences s'étendant sur une période de plus de 15 ans. Ces expériences ont porté sur le classement de voyelles naturelles sur des "axes", imposés par l'expérimentateur, portant un label sémantique tel que "clair"-"sombre", "étroit"-"large", selon le principe du différentiel sémantique (OSGOOD et al. 1957, SNIDER & OSGOOD 1969). L'idée d'une telle expérience remonte à R. JAKOBSON (1939, 1941), qui présente les "découvertes géniales de KOEHLER (1909) et de STUMPF (1926),

maîtres de l'acoustique moderne" (p. 374). Les voyelles possèdent selon ces auteurs, deux caractères irréductibles : l'un est le caractère chromatique-achromatique, l'autre le caractère clair (aigu) - sombre (grave). Les sons deviennent de plus en plus susceptibles d'être opposés au fur et à mesure que le chromatisme baisse : [a], la voyelle la plus chromatique, est la moins apte à l'opposition clair-sombre. Les voyelles étroites [i] et [u] (les moins chromatiques) sont les plus aptes à l'opposition clair/sombre.
On a donc le schéma suivant :

chromatique
↕
achromatique



Par analyse comparative des réponses sur les très nombreux axes présentés, FJ essaie de regrouper ces axes en un petit nombre de dimensions psychologiquement significatives qui doivent permettre une description compacte de la structure des systèmes vocaliques, une explication des mutations diachroniques et une explicitation du symbolisme associé aux différentes voyelles. FJ ne présente aucun critère permettant le regroupement des axes en dimensions.

La première dimension est le regroupement des axes "clair-sombre", "léger-lourd", "dur-mou", "plein-non-plein" parmi d'autres. Les voyelles s'ordonnent de la façon suivante sur cette dimension :

clair ← i y ɛ - œ ɔ - u - a → sombre

Le tiret indique que les voyelles ainsi regroupées ont été placées sensiblement au même endroit de l'axe. La figure 1-II indique la position approximative de cette dimension sur le triangle vocalique.

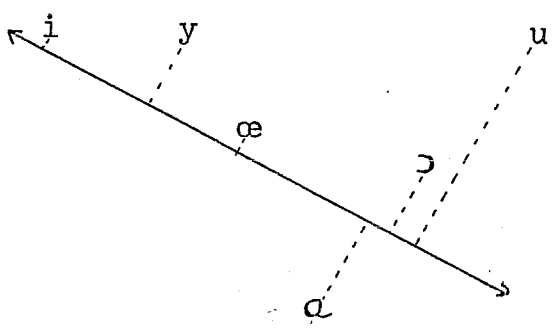


fig. I - II

Selon FJ, cette dimension pourrait représenter, en termes de fréquence formantique, une dimension F2-FI (cf. table I-II).

Table I-II

F2	-	FI	=	
2200		250	1950	Hz
1900		250	1650	
1600		500	1100	
1500		500	1000	
1100		700	400	
1000		500	500	
800		250	550	

N.B. : Les valeurs de cette table ne sont là qu'à titre d'exemple. Elles ne représentent pas les valeurs des voyelles utilisées par F.J.

En ce qui concerne cette dimension, on peut noter que LADEFOGED (1971), s'appuyant sur FANT (1959), préconise une dimension F'2-FI, sans référence d'ailleurs aux expériences de F.J. F'2 est une moyenne pondérée de F2 et F3. LINDAU (1975), une élève de LADEFOGED, a publié, à partir des données de LINDBLOM & SUNDBERG (1969) sur les fréquences des formants des voyelles cardinales de D. JONES (disque Linguaphone, Londres, 1956), un graphe formantique F2-FI /FI sur lequel on voit nettement que [a, ɔ, o, u] ont la même valeur sur F2-FI. En ce qui concerne les fréquences des formants des voyelles cardinales de D. JONES, DEBROCK & FORREZ (1976) ont publié des valeurs assez différentes de celles de LINDBLOM & SUNDBERG (1969) pour certaines voyelles ([a, u, o, ɛ]) à partir d'une méthode de prédiction linéaire. Ils n'ont pu également retrouver le premier formant des voyelles fermées ou mi-fermées. Ces écarts pourraient s'expliquer par l'utilisation d'un enregistrement plus ancien. [i] et [y] sont néanmoins trop proches. Les expériences de FJ de 1949-1951 sur 200 étudiants donnaient pour cette dimension une division tripartite /i — y — e / ɛ — œ — ø / u — œ — o — ɔ — a/ avec le deuxième groupe plus proche du premier que du troisième. Mais alors [ø] est trop éloigné de [e]. On remarquera que ce sont les voyelles labialisées d'avant qui font problème.

La deuxième dimension correspond aux axes "étroit-large", "petit-gros", "tendu-relâché", "ouvert-fermé".

On obtient la quadruple partition suivante :

étroit i y u - ɛ - œ - a ɔ large

Le problème essentiel est de savoir si cette dimension est différente de la première. La seule différence est que [u] et [a] ont rejoint [ɛ] et [œ] en s'opposant à [ɔ] . On peut tenter une description articulatoire de cette dimension : le centre de gravité de la langue (sauf pour [y]) se déplace d'avant en arrière quelque soit sa hauteur. A titre d'exemples, on peut prendre les tracés radiologiques de LADEFOGED (1964) ou BRICHLER (1970). On voit que [u] est, de ce point de vue, une voyelle centrale, comme le confirme d'ailleurs tous les tracés radiologiques publiés à ce jour. Nous proposerons plus bas une autre explication.

La troisième dimension, sous-tendue par les séries "serré-lâche", "compact-diffus" se présente ainsi :

serré i y u ɛ - œ - ɔ a lâche

Pour une autre expérience, l'axe "serré-lâche" a donné une répartition en deux groupes [i - y - u - o / ɛ - ø - a]
Il pourrait s'agir de la dimension articulatoire "ouvert-fermé" dont on sait d'ailleurs qu'elle corrèle de façon précise avec la valeur du premier formant.

Nous ne discuterons pas en détail les expériences portant sur les couleurs, en liaison avec la valeur symbolique de la voyelle. Notons seulement que la couleur jaune correspond au pôle "clair" de la première dimension, alors que le bleu correspond au pôle "sombre". Les voyelles [u] et [ø] sont très variables et [a] est la plus sombre de toutes.

A ce point, il n'apparaît même pas clairement si les dimensions sous-tendent des positions articulatoires, des fréquences formantiques, une combinaison des deux ou autre chose encore... Le nombre même de dimensions est sujet à caution. Aucune dimension, par exemple, ne sépare les voyelles labiales des non-labiales. On pourrait cependant voir la trace

de cette dimension dans le fait que les voyelles [i] et [y] sont toujours éloignées. Quant à la non-séparation des labialisées d'arrière, elle pourrait être due au fait qu'une voyelle d'arrière est non-marquée si elle est labialisée : Non-marqué "arrière" (au sens de CHOMSKY-HALLE) implique "labialisation".

Il nous semble possible de rendre compte des deux premières dimensions. Si l'on postule que les voyelles se répartissent dans un espace bi-dimensionnel FI et F'2, la méthode monodimensionnelle force le sujet à projeter les voyelles sur un axe unique. Il peut le faire d'une infinité de manières, à partir sans doute de [i] qui occupe une position extrême à la fois sur FI et F'2 (fig. 2-II), ce qui n'est pas le cas pour [ɑ] et [ɯ]. Si l'axe retenu par le sujet est proche de l'horizontale (α), on aura la série

i e y - ɛ ø - œ - a u o ɔ

et si l'axe est fortement incliné (β), on pourrait obtenir

i y - e ɛ - ø - u œ - o ɔ - a

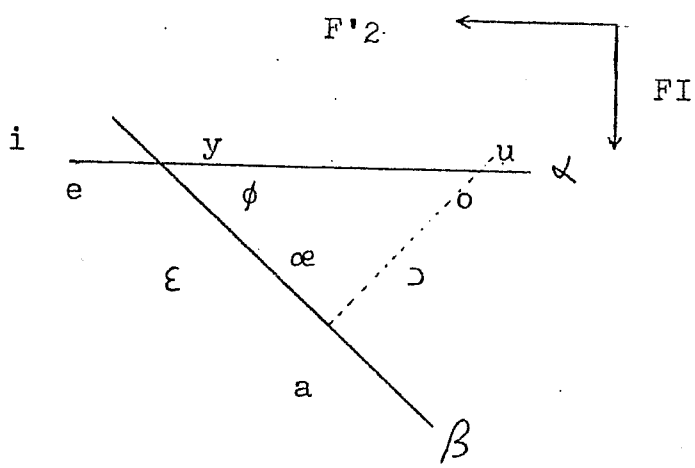


fig. 2 - II

La différence est que [ɯ] est beaucoup plus central dans ce cas. Si le sujet sent que [ɯ] est difficile à placer, il peut hésiter et placer [ɯ] en position centrale pour certains axes et à l'extrémité pour d'autres. Ceci conduira le chercheur à distinguer deux dimensions. Ce point sera repris plus complètement dans le chapitre suivant à la lumière d'un article récent de MILLER & CARTERETTE (1975).

III - JERRENTRUP et THÜRMANN (1974 a)

Les expériences similaires réalisées par d'autres auteurs ne semblent pasconfirmer entièrement les résultats deFJ. De plus, il apparaît que la plus importante d'entre elles (JERRENTRUP & THÜRMANN 1974 a, J & T) souffre d'un artefact qui infirme l'ensemble des conclusions. Comme il est difficile de démontrer la présence d'un artefact, nous présenterons d'abord les conclusions des deux auteurs. Puis nous réanalyserons leurs résultats bruts en tentant de mettre en évidence l'artefact dû à l'utilisation de voyelles synthétiques.

Le travail de J & T envisage en effet, à l'inverse du précédent, le classement de voyelles synthétiques à 2 formants sur 18 axes sémantiques. La table II-II précise les fréquences formantiques employées.

	i	e	ɛ	a	ɔ	o	u	y	ø	ə	
F1	360	430	510	660	450	370	260	260	390	510	Durée : 600msec
F2	2530	2240	1860	1260	930	810	770	1750	1420	1390	F_0 = 110 Hz

Table II-II

On remarquera que la synthèse à 2 formants exige que le deuxième formant de [i] et [e] soit plus élevé que la valeur naturelle moyenne. Le prétest de deux séries d'axes permet aux auteurs de dégager par analyse factorielle classique 3 facteurs ou dimensions. Les corrélations entre les axes et les facteurs extraits (op. cit., table 4, p. 398) conduisent J & T à interpréter les 3 facteurs comme étant les dimensions de "clarté" (Helligkeit), de "valence" (Valenz) et de "puissance" (Potenz, Stabilität, Form). Ces dénominations abstraites recouvrent les paires suivantes : "clarté" - étouffé-perçant, émoussé-pointu, large-étroit, sombre-clair ; "valence" - laid-beau, son non-humain-son humain, rigide-mouvant, haïssable-agréable ; "puissance"-impur-pur, simple-complexe, structuré-informe, fort-faible.

L'expérience principale, test et retest, porte sur 6 paires pour chacune des 3 dimensions extraites précédemment. 20 sujets étudiants ont été testés. En utilisant un algorithme mis au

point par JOHNSON (1967), J & T extraient plusieurs partitions des distances entre voyelles, distances calculées par 2 méthodes : distance euclidienne inter-stimuli et coefficient de corrélation. On obtient le groupement moyen suivant :

((a , ɛ) ɔ) ((y (o , ø)) u) (ə , e , i)

[u] est assez variable, ainsi que [ə] . Les parenthèses permettent de reconstruire la hiérarchie des distances et s'interprètent comme en syntaxe. Notons que le même algorithme sur les distances euclidiennes calculées à partir des fréquences des formants ne conduit pas à la même partition. Pour l'analyse factorielle des intercorrélations d'axes, 3 facteurs rendent compte de plus de 91 % de la variance. Les auteurs indiquent que les dimensions sont fortement corrélées, bien que, pour nous, la signification à donner à ce résultat reste obscure dans le modèle envisagé. L'analyse factorielle des corrélations de stimuli révèle à nouveau 3 facteurs qui rendent compte de 98 % de la variance (dont plus de 60 % pour le premier). La première dimension oppose [ɑ, ɛ, ɔ] aux autres voyelles alors que la deuxième regroupe [y, ø, u, o] à une extrémité de l'axe. L'accord entre la méthode de classification hiérarchique de JOHNSON et l'analyse factorielle est donc total. Les voyelles se répartissent de la manière suivante sur les trois dimensions :

I - clarté Helligkeit	a ɛ-ɔ i - e a y - o - ø - u		clair-sombre pointu - émoussé
	clair	sombre	strident - étouffé
II - valence Valenz	ø o e y u ɔ ə i a ɛ		beau - laid méchant - bon
	positif	négatif	dégoûtant - attirant
III - puissance Potenz	ɔ - e a - ə ɛ u o ø - i y		stable - instable impérieux - humble
	fort	faible	fort - faible

Table III - II

I	ɑ	ɛ	ɔ	i	e	ə	y	o	ø	u	c	hell - dunkel clair - sombre
2	ɔ	ɑ	ɛ	i	e	ə	o	y	ø	u	c	spitz - stumf pointu - émoussé
3	ɑ	ɔ	ɛ	e	i	ə	o	y	u	ø		lebhaft - lahm animé - mobile
4	ɑ	ɛ	e	ɔ	i	ə	o	u	y	ø	c	schrill - dumpf strident - étouffé
5	ɔ	ɑ	ɛ	e	ə	u	i	o	y	ø	p	herrisch - unterwürfig impérieux - humble
6	i	ɑ	ɛ	ɔ		ə	y	o	e	u	p	labil - stabil instable - stable ([ø] manque)
7	ɛ	ɑ	ɔ	ə	i	u	o	e	y	ø		giftig - harmlos toxique - non-toxique
8	ɑ	ɔ	ɛ	e	i	u	o	ø	ə	y		ausgeprügt - schlicht ampoulé - simple
9	ə	ɔ	ɛ	ɑ	e	u	o	i	ø	y	p	stark - schwach fort - faible
I0	ə	ɑ	ɛ	ɔ	e	o	u	ø	i	y		geformt - zefließend structuré - informe
II	ɛ	ɑ	ə	ɔ	e	u	i	o	ø	y		heftig - sanft violent - doux
I2	ɛ	ɑ	i	ə	ɔ	u	y	e	o	ø	v	häßlich - schön laid - beau
I3	ɛ	ɑ	ə	i	ɔ	u	e	o	ø	y	v	boshaft - liebevoll méchant - bon
I4	ɛ	ɑ	ə	i	u	ɔ	e	y	ø	o	v	abstoßend - anziehend dégoûtant - attirant
I5	y	ɛ	i	ɑ	ɔ	ə	ø	o	e	u		schmal - breit étroit - large
I6	e	ɔ	ə	ɑ	i	u	ø	y	o	ɛ		rein - gemischt pur - impur
I7	ɛ	i	ə	ɑ	u	ɔ	o	y	e	ø		künstlich - natürlich artificiel - naturel
I8	e	o	i	ɑ	ø	ɛ	y	ɔ	u	ə		sprachlich - nichtsprachlich linguistique - non-lg.

Ces dimensions ne sont pas indépendantes : le coefficient de corrélation entre "valence" et "clarté" vaut 0.87, ce qui est très important. Il est encore de 0.61 entre "clarté" et "puissance" et de -0.52 entre "valence" et "puissance". Plus important encore, les coefficients de corrélation entre F2 ou F2-F1 et les 18 axes sont très faibles : tous (sauf un) sont compris entre -0.38 et 0.38. Par contre, on observe de fortes corrélations positives ou négatives entre les axes et F1 : par ex. strident-F1 = -0.78 ; bon-F1 = 0.77 ; humble-F1 = -0.75. Il est cependant curieux (et inquiétant) de constater que ces 3 axes représentent 3 dimensions différentes et ont pourtant des coefficients de corrélation identiques.

Cet ensemble de faits nous a conduit à tenter une nouvelle analyse des données. A partir du tableau 8, p. 410 de l'article de J & T, nous avons construit la table III-II. Cette table présente sur la première ligne l'axe présentant la plus grande variance, c'est-à-dire la meilleure séparation des voyelles. Il se trouve de plus que cet axe "clair - sombre" présente les voyelles dans l'ordre indiqué par l'algorithme de JOHNSON et l'analyse factorielle (première dimension extraite). Nous avons ensuite placé les autres axes dans un ordre approximatif de similitude par rapport à la première ligne. La ligne 18 présente l'axe le plus différent de l'axe placé sur la première ligne. Ce tableau reprend donc les moyennes de groupe au niveau ordinal (en inversant les axes si nécessaire ; leur orientation est en effet arbitraire).

On peut faire, à partir de ce tableau, les remarques suivantes : Une majorité d'axes présentent la division tripartite [a - ɛ - ɔ i - e - ə y - o - ø - u]

Le tiret indique qu'à l'intérieur des trois groupes, l'ordre n'est pas pertinent. Ceci est vrai pour les axes associés par J & T à la dimension "clarté" mais aussi pour les axes associés à la dimension "puissance" (lettres c ou p dans la dernière colonne de la table III-II). Les 8 premiers axes présentent cet ordre avec des intrusions aléatoires des voyelles [e, i , ə]

On peut poser le problème de l'influence de la forme phonique du mot allemand désignant l'axe au vu de l'axe 6 : stabil/labil provoque un déplacement important de la voyelle [i]. Cet ordre sépare nettement les voyelles ouvertes des voyelles labialisées. Les axes 12, 13 et 14 (ainsi que 11 dans une moindre mesure) présentent un ordre légèrement différent. [e] est systématiquement repoussé à droite, ainsi que [ɔ], alors que [ə] et [i] se déplacent vers la gauche. On a donc l'ordre

ɛ - ɑ i - ə ɔ - u e o - ø - y

qui n'est pas très éloigné de l'ordre de l'axe 17, naturel - artificiel. Ces axes semblent correspondre à une dimension affective. L'apparition d'une dimension certes intéressante pour le poète, mais d'un intérêt relatif pour le phonéticien, montre bien une des limites de cette approche. De plus, il faut souligner que cette dimension affective a été obtenue à partir de stimuli synthétiques à 2 formants dont on connaît le caractère peu naturel. Il nous paraît dangereux de vouloir tirer de ceci une conclusion quant à l'ordre des voyelles naturelles sur une telle dimension affective. Mais il y a plus grave. La dimension primordiale est une dimension "claire - sombre", également obtenue par FJ. Or l'ordre des voyelles n'est absolument pas le même pour ces deux auteurs. L'absence de coefficient de corrélation significatif avec F2 ou F2-F1 permettait déjà de présager de la chose.

FJ :	i	y	ɛ - œ		ɔ	u	ɑ			
J & T	ɑ	ɛ	ɔ	i	e	ə	y	o	ø	u

Il nous semble que nous sommes en présence d'un artefact majeur dû à l'utilisation de voyelles synthétiques. Notre expérience en synthèse nous suggère que le caractère "éclatant" et "métallique" de [ɑ] a pu fausser les résultats. Il est peut être significatif que J & T ne discutent pas les résultats antérieurs de FJ, qu'ils citent néanmoins dans la bibliographie. On doit également s'interroger sur l'existence de la dimension "puissance". Celle-ci est obtenue par "l'addition" des axes

"stable - instable"	u	e	o	y	ə		ɔ	ɛ	a	i
"impérieux - humble"	ɔ	a	ɛ	e	ə	u	i	o	y	o
"fort - faible"	ə	ɔ	ɛ	a	e	u	o	i	o	y

Il est évident que l'axe stable-instable est inversé par rapport aux deux autres. C'est pour cette raison que [i] vient s'intercaler dans la série [u, o, ø, y]. De plus il est surprenant que J & T ne publient pas pour le test principal une table identique à la table 4 pour le pré-test, qui permettrait de connaître les corrélations entre les axes et les trois dimensions. Le pré-test indique en effet que l'axe 8 "ampoulé - simple" corrèle maximalement avec la dimension "puissance" (0.91). Or ce même axe 8 n'est pas utilisé pour calculer l'ordre des voyelles sur la troisième dimension dans le test principal. S'il avait été utilisé, la dimension 3 aurait été confondue avec la dimension 'clarté'.
Les axes 15, 16 et surtout 18 du tableau III-II présentent les voyelles dans un ordre encore différent, dont il faudrait rendre compte.

En conclusion, notre analyse du travail de J & T nous a permis de dégager deux dimensions. La première est une dimension "clair-sombre", qui ne correspond pas à celle mise en évidence par FJ pour des voyelles naturelles. L'emploi de voyelles synthétiques de qualité moyenne (deux formants) y est sans doute pour quelque chose. La deuxième dimension est soit une dimension "affective", soit une dimension résultant de l'interaction d'une dimension affective et de la première dimension. On ne peut exclure que l'utilisationde voyelles synthétiques soit à l'origine de cet artefact. Il est possible que d'autres dimensions soient nécessaires pour rendre compte des axes "linguistique - non-linguistique" et "pur - mélangé". De plus, la durée fixe adoptée pour la synthèse (600 msec) ne permet pas de synthétiser l'ensemble des voyelles de l'allemand comme l'a montré BENNETT (1968) dans son étude de l'influence du timbre et de la durée sur la perception des voyelles de l'anglais et de l'allemand. On ne peut exclure une interaction entre le nivellement artificiel de la durée et le placement des voyelles sur les différents axes, conséquence de l'utilisation phonologique de la durée en allemand.

Cet ensemble de faits nous a conduit à ne pas poursuivre notre analyse des résultats d'un travail ultérieur de J & T (1974 b), comparant les performances de phonéticiens et de sujets naïfs pour le même ensemble de voyelles synthétiques.

IV - JOHANSSON (1975)

Le travail de JOHANSSON (1975), qui semble ignorer la contribution de J & T, a pour but explicite de poursuivre les travaux de FJ. 10 voyelles suédoises naturelles ont été utilisées. La table IV-II précise les fréquences formantiques mesurées au sonagraphe.

Table IV-II

	i	e	ɛ	ɑ	a	o	u	y	ʉ	ø
F1	200	350	500	800	500	350	250	225	250	400
F2	2300	2200	2000	1250	900	650	600	2200	1850	1600
F3	2900	2650	2300	-	-	-	-	2400	2400	2200

14 sujets non-phonéticiens ont passé le test 1 qui demandait d'indiquer pour chacune des voyelles un ou plusieurs adjectifs tirés de la liste reproduite à la table V-II (test 1). Les résultats, basés sur la fréquence de sélection des adjectifs, sont présentés à la même table. 4 adjectifs, "clair", "sombre", "étroit" et "étouffé", donnent lieu à une répartition en 3 groupes / i e ɑ / ɛ a y / ʉ u o ø /
Si l'on tient compte du fait que [y], à l'inverse de [ø], est faiblement labialisé en suédois (cf. F3 [y] = 2400 Hz), cette dimension oppose les voyelles palatales aux voyelles arrondies, avec, entre les deux, un "marais" comprenant [ɛ, a] et [y].
Cet ordre, y compris le caractère relativement sourd de [ɛ], se retrouve pour la première dimension de FJ.
La dimension "doux" pourrait correspondre à la deuxième dimension de FJ : [a] et [u] se trouvent plus près du centre. La dimension "brillant" oppose les voyelles arrondies [u o ʉ ø] à [a e ɛ ɑ] . On peut, bien sûr, proposer une interprétation phonétique, mais il faut noter que l'adjectif est "klar", ce qui pourrait expliquer la position extrême de [a]. Pour 4 axes

sur 6, d'ailleurs, une voyelle extrême correspond à une voyelle de l'adjectif.

Le deuxième test, qui exigeait un choix libre d'adjectifs, ne donne pas de réponses utiles en raison de la dispersion des réponses. Trois points sont cependant à noter : l'influence de la structure phonique de l'adjectif est évidente (ex : [ɛ] - bräkig (bêlement de chèvre), [i] est très souvent décrit par "pointu" (spetsig), les très rares adjectifs "articulatoires" ne correspondent pas à une réalité phonétique (ex : [u] est mi-ouvert).

Le test III, identique au test I quant à la méthode, a été présenté à un groupe assez hétérogène de 33 sujets. Les résultats sont reproduits à la table V-II (test III). Ce test confirme l'importance de la dimension "clair - sombre" où [i, a, ɛ] s'opposent à [ʉ u ø o] (axes "clair", "perçant", "sombre", "dur", et sans doute "sombre (ténébreux)" malgré la position de [ɛ]. [y] et [a] sont relativement flottants. Pour la dimension "brillant" (klar), on ne peut à nouveau éliminer l'hypothèse d'une influence phonique en raison de la position extrême de [a] . Mais cet ordre se retrouve pour l'axe antonyme "étouffé" (dov), ce qui rend moins certain la présence d'un artefact. Quant à la dimension "doux (mou)", nous ne discernons pas la raison de l'ordonnancement. Serait-ce une dimension "affective" ?

En conclusion, le travail de JOHANSSON confirme l'existence d'une dimension "clair - sombre" déjà mise en évidence par FJ. La variabilité de certaines voyelles pourrait amener à scinder cette dimension. On peut discerner la trace d'une dimension "brillant - étouffé (mat)" qui n'apparaît pas chez FJ. La troisième dimension éventuelle ne correspond pas non plus à la troisième dimension de FJ.

Si l'on excepte le travail de UNGEHEUER (1965), qui est trop succinct pour permettre une analyse critique, ces trois études sont les seules analyses monodimensionnelles publiées à ce jour. Il convient néanmoins d'examiner quelques travaux dont la méthodologie est proche de celle employée par les trois auteurs précédents.

Test I/III

Table V - II

I	i	e	ɐ	ɛ	a	y	ʉ	u	o	ø	ljus	clair
IIII	i	y	ɐ	ɛ	e	a	u	ʉ	ø	o	ljus	clair
III	i	y	ɐ	ɛ	e	ø	ʉ	a	o	u	spetsig	perçant
I	e	i	ɐ	y	ɛ	ʉ	a	o	u	ø	mörk	sombre
III	i	ɐ	y	e	ɛ	ʉ	a	o	ø	u	mörk	sombre
I	ɐ	i	e	y	ʉ	ɛ	a	ø	u	o	tunn	étroit (grèle)
I	e	i	ɐ	o	ʉ	u	ø	a	ɛ	y	fyllig	doux (agréable)
I	a	e	ɛ	ɐ	i	y	ø	ʉ	o	u	klar	brillant
III	a	ɐ	e	ø	o	y	ʉ	i	ɛ	u	klar	brillant
I	a	e	ɛ	ɐ	i	y	ø	ʉ	o	u	dov	étouffé (mat)
III	ɐ	e	i	ɛ	y	a	u	o	ʉ	ø	hård	dur
III	i	ɐ	a	e	u	y	o	ʉ	ɛ	ø	dyster	sombre ténébreux
III	ɐ	ø	i	a	e	y	ɛ	u	ʉ	o	mjuk	doux mou

V - CALBRIS (1974)

L'étude de CALBRIS (1974) porte sur la perception des voyelles orales françaises par des français et des étrangers. 41 sujets français sans formation phonétique devaient juger 14 paires d'énoncés français (ex : du pain - deux pains) opposant les 10 voyelles orales. Seules les voyelles adjacentes horizontalement ou verticalement sur le trapèze vocalique ont été testées. Les sujets devaient simplement indiquer la voyelle la plus grave. CALBRIS a obtenu l'ordre suivant sur la dimension aigu - grave

i ɛ e y œ ø ɔ a u o

aigu ←——————————→ grave

L'axe est faiblement incliné sur l'horizontale (cf. fig. 2-II). La voyelle [ø] est plus aiguë que dans l'expérience de JOHANNSON. L'ordre des voyelles sur cette dimension est intermédiaire entre ceux des deux premières dimensions de FJ.

L'expérience a été répétée pour la dimension "ouvert - fermé" par 49 sujets français. Notons qu'une dimension de ce type n'a pas été mise clairement en évidence dans les études que nous avons analysées. L'ordre est conforme à ce que prévoit une analyse articulatoire. Pourtant [y] et [œ] sont sentis comme plus fermés que [i] et [ɛ]. La labialisation qui les caractérise pourrait expliquer le comportement des sujets. Si l'on combine les deux dimensions de façon orthogonale, on obtient le schéma 3-II (CALBRIS (1974) p. 107). La clarté du schéma exige que l'on porte trois [ɑ] pour les trois séries verticales. Ceci indique clairement que les axes ne peuvent être orthogonaux.

L'essai d'interprétation articulo-perceptive qui suit, fondée sur BRICHLER (1970), est moins convaincant. Il n'y a pas de raison particulière de privilégier une interprétation articulatoire au détriment d'une interprétation acoustique.

L'étude de CALBRIS est la seule à mettre en évidence une dimension "ouvert - fermé". On peut se demander dans quelle mesure les réponses n'ont pas été induites par une présentation trop directive.

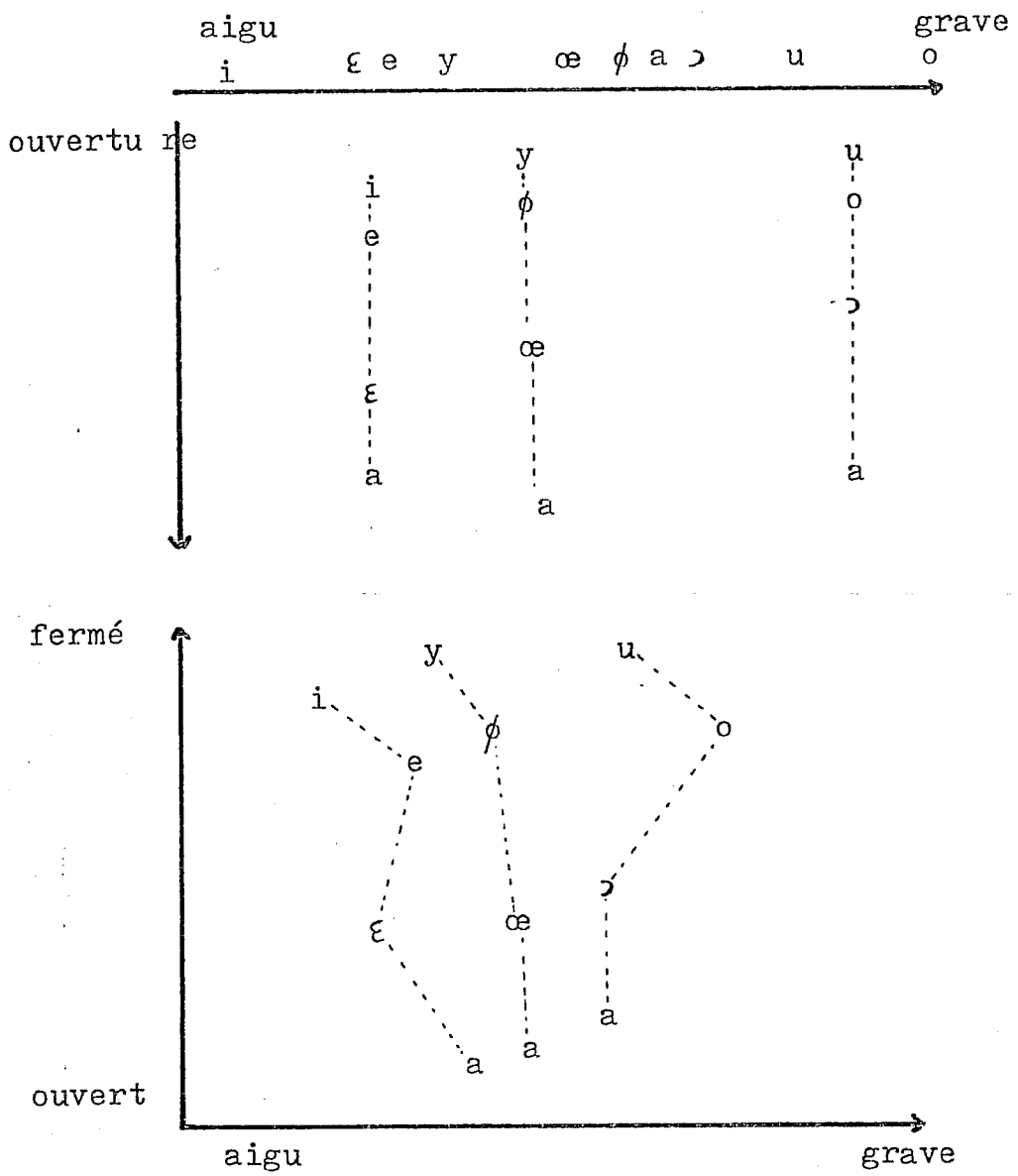


fig. 3 - II
(CALBRIS 1974 p. 107)

Pour ce qui est de la perception des sujets étrangers, si tant est qu'il soit acceptable de tester une population aussi hétérogène, on retrouve une dimension "grave - aigu ". La seule modification est que [ɔ] a rejoint [o]. Quant à la dimension "ouvert - fermé", à la différence des sujets français, [y] et [œ] ne sont pas sentis comme plus ouverts que [i] et [ɛ].

La conclusion d'une "perception plus phonologique que phonétique et plus proprioceptive qu'auditive" pour les sujets français nous paraît pour le moins prématurée. Quant aux conclusions, elles présupposent que les deux dimensions définies plus haut sont effectivement les deux seules dimensions utilisées par les sujets, ce qui n'a pas été démontré . On perçoit nettement les limites de ce genre d'études qui imposent les dimensions de réponse et qui, par un raisonnement circulaire, concluent que ces dimensions sont bien les dimensions utilisées par les sujets.

VI - LES EXPERIENCES D'ASSOCIATION DE VOYELLES ET DE SONS PURS

Il nous semble que l'on peut considérer les expériences d'association de voyelles et de sons purs comme des analyses monodimensionnelles. La dimension (implicite) est une dimension "clair - sombre". La psychophysique nous apprend en effet que cette dimension est sans conteste celle qui est le plus souvent associée à une variation de fréquence pour un son pur. De très nombreuses expériences de ce type ont été réalisées depuis le milieu du XIXe siècle. Leur analyse détaillée serait fastidieuse. Nous ne prendrons que la dernière d'entre elles, réalisée par FANT (1959, p. 87-89). La table suivante résume les résultats.

fréquence (Hz)	Voyelle	fréquence (Hz)	Voyelle
4000	i	1250	æ
2500	e	1000	a
2500	y	625	o
1750	ɨ	400	u
1750	ø		

Ces résultats confirment les résultats antérieurs tels ceux de FARNSWORTH (1937). 12 sinusoïdes dont la fréquence variait de 250 à 4000 Hz ont été présentées à 3 groupes de 20 sujets, dans une tâche d'identification, en choix forcé, portant sur les 9 voyelles ci-dessus. On a observé une variabilité importante des réponses. La moyenne montre pourtant une bonne corrélation avec F'2.

L'ordre des voyelles sur cette dimension fréquentielle ne correspond que partiellement à celui des expériences d'E. FISCHER-JØRGENSEN pour la dimension "grave - aigu ".

i ey ʉø æ ɑ o u

aigu ⟵⟶ grave

On notera en particulier la séparation entre [ɑ] , [o] et [u] .

Les données ci-dessus peuvent être complétées par le travail de DELATTRE, LIBERMAN, COOPER et GERSTMAN (1952). Les auditeurs devaient associer une voyelle synthétique à un formant variant de 240 à 3000 Hz à l'une des voyelles suivantes : i e ɛ ɐ ɑ ɔ u . A partir de ces réponses préliminaires, 7 voyelles furent présentées dans un test d'identification, dont les résultats sont présentés dans la table suivante.

Fréquence du formant unique (Hz)	Voyelle	% d'identification
240	u	86
480	ɔ	82
720	ɑ	75
1200	ɐ	89
2160	ɛ	61
2520	e	64
3000	i	89

On remarquera que l'ordre est le même que celui obtenu par FANT (1959). Les scores d'identification pour [e] et [ɛ] sont nettement plus faibles que la moyenne, ce qui indique qu'une voyelle à un seul formant n'est pas une bonne approximation. Cette réduction d'information a été utilisée en reconnaissance de la parole (TEACHER, KELLETT et FOCHT 1967) sous l'appellation de "formant unique équivalent" (BUTCHER 1974).

L'intérêt de ces études reste limité. Elles prouvent seulement que certaines caractéristiques physiques du timbre peuvent être ordonnées sur un axe de manière à rendre compte de leur étagement fréquentiel. Le timbre est donc un "objet" partiellement analysable.

VII - CONCLUSIONS

Les techniques d'analyses monodimensionnelles nous paraissent peu intéressantes. Elles se heurtent à de nombreuses difficultés aussi bien pratiques que théoriques :

- l'hypothèse que chaque dimension peut être associée à un (ou plusieurs) label (s) sémantique (s) est une hypothèse gratuite. Elle ne peut être ni infirmée ni prouvée.
- Aucun test quantitatif ne permet de connaître le nombre de dimensions. Ceci est la conséquence directe de l'utilisation d'un modèle qui ne permet pas d'exprimer la notion d'indépendance des dimensions. Il apparaît souvent que l'ordre des voyelles n'est que "légèrement" différent sur deux axes : une interversion de 2 ou 3 voyelles sur 10, par exemple. Il est alors impossible de décider si l'on est en présence d'une ou de deux dimensions.
- Il n'est pas certain que les dimensions extraites soient toutes de type "linguistique". Plusieurs analyses semblent mettre en évidence des dimensions "affectives" qui reflètent une perception "esthétique".

Sur le plan pratique, les défauts principaux proviennent :

- de la nécessité pour l'expérimentateur de proposer la liste des "labels".
- de la sensibilité importante de la méthode à la forme phonétique du label choisi pour une dimension.
- de la sensibilité de la méthode à la qualité de la voyelle. Des différences de qualité peuvent introduire des réponses "esthétiques".

L'absence d'accord entre les différentes études, ainsi que les nombreux artefacts que l'on peut mettre en évidence, ne militent pas en faveur de la méthode. Nous avons rejeté ce type d'approche au profit des techniques multidimensionnelles qui seront présentées dans le chapitre suivant.

C H A P I T R E III

ANALYSES MULTIDIMENSIONNELLES DE L'ESPACE PERCEPTIF DES VOYELLES

I - INTRODUCTION

Le chapitre précédent a montré que les analyses monodimensionnelles sont inadéquates pour rendre compte de la structuration de l'espace perceptif. Le modèle implicite est très contraignant et la forme-même de l'analyse fait naître un certain nombre d'artefacts. Ce chapitre traite des analyses multidimensionnelles. Il présente, après une analyse des fondements théoriques et une présentation détaillée du modèle mathématique, une revue des différents travaux ainsi qu'une contribution originale à l'étude de la perception des voyelles du français. En raison de la diversité des méthodes, une présentation des travaux antérieurs n'est pas chose facile. Aussi, lorsque les résultats détaillés étaient disponibles, avons-nous procédé à une réanalyse par la méthode INDSCAL (CARROLL et CHANG 1970). Cette unification de la présentation des résultats permet des comparaisons plus fructueuses.

II - FONDEMENTS THEORIQUES DE L'ANALYSE MULTIDIMENSIONNELLE

A/ Hypothèses générales

L'analyse multidimensionnelle a pour but de mettre en évidence les dimensions psychologiques qui sous-tendent les mesures de similarité entre des stimuli. On suppose que le système perceptif transforme chaque percept auditif (issu de la présentation des stimuli) en un point dans un espace à plusieurs dimensions. Cette transformation s'effectue de manière à ce que les distances entre les points correspondent approximativement aux similarités entre les stimuli.

L'analyse multidimensionnelle repose sur les hypothèses suivantes :

a)- les stimuli peuvent être représentés par des points dans un espace.

b)- un espace comporte plus d'une dimension.

c)- Cet espace est relativement invariant par rapport aux choix des sujets, des conditions expérimentales, du contexte des stimuli, etc...

d)- Les sujets sont capables de fournir des indices de similarité (ou de dissimilarité), lorsqu'on leur présente successivement deux stimuli.

e)- Ces indices de similarité sont relativement stables pendant la durée de l'expérience (quelques dizaines de minutes).

f)- Les fluctuations temporelles et quantitatives de ces indices sont aléatoires.

g)- Les indices de similarité sont fonction des différences sur chacune des dimensions de l'espace. Elles sont en outre fonction de toutes les différences physiques utilisées "normalement" par le sujet, et non d'un aspect choisi arbitrairement par le sujet pour la durée de l'expérience.

Il existe également des contraintes plus particulières : Pour la forme de la fonction liant les indices de similarité aux distances, il faut supposer pratiquement que les distances sont des distances de MINKOWSKI.

$$d(x,y) = \left[\sum_{i=1}^{n} |x_i - y_i|^R\right]^{1/R}$$

avec, pour cas particulier important, les distances euclidiennes où r = 2 ; (d(x, y) : distance entre x et y ; x_i , y_i : coordonnées de x et y sur la ième dimension).

D'autres fonctions plausibles sont exclues : l'indice de similarité ne peut être, par exemple, le produit des différences sur chaque dimension (TVERSKY et KRANTZ 1970).

On suppose souvent, pour des raisons de simplicité mathématique, que les dimensions sont indépendantes (orthogonales). Beaucoup d'utilisateurs insistent également sur le fait que les dimensions extraites doivent correspondre à des attributs physiques des stimuli. Il ne faut pas s'attendre à une correspondance bi-univoque entre les attributs physiques et les dimensions extraites. On peut cependant exiger une relation ordinale (c'est-à-dire conservant au moins l'ordre des paramètres) entre certaines dimensions psychologiques et

certaines dimensions (ou groupes de dimensions) physiques. Ce n'est d'ailleurs qu'au prix de cette hypothèse supplémentaire que l'on pourra interpréter les dimensions.

B/ Problèmes actuels

Certaines de ces hypothèses ont récemment été remises en question par CHIPMAN et CAREY (1975) dans une étude concernant les bandes de bruit. Il nous paraît indispensable de présenter brièvement cette étude en raison des perspectives importantes ouvertes. Nous retrouverons ces problèmes dans la discussion de notre expérience.

Le problème est abordé par le biais du rapport entre les analyses uni- et multidimensionnelles d'un même ensemble. L'hypothèse de base est que l'espace mulditimensionnel devrait avoir le même nombre de dimensions que le nombre de dimensions monodimensionnelles indépendantes. Pour les bandes de bruit, de nombreuses études ont montré la possibilité de réaliser des analyses monodimensionnelles, ce qui distingue ce type de stimuli des voyelles. D'autre part, comme les dimensions extraites des analyses monodimensionnelles ne sont pas liées à une seule variable physique, la solution multidimensionnelle ne devrait pas forcément être liée aux variables physiques. On sait par exemple que l'intensité d'une bande de bruit est liée à la fréquence. Les dimensions des bandes de bruit révélées par analyse monodimensionnelle sont : l'intensité subjective ou "loudness", la densité, le volume et la hauteur subjective. La solution multidimensionnelle ne semble posséder que deux dimensions. L'une d'entre elles est très certainement l'intensité subjective. Mais l'autre ne correspond pas à la fréquence ou à la hauteur subjective. Il semble y avoir interaction entre la fréquence et l'intensité. Aux faibles intensités, les différences fréquentielles ont beaucoup moins d'effet sur l'indice de similarité qu'aux fortes intensités. Mais les axes peuvent également être interprétés comme des axes de "densité" ou de "volume". Il apparaît que densité et intensité subjective sont linéairement dépendantes, par exemple. Le fait que la hauteur

n'apparaisse pas clairement est gênant, en raison de la facilité avec laquelle cette dimension est mise en évidence dans une analyse monodimensionnelle. On observe seulement une relation ordonnée entre la hauteur et la position des points ; cette relation n'est valable qu'à une intensité subjective donnée.

Cette étude pose deux questions fondamentales :

- Est-il possible de projeter les points sur une dimension intéressant la totalité de l'espace perceptif ?

- La notion même de "dimension" est-elle pertinente ? Une réponse élégante, mais révolutionnaire, à cette question consiste à affirmer la réalité psychologique de la configuration des points et non des dimensions. Les points ne peuvent être regroupés que sur la base de leur proximité spatiale. Dans ce cas, la notion de dimensions indépendantes disparaît au profit de la forme de la configuration. Dans le domaine de la perception des couleurs, par exemple, cette hypothèse permet de représenter les couleurs sur un cercle (SHEPARD 1962). Dans CHIPMAN et CAREY (1975) les points se projettent sur un arc de cercle d'une manière qui rend parfaitement compte des différences de hauteur. Le problème essentiel est de savoir si ces configurations sont le résultat ou non de l'utilisation d'une méthode particulière de "scaling". Quant à la première question, elle interroge la notion d'homogénéité de l'espace. Des stimuli très différents s'inscrivent-ils dans le même espace perceptif ? Bien qu'il soit trop tôt pour discuter la valeur de ces questions, il s'agit là de précieux axes de recherche qu'il faut garder à l'esprit au moment de l'interprétation des résultats.

Une deuxième étude pose des problèmes d'interprétation d'une grande importance méthodologique (MILLER et CARTERETTE 1975). Il s'agit à nouveau de l'absence d'une dimension unique de hauteur dans une expérience portant sur l'espace perceptif des sons musicaux. Ces sons musicaux synthétiques à une même fréquence fondamentale s'opposaient par le nombre d'harmoniques (3, 5 ou 7), la variation d'intensité des harmoniques en fonction du temps, et le délai d'apparition des

harmoniques en fonction de leur rang. Or il apparaît que deux dimensions perceptives recouvrent la même dimension physique (nombre d'harmoniques), dont on pouvait penser à priori qu'elle était linéaire (fig. 1-III).

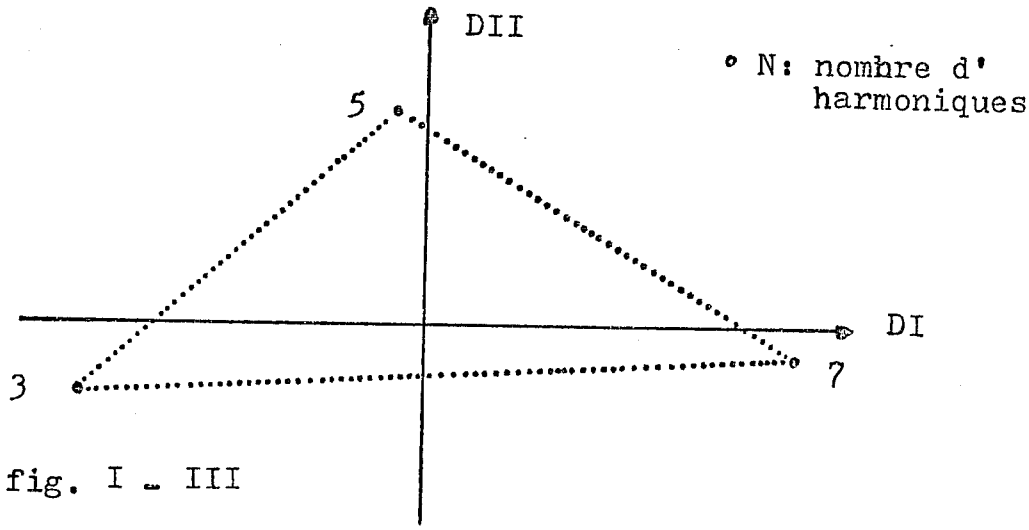


fig. I - III

Plusieurs explications sont possibles. La deuxième dimension peut être artefactuelle. Elle ne serait due qu'à l'imprécision des mesures fournies par les sujets. Mais une autre interprétation est possible : lorsque le domaine fréquentiel est trop étendu, les distances entre les stimuli extrêmes sont sous-estimées par rapport aux distances entre les stimuli contigus. L'espace ne serait plus globalement linéaire. Les jugements de similarité ne seraient valables que dans une portion de l'espace seulement. En ce qui concerne les voyelles, les seuls jugements valides seraient ceux portant sur les voyelles adjacentes du trapèze vocalique ; les jugements sur les voyelles extrêmes ([i, ɑ, u]) ne seraient pas valables.

Même si ce type de distorsion ne paraît pas se produire systématiquement, il n'est pas à exclure pour les dimensions de rang élevé.

Une dernière étude théorique mérite de retenir notre attention (BERNSTEIN 1975). Cette étude jette, elle aussi, un doute sur la possibilité d'effectuer des jugements de similitude

sur l'ensemble de l'espace vocalique. Bernstein part d'une définition très formelle de la décomposabilité d'un espace, qui suppose que les distances entre deux voyelles peuvent être exprimées par une fonction monotone, réelle et croissante de 3 variables, elles-mêmes fonctions réelles de F1, F2 et F3. Il y aura violation de la décomposabilité dans le cas suivant : soient deux ensembles de 4 voyelles ; pour les deux ensembles, les valeurs de F2 se trouvent à 1000, 1250, 1500 et 1750 Hz. Pour le premier groupe, F1 est fixe à 250 Hz. Pour le deuxième groupe, F1 est égal à 500 Hz. F3 vaut toujours 2500 Hz. Il y aura violation de la décomposabilité si, la distance entre les voyelles (250, 1000) et (250, 1250) ayant été jugée plus grande que la distance entre la voyelle (250, 1500) et la voyelle (250, 1750), l'inverse est vrai pour les voyelles (500, 1000) et (500, 1250) par rapport à la paire (500, 1500) et (500, 1750). Une expérience portant sur plus de 2000 paires de voyelles synthétiques à 3 formants indique que l'espace vocalique doit être divisé en 2 zones de tailles inégales pour respecter le critère de décomposabilité. La plus petite contient les voyelles ([a, ɔ]), tandis que la plus grande comprend les autres voyelles. Il est probable, en outre, qu'aucun jugement de similitude sur des voyelles qui diffèrent de plus de 200 ou 300 Hz sur F1 n'est compatible avec l'hypothèse de décomposabilité.

Après ces remarques théoriques, il convient de présenter en détail la technique d'analyse multidimensionnelle de CARROLL et CHANG (1970).

III - LA METHODE INDSCAL DE CARROLL ET CHANG (1970)

A/ Méthode et résultats

La référence principale est l'article de CARROLL et CHANG (1970). On peut également consulter CARROLL (1972). Une courte introduction, peu pédagogique, est parue en français (CAILLIEZ et PAGES 1976, p. 493-494). Pour une description technique de l'algorithme et un listing du programme Fortran, on consultera LONCHAMP (1974). De nombreux programmes

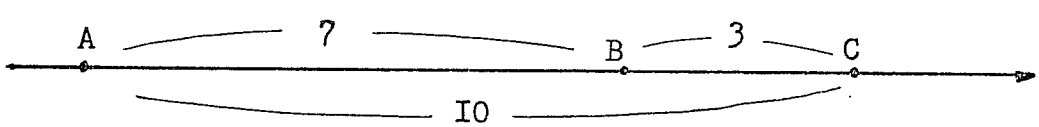
auxiliaires ont été mis au point au Laboratoire de Phonétique de l'Université de Nancy II par l'auteur. Ceux-ci effectuent en particulier le calcul des constantes additives, le calcul des distances moyennes, l'affichage des résultats et le calcul du coefficient de corrélation entre les données initiales et les résultats.

La méthode INDSCAL (= INDividual SCALing ; "scaling" individuel), comme la plupart des méthodes d'analyse multidimensionnelle, a pour but de décomposer un tableau de distances inter-stimuli en un nuage de points dans un espace à m dimensions orthogonales. Chaque stimulus est défini comme un point de cet espace. La distance euclidienne entre deux stimuli est une fonction monotoniquement décroissante de la similitude subjective. Deux exemples très simples permettent d'imaginer le mécanisme, sans entrer dans les détails techniques.

	A	B	C
A	0	7	10
B	7	0	3
C	10	3	0

Par hypothèse, ce tableau est symétrique, c'est-à-dire que la distance entre les stimuli A et B est égale à la distance entre les stimuli B et A. Par définition, la distance entre A et A est nulle. On vérifie que l'on a $n(n-1)/2$ distances (ici $n = 3$; $n(n-1)/2 = 3$) différentes non-nulles.

Ce tableau est en réalité le cas particulier où les distances entre les stimuli impliquent que ces 3 stimuli se placent sur une droite (dimension).



Pour le tableau suivant, au contraire, les points ne peuvent se trouver sur une seule dimension. Dans cet exemple théorique, les distances satisfont au théorème de Pythagore ($a^2 + b^2 = c^2$) pour un triangle rectangle (fig. 2 - III).

	A	B
A	3	4
B	0	5

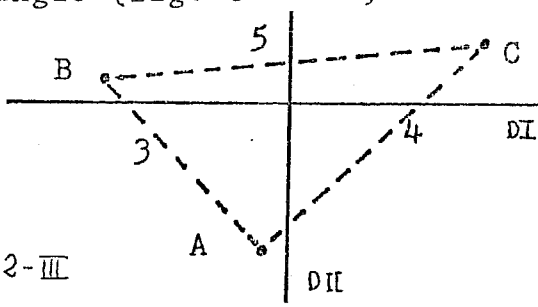


Fig. 2-III

On voit intuitivement comment les distances inter-stimuli permettent d'estimer le nombre de dimensions nécessaires pour placer les points dans un espace, de telle manière que les distances dans l'espace approchent au mieux les indices de similarité fourni s par les sujets. Pour un plus grand nombre de points, il faut une méthode mathématique complexe.

Les données initiales (distances dérivées des indices de similarité comme nous le verrons plus loin) sont contenues dans un tableau de $n(n-1)/2$ cases, où n est le nombre de voyelles. La méthode INDSCAL est, à strictement parler, "métrique", car elle exige comme données initiales, non pas des données issues d'échelles ordinales (où les seules conditions sont : A, B et C étant des distances, $A = B \Rightarrow B = A$; $A < B$, $B < C \Rightarrow A < C$), mais des données issues d'échelles d'intervalles (ou à intervalles équirépartis), où il est légitime de comparer les différences numériques entre les paires ; c'est-à-dire où la phrase "la distance x est deux fois plus grande que la distance y" a un sens. (sur ces types d'échelles, cf. REUCHLIN (1976)). Mais CARROLL et CHANG (1970) indiquent que cette restriction n'est pas fondamentale dans la pratique.

Si la violation des postulats métriques n'est pas trop forte, le résultat de l'analyse ne sera pas affecté.

INDSCAL possède une particularité unique qui assure sa supériorité sur les autres techniques d'analyse multidimensionnelle. L'orientation des axes (dimensions) est unique

et ne nécessite aucune rotation à postériori des résultats. Il faut cependant préciser, et il est surprenant que cette information capitale ne se trouve qu'en note dans un exemple d'application de la méthode (WISH, DEUTSCH et BIENER 1972), que cette orientation unique n'est assurée que lorsque les "poids" de chaque dimension (cf. infra) sont suffisamment différents pour chaque sujet. L'analyse d'un sujet unique, telle que nous la pratiquerons dans certain cas, ne garantit pas la bonne orientation des axes et rend la méthode approximativement équivalente à la variante métrique de la méthode de KRUSKAL (1964a, 1964b).

Les résultats se présentent sous la forme

a) d'un tableau des coordonnées des points (voyelles) sur les dimensions. Ces coordonnées sont "normalisées" de telle manière que la moyenne soit égale à zéro, et que la variance soit égale à l'unité. Ces points se placent dans l'espace perceptif du groupe, c'est-à-dire sur les dimensions communes à l'ensemble du groupe.

b) d'un tableau des "poids" des dimensions pour chacun des sujets.

A chaque sujet correspond un espace spécifique, ce qui justifie le nom de "scaling" individuel. Cet espace spécifique dérive de l'espace normalisé par une contraction ou une expansion des différentes dimensions. La méthode ne permet pas les rotations individuelles des configurations. La position des points (voyelles) dans l'espace est donnée par les coordonnées pondérées ; ces nouvelles coordonnées s'obtiennent en multipliant les coordonnées normalisées par la racine carrée des "poids" pour chaque dimension. Ces poids indiquent l'importance de la dimension dans l'espace spécifique du sujet. Nous empruntons la fig. III-1 à BEECKMANS (1974, p. 116).

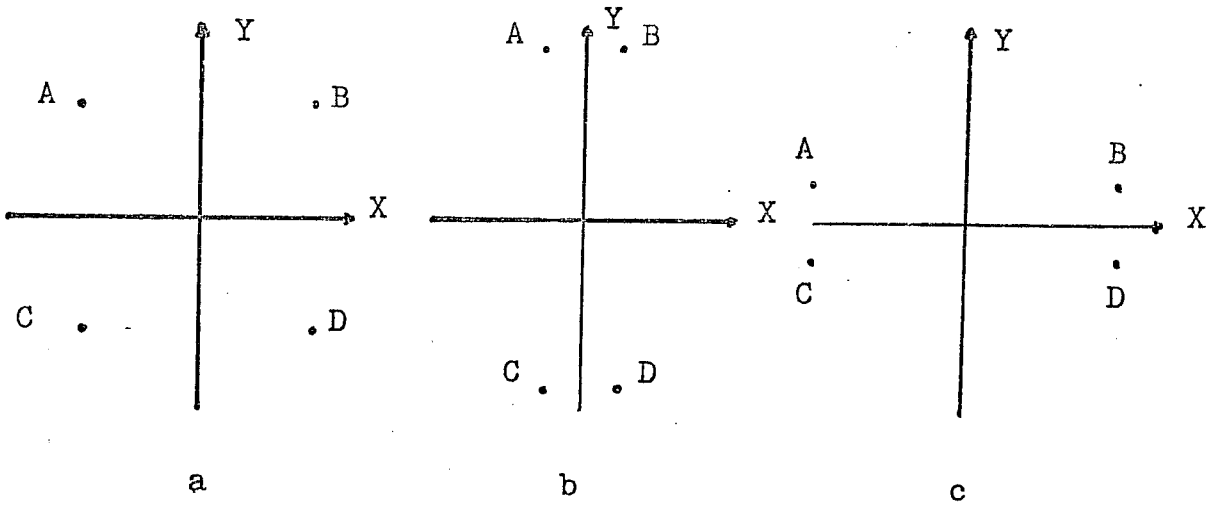


fig. 3 - III

La figure a présente l'espace perceptif normalisé commun à 2 sujets. Les deux autres figures (b et c) présentent l'espace spécifique des deux sujets qui n'accordent pas la même importance à chaque dimension. Les résultats du sujet de la figure b indiquent que celui-ci privilégie le paramètre physique du stimulus représenté par la dimension y ; à l'inverse, l'autre sujet (figure c) accorde une bien plus grande importance au paramètre physique apparaissant dans la dimension x.

B/ Interprétation des résultats

a) Critères internes

Le point essentiel est de savoir si la configuration des points de l'espace à m dimensions reflète bien les distances initiales. Un critère couramment utilisé pour juger de cette concordance est la "quantité de variance expliquée" ("variance accounted for"). Si l'étude analytique de cette notion est trop complexe pour être entreprise ici, on peut considérer, en simplifiant, que les distances initiales sont réparties de façon normale (moyenne x, écart type s). On comparera ce dernier paramètre à celui issu de la distribution des distances calculées; techniquement, il s'agit de la

corrélation entre les produits scalaires des distances initiales et des distances calculées. Cette comparaison n'est pas faite explicitement, mais est un sous-produit de la méthode matricielle utilisée. On montre d'autre part que cette quantité n'est autre que la somme des distances euclidiennes entre la position des poids et l'origine des axes (CARROLL et CHANG 1970, p. 316). Une restriction importante est que le calcul du pourcentage de variance expliquée suppose que les distributions des points sur les dimensions sont non-corrélées. Dans le cas contraire, on obtient une sous-estimation du taux de variance expliquée. Cette "marge de sécurité" explique l'utilisation de ce critère, bien que l'hypothèse de non-corrélation ne soit pas toujours vérifiée pour un nombre de points de l'ordre de la dizaine.

Il semble d'autre part que, pour des données de distances aléatoires fournies par des expériences de simulation (méthode de Monte-Carlo), le pourcentage de variance expliquée pour m dimensions soit égal approximativement à $m/(n - 1) \times 100$ % (n : nombre de points ; m : nombre de dimensions).

Pour la simulation sur 25 points effectuée par CARROLL et CHANG on obtient les résultats suivants :

nombre de dimensions	% variance expliquée par simulation	% variance calculée par $m/(n - 1) \times 100$
1	4.864	4.2
2	9.653	8.3
3	13.302	12.5
4	16.471	16.7
5	19.232	20.8

L'accord est très satisfaisant. Une étude de HOWARD (1977) pour 8 stimuli donne les résultats suivants :

nombre de dimensions	% variance expliquée (simulation)	% variance calculée
1	15	14.3
2	28	28.6
3	41	42.9

On sait par ailleurs qu'il est toujours possible de placer n points dans un espace à n - 1 dimensions de manière à rendre compte exactement des distances. On vérifie que, dans ce cas, le pourcentage de variance expliquée est bien égal à 100 % : (n - 1) x m / (n - 1) x 100 = 100.

En théorie donc, une dimension n'est significative que si elle explique plus de 1/(n - 1) x 100 % de la variance. Malheureusement, en cas de corrélation entre les positions des points sur deux dimensions, il est théoriquement possible qu'une dimension soit significative alors qu'apparemment elle explique moins de variance que le critère défini plus haut. C'est pourquoi nous avons utilisé dans cette étude, à titre de vérification supplémentaire, le calcul direct du coefficient de corrélation entre les données initiales et les résultats de l'analyse.

D'autres critères sont également disponibles pour juger de l'acceptabilité d'une dimension :

- Les différentes analyses multidimensionnelles que nous avons réalisées sur des voyelles ou des contours intonatifs nous ont montré que l'obtention de poids négatifs signalait toujours l'apparition d'une dimension artefactuelle. Nous appelons dimension artefactuelle une dimension sur laquelle l'ordre des voyelles est déterminé uniquement par les variations aléatoires dues au "bruit" (variance résiduelle). L'obtention de poids négatifs indique une violation grave des axiomes métriques des données initiales.

- Dans ce cas, on observe toujours une ou plusieurs des caractéristiques suivantes :

  - L'utilisation d'une configuration de départ différente pour cette dimension ordonne différemment les points. Cet effet s'observe également, comme les suivants, dans des cas où les poids ne sont pas négatifs. Une méthode commode est d'utiliser un nouvel ordre, dont le coefficient de corrélation avec le précédent soit nul. On vérifie que le programme est capable d'amener les points dans l'ordre observé précédemment. L'échec signale une dimension artefactuelle.

∘ Un pourcentage de variance expliquée inférieur ou égal au minimum théorique d'accroissement (1/(n - 1) x 100%). Ce critère, pour les raisons exposées plus haut, est moins décisif que le précédent. Par contre, la diminution ou la stagnation du pourcentage rend la dimension très "suspecte". Dans ce cas, la diminution est le signe d'une forte corrélation entre deux dimensions.

∘ Le dernier critère, d'un caractère beaucoup plus subjectif, est l'interprétabilité de la dimension. Celle-ci dépend de l'ingéniosité du chercheur.

En cas de doute, il est recommandé de calculer directement le coefficient de corrélation entre les distances initiales et les distances dans l'espace pour les configurations à m et $m + 1$ dimensions. L'intégration d'une nouvelle dimension doit augmenter la différence entre les deux coefficients $(\sqrt{r_{m+1}} - \sqrt{r_m})$ d'une quantité au moins égale à $1/(n - 1)$, où n est le nombre de points.

Ces conclusions, fondées en partie sur l'expérience acquise, ne correspondent pas à celles que tire TERBEECK (1977). Mais la différence est peut-être à mettre au compte des différences entre les deux modèles utilisés

b) Critères externes

Sur un plan plus pratique, l'interprétation des dimensions est fonction de l'indépendance des caractéristiques physiques qui décrivent les stimuli.

Considérons un espace à deux dimensions décrivant l'espace perceptif des stimuli composés de 3 paramètres physiques variables. Sans entrer dans l'ensemble des détails techniques, plusieurs cas doivent être envisagés :

- dans le cas le plus favorable, chacune des dimensions perceptives est fortement corrélée avec une dimension physique et une seule. Ceci revient à dire qu'une dimension physique n'est pas pertinente du point de vue perceptif. L'interprétation des dimensions se fait sans ambiguïté !

- dans tous les autres cas, une interprétation stricte des dimensions est impossible. Supposons une forte corrélation entre l'une des dimensions et deux paramètres physiques. En anticipant quelque peu sur les résultats, considérons l'exemple théorique suivant : soit une dimension sur laquelle les voyelles se répartissent de la manière suivante :

[i e ɛ y ø œ a ɔ o u]

Ces voyelles possèdent les fréquences formantiques F2 et F3 présentées ci-dessous.

Voyelle	F2	F3
i	2300	2900
y	1850	2300
e	2000	2750
ø	1600	2300
ɛ	1800	2600
œ	1500	2400
a	1300	2500
ɔ	1000	2500
o	850	2350
u	900	2150

Le coefficient de corrélation de rang entre F2 et la dimension est de 0,975. Il est encore de 0.897 entre F3 et la dimension perceptive. L'application de la formule F2 + 1/3 F3 (choisie arbitrairement) conduit à une corrélation de rang parfaite (r = 1.00). Le problème est qu'il existe un très grand nombre de combinaisons linéaires de F2 et F3, qui donnent une corrélation de rang plus élevée que F2 ou F3 seuls. Dans ce cas, les données fournies par l'analyse multidimensionnelle ne sont pas suffisantes pour interpréter complètement le rapport entre une dimension perceptive et les paramètres physiques. Cet exemple théorique est représentatif des résultats empiriques obtenus. L'interprétation des rapports entre dimensions perceptives et paramètres physiques ne pourra être faite que par d'autres types d'expériences (cf. chapitre IV).

Dans l'analyse détaillée des résultats, nous nous limiterons aux remarques suivantes :

- si une dimension perceptive présente les voyelles dans un ordre (relativement) semblable à leur étagement sur F1, nous dirons que cette dimension corrèle avec F1. POLS et *al.* (1969) ont en effet montré, sur un large échantillon, l'indépendance des valeurs de F1 et F2 (r = - 0,03).

- si les coordonnées des voyelles sur une deuxième dimension sont conformes aux valeurs de F2, nous dirons que cette dimension corrèle avec F2 ou F'2. La notion de F'2 ou "deuxième formant équivalent" a été développée par FANT (1959) pour rendre compte des expériences de perception de voyelles synthétiques à 2 formants. F'2 est une combinaison linéaire de F2 et F3 (et éventuellement F4). POLS et *al.* (1969) ont obtenu un coefficient de corrélation de 0,43 entre les valeurs de F2 et F3, ce qui indique que la dimension perceptive peut être prédite par F2, F3 ou une combinaison linéaire de F2 et F3. L'étude de F'2 fera l'objet d'une étude détaillée au chapitre IV.

C/ Le pourcentage de variance expliquée

Il convient de préciser les raisons pour lesquelles cette quantité peut être faible.

- Les données fournies par les sujets peuvent être de "mauvaise qualité". C'est en particulier le cas lorsque les réponses pour un même couple de stimuli présentent une variabilité non-négligeable. Ceci indique souvent que la tâche proposée au sujet est (trop) difficile. On fait généralement l'hypothèse que cette variabilité est aléatoire ; la moyenne est alors un meilleur estimateur que les données individuelles. On parle alors métaphoriquement de "bruit". Cette variabilité introduit un pourcentage important de variance résiduelle. Aucune dimension supplémentaire ne rend compte dans ce cas de plus de 1/(n - 1) x 100 % de la variance. Un "bruit" important ne signifie nullement que les dimensions extraites soient elles-mêmes entachées de bruit. Par contre un "bruit" important

peut rendre impossible l'extraction d'une dimension significative si celle-ci ne rend compte que d'un taux de variance légèrement supérieur au minimum théorique. Si le "bruit" est effectivement responsable du faible taux de variance expliquée, l'analyse des moyennes de groupe doit provoquer une élévation importante de ce taux pour un même nombre de dimensions, car le moyennage (lissage) tend à réduire le "bruit" dans une très large proportion. Notre expérience en fournit un bon exemple.

- Un petit nombre de sujets peuvent utiliser des dimensions spécifiques. Un sujet jugeant la durée des voyelles dans une expérience portant sur le timbre serait un exemple extrême. C'est pourquoi il est important de contrôler les paramètres non-pertinents des stimuli. Cet artefact majeur est difficilement détectable.

- Une dimension ne peut être pertinente pour un très petit nombre de stimuli. Ceci pourrait être le cas si l'on incluait une voyelle nasale dans un ensemble de voyelles orales. Le corpus doit être le plus homogène possible.

On note quelquefois des variations importantes de la quantité de variance expliquée selon les sujets. Le travail de MILLER et CARTERETTE (1975) indique en particulier que certains sujets obtiennent un taux important de variance expliquée en réduisant le nombre de dimensions qu'ils utilisent pour juger des similarités. Pour les voyelles, ceci consisterait à répondre par exemple "identique" pour les couples de voyelles antérieures, ou postérieures, et "différent" pour les couples de voyelles s'opposant par cette caractéristique. Une seule dimension rend alors parfaitement compte du comportement du sujet, en raison de la simplification apportée à la tâche. Un "bon" sujet n'est pas toujours celui qui obtient le taux de variance expliquée le plus élevé, mais celui qui obtient un taux élevé avec des poids importants pour chaque dimension. Un tel comportement révèle la difficulté de la tâche, et n'est sans doute pas le signe d'une volonté consciente de simplification.

D/ Méthodes d'estimation des similarités

Trois méthodes sont couramment utilisées pour obtenir des estimations de (dis)similarité. La qualité des résultats dépend dans une large mesure de la méthode retenue.

1 - Estimation à partir d'une matrice de confusions

Cette méthode a été employée par différentes équipes (BEECKMANS, VAN DER KAMP ET POLS, etc.). Deux stimuli sont d'autant plus proches ou semblables qu'ils sont plus souvent confondus dans une tâche d'identification. Le meilleur argument en faveur de cette méthode est la simplicité et le caractère naturel de la tâche que le sujet doit effectuer. Mais elle soulève des problèmes méthodologiques importants.

Alors que, par définition, la valeur d'un indice ne dépend pas de l'ordre des stimuli, les matrices de confusions sont asymétriques. La probabilité que la voyelle i soit confondue avec la voyelle j, (Pi/j), n'est pas identique à la probabilité que la voyelle j soit confondue avec la voyelle i , (Pj/i). L'exemple suivant est tiré de van der Kamp et Pols (1971 : dans leur expérience, la probabilité que [oe] soit confondu avec [e] est de 9 %, alors que la probabilité inverse est de 24,5 %. Comme l'indique CARTERETTE dans sa discussion d'un exposé de POLS (1970) "... une des raisons pour lesquelles je n'ai jamais publié d'analyses multidimensionnelles à partir des matrices de confusions est que je n'ai jamais pu résoudre le problème de la symétrie" (p. 472). Diverses méthodes de symétrisation ont été envisagées (VAN DER KAMP et POLS 1971). La première, qui consiste à calculer deux configurations différentes basées sur les réponses et les stimuli, est de l'aveu même des auteurs ininterprétable (KLEIN, PLOMP et POLS 1970). Plusieurs méthodes de calcul de la moyenne (arithmétique, géométrique, avec ou sans pondération) sont disponibles (SHEPARD 1957, 1972).

$$Sij = \left[\frac{P_{i/j} \cdot P_{j/i}}{P_{i/i} \cdot P_{j/j}}\right] \qquad Sij = \left[\frac{P_{i/j} + P_{j/i}}{P_{i/i} + P_{j/j}}\right]$$

On peut également éliminer le "biais de réponse" (WAGENAAR 1968). Comme JANSON (1977), nous ne sommes pas persuadés que le problème doive être abordé sous l'angle du biais de réponse. L'analyse des matrices de confusions montre que l'asymétrie n'est pas une caractéristique personnelle du sujet (ou de la paire, si l'on change les stimuli physiques). L'asymétrie n'est pas valable pour l'ensemble de la matrice de confusions. Pour une langue comme le néerlandais, qui associe des différences de timbre et de durée ([ø:-œ ; e:-ɛ]) pour distinguer des paires de voyelles physiquement proches, on observera un véritable "biais de réponse" si l'on présente des voyelles longues abrégées, ou des voyelles brèves allongées artificiellement. Mais dans la plupart des cas, l'asymétrie a des causes différentes. Si l'échantillon choisi est, par exemple, légèrement trop ouvert, la probabilité de confusion avec la voyelle d'aperture immédiatement supérieure sera largement augmentée. Par contre, la probabilité inverse ne sera pas modifiée. JANSON (1977) présente une analyse détaillée de la matrice de confusions de KLEIN et al. (1970) qui démontre que la majeure partie de l'asymétrie est due à une déviation physique de certains stimuli par rapport à la norme. L'estimation non-biaisée des indices à partir des matrices de confusions ne nous semble possible qu'à partir des résultats de l'identification d'un très grand nombre d'échantillons de chaque voyelle (cf. PETERSON et BARNEY 1952). Mais se pose alors le problème des cases vides.

Le deuxième problème méthodologique est qu'un trop grand nombre de cases vides dans la matrice de confusions conduit à une estimation identique des distances pour un grand nombre de couples. Si la probabilité que [a] soit perçu comme [o] ou [u] est nulle (absence de confusions), les distances entre [a - o] et [a - u] seront considérées comme identiques. Or il est probable que le nombre plus important de confusions entre [o] et [u] qu'entre [ɔ] et [u] conduira à une information contradictoire avec la précédente : les voyelles [o] et [u] sont à une distance non-négligeable l'une de l'autre. Ceci contribuera à accroître le "bruit" des

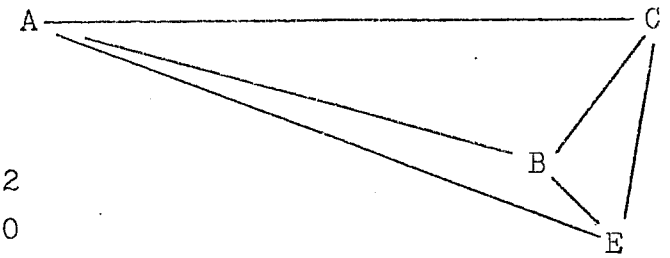
estimations. Pour éviter ce problème, certains auteurs (par ex. VAN DER KAMP et POLS 1971) sont amenés à dégrader la qualité des stimuli en réduisant leur durée à 1 période. Mais on ne peut plus parler de tâche naturelle. L'information spectrale contenue dans une période unique n'est pas forcément représentative de la voyelle entière, en raison des variations à court-terme de la source vocale. Il n'est pas non plus certain que l'oreille soit capable d'extraire les informations spectrales avec la même précision sur des durées aussi brèves.

L'utilisation des matrices de confusions ne nous semble envisageable que dans le cas où elles sont le résultat de l'identification d'un grand nombre d'échantillons par un nombre important de sujets (cf. PETERSON et BARNEY 1952). La mise en oeuvre est alors très lourde, et il n'est plus possible de spécifier exactement les stimuli employés.

2 - La méthode triadique

Le sujet doit indiquer, pour toutes les triades de stimuli, le couple le plus semblable et le couple le plus dissemblable, sans indiquer la grandeur de ces différences. Le couple le plus semblable est crédité de 2 points, le couple le plus différent de 0 point et le troisième couple de 1 point. La matrice de similarité est constituée par l'addition des résultats partiels. Le score pour chaque paire est donc proportionnel au nombre de fois que cette paire a été jugée plus semblable que les autres paires. Cette méthode suppose pratiquement un dispositif de présentation et de réponse automatique couplé à un ordinateur, afin de permettre l'écoute "à la demande" de chaque élément de la triade. Il est nécessaire de minimiser l'effet d'ordre et la charge mémorielle. Un tel système est décrit dans POLS et al. (1969).

On peut montrer que cette méthode conduit à une sous-estimation systématique des distances les plus grandes. Il s'agit là d'un artefact majeur dont il est difficile de prévoir l'effet dans une analyse multidimensionnelle. L'application de la méthode triadique conduit aux indices de similarité suivants pour la figure ci-dessous:



A - B: 2
A - C: 0
A - E : I
B - C : 3
B - E : 4
C - E : 2

On voit que la distance AB, d'après la méthode triadique, est identique à la distance EC. On trouvera dans TERBEECK (1977) une discussion plus complète. En conclusion, l'utilisation de la méthode triadique provoquera une certaine distorsion de l'espace perceptif calculé.

3 - Echelle d'intervalle ("Rating scale") et estimation numérique directe ("free magnitude estimation")

Nous regroupons dans ce paragraphe deux techniques voisines, utilisées par MOHR et WANG (1968), SINGH et WOODS (1971) et nous-mêmes (CARTON et LONCHAMP 1977 par ex.) pour la première, et HANSON (1967) pour la seconde.

Le sujet doit indiquer, pour une échelle d'intervalles, le degré de (dis)similarité en cochant l'une des 7 ou 9 cases d'une ligne de réponses. On transforme l'indice obtenu pour chaque couple en une estimation numérique en affectant chaque case d'une valeur variant généralement de 1 à 7 (ou 9). La tâche est beaucoup plus difficile que celle requise par les méthodes présentées précédemment. Les réponses ne se font plus sur une échelle ordinale, mais sur une échelle d'intervalles, où les notions de "double" et de "moitié" ont un sens. Le problème de la symétrie ne se pose pas, car on peut faire la moyenne des estimations pour les couples i, j et j,i.

Le principal reproche (POLS et al. 1969, p. 460) est la variabilité et le "bruit" important des réponses, à la

fois pour les paires et entre les sujets. Ce reproche est certainement vérifié pour les données de HANSON (1967) et celles que nous présenterons. Il semble pourtant que l'utilisation des moyennes de groupe donne de bons résultats. Cette technique est facile à mettre en oeuvre sur le plan matériel. Mais il se produit sans doute un artefact non-négligeable qui n'a jamais été mentionné à notre connaissance : un effet de "plafonnement" ("ceiling" effect) biaise les estimations des distances les plus grandes. La deuxième critique de Pols et al. (1969), concernant la transformation en produits scalaires, pourrait avoir un lien avec cet artefact.

La méthode d'estimation numérique directe consiste à proposer un chiffre arbitraire pour l'une des paires, et à demander au sujet d'indiquer un chiffre proportionnel à la différence perçue pour les autres paires. HANSON (1967) a montré que ces deux méthodes donnaient des résultats équivalents.

Il n'existe donc pas de techniques parfaites. SINGH, WOOD et BECKER (1972) ont comparé 3 méthodes (échelle d'intervalles sur 7 points, estimation numérique directe, méthode ABX (triadique)) lors d'une étude portant sur des consonnes. Aucune d'elles ne s'est révélée supérieure. Toutes les méthodes d'estimation se heurtent à des difficultés. Celles-ci devront toujours être présentes à l'esprit dans l'interprétation des résultats. Les difficultés de l'analyse multidimensionnelle résident, pour nous, plus dans les méthodes d'estimation des dissimilarités, que dans le modèle mathématique sous-jacent.

E/ Transformations des indices en distances

Le dernier point de notre étude méthodologique concerne un problème qui est souvent négligé : la transformation des indices de (dis)similarité en "distances". Le terme "distance" a ici son sens mathématique. Dans nos analyses précédentes (par ex. CARTON et LONCHAMP 1977) nous avions implicitement considéré que les indices vérifiaient les

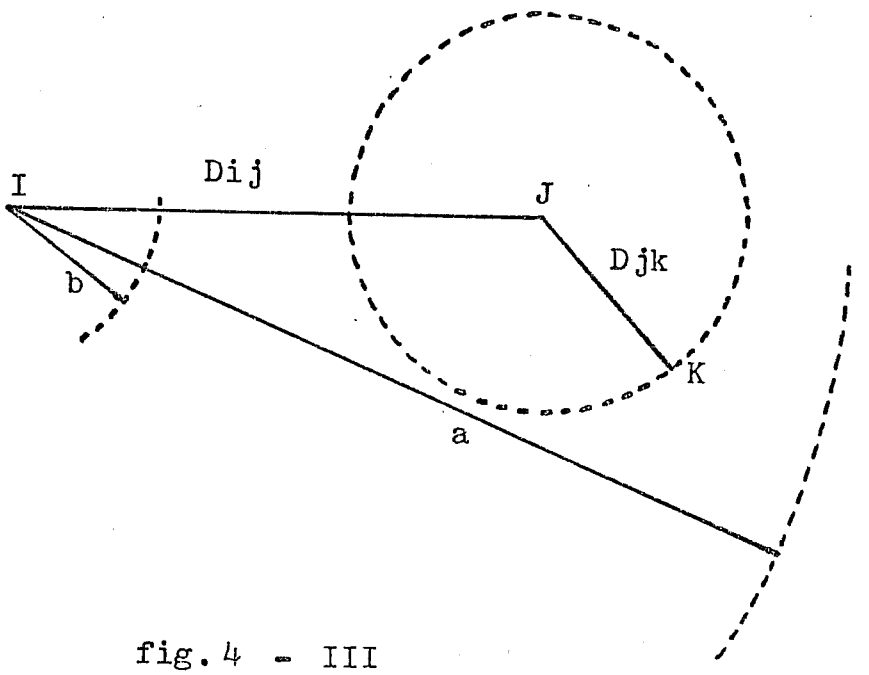


fig. 4 - III

propriétés mathématiques des distances. Les données initiales du programme INDSCAL doivent, en effet, être des distances. Or ceci n'est pas forcément vrai.

Par définition, nos indices de dissimilarité (d) possèdent les propriétés suivantes:

a) $d(i,j) = d(j,i)$: l'indice de dissimilarité de i par rapport à j est égal à celui de j par rapport à i.

Dans notre programme, cette égalité est obtenue par le calcul de la moyenne

$$d(i,j) = d(j,i) = (d(i,j) + d(j,i))/2$$

b) $d(i,i) = d(j,j)$. L'indice de dissimilarité de i par rapport à lui-même est inférieur à celui de tous les autres couples où $i \neq j$.

Dans le programme, cette inégalité est satisfaite en posant $d(i,i) = 0$.

L'analyse multidimensionnelle exige que l'on transforme les indices d en distances (D). Les distances ont deux propriétés supplémentaires :

c) $D(i,i) = D(j,j,) = 0$. Cette égalité est satisfaite.

d) $D(i,j) < D(i,k) + D(k,j)$: "inégalité triangulaire"

Le schéma suivant fera saisir le sens de l'inégalité triangulaire. Soient i et j 2 points d'un espace à deux dimensions à une distance $D(i,j)$. Soit k un troisième point à une distance $D(j,k)$ de j. k ne peut se trouver que sur un cercle de rayon $D(j,k)$ centré sur j. Dans un espace réel, ceci implique que la distance $D(i,k)$ est au plus égale à $D(i,j) + D(j,k)$, et au moins égale à $D(i,j) - D(j,k)$. Soient a et b deux distances inférieures et supérieures respectivement à ces limites. Il n'existe pas de point k à une distance a ou b de i qui satisfasse l'inégalité triangulaire entre $D(i,j)$, $D(j,k)$ et $D(i,k)$ (fig. 4-III)

Nos indices étant de grandeur arbitraire, il est permis de les augmenter d'une quantité x ("constante additive") arbitraire. On choisit une valeur x telle que l'inégalité triangulaire soit satisfaite. Une méthode simple (approximative) est donnée par TORGESON (1958, p. 268-277). On peut également consulter CAILLIEZ et PAGES (1976, § 722.1). Les indices sont alors transformés en distances. On observera dans nos résultats l'amélioration de la quantité de variance expliquée résultant de cette transormation.

Cette présentation des méthodes et des problèmes de l'analyse multidimensionnelle va nous permettre une évaluation critique de l'ensemble des analyses publiées jusqu'en 1977. Puis nous présenterons une contribution originale qui tient compte de l'ensemble des résultats et critiques ainsi dégagés.

IV - REVUE CRITIQUE ET ANALYSE DES TRAVAUX ANTERIEURS PAR LA METHODE INDSCAL

A/ Peters (1963)

La première analyse multidimensionnelle a été réalisée par PETERS (1963) à l'aide du modèle métrique euclidien mis au point par TORGESON (1958, p. 247-297). C'est un des premiers exemples de l'utilisation d'un ordinateur dans ce domaine. Les mesures de similarité ont été obtenues par une échelle d'intervalles à 9 positions. Cette étude est unique en ce qu'elle porte sur des différences "intraphonologiques". 100 sujets devaient juger des similarités entre 496 paires de voyelles [æ] produites par 32 locuteurs différents (F_0 = 130 Hz, durée : 1,5 s ; intensité sonore égalisée). L'analyse perceptive permet, selon l'auteur, d'extraire 5 facteurs, dont l'interprétation est recherchée par le calcul de coefficients de corrélation avec 10 mesures physiques. L'analyse physique fait malheureusement preuve d'une grande naïveté et d'une erreur méthodologique de première magnitude. Les 10 facteurs acoustiques sont :

- Les "centroïdes" de F1, F2 et F3. Aucune méthode de calcul n'est donnée. D'après les chiffres, il pourrait

s'agir d'une moyenne pondérée des harmoniques les plus intenses. Pour F1 et F2, les valeurs tombent dans 3 groupes distincts ; pour F1, une valeur à 510 Hz, 22 valeurs entre 622 et 658 Hz, et 9 valeurs entre 750 et 770 Hz. L'utilisation du coefficient de corrélation est une erreur importante car la distribution n'est absolument pas gaussienne.

- La différence F1 - F2.

- L'intensité de F1, F2 et F3. Pour F1, la différence ne dépasse pas 4 dB. Le calcul d'un coefficient de corrélation à partir de ces valeurs n'est pas possible.

- La différence d'intensité entre F1 et F2.

- "L'écart en dB entre l'intensité du centroïde et la fréquence inférieure la moins intense du formant" (op. cit. p. 243). Ceci est pour nous incompréhensible.

L'erreur méthodologique est que PETERS néglige de vérifier l'indépendance de ces différentes mesures physiques. On sait par exemple que l'intensité d'un formant dépend dans une large mesure de sa fréquence.

L'interprétation des facteurs semble la suivante, si l'on ne retient que les fréquences formantiques :

facteur I : fréquence de F1
facteur II : fréquence de F3 (?)
facteur III : ininterprétable
facteur IV : fréquence de F2.
facteur V : corrèle avec le facteur II (r = 0.617).

La seule conclusion est l'importance des fréquences formantiques pour la perception dans une même zone phonologique. Ce type de travail devrait être repris sur des bases méthodologiques plus saines.

B/ MOHR et WANG (1968)

Le travail de MOHR et WANG (1968) n'est pas à proprement parler une analyse multidimensionnelle. Mais les résultats présentés peuvent être analysés par la méthode

INDSCAL. Ils présentent un tableau (table VII, p. 38, op. cit.) d'indices de similarité entre 9 voyelles ([i, e, æ, y, ø, œ, u, o, ɔ]) perçues par 80 sujets de langue maternelle anglaise, connaissant une langue étrangère à un degré variable et possédant quelques notions de phonétique. Une échelle de similarité à 6 positions a été utilisée. Ce tableau est intéressant dans la mesure où il teste des voyelles qui n'existent pas dans l'inventaire phonologique de l'anglais. Une autre partie de l'étude concerne les voyelles nasales et les consonnes. L'analyse de MOHR et WANG porte uniquement sur des comparaisons entre certains groupes de voyelles définis par les traits distinctifs de CHOMSKY et HALLE (1968). Parmi les constatations les plus importantes, on peut citer : les membres d'une catégorie marquée (au sens phonologique du terme) sont en moyenne plus proches entre eux que les membres d'une catégorie non-marquée. Pour les traits fermé/ouvert, aperture moyenne ("mid")/aperture extrême, palatal/non-palatal, les distances moyennes entre les paires pour lesquelles les deux membres ont en commun le même trait sont identiques aux distances entre les paires ayant un trait différent. On ne peut donc parler de marque pour ces trois traits. Par contre, la distance moyenne entre les voyelles nasales est beaucoup plus faible que la distance moyenne entre les voyelles non-nasales. [+ nasal] est donc un trait marqué pour cette population. Le trait labial n'est marqué que pour les voyelles antérieures (palatales) [i, e, æ]

Mais ces remarques n'épuisent pas la matrice des similarités. Nous avons procédé à une analyse multidimensionnelle des données par la méthode INDSCAL en considérant les moyennes de groupe (seules fournies) comme provenant d'un sujet unique "moyen". Les coordonnées des voyelles dans l'espace à 3 dimensions (non-pondérées par les poids) sont présentées dans le tableau suivant.

	DI	DII	DIII
i	-.93	1.02	-1.36
e	-.92	.34	-1.41
æ	-.41	-1.37	-1.30
y	-.55	1.07	.79

	DI	DII	DIII
ø	-.71	.12	.58
œ	-.65	-.85	.17
u	1.17	1.36	1.15
o	1.61	-.23	1.03
ɔ	1.40	-1.48	.35

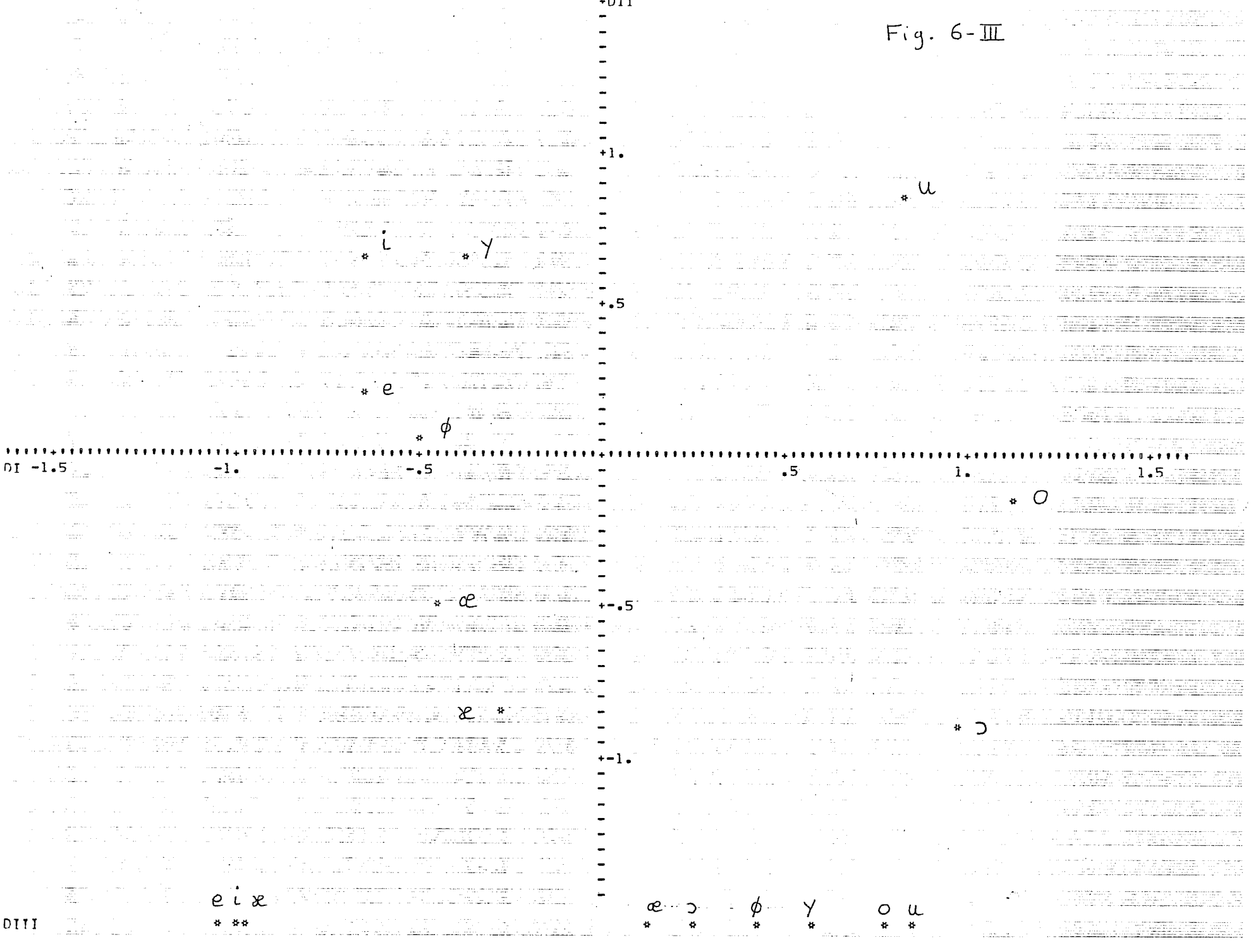
Le tableau des poids est le suivant :

DI	DII	DIII	% variance expl.
0.49	0.38	0.55	68.7 %

La quantité de variance s'établit donc à 69 % pour trois dimensions. Les indices bruts ont été considérés comme des distances. Les voyelles sont présentées graphiquement à la fig. 5-III

L'interprétation des résultats ne pose pas de problème majeur. La dimension 1 reflète une distinction antérieure-postérieure. La corrélation avec F2 ou F'2 est bonne. Les voyelles antérieures arrondies [y, ø, œ] sont très proches des voyelles antérieures non-arrondies. [œ] s'intercale entre [e] et [æ] à la place de [ɛ]. La deuxième dimension correspond sans doute à F1. Quant à la troisième dimension, elle sépare parfaitement les voyelles antérieures non-arrondies des autres voyelles arrondies. Ceci revient à écarter considérablement les voyelles arrondies et non-arrondies antérieures sans augmenter la distance entre les voyelles postérieures et les voyelles antérieures arrondies. Cette solution n'est pas conforme au résultat précédent, qui considérait le trait labial comme marqué pour les voyelles antérieures. Rappelons que les sujets ne possèdent pas [y, ø, œ] dans l'inventaire de leur langue maternelle. On peut se demander également dans quelle mesure les connaissances de phonétique ne les ont pas incités à utiliser la représentation du triangle articulatoire. TERBEECK (1977) a analysé les mêmes données à l'aide d'une méthode voisine. L'utilisation des moyennes de groupe exclut que l'on puisse fixer la position des axes dans l'espace. Les rotations sont donc possibles. Sa figure 2-12 présente les voyelles antérieures arrondies comme nettement centrales. Par contre sa troisième dimension sépare

Fig. 6-III



les voyelles d'avant arrondies des 6 autres voyelles. Cette différence semble indiquer que l'espace perceptif ne contient en fait que deux dimensions. L'introduction d'une troisième dimension permet plusieurs positions des groupes de voyelles :

- notre solution où les voyelles [y, ø, œ] et [i, e, æ] sont proches sur les deux premières dimensions, et séparées par la troisième dimension.
- la solution de TERBEECK où les voyelles sont équidistantes des voyelles antérieures et postérieures sur les deux premières dimensions et nettement opposées à celles-ci sur la troisième.

Ces solutions semblent bien équivalentes en termes de distance .

La comparaison des deux analyses multidimensionnelles suggère fortement que l'espace perceptif des sujets de MOHR et WANG ne dépasse pas deux dimensions. Pour trois dimensions, la solution obtenue n'est pas stable, et dépend sans doute des conditions initiales et de l'algorithme utilisé.

C/ HANSON G. (1967)

La thèse de HANSON (1967) reste à ce jour le travail le plus volumineux sur l'analyse multidimensionnelle des voyelles. Un bref résumé a été publié (HANSON 1968). Cet ouvrage touffu ne comporte pas moins de 74 expériences utilisant 5 ensembles de stimuli, 2 types d'analyse (analyse factorielle par composantes principales, méthode de KRUSKAL), 4 techniques de test, 3 types de sujets (enfants, étudiants, mal-entendants). Le but de HANSON est "d'améliorer et d'étayer un modèle de perception fondé sur la théorie des traits distinctifs de JAKOBSON, FANT et HALLE (1952)... et de remplacer la description unidimensionnelle des sons de la parole par une description multidimensionnelle" (op. cit., p. 8). On peut noter la confusion entre le plan de l'analyse phonologique et le plan de la modélisation de la perception que dénonce à juste titre ROSSI (1977).

Il n'est pas possible d'analyser l'ensemble de ce travail considérable. Nous n'examinerons que ce que l'auteur lui-même appelle les "expériences principales" (n° 22 à 25). Nous avons procédé à une réanalyse des données par la méthode INDSCAL, qui nous conduit à des conclusions fort différentes de celles de l'auteur.

Les expériences 1 à 10 conduisent à des résultats méthodologiques d'un intérêt considérable. Ces expériences justifient l'application de l'analyse multidimensionnelle à la perception des voyelles.

- Les informations quantitatives fournies par les indices de dissimilarité permettent d'extraire des dimensions qui correspondent aux traits distinctifs caractérisant les voyelles. Selon l'auteur, un trait purement perceptif semble également présent.

- La méthode est relativement insensible à des variables externes comme l'âge des sujets, les variations dialectales et les techniques d'obtention des indices de dissimilarité (tracer un segment de droite correspondant à la différence perçue ; cocher un point sur une ligne matérialisant le continuum "semblable - non semblable", estimation numérique directe). On peut donc faire l'hypothèse que la technique que nous avons employée est valide, car elle n'est que la combinaison de deux techniques testées par Hanson, à savoir, cocher une zone (case) sur une ligne matérialisant la magnitude des différences perçues. Pour des raisons de simplicité HANSON préconise et utilise l'estimation numérique directe.

- La motivation des sujets apparaît comme un facteur qui conditionne le succès ou l'échec de la tentative. (op. cit., p. 50).

Les expériences 11 à 17 avaient pour but de vérifier l'équivalence des indices de similarité et de dissimilarité.

D'autre part, le fait que certains sujets n'aient pas le sentiment de pouvoir fournir une estimation précise ne se manifeste pas dans les résultats. Les expériences 18 à 21 montrent que le choix d'une paire particulière comme point d'ancrage de la valeur numérique de dissimilarité, ainsi que le choix de cette valeur, n'ont pas d'influence sur les résultats. Cette technique n'est d'ailleurs plus utilisée.

Les 9 voyelles suivantes, synthétisées par un OVE II, forment le corpus des expériences principales : [i, e, ɛ, a, o, u, y ø, ɯ]. Aucune indication n'est donnée sur la qualité de ces voyelles. Il est fort improbable que celles-ci aient été identifiées sans ambiguïté. Ces voyelles sont présentées graphiquement à la fig. 6-III et dans la table ci-dessous :

	FI	FII	FIII	FIV	(Hz)
i	220	2150	3100	3500	
e	340	2100	2700	3500	durée : 500 msec
ɛ	440	2000	2650	3500	Fo : constant
a	625	950	2550	3250	
o	410	750	2450	3250	
u	280	600	2250	3250	
y	240	2000	2450	3250	
ø	375	1700	2300	3500	
ɯ	240	1600	2150	3250	

5 sujets étudiants ont estimé numériquement les dissimilarités entre les 72 couples de voyelles. L'expérience a été renouvelée 4 fois pour la même série de sujets. L'estimation de la constante additive a été faite à l'aide de la méthode la plus précise connue (op. cit., p. 56).

Nous présenterons d'abord les résultats obtenus par l'analyse factorielle en composantes principales (HARMAN 1960). D'après l'auteur, les deux premières dimensions correspondent aux deux premiers formants.

Une troisième dimension "interprétable" est présentée, sans justification, comme nécessaire. Elle ne correspond pas à une variable physique ou à un trait distinctif. Cette

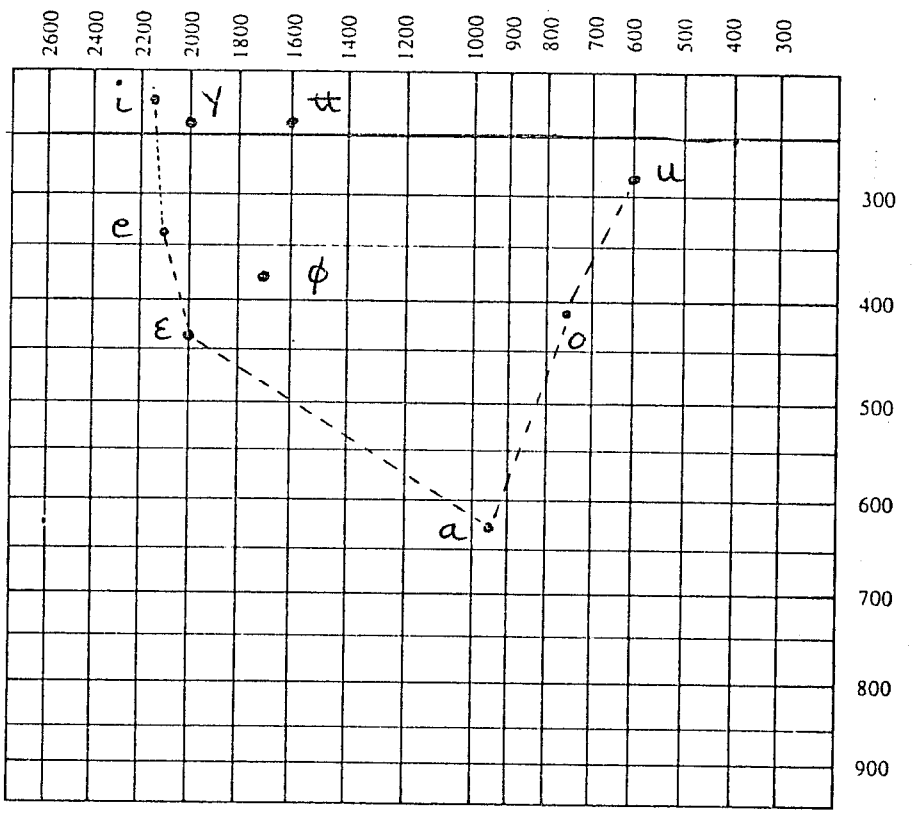

2600
2400
2200
2000
1800
1600
1400
1200
1000
900
800
700
600
500
400
300
300
400
500
600
700
800
900
i
y
ɨ
u
e
ø
ɛ
o
a

Fig. 6 - III

troisième dimension oppose 3 groupes de voyelles [i, a], [y e o u ɯ] et [ɛ ø] . Hanson interprète ce facteur comme un facteur de contraste perceptif. La discussion de cette interprétation est extrêmement succinte. Notre interprétation de cette dénomination est que cette dimension augmente la distance entre certaines voyelles non-extrêmes ([ɛ, ø]) et deux voyelles extrêmes ([i, a]). Une deuxième interprétation est brièvement évoquée : cette dimension serait corrélée avec une mesure dérivée de la fréquence formantique du 3ème formant par la formule

$$F3\ rel. = \frac{F_3^2 - \overline{F}_3^2}{\overline{F}_3^2} \qquad \overline{F3} : \text{moyenne de } F3.$$

Cette formule donne l'ordre suivant :

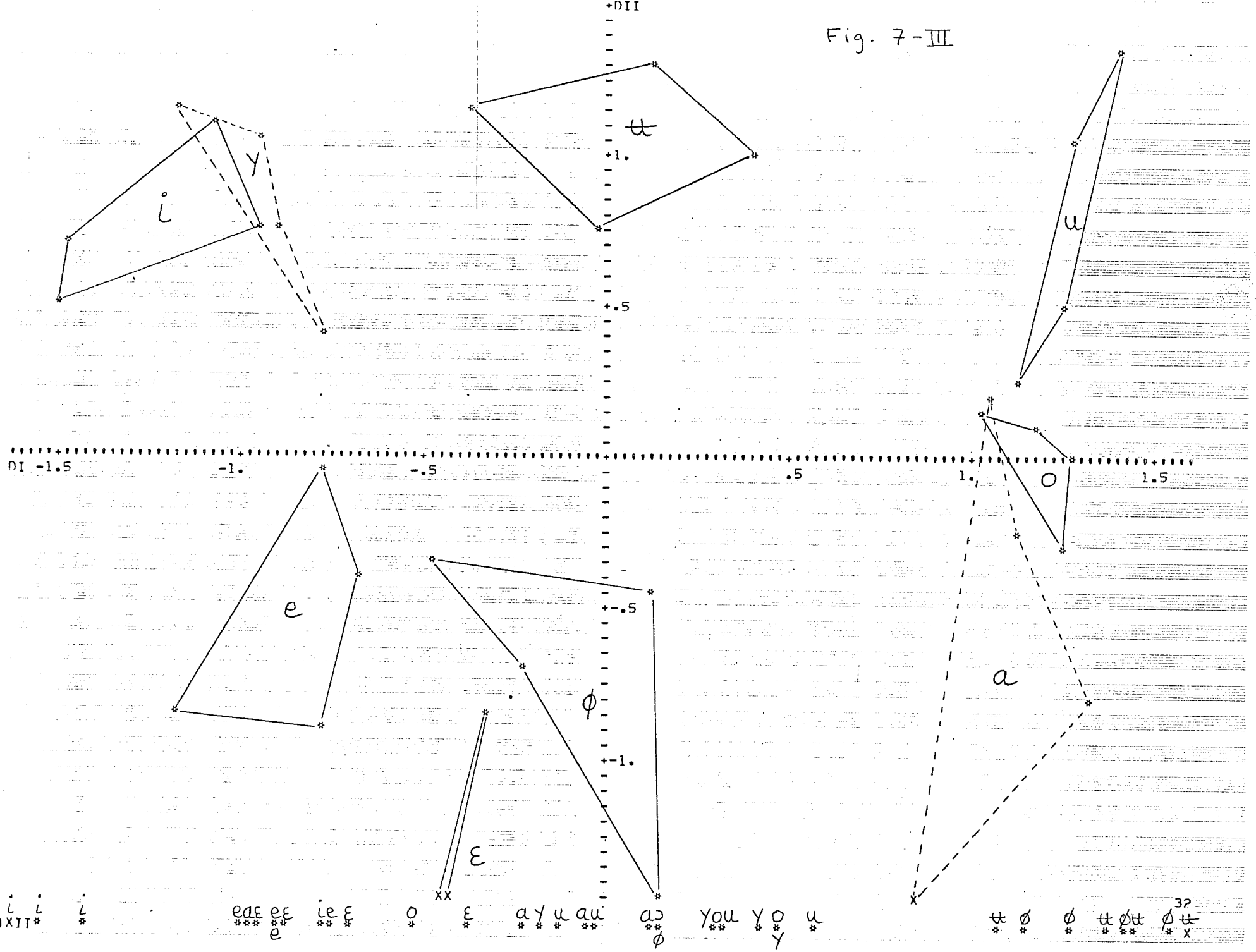
i e ɛ a o-y u ø ɯ
/ i a / y e u o ɯ / ɛ ø /

On a noté sur la ligne suivante l'ordre des voyelles sur DIII. La correspondance ne peut être décrite comme satisfaisante. Quant à la suggestion de ÖHMAN, présentée en trois lignes p. 93, elle ne nous paraît pas compréhensible. Il faut insister sur l'indigence de la discussion des résultats par rapport au luxe des considérations méthodologiques.

L'analyse par la méthode de Kruskal donne des résultats identiques pour les deux premières dimensions, mais l'accord est globalement inférieur en ce qui concerne la corrélation avec le plan F1-F2. Par contre la troisième dimension corrèle parfaitement avec F3. Ce résultat est surprenant dans la mesure où les valeurs de F3 ont été choisies d'une façon relativement arbitraire pour ces voyelles synthétiques. La divergence entre les résultats fournis par les deux techniques d'analyse n'est pas discutée.

Notre réanalyse a pour but de clarifier le problème de l'interprétation de la troisième dimension. Beaucoup d'auteurs (par ex. POLS, VAN DER KAMP et PLOMP (1969) ; TERBEEK (1977, p. 56)) ont noté les variations individuelles importantes entre les sujets (table 9.I, p. 43, op. cit.). Le pourcentage de variance expliquée ne sera pas, à priori, élevé. Les 4 expériences successives ont été analysées par 4 analyses distinctes. La fig. 7-III présente les résultats tri-dimensionnel

Fig. 7-III


+DII
+1.
+.5
+-.5
+-1.
DI -1.5
-1.
-.5
.5
1.
1.5
i
y
ɨ
u
e
o
ɛ
ø
a
ɛ
ø
ɨ
ø
ø
ɨ
ø
ɨ
ø
ɨ
DXII
eɑɛ eɛ
e
ie ɛ
o
ɛ
a y u au
ɑɔ
ø
you y o
y
u
32

cumulés, alors que les figs. 8- , 9- , 10- , et 11-III présentent les résultats pour les 4 replications. Les dissimilarités brutes ont été directement utilisées comme indices de distance . Pour des raisons pratiques, et compte-tenu des fortes variations observées pour les poids, les résultats présentés sont les résultats non-pondérés multipliés par 0.75. La table suivante présente les poids et le pourcentage de variance expliquée de chaque sujet pour les 4 répétitions.

Sujet	Poids	DI	DII	DIII	% variance expliquée
1		.82	.21	.34	84
		.50	.45	.40	61
		.66	.24	.30	58
		.80	.34	.63	76
2		.71	.52	.22	82
		.52	.57	.39	75
		.83	.20	.23	78
		.54	.33	.57	73
3		.40	.44	.41	52
		.40	.27	.52	50
		.33	.56	.47	64
		.41	.56	.47	70
4		.70	.19	.20	57
		.82	.29	.15	78
		.80	.22	.12	70
		.69	.37	.30	70
5		.48	.38	.52	65
		.68	.35	.33	69
		.66	.30	.33	63
		.55	.35	.50	68

On observe de brutales variations dans le poids des dimensions, bien que subsiste (cf. DIM I, sujet 3) une certaine cohérence pour chaque sujet.

Fig. 8-III

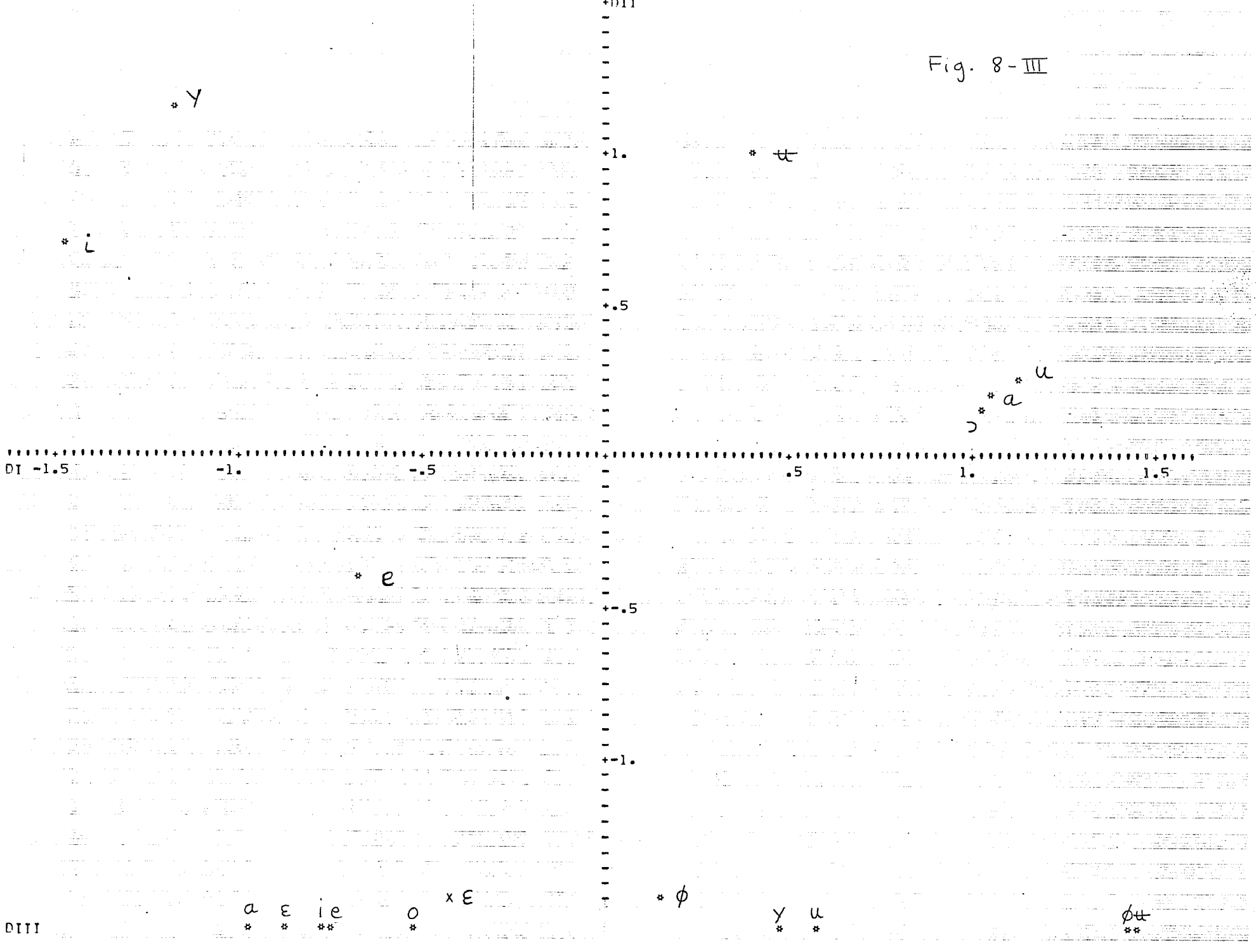


Fig. 9-III

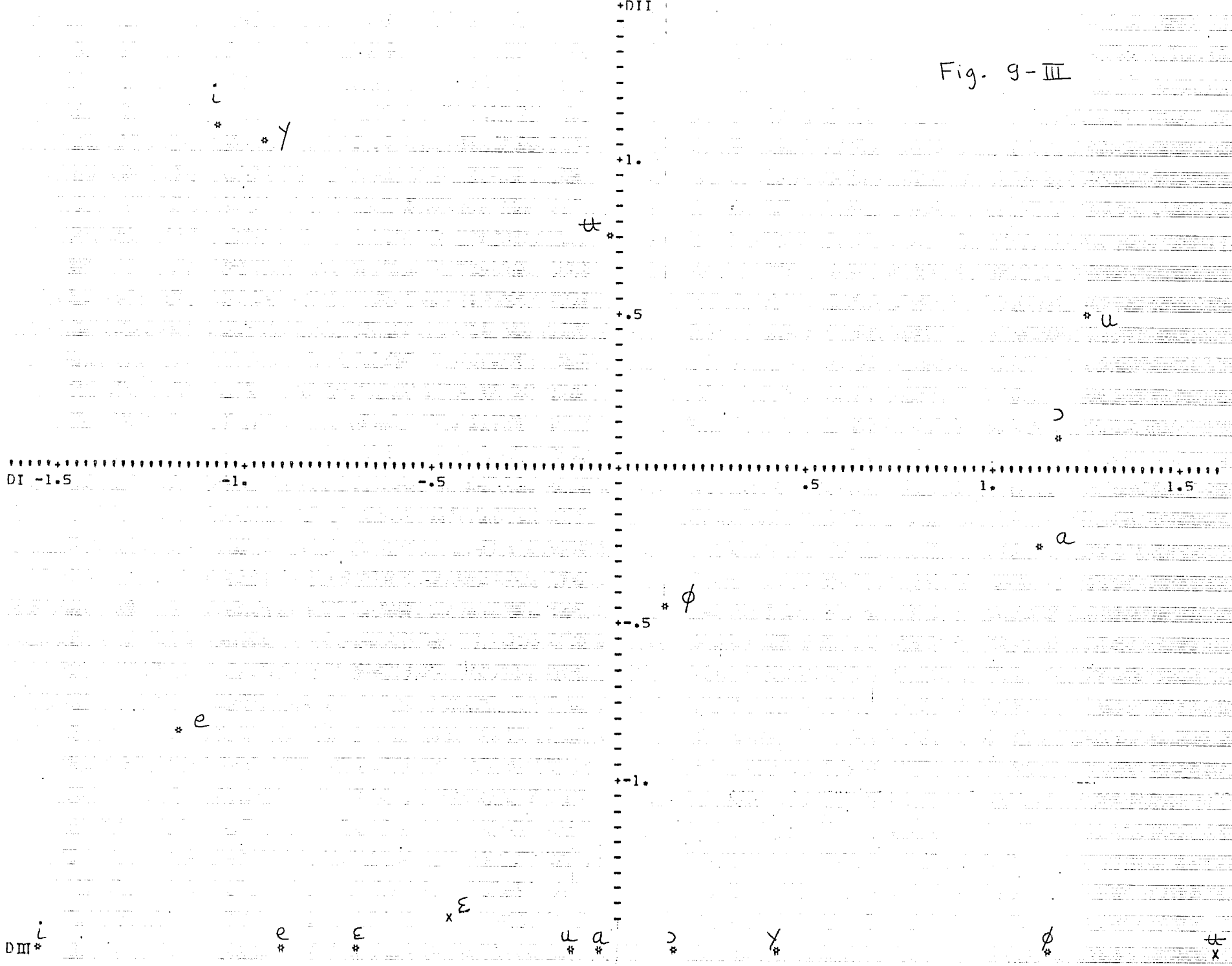


Fig. 10-III

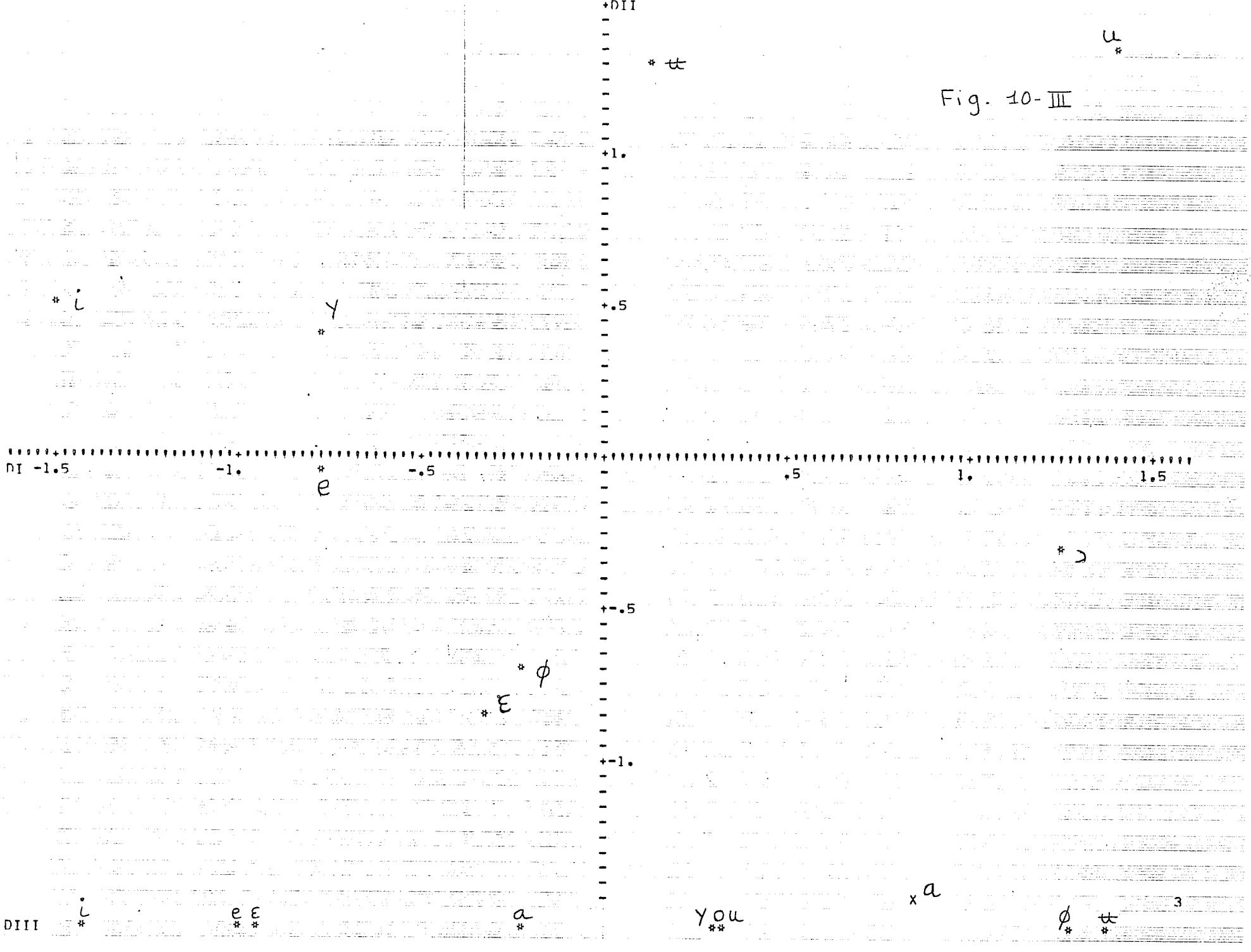
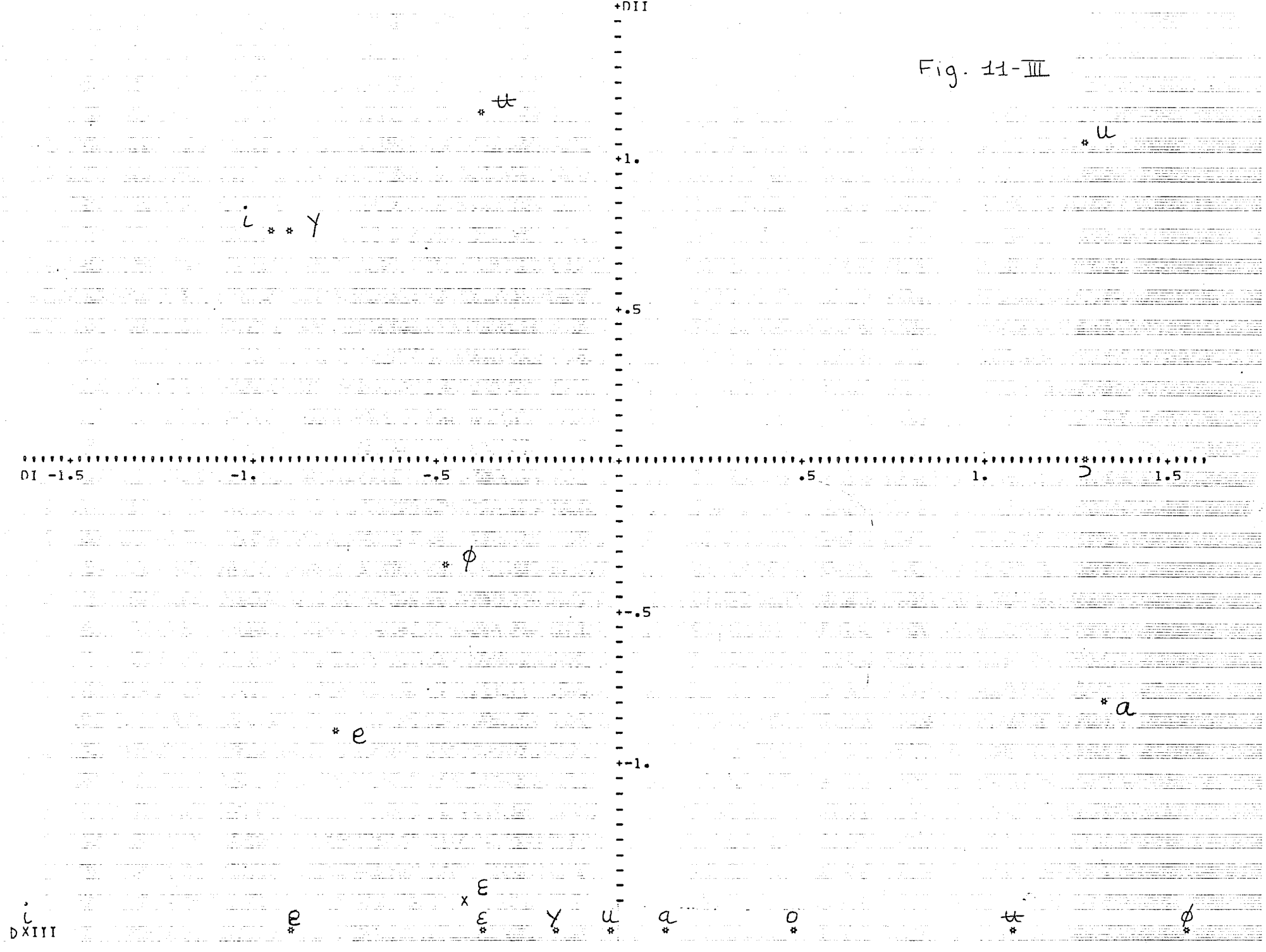


Fig. 11-III



Les coordonnées des voyelles dans l'espace tridimensionnel sont données dans les 4 tableaux suivants :

exp. 22	DI	DII	DIII
i	-1.59	.76	-.84
e	-.72	-.43	-.81
ɛ	-.45	-1.73	-.94
a	1.13	.21	-1.04
o	1.11	.16	-.57
u	1.20	.29	.62
y	-1.27	1.23	.50
ø	.17	-1.56	1.52
ɯ	.42	1.06	1.55

exp. 23	DI	DII	DIII
i	-1.14	1.18	-1.67
e	-1.26	-.92	-.96
ɛ	-.47	-2.07	-.74
a	1.20	-.26	-.06
o	1.26	.09	.16
u	1.35	.53	-.14
y	-1.03	1.12	.45
ø	.13	-.48	1.23
ɯ	-.04	.81	1.73

exp. 24	DI	DII	DIII
i	-1.60	.53	-1.54
e	-.83	-.05	-1.06
ɛ	-.36	-.89	-1.02
a	.90	-1.74	-.25
o	1.34	-.34	.35
u	1.50	1.46	.36
y	-.83	.43	.31
ø	-.25	-.77	1.36
ɯ	.13	1.37	1.48

exp. 25	DI	DII	DIII
i	-1.03	.79	-1.88
e	-.84	-.98	-.96
ɛ	-.46	-1.75	-.41
a	1.42	-.85	.14
o	1.38	-.01	.51
u	1.36	1.11	-.02
y	-.96	.83	-.18
ø	-.52	-.39	1.67
œ	-.41	1.25	1.14

Le résultat le plus marquant de notre réanalyse est que l'espace perceptif optimal n'est sans doute pas tri- mais bi-dimensionnel.

Cette conclusion repose sur les indices suivants :

- On constate que l'ordre des voyelles sur la 3ème dimension n'est autre que l'ordre (choisi arbitrairement) de la configuration initiale (à partir de laquelle le programme INDSCAL cherche la configuration optimale). Nous avions choisi arbitrairement l'ordre suivant pour l'initialisation de la 3ème dimension : [i e ɛ a o u y ø œ]
Or le calcul du rang moyen des 36 voyelles sur cet axe donne le résultat suivant : i = 4, e = 7.75, ɛ = 10, a = 14.5, o = 21, u = 22, y = 22.25, ø = 32.25, œ = 32.75

On constate la totale conservation de l'ordre. Deux conclusions sont possibles : Cet ordre correspond par hasard à l'ordre des voyelles sur la troisième dimension. Dans ce cas, l'analyse INDSCAL fournit des résultats différents des deux autres méthodes employées par Hanson. Mais une autre conclusion est plus probable : Cette troisième dimension est purement artefactuelle. Le programme tente de découvrir une structure quelconque dans le bruit résiduel. Cette conclusion est renforcée si l'on considère les résultats de l'exp. n° 22 (fig. 8-III)
Les voyelles [a, o] et [u] sont beaucoup plus proches, dans l'espace bi-dimensionnel, que ce que l'on observe à l'exp. n° 24. Cette variabilité s'explique aisément si l'on considère que les

distances intervocaliques sont représentées, de façon aléatoire, par les positions des points sur la troisième dimension dans l'expérience 22, ou la deuxième dimension dans l'expérience 24.

L'hypothèse la plus probable est que les données de HANSON ne nécessitent qu'un espace bidimensionnel, fortement isomorphique avec le plan F1-F2 ou F1-F'2. La troisième dimension ne rend compte que de la variance résiduelle, ou "bruit", qui entache les mesures de dissimilarité. La fraction des distances "prises en charge" par la troisième dimension vient en déduction des distances associées à la deuxième dimension. La "fiabilité" de cette dernière en est réduite d'autant. Cette conclusion ne peut être que provisoire en raison du "bruit" important des indices de similarité. L'hypothèse d'une troisième dimension noyée dans le "bruit" n'est pas à exclure totalement. On notera enfin la supériorité de l'analyse INDSCAL par rapport aux méthodes employées précédemment.

En résumé, cette étude, contrairement aux conclusions de HANSON, illustre plutôt l'échec d'une représentation tridimensionnelle des dissimilarités inter-vocaliques. Cette conclusion est renforcée par la tentative de réanalyse des mêmes données par TERBEEK (1977, p. 56-57). "... Il est décevant de constater que les solutions employant plus de 3 dimensions ne sont pas interprétables... sans doute à cause du "bruit"..." Pour la solution à 3 dimensions, "une seule des trois est facile à interpréter - une dimension "antérieur-postérieur " (=F2)... les deux autres dimensions... mélangent des informations relatives à la hauteur (=F1) et à la labialisation mais ces 2 dimensions ne sont pas interprétables..." Cet échec nous conforte dans l'idée que la structure perceptive sous-jacente est seulement bidimensionnelle.

D/ VAN DER KAMP ET POLS (1971)

Cette étude nous semble être un exemple des artefacts provoqués par l'utilisation de la méthode employant des voyelles dont la qualité a été dégradée afin de diminuer le nombre de cases nulles de la matrice de confusions.

Nous ne pensons pas que l'on puisse généraliser les résultats obtenus à l'espace perceptif des voyelles d'une langue donnée. Le corpus comprenait les 11 voyelles néerlandaises suivantes :

[i e ɛ ɑ a ɔ o u y ø œ]

Pour chaque voyelle, une période choisie au centre de la syllabe [hVt] a été extraite numériquement par conversion analogue-digitale et stockée dans la mémoire d'un ordinateur après normalisation du fondamental (F_o = 125 Hz), de l'attaque et de l'intensité subjective. Cette période numérisée a été présentée 6 fois à 15 locuteurs dans une tâche d'identification afin d'obtenir une matrice de confusions globale. La faible durée (8 msec) d'un stimulus explique le faible pourcentage d'identifications correctes : 30,8 %.

La majeure partie de la discussion traite du problème de la non-symétrie de la matrice de confusions et des techniques de symétrisation. Nous avons résumé brièvement cette discussion dans le paragraphe consacré aux techniques d'estimation des similarités. Les tableaux symétrisés ont été analysés par la méthode de KRUSKAL. Les résultats indiquent que les diverses techniques de symétrisation sont plus ou moins équivalentes.

Nous avons reproduit les résultats aux fig. 12- et 13-III pour les trois dimensions extraites. Les différents symboles pour chaque voyelle indiquent la technique de symétrisation.

La position des voyelles dans l'espace reflète les résultats de la matrice de confusions. Par rapport à leur position sur le trapèze vocalique, 5 voyelles se trouvent déplacées d'une façon importante : [e, u, ɛ, ø, œ] . Pour [e] la matrice de confusions indique que cette voyelle a été identifiée comme une des voyelles antérieures arrondies [y, ø] ou [œ] dans 50 % des cas, comme une voyelle antérieure ou postérieure dans 20 % des cas respectivement, et comme elle-même dans 10 % seulement des présentations. Il n'est donc pas étonnant de trouver la voyelle [e] très proche de [y, ø] et [œ] dans l'espace perceptif. Pour [u], la voyelle se trouve pratiquement

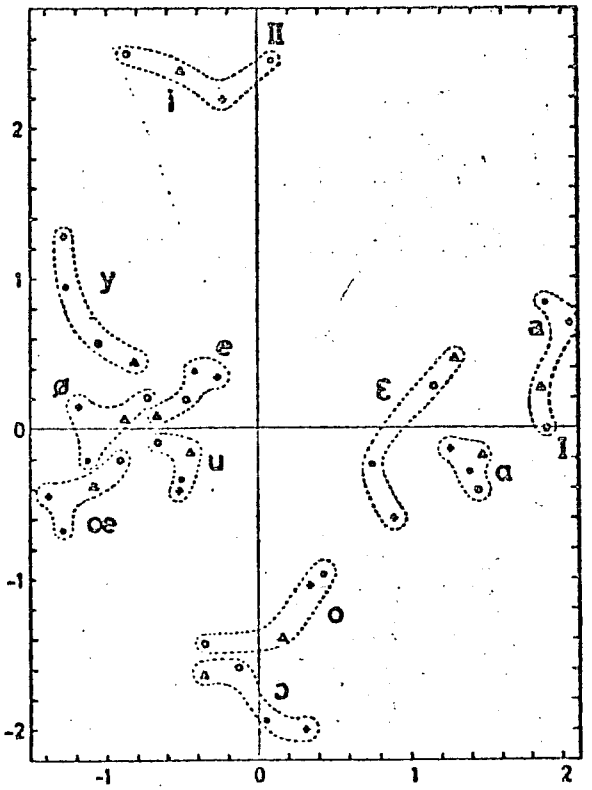


Fig. I2 - III (Van der Kamp et Pols I97I)

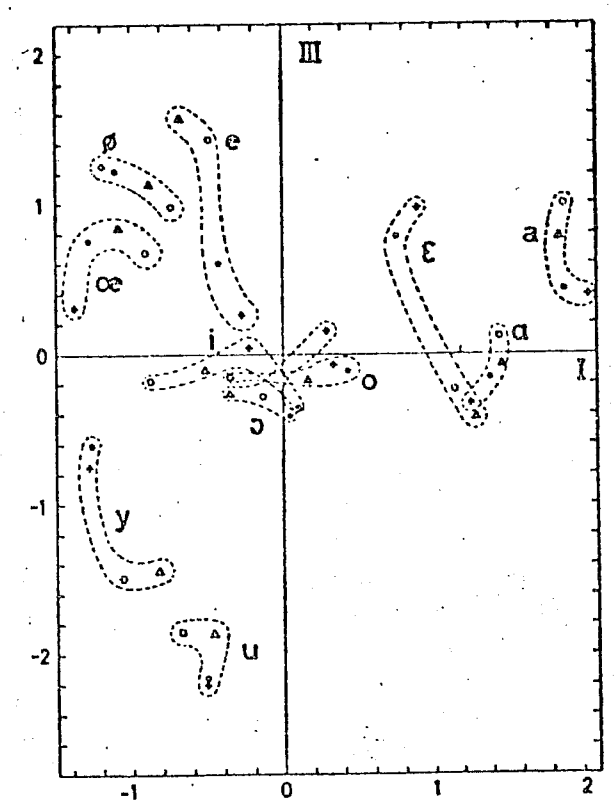


Fig. I3 - III (Van der Kamp et Pols I97I)

à l'origine des axes I et II, tout en occupant une position extrême sur la 3ème dimension. La matrice de confusions indique que cette voyelle présente un taux de confusions faible mais presque identique pour toutes les voyelles. Cette configuration spatiale permet à [u] de se trouver à égale distance de toutes les voyelles, tout en étant éloigné de chacune d'elles.

La position des voyelles dépend donc de façon cruciale des données de la matrice de confusions. Mais il est impossible de prétendre que les données de confusions sont "typiques" pour les voyelles du néerlandais. Elles dépendent du stimulus particulier choisi pour le test. Comme nous l'avons montré plus haut, une légère variation de l'aperture d'une voyelle peut provoquer une augmentation spectaculaire des confusions avec les voyelles adjacentes. Pour [e], par exemple, son comportement s'explique facilement si l'on considère que le deuxième formant était à une fréquence plus basse que la moyenne, ce qui augmente le nombre de confusions avec les voyelles antérieures arrondies. Cet écart pourrait être dû au phénomène de centralisation des voyelles en contexte. La présentation d'une période isolée ne permet plus aux auditeurs de compenser cet effet articulatoire. L'absence de mesures fréquentielles ne permet pas de vérifier cette hypothèse.

Cette étude confirme simplement que F1 et F2 (ou F'2) jouent un rôle essentiel dans le positionnement des voyelles dans l'espace auditif. La première dimension corrèle fortement avec la fréquence du premier formant, alors que la dimension présente les voyelles dans un ordre compatible avec la fréquence du deuxième formant (F2) ou une somme pondérée des formants supérieurs, incluant F2 pour une large part. (F'2 : FANT (1959), p. 80).

Quant à la troisième dimension, nous suggérons la présence d'une dimension artefactuelle. Cette troisième dimension aurait pour but de rendre compte des distorsions qu'implique la méthode utilisée. Il paraît peu probable en effet que [u] soit aussi proche de [o] que de [ɛ] comme l'indique la matrice de confusions. De plus, cette dimension ne corrèle

avec aucun paramètre physique "classique". Son acceptation se heurte à un grave problème d'interprétation.

La technique utilisée permet certainement de représenter graphiquement les résultats d'une matrice de confusions. Mais celle-ci, basée sur l'audition d'une période d'un unique échantillon, ne nous paraît pas être représentative des confusions observées par d'autres techniques, (PETERSON et BARNEY 1952 par exemple). La comparaison avec la matrice de confusions présentée par KLEIN, PLOMP et POLS (1970) en apporte une preuve certaine.

E/ POLS, VAN DER KAMP ET PLOMP (1969)

Ce travail, contemporain du précédent, mais publié plus rapidement, a été réalisé par la même équipe de chercheurs. Son importance nous a conduit à réaliser une nouvelle analyse des résultats par la méthode INDSCAL. La série de 11 voyelles décrites plus haut a été utilisée ; mais dans ce cas, les stimuli étaient constitués par la répétition d'une période unique pendant 405 msec. La fig. 14-III présente les fréquences des 3 premiers formants (op. cit., table III, p. 464). L'article contient des remarques contradictoires sur la qualité de ces voyelles "pseudo-synthétiques". "... en dépit des modifications et de la présentation hors-contexte, 10 des 11 voyelles ont été identifiées presque unanimement par les 15 sujets.... les opinions sur la 11ème ([e]) étaient partagées" (p. 464). "... La plupart des sujets ne se sont pas aperçus que les stimuli étaient en réalité des voyelles ; on peut en conclure qu'ils n'ont pas utilisé de critères linguistiques pour former leur jugement". Les auteurs soulignent que les stimuli ne sont pas forcément des échantillons représentatifs des voyelles néerlandaises. Certaines valeurs formantiques sont relativement éloignées des moyennes calculées par POLS, TROMP et PLOMP (1973).

F 3	[y]	2720 Hz : 2208 Hz	Dans ces valeurs, la pre-
F 1	[ɛ]	740 Hz : 583 Hz	mière valeur est la fré-
F 3	[e]	3300 Hz : 2553 Hz	quence de la période

retenue, la deuxième étant la valeur moyenne de ce paramètre pour 50 locuteurs masculins.

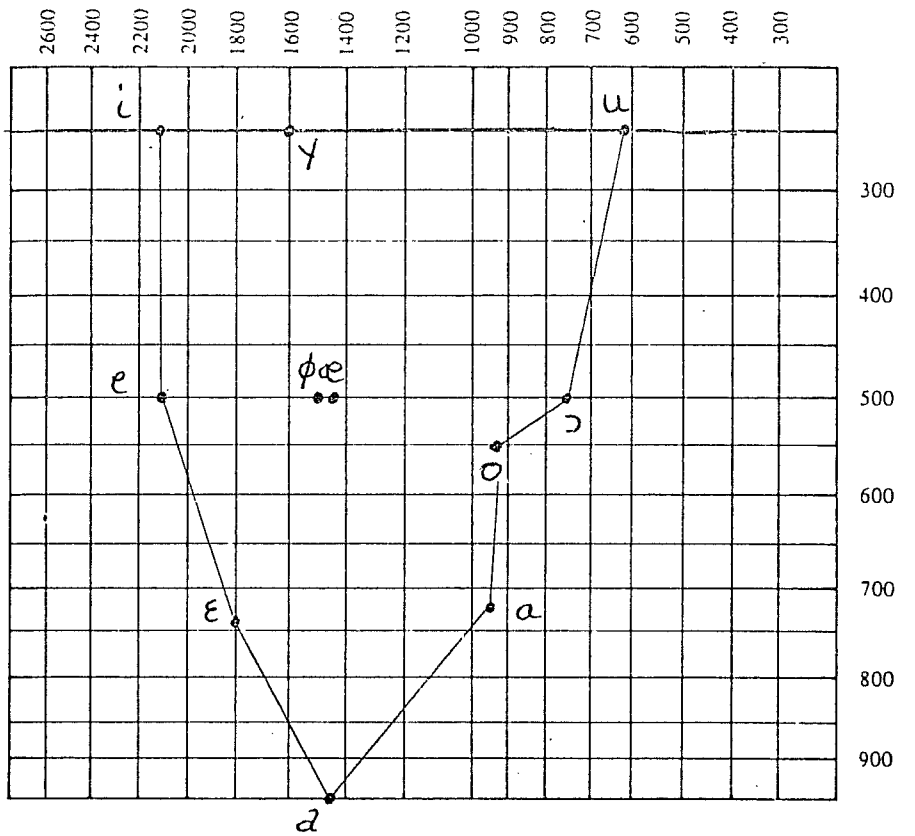


Fig. I4 - III

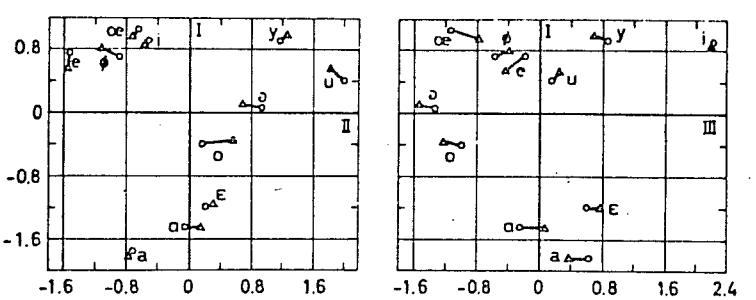


Fig. I5 - III (Pols et al. I969)

La fig. 15-III reproduit les résultats obtenus par la méthode de KRUSKAL (op. cit., fig. 2, p. 462). L'estimation des indices a été faite par la méthode triadique. Pour notre réanalyse, nous avons transformé les indices de similarité en indices de dissimilarité à l'aide de la formule di = 300 - si, où di et si sont les deux indices pour le ième couple. La valeur de la constante numérique a été choisie de telle manière que la moyenne des indices de dissimilarité soit égale à celle des indices de similarité ($x = 134$, $\sigma = 46$). Aucune correction de constante additive n'a été effectuée.

Les coordonnées des voyelles dans un espace à 3 dimensions sont présentées dans le tableau suivant et à la fig. 16-III

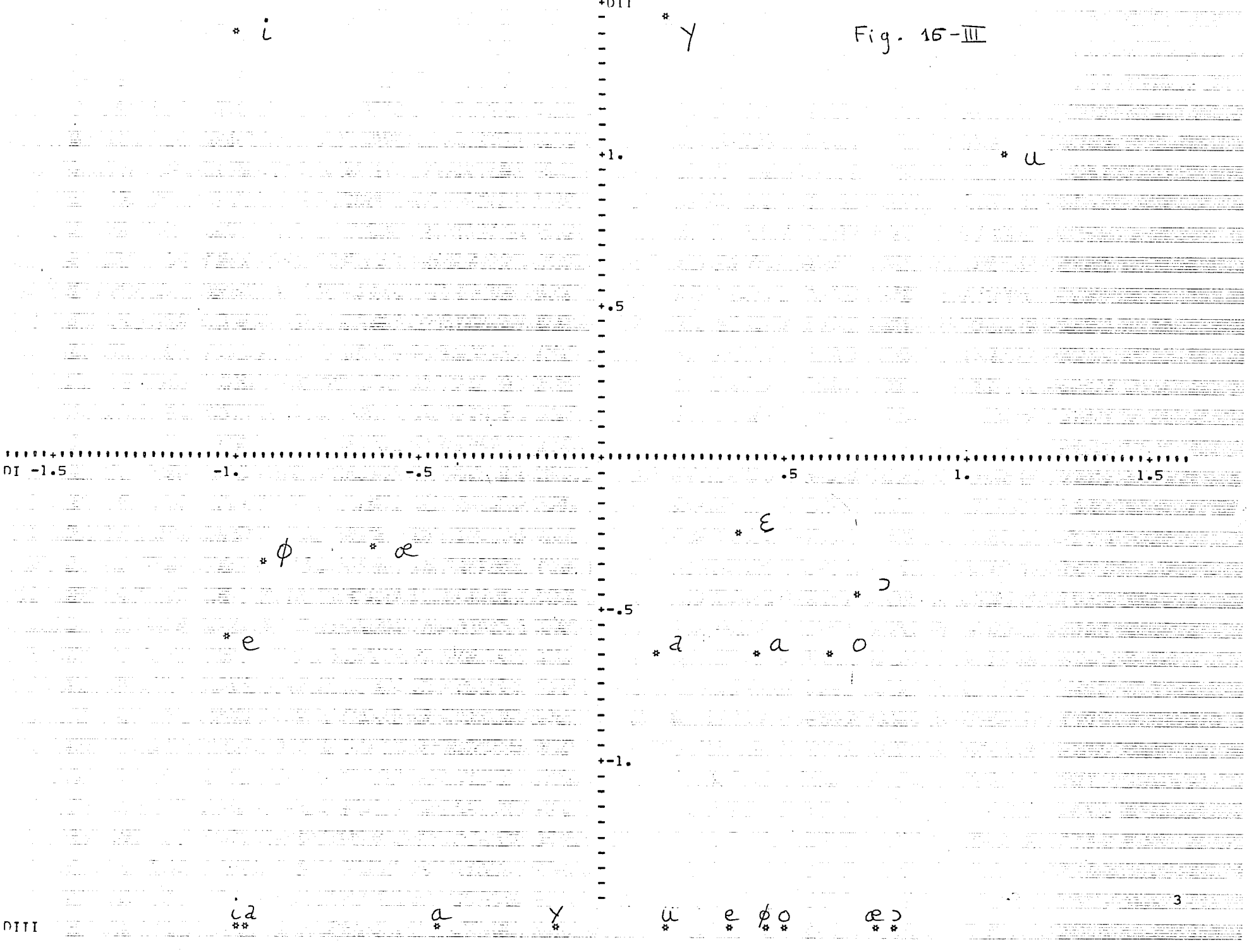
	DI	DII	DIII
i	-1.38	1.73	-1.61
e	-1.42	-.75	.55
ɛ	.53	-.30	-.73
ə	.20	-.81	-1.60
a	.60	-.80	-.74
ɔ	.97	-.55	1.29
o	.86	-.77	.82
u	1.51	1.24	.28
y	.26	1.81	-.22
ø	-1.29	-.43	.74
œ	-.85	-.38	1.21

Les poids associés à chaque dimension, ainsi que le pourcentage de variance expliquée, sont présentés dans le tableau ci-dessous :

Poids DI	Poids DII	Poids III	% variance expliquée
0.52	0.66	0.38	85 %

Les résultats de la fig. 16-III tiennent compte des poids. L'orientation des axes a été choisie de telle manière que la première dimension de notre réanalyse corresponde à F2 et la deuxième à F1. Pour comparer les deux configurations, il est

Fig. 15-III



nécessaire de procéder à une réorientation des axes de POLS et al.. Cette opération est légitime car aucune des deux méthodes n'assure une orientation définie des axes dans l'espace.

On observe un bon accord entre les deux analyses pour les deux premières dimensions. La voyelle [ɛ] est plus proche dans notre analyse du groupe [ɑ ɑ o ɔ] . D'autre part, le groupe [e ø œ] est beaucoup moins central. La correspondance entre le plan des deux premières dimensions et le plan F1 - F2 est plus qualitatif que quantitatif. Les résultats suggèrent fortement que les voyelles n'ont pas été reconnues individuellement mais comme appartenant à 3 grands groupes. Le premier groupe comprend les voyelles [i, y, u], caractérisées par un premier formant ayant une fréquence basse ; le deuxième groupe comprend les voyelles possédant un deuxième formant élevé. [e ø œ] L'intensité extrêmement faible du troisième formant de [e] n'est sans doute pas étrangère à sa classification parmi les voyelles arrondies. Ou alors, la fréquence très basse de F3 pour les voyelles [ø] et [œ] a pu être prise pour le deuxième formant d'une voyelle non-labialisée. Quant au troisième groupe, il se compose des voyelles [ɑ ɑ o ɔ] , qui ont toutes des fréquences F1 et F2 proches. Ces déviations par rapport au spectre moyen des voyelles naturelles sont la conséquence de l'utilisation d'une période répétée.

Les auteurs concluent que la troisième dimension est ininterprétable. Nous ne sommes pas certains de la stabilité de l'ordre des voyelles sur cette dimension. Il est probable que cette dimension ne rend compte que des erreurs résiduelles dues au "bruit". Il ne faut pas perdre de vue également que la méthode triadique sous-estime systématiquement certaines distances, ce qui peut faire apparaître une dimension artefactuelle supplémentaire.

L'article de POLS et al. contient également une analyse de la configuration physique obtenue par analyse en composantes principales des spectres tiers d'octave. Celle-ci permet d'extraire 3 paramètres physiques ; F1, F2 et L2, l'intensité du deuxième formant. L'apparition de L2 à la place de F3 est

surprenante. D'autre part, L2 est fortement corrélé avec F2 : L2 ne constitue pas une dimension indépendante. L'essai de corrélation entre les dimensions perceptives et les dimensions physiques doit être considéré comme un échec car seuls F1 et F2 paraissent interprétables.

En conclusion, on peut dire que malgré l'utilisation de techniques sophistiquées, les résultats sont décevants. Les stimuli choisis ne sont pas représentatifs des voyelles naturelles. La technique de "synthèse" des voyelles a pour effet de perturber les fréquences et les amplitudes des formants. Il est probable que la méthode triadique d'estimation des distances provoque l'apparition d'une dimension ininterprétable.

La minceur des résultats, qui contraste avec l'importance des moyens utilisés, a conduit TERBEECK à procéder à une nouvelle analyse des données individuelles (TERBEECK 1977 p. 46-55). Le critère utilisé dans la méthode d'analyse PARAFAC suggère une interprétation bidimensionnelle. Des poids négatifs apparaissent pour les dimensions de rang supérieur. Les solutions sont pourtant uniques jusqu'à 4 dimensions. Mais seul F1 est interprétable dans la solution à 4 dimensions que préconise TERBEECK. Les trois autres dimensions sont interprétées comme a) une dimension antérieure non-arrondi postérieur arrondi ; b) une dimension opposant les voyelles fermées ou ouvertes aux voyelles d'aperture moyenne ; c) une opposition (e-ɑ) (?). La solution en deux dimensions est facilement interprétable et correspond à celle que nous avons présentée. De plus, les voyelles dans la solution quadri-dimensionnelle semblent se répartir sur deux courbes de la manière suivante :

1) ɑ a ɛ ɔ o e ø œ u i y
2) u o ɔ y ɛ a œ ø ɑ i e

La deuxième courbe correspond approximativement à F2, avec un placement aberrant de la voyelle [ɛ], alors que la première courbe correspond à F1. L'utilisation de la méthode triadique pourrait être à l'origine de cette "curvature" des dimensions en raison de la sous-estimation des distances longues. Cette

curvature peut être comparée à la torsion que l'on applique à une baguette pour la transformer en un arc : l'effet de la sous-estimation des distances extrêmes a le même effet que la corde tendue entre les deux pointes de l'arc. Cette corde est plus courte que la longueur de l'arc. Cette curvature transforme une droite en un plan et détermine donc l'apparition d'une dimension supplémentaire artefactuelle.

Nous ne retiendrons de cette étude que l'impossibilité d'interpréter une troisième dimension.

F/ KLEIN, PLOMP et POLS (1970)

Ce travail fait suite aux deux précédents. La partie de cette expérience traitant de l'espace perceptif des voyelles utilise des segments de voyelles naturelles d'une durée de 100 msec, présentés à 10 locuteurs dans une tâche d'identification. 12 voyelles néerlandaises (i e I ɛ ə a ɔ o u y ø œ), produites par 50 locuteurs ont été utilisées pour construire les matrices de confusions.

L'examen de la matrice globale de confusions révèle, contrairement à ce qu'affirment les auteurs, un "biais de réponse" important (JANSON 1977). [a], par exemple, est perçu comme [ɑ] dans 37 % des cas, alors que l'inverse n'est vrai que dans 8 % des cas. La méthode de symétrisation par moyennage n'est donc pas justifiée. La matrice des indices de similarité comporte 15 cases (sur 66) égales à 0, 1 ou 2, alors que la valeur la plus élevée atteint 283. Le nombre important de valeurs très faibles implique qu'un nombre équivalent de distances seront égales. Une distorsion est à prévoir, sans que l'on puisse préciser sa forme ou son ampleur. Du fait de l'égalisation des durées, les confusions entre [ɔ-o, œ-ø, I-e] sont très importantes, ce qui indique que la durée est un paramètre important pour leur identification.

L'analyse du tableau des similarités a été effectuée par la méthode de KRUSKAL. Le critère de "stress" suggère une interprétation bi- ou tri-dimensionnelle. Les dimensions I et II (cf. fig. 17-III reproduite de la fig. 9, p. 1008 de

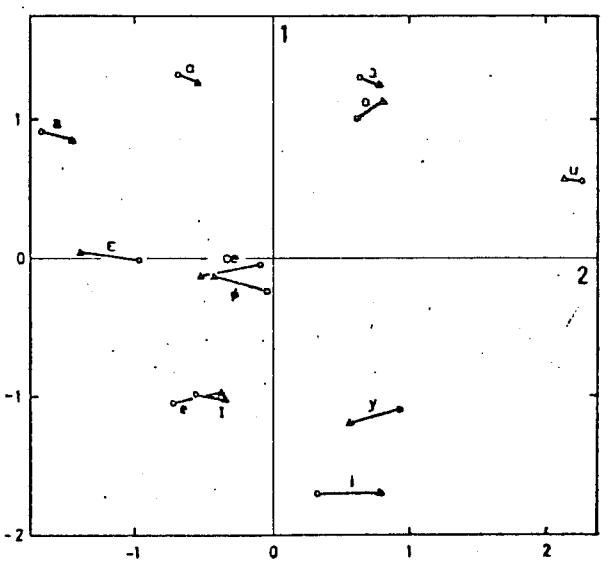


FIG. 6. F_1–F_2 configuration (O) and I–II configuration (Δ) after rotation to optimal congruence.

Fig. I7 - III (KLEIN et al. I970)

l'article original) corrèlent de façon satisfaisante avec le plan F1 - F2. Les "∘" représentent les voyelles dans l'espace perceptif, et les points les voyelles dans l'espace physique, après rotation au maximum de congruence. L'espace physique est le résultat d'une analyse en composantes principales des spectres tiers d'octave. On notera cependant les positions aberrantes de [y] et de [œ - ø]. La troisième dimension sépare les voyelles antérieures arrondies [y - ø - œ] ainsi que [u] (mais pas [o]). Cette troisième dimension perceptive corrèle fortement avec la troisième dimension physique (r = 0.907). Mais on remarquera que la troisième dimension a simplement pour objet d'augmenter les distances entre les voyelles antérieures arrondies et les antérieures non-arrondies, ainsi qu'entre [i] et [u]. Ces voyelles sont précisément les voyelles "aberrantes" dans le plan des deux premières dimensions. Mais un problème important demeure : la troisième dimension physique n'est pas interprétable. Il est également impossible de préciser l'influence de la transformation des indices en distances sur le nombre de dimensions : on sait que cette transformation conduit à de sérieuses distorsions. On ne peut donc écarter la possibilité d'une dimension artefactuelle. On constatera que toutes les études utilisant des matrices de confusions font apparaître une dimension ininterprétable.

La conclusion essentielle que nous tirerons de cette étude est que seules les deux premières dimensions de l'espace perceptif sont interprétables et qu'on ne peut rejeter la possibilité d'une troisième dimension artefactuelle résultant de la méthode employée.

G/ SINGH ET WOODS (1971)

Cette étude reste l'une des meilleures publiées à ce jour. Elle porte sur 12 voyelles de l'anglais américain [i ɪ e ɛ æ ɑ ɔ o ʊ u ʌ ɝ] prononcées par un sujet féminin. Toutes les paires ordonnées ont été enregistrées séparément. Cette procédure présente un avantage et un inconvénient : les voyelles de chaque paire ne sont plus strictement comparables,

ce qui ne permet pas la détermination de la position acoustique de la voyelle sur le plan F1-F2 ; mais le problème du choix d'un échantillon représentatif ne se pose plus. Ces paires ont été présentées à 18 sujets dans une tâche de scaling sur 7 positions. Les auteurs n'indiquent pas si les indices ont été transformés en distances. SINGH et WOODS ont utilisé la méthode de KRUSKAL. Pour obtenir une estimation du pourcentage de variance expliquée, nous avons réanalysé les données par la méthode INDSCAL, sans calcul d'une constante additive. Il y a peu de différence entre la solution originale et la nôtre : aussi ne présenterons-nous que nos propres résultats.

Les coordonnées des voyelles sont les suivantes dans un espace à 3 dimensions :

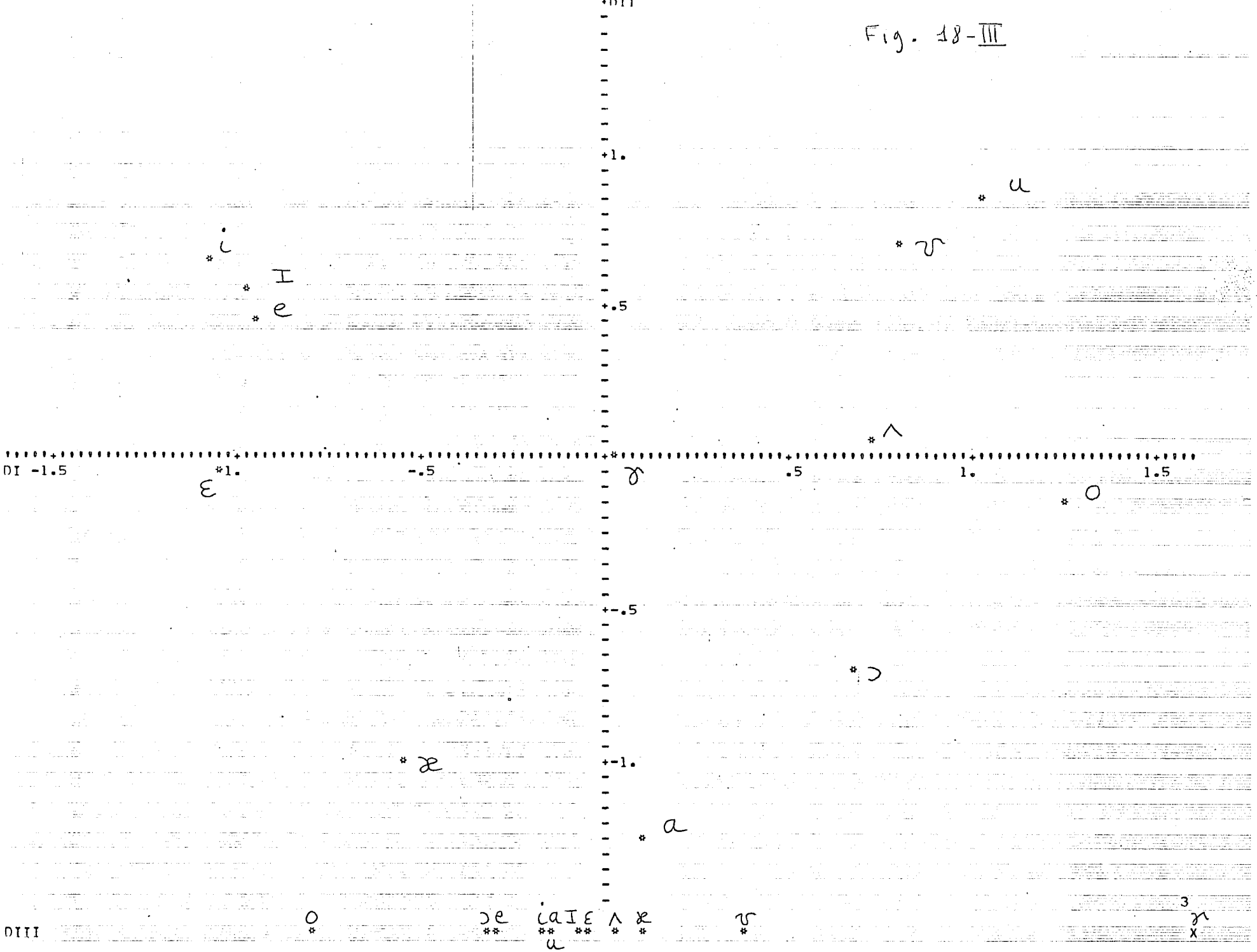
	DI	DII	DIII
i	-1.27	.99	-.52
ɪ	-1.14	.83	-.15
e	-1.12	.67	-.53
ɛ	-1.23	-.08	-.10
æ	-.66	-1.52	.19
a	.12	-1.91	-.26
ɔ	.79	-1.08	-.57
o	1.48	-.25	-1.38
ʊ	.93	1.03	.66
u	1.22	1.29	-.29
ʌ	.84	.05	.03
ɚ	.03	-.02	2.92

La figure 18-III présente ces mêmes résultats sous une forme graphique. Le tableau des poids révèle que le pourcentage de variance expliquée atteint 82 %.

DI	DII	DIII	% var. expl.
.72	.44	.33	82 %

Ce dernier résultat est très satisfaisant. Nous avons tenté d'extraire une quatrième dimension. Mais le pourcentage de variance expliquée retombe à 61 %, ce qui indique que cette

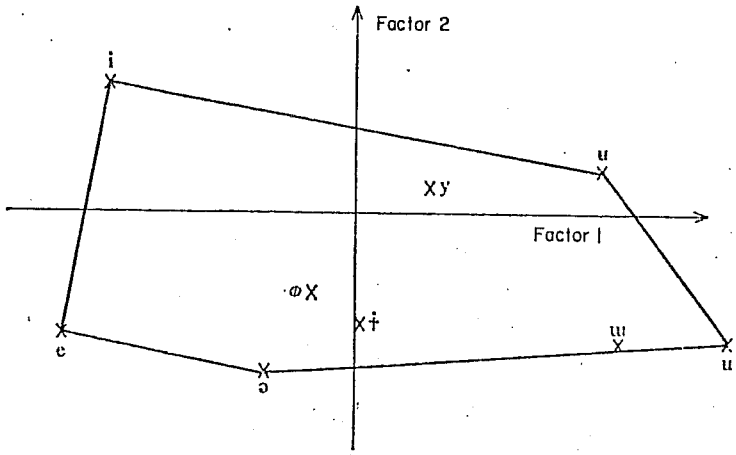
Fig. 18-III



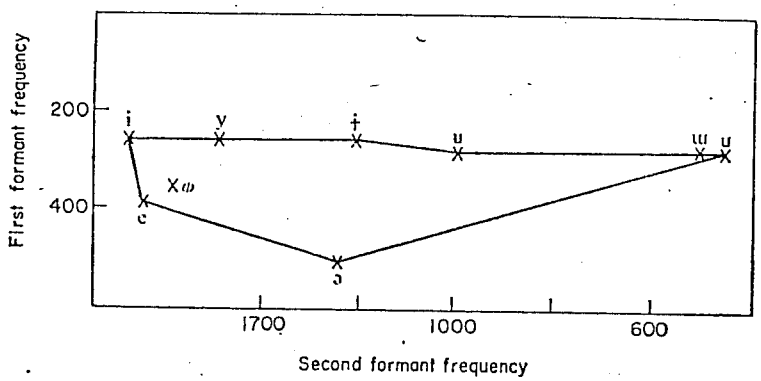
quatrième dimension corrèle fortement avec les trois premières. Ceci indique sans ambiguïté que l'espace perceptif ne comporte pas plus de 3 dimensions. Les deux premières dimensions corrèlent avec F1 et F2 (ou F'2). Quant à la troisième dimension, elle oppose la voyelle [ɚ] à toutes les autres voyelles. L'interprétation de cette dimension est donc simple : il s'agit d'une dimension particulière à la voyelle rétroflexe ou "rhotacisée" (LADEFOGED 1975). A l'exception de la voyelle [ʌ], la position des voyelles dans l'espace est proche de leur position dans le plan F1-F2. On notera en particulier la position extrême de [o] par rapport à [u]. [ʊ] apparaît bien comme relativement centralisé par comparaison à [ɯ]. Pour les voyelles antérieures, la correspondance avec les données formantiques aurait exigé que [i] soit un peu plus extrême. On notera que le trait distinctif de tension, qui oppose les couples [i-ɪ, o-ɔ, ɯ-ʊ] n'apparaît pas comme une dimension perceptive. La conclusion essentielle est qu'un espace tri-dimensionnel permet de représenter de façon satisfaisante les indices de similarité entre les couples de voyelles de l'anglais américain. L'une de ces trois dimensions semble spécifique à la voyelle rétroflexe [ɚ]. L'absence de voyelles arrondies antérieures dans cette langue laisse en suspens la généralisation de cette conclusion à d'autres langues. D'un point de vue méthodologique, notons que l'emploi d'un "scaling" à 7 positions a donné des résultats interprétables sans difficulté.

H/ SHEPARD (1972)

SHEPARD (1972) a procédé à une analyse multidimensionnelle du tableau de confusions de PETERSON et BARNEY (1952). Ce tableau présente les confusions entre 10 voyelles de l'anglais américain : [i ɪ ɛ æ ɑ ɔ ʌ ʊ ɯ ɚ]. La matrice de confusions a été obtenue par un test d'identification de 1520 mots présentés à 70 auditeurs. C'est de loin le test d'identification le plus important ayant été réalisé à ce jour. A partir des probabilités de confusions de chaque paire, SHEPARD a obtenu une distance symétrique par la formule suivante :



Perceived positions of stimulus vowels in two-dimensional perceptual space generated in experiment 2.



Formant structure of stimulus vowels used in experiments 2 and 4.

Fig. 20 - III (BUTCHER 1974)

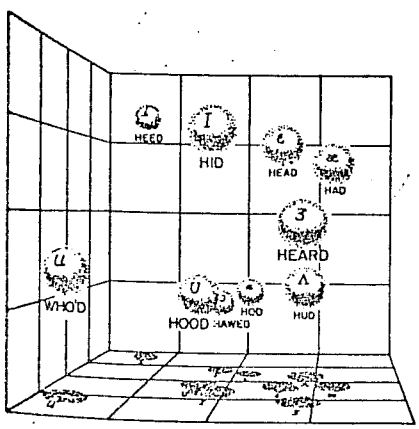


Fig. 4.3. Three-dimensional spatial representation for 10 vowel phonemes (based on the data of Peterson and Barney).

Fig. 19 - III (SHEPARD 1972)

$$D(i,j) = \frac{pi,j + pj,i}{pi,i + pj,j}$$

D(i,j) vérifie les axiomes métriques alors que les probabilités de confusions pi,j et pj,i ne les vérifient pas. Il reste néanmoins de nombreuses cases vides, d'où un nombre important de distances identiques. L'analyse multidimensionnelle a été effectuée par une variante de la méthode de KRUSKAL (1964a, 1964b), utilisant une fonction exponentielle entre les distances subjectives et les distances dans l'espace perceptif. Cette méthode ne permet qu'une analyse des moyennes de groupe, et n'assure donc pas une orientation unique des axes dans l'espace. Elle ne permet pas non plus d'estimer la quantité de variance expliquée. Pour connaître cette dernière valeur, il faut calculer le coefficient de corrélation entre les distances initiales et les distances dans l'espace. La méthode de KRUSKAL est une méthode itérative qui cherche à minimiser, par une méthode du gradient, l'écart entre l'ordre des distances initiales et l'ordre des distances dans l'espace ; (cf. également (en français) CAILLIEZ et PAGES 1976, p. 256-260).

Les résultats ne sont pas faciles à interpréter. (cf. fig. 19-III reproduite de l'article original) Si les deux premières dimensions (après rotation) corrèlent fortement avec F1 et F2 (r = 0.88 et 0.98) dans un système d'axes pratiquement orthogonaux (99°), on peut également obtenir une bonne corrélation de la troisième dimension avec F3 (r = 0.82) mais cette dimension est statistiquement liée à la deuxième. Les trois dimensions ne sont donc pas orthogonales (indépendantes). SINGH (1976, p. 151) propose d'interpréter cette dimension comme séparant les voyelles tendues des voyelles relâchées (tense/lax). Cette interprétation n'est pas conforme aux résultats. Notons au passage que le livre de SINGH (1976) n'est par endroit qu'une caricature d'un examen sérieux de l'apport de l'analyse multidimensionnelle à la théorie des traits distinctifs, et qu'il mérite parfaitement les critiques publiées. La position extrême de la voyelle [ɚ] suggère une dimension de "rétroflexion" conformément aux résultats de

SINGH et WOODS (1971). Les autres voyelles se répartissent de façon aléatoire sur cette dimension supplémentaire, qu'elles utilisent pour diminuer les écarts qui pourraient naître du biais de réponse affectant les voyelles tendues et relâchées. Mais cette dimension ne les sépare pas, à preuve la position de [u] par rapport à [i] et [ʊ]. Le pourcentage de variance expliquée s'établit à 97 % pour deux dimensions, et à 99 % pour trois dimensions. Cette augmentation excessivement faible semble bien prouver que la plupart des voyelles ne sont pas "concernées" par cette troisième dimension. Ce taux de variance expliquée très élevé s'explique par le grand nombre de réponses individuelles (106 400), qui assurent des pourcentages particulièrement fiables. Ce nombre de réponses gigantesque est nécessaire pour utiliser l'analyse multidimensionnelle à partir d'un tableau de confusions.

Les résultats de SHEPARD (1972) confirment que, dans une très large mesure, les distances associées à une matrice de confusions peuvent être représentées dans un espace bidimensionnel. Plusieurs expériences semblent confirmer le caractère particulier de la voyelle [ɝ] de l'anglais américain.

I/ ANGLIN (1971)

Ce travail non-publié se trouve résumé dans SINGH et WOODS (1971) et SINGH (1976, p. 148-150). 12 voyelles américaines ont été étudiées dans un contexte de mot (b-t) possédant une signification : [i ɪ e æ a ɔ o ʊ u ʌ ɝ] . Une interprétation portant sur 4 dimensions est proposée. Les deux premières dimensions correspondent à F1 et F2 ou F'2. La troisième dimension sépare la voyelle rétroflexe des autres voyelles. Quant à la quatrième dimension, les auteurs suggèrent une opposition tendu/relâché. La perception d'une telle dimension n'est pas surprenante dans un contexte de mot, alors que cette dimension n'apparaît pas pour les expériences portant sur des voyelles isolées. Cette opposition s'appuie sur un critère de durée autant que sur les différences spectrales. La présence d'un contexte permet d'apprécier les durées relatives de la voyelle et des deux consonnes. Cette expérience

pourrait être interprétée comme une preuve importante de la différence entre la perception des voyelles en contexte et la perception des voyelles isolées. Pourtant, l'examen détaillé de la position des voyelles sur cette dimension ne confirme pas cette interprétation. La figure suivante présente les voyelles tendues et relâchées sur cette dimension.

ʌ ɛ ʊ ɪ œ u a ɔ i o

Une flèche relie les couples correspondants. S'il est certain que voyelles tendues et relâchées sont éloignées (bien que ceci soit discutable pour la paire [ɔ - o]), on n'observe pas un regroupement des voyelles selon le trait distinctif proposé. Nous interprétons cette dimension comme une dimension d'interaction, où les écarts résiduels entre les voyelles (qui peuvent provenir pour une faible part de leur caractère tendu ou relâché) sont pris en compte. L'absence des résultats détaillés ne permet pas de connaître la quantité de variance expliquée par cette dimension. Il faudrait de plus, connaître la stabilité de cet espacement. Rappelons que pour une dimension "supplémentaire", le programme ne modifie pas l'ordre des voyelles défini par les conditions initiales. On peut donc "vérifier" une hypothèse quelconque de cette manière.

Pour autant qu'il soit possible de conclure en l'absence des résultats précis, on peut dire que l'espace perceptif mis en évidence par ANGLIN (1971) possède les deux dimensions extraites par toutes les expériences, ainsi qu'une dimension séparant la voyelle rétroflexe [ɹ]. L'existence d'une quatrième dimension n'est pas confirmée par les résultats qui ont été publiés.

J/ GRANT (1971)

Ce travail non -publié est résumé, comme le précédent, dans SINGH (1976, p. 151-152). L'auteur a utilisé 8 voyelles communes à l'anglais américain et au hindi dans un contexte / bV/ : [i ɪ ɛ a o ʌ ʊ u] de INDSCAL. L'espace perceptif est pratiquement identique pour les deux groupes de sujets. Les voyelles se placent dans un espace à deux dimensions, qui correspondent à F1 et F2 (ou F'2). Notons que la voyelle [ɹ] ne fait pas partie de l'inventaire.

K/ BUTCHER (1974)

Le travail de BUTCHER (1974) a pour but de mettre en évidence les corrélats acoustiques ou articulatoires des voyelles fermées ou mi-fermées. Il utilise une méthode d'analyse multidimensionnelle "primitive" (TORGERSON, 1958). FONAGY (1963) a montré que, dans les descriptions antérieures au XXème siècle, une dimension "clair-sombre" est très souvent utilisée. BUTCHER présente au préalable plusieurs analyses monodimensionnelles (cf. chapitre II) que nous avons préféré placer ici, pour pouvoir les comparer avec les analyses multidimensionnelles du même auteur. 16 sujets anglais ont placé deux ensembles de voyelles cardinales sur la dimension clair-sombre de la manière suivante :

1) i ɨ y ʉ ɯ u

2) i e y ø œ ɛ a

L'expérience suivante remplace le classement par paire par une méthode triadique. On obtient l'ordre suivant :

i e ə ø ɨ y ʉ ɯ u

BUTCHER ne tire aucune conclusion des analyses monodimensionnelles. Ceci illustre bien le faible intérêt de cette méthode. Il traite ensuite les résultats par analyse multidimensionnelle. Nous reproduisons les résultats à la fig. 20-III ; ils sont accompagnés d'un graphe des fréquences formantiques. L'auteur interprète la position de [ʉ], [ɯ] et [u] comme le résultat de l'utilisation de stimuli ne couvrant pas régulièrement tout l'espace. Cette critique est certainement justifiée. Il note ensuite la bonne correspondance entre l'analyse acoustique et perceptive. BUTCHER interprète la position basse de [ɨ] comme la preuve que les auditeurs anglais perçoivent toute voyelle centrale comme ouverte en raison de la liaison ouverture - centralisation dans leur langue maternelle. De la même façon [y] serait groupé avec les voyelles postérieures car la labialité n'est pas compatible avec l'antériorité en anglais. On notera la position toute différente de [y] dans notre analyse des résultats de MOHR et WANG.

[y] est sans doute perçu comme une voyelle centrale fermée. Nous avons analysé les résultats par la méthode INDSCAL.

Les coordonnées des voyelles sur les deux dimensions sont les suivantes :

	DI	DII
i	-2.26	1.08
ʏ	-.33	.80
ɨ	.25	-.56
ʉ	.15	1.40
ɯ	1.25	.03
u	1.29	.69
e	-.62	-1.60
ø	-.03	-.65
ə	.31	-1.20

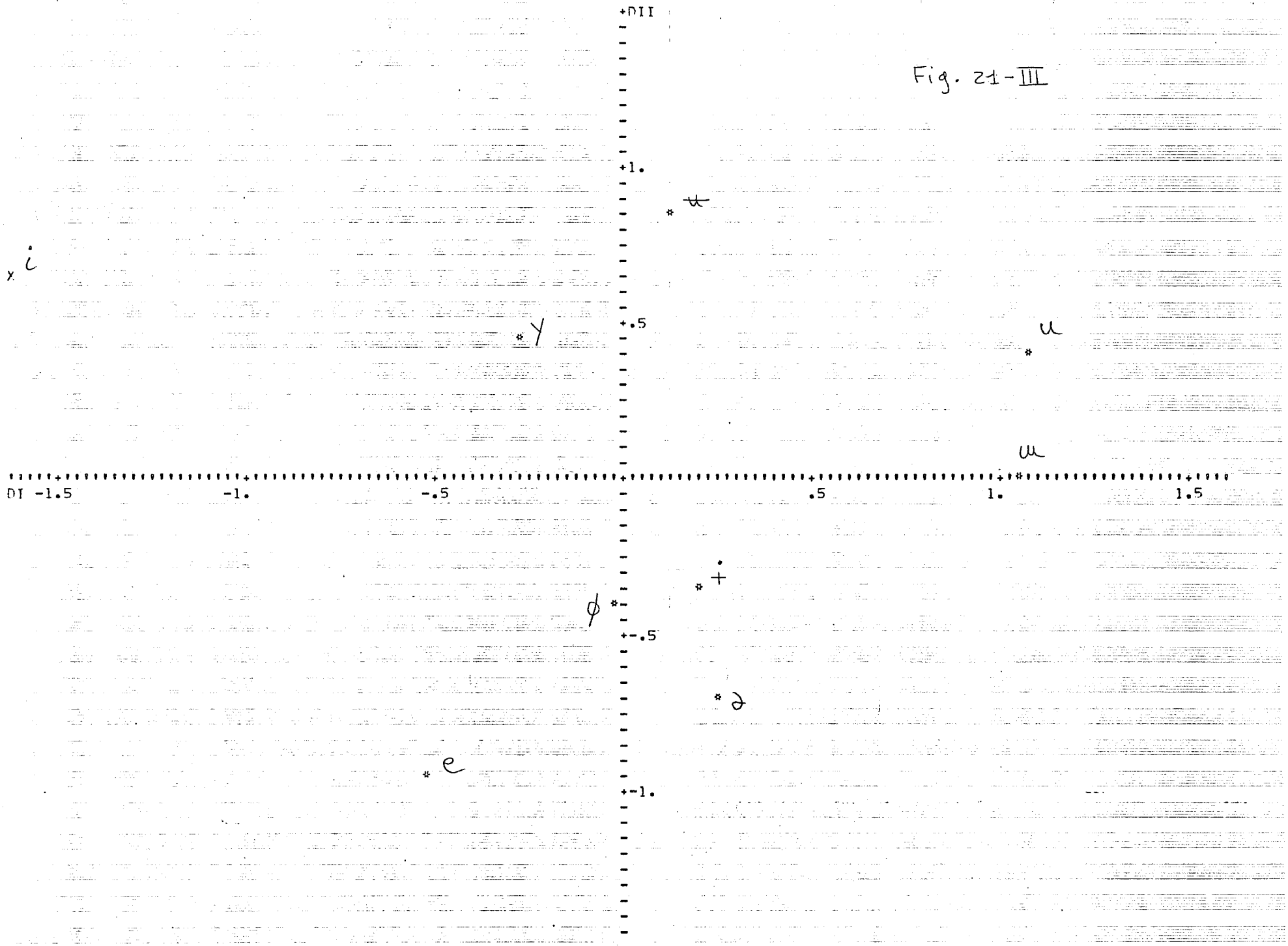
La première dimension explique un pourcentage de variance bien plus important que la deuxième, ce qui reflète le déséquilibre de la répartition des voyelles dans l'espace.

DI	DII	% var. expliquée
0.70	0.35	62 %

La figure 21-III présente les voyelles dans l'espace bidimensionnel.

On note peu de différence avec les résultats de BUTCHER. Mais il apparaît que la mauvaise répartition des voyelles dans l'espace exagère la distance entre [i] et [e]. Les voyelles fermées se répartissent en 3 groupes [i - ʉ - ʏ] . Les voyelles [ø ɨ ə] forment un groupe compact. Nous avons quelques doutes sur la validité de l'analyse acoustique. Un essai d'analyse en 3 dimensions s'est révélé impossible. Il semble donc que deux dimensions perceptives soient suffisantes. Ce résultat ne s'applique qu'à la population étudiée. Mais la relative correspondance avec les caractéristiques acoustiques suggère que celles-ci ont été utilisées bien que de nombreuses voyelles ne fassent pas partie de l'inventaire phonologique de la langue maternelle des sujets. L'accord entre les analyses mono- et multidimensionnelles n'est pas parfait. On notera surtout l'inversion de [œ] et [ø] par rapport à [ʏ] .

Fig. 21-III



L/ WILSON ET BOND (1977)

Le but de cette étude est double : rechercher les dimensions de l'espace perceptif pour 10 voyelles (anglais américain) et étudier l'effet de 4 contextes consonantiques sur le poids des dimensions. Notons dès à présent que cette deuxième partie nous paraît prématurée, devant l'imprécision des estimations subjectives de distances. L'outil nous paraît bien grossier pour une étude aussi "fine". Il n'est pas certain que l'analyse multidimensionnelle soit un instrument adéquat pour l'étude des variations dues à l'introduction d'un contexte. Il faudrait au préalable montrer que l'estimation des poids est stable pour les voyelles isolées, ce qui ne semble pas être le cas. 49 mots du type $C_1V\ C_2$ ont été enregistrés par 6 locuteurs masculins pour C_1 = /h/ ou /n/ et par 4 locuteurs pour C_1 = : /l/ ou /r/. C_2 était /n/ dans le cas d'une consonne initiale nasale et /d/ ailleurs. La séquence* [r ɹ] , impossible linguistiquement, a été omise. Une période répétée pendant 80 msec a été présentée à 20 sujets pour le contexte /h/ et /n/, et à 22 sujets pour la deuxième condition. Les matrices de confusions d'un test d'identification ont été analysées par la méthode INDSCAL.

Pour les auteurs, trois dimensions sont nécessaires pour rendre compte de la perception des voyelles. [ɹ] exige une quatrième dimension. Le pourcentage de variance expliquée s'élève à 87.7 % pour les trois dimensions, et à 89 % pour 4. Deux d'entre elles correspondent à F1 et F2 ou F'2 comme dans les études précédentes. La troisième serait une dimension "antérieure/postérieure" pour les voyelles ouvertes . Cette dernière regroupe les voyelles de la manière suivante :

ɛ æ ɪ ʊ u i ʌ ɑ ɔ

Son poids n'est pas négligeable (30 % de la variance expliquée). Elle sépare les voyelles ouvertes [ɛ] et [æ] de [ɑ] et [ɔ] L'analyse est difficile dans la mesure où les matrices de confusions ne sont pas publiées. L'emploi de matrices de confusions suggère néanmoins un artefact (cf. fig. 22-III).

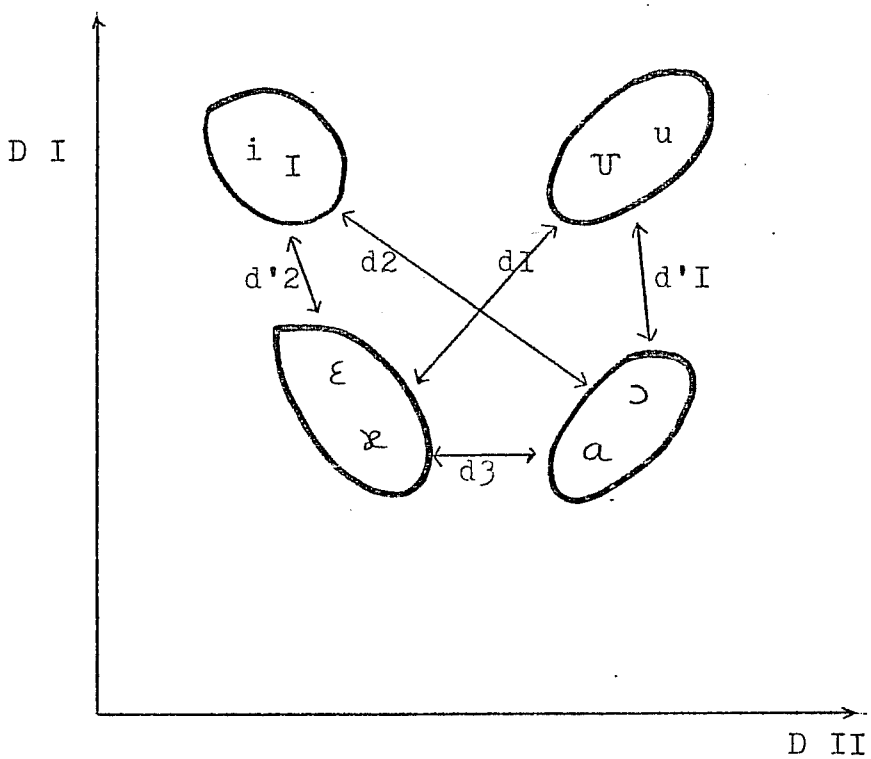


Fig. 22 - III

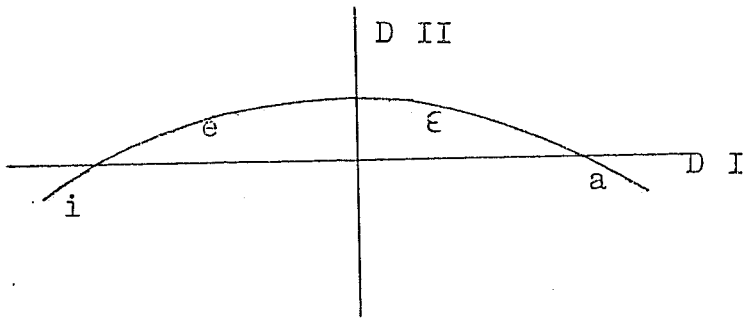
Il suffit de supposer que le nombre de confusions est approximativement le même pour [ɛ-ʊ, ɛ-u, ɜ-ʊ, ɜ-u] comparé à [ɔ-I, ɔ-i, a-I, a-i] (d_1 et d_2 fig. 22-III). Dans ce cas [ɛ,ɜ] et [ɔ,a] se trouvent à égale distance de [i,I] et de [u,ʊ] Si l'on suppose de plus que les confusions entre [ɛ,ɜ] et [I,i] ainsi qu'entre [u,ʊ] et [a,ɔ] ne sont guère plus importantes, ceci explique géométriquement que [ɛ,ɜ] et [ɔ,a] sont très proches dans le plan F1-F2. ($d_1 = d_2$; $d'_1 = d'_2$) Si enfin les confusions entre [ɛ,ɜ] et [a,ɔ] ne sont pas très grandes (d_3), cette dernière information est incompatible avec les précédentes. L'analyse multidimensionnelle ne peut résoudre ce conflit qu'en établissant une 3ème dimension pour séparer ɛɜ et a,ɔ.

Ce qui précède implique à nouveau que l'estimation des distances à partir des matrices de confusions provoque un artefact majeur. De plus, il faut noter qu'il suffit de quelques réponses "aberrantes" pour changer de façon appréciable la distance mesurée. L'absence de matrices de confusions ne permet pas de vérifier notre hypothèse. Mais les résultats précédents, ainsi que ceux que nous verrons plus loin, montrent que cette dimension n'apparaît nulle part ailleurs. Nous portons donc un jugement très réservé sur la réalité de cette troisième dimension. En ce qui concerne l'influence du contexte consonantique, l'absence des résultats détaillés ne permet aucune conclusion. Mais l'incertitude sur le nombre même des dimensions suffit, à nos yeux, pour affirmer qu'une telle analyse est prématurée.

M/ TERBEECK (1977)

Le travail de D. TERBEECK (1977), que nous avons reçu lorsque ce chapitre était en cours de rédaction, tire des conclusions très différentes des nôtres et mériterait de ce fait une analyse détaillée. TERBEECK s'est heurté à l'ensemble des difficultés que nous avons rencontrées, mais a choisi, dans la plupart des cas, des interprétations différentes. Critiquer en quelques lignes une thèse de 270 pages d'une densité exceptionnelle n'est pas commode. Mais nous

sommes persuadés que les conclusions de TERBEECK sont erronées et nous voudrions esquisser une démonstration. Ceci nous semble d'autant plus urgent que des phonologues peu avertis pourraient utiliser les résultats de TERBEECK pour présenter un système de traits distinctifs vocaliques qui serait "fondé" sur des données expérimentales. Beaucoup de résultats quantitatifs que nous utilisons pour juger de la validité des résultats ne sont pas présentés dans cette thèse. Il conviendrait de faire une analyse multidimensionnelle sur les données brutes, ce que le temps ne nous a pas permis de faire. L'analyse "globale" porte sur 10 voyelles naturelles [i, e, æ, a, o, u ø y ɨ ɤ] prononcées par un sujet américain et soumises au jugement de 35 sujets de 6 langues maternelles différentes (allemand, anglais, américain, turc, suédois et thaï). L'acceptabilité des échantillons n'a pas été testée et comme le reconnaît honnêtement l'auteur, il n'est pas sûr que les symboles phonétiques correspondent à la catégorisation que les sujets ont pu faire. L'analyse utilise la méthode PARAFAC (HARSCHMAN 1970) qui selon l'auteur, est "équivalente" à la méthode INDSCAL. Dans la présentation des résultats, la somme des carrés des poids a été normalisée (à 1), ce qui ne permet plus de retrouver l'estimation de la variance expliquée pour chaque dimension. Comme chez POLS et al. (1969), l'acquisition des données a été faite par la méthode des comparaisons triadiques. On a vu que celle-ci conduit à une sous-estimation systématique des distances les plus importantes.
Si, par hypothèse, [i, e, ɛ] et [a] sont sur une droite de l'espace perceptif, la sous-estimation de la distance [i - a] par rapport à la somme des distances [i-e , e-ɛ , ɛ-a] conduira les voyelles à se placer sur une courbe. Celle-ci (cf. schéma ci - après) ne peut être analysée que par deux dimensions, ce qui provoque l'apparition d'une dimension artefactuelle. TERBEECK note à de nombreuses reprises que les voyelles semblent se placer sur de telles courbes lorsqu'il augmente le nombre de dimensions.



Sur notre schéma, la dimension I rend compte d'une très grande partie de la variance. L'artefact ne se manifestera que si l'on cherche à extraire l'intégralité de la variance, c'est-à-dire si l'on augmente le nombre de dimensions.

Le chapitre 2 du travail de TERBEECK présente une nouvelle analyse de certaines études antérieures. Les analyses ont été discutées dans les paragraphes correspondants de ce chapitre.

L'auteur cherche à démontrer que 6 dimensions sont nécessaires pour décrire l'espace perceptif. Ce parti pris en faveur d'un grand nombre de dimensions s'explique en partie par la volonté de comparer les dimensions aux traits distinctifs de la phonologie. L'examen du choix des critères retenus par TERBEECK pour estimer le nombre de dimensions appelle les remarques suivantes :

- Les dimensions au-delà de la deuxième n'apportent qu'une très faible contribution à l'accord entre les distances initiales et les distances de l'espace (critère de "fit" p.146).

- Au-delà de la troisième dimension, certains poids deviennent négatifs.

- Les voyelles ne se répartissent pas en "nuage" sur les six dimensions, mais sur trois courbes particulières.

Ce comportement des voyelles dans les espaces à plus de trois dimensions nous semble la preuve que les dimensions supérieures à la troisième sont des artefacts. L'examen des six dimensions révèle les points suivants :

- La dimension I sépare les voyelles postérieures des voyelles plus antérieures [i, e, æ, y, ø, ɨ] . La position de la voyelle [ɤ] est intermédiaire.

- La dimension II est essentiellement la même que la première. [ɨ] , néanmoins, a rejoint le groupe postérieur, alors que [ɤ] est passé à l'extrêmité de l'axe. Les voyelles se placent sur une courbe, qui, redressée, présente les voyelles dans l'ordre : [a, o, u, ɨ, æ, i, e, y, ø, ɤ]

Nous proposons, en accord avec les conclusions de SINGH et WOODS (1971), une interprétation différente de celle de TERBEECK : la position de la voyelle suggère une dimension rétroflexe/non rétroflexe.

- La troisième dimension oppose en trois groupes distincts, les voyelles d'aperture extrême (ouvertes ou fermées) aux voyelles d'aperture moyenne :

[i y u] [a æ] [ø e i ɤ o]

- La quatrième dimension oppose les voyelles fermées ou d'aperture moyenne aux voyelles ouvertes en deux groupes :

[æ a] [o ɤ i e ø u y i]

A nouveau, les voyelles se répartissent sur une courbe dans l'espace des dimensions DIII - DIV. L'ordre des voyelles est le suivant :

[æ a o ɤ ɨ e ø i y u]

Cet ordre suggère une dimension corrélant fortement avec la fréquence du premier formant.

- Les deux dernières dimensions opposent les voyelles centrales aux voyelles périphériques,

[ø ɨ ɤ y u æ a i o e]

et les voyelles arrondies aux voyelles non-arrondies :

[e a æ ø ɨ ɤ o u y]

Sur la courbe formée par ces deux dimensions dans l'espace, l'ordre est le suivant :

[e i a æ ø ɨ ɤ y o u]

Cet axe unique sépare les voyelles antérieures non-arrondies, les voyelles centrales (à l'exception de [y]) et les voyelles postérieures ainsi que [y] . L'influence de F2 est certaine, bien que l'ordre ne soit pas identique à l'étagement fréquentiel de ce dernier. Notre interprétation est confirmée par la description de la position des voyelles pour les analyses de 1 à 6 dimensions. La fig. 23-III a été tracée à partir des figures publiées par l'auteur p. 149-153. Nous avons retenu une interprétation en trois dimensions. Pour les deux premières, l'ordre est bien conforme aux fréquences formantiques F1 et F2, avec cependant une distorsion importante pour [i] et [e]. La troisième dimension oppose les voyelles [ɤ ɨ ɸ] . Il est vraisemblable que les voyelles [ɸ, ɨ] ont été rapprochées des rétroflexes sur la base d'une similitude acoustique.

Notre interprétation des données de TERBEECK nous semble plus satisfaisante que celle de l'auteur pour les raisons suivantes :

- les dimensions extraites sont conformes aux résultats obtenus par les autres chercheurs.
- aucun poids n'est négatif.
- les dimensions de rang supérieur n'apportent qu'une faible augmentation de la variance expliquée.
- la notion de dimension "courbe" n'est plus nécessaire : celle-ci est sans doute liée à un artefact provoqué par la méthode employée.

Nous ne passerons pas en revue les résultats pour les différentes familles linguistiques. Les dimensions proposées, dont le nombre varie entre 3 et 5 selon la langue, sont d'une interprétation effroyablement complexe. Il est certain qu'une réduction du nombre de dimensions donnerait des résultats plus interprétables.

En conclusion, le travail de TERBEECK n'apporte aucune preuve de l'existence de plus de trois dimensions. Notre interprétation montre au contraire que les résultats obtenus sont parfaitement compatibles avec ceux des études antérieures.

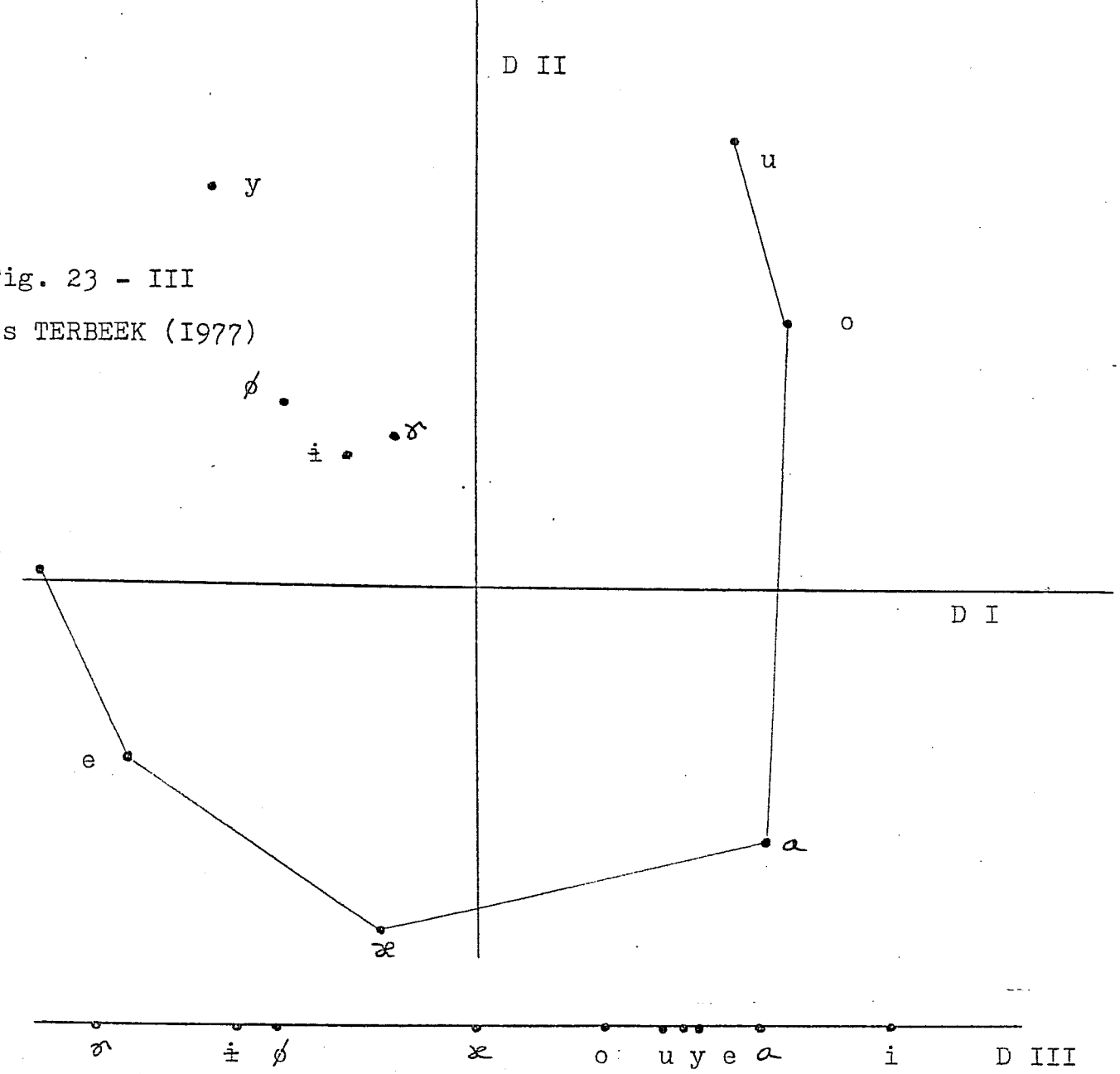


Fig. 23 - III

d'après TERBEEK (1977)

N/ BEECKMANS (1974)

La première analyse multidimensionnelle portant sur le français a été réalisée par BEECKMANS (1974). Cette étude analyse l'influence de la fréquence fondamentale sur l'espace perceptif. Ce problème étant relativement marginal par rapport au propos de ce chapitre, nous nous contenterons d'analyser l'espace vocalique global. 49 stimuli synthétiques, produits à l'aide de l'analogue du conduit vocal de l'E.N.S.E.R. de Grenoble, ainsi que des voyelles naturelles prononcées par deux locuteurs, l'un masculin et l'autre féminin, ont été utilisés. Les 49 stimuli synthétiques se composaient des 7 voyelles [i e ɛ a ɔ o u] à 7 fréquences fondamentales différentes : 80, 100, 120, 150, 200, 250 Hz. On notera l'absence de voyelles antérieures arrondies. La durée était de 500 msec. Les fréquences F1 et F2 étaient copiées sur celles du locuteur masculin.

6 répétitions de ces 63 voyelles ont été présentées à 8 sujets dans une tâche d'identification. Le pourcentage de réponses correctes diminue avec l'augmentation de F_0, sauf pour [a] et [i]. Le taux de reconnaissance global ne dépasse pas 78 % pour le locuteur masculin. La matrice de confusions a été analysée par la méthode INDSCAL. Le nombre de confusions (c) pour chaque paire a été transformé en distance relative par la formule $d' = -c$. Ces distances relatives sont ensuite converties en distances absolues par $d = d' + a$, où a est une constante additive. Celle-ci a été prise égale à 48, de telle manière que le nombre maximal de confusions corresponde à une distance nulle. Le terme de constante additive dans ce cas est un abus de langage. Cette méthode de calcul ne correspond en rien à la méthode présentée dans la première partie de ce chapitre. Les distances absolues ont été rendues symétriques par calcul de la moyenne arithmétique $Djk = Dkj = (djk + dkj)/2$, où j et k sont les indices de 2 voyelles quelconques.

On ne constate pas de différence entre l'analyse des voyelles naturelles et synthétiques. L'accroissement du pourcentage de variance expliquée en fonction du nombre de dimensions est le suivant :

nombre de dim.	% var. expl.	accroissement relatif
1	33	33
2	51	18
3	70	19
4	81	11
5	90	9

Au-delà de la troisième dimension, l'accroissement n'est pas supérieur à l'accroissement théorique pour des données aléatoires. Le nombre de dimensions est donc, au maximum, égal à 3, comme le note également BEECKMANS.

L'interprétation des dimensions I et II est parfaitement claire. Celles-ci correspondent à F1 et F2 (ou F'2). L'ajustement par la méthode des moindres carrés de l'espace perceptif et de l'espace physique F1-F2 est excellent (op. cit. fig. 9). La qualité de l'accord élimine F2-F1 comme base physique du corrélat psychologique. Il est dommage que l'ajustement n'ait pas été tenté avec F'2 à la place de F2.

Mais pour la troisième dimension "aucune explication n'apparaît en termes de variables physiques" (p. 124). BEECKMANS note que cette troisième dimension "peut être due à la déformation des informations au passage des données de confusions aux données de distances". Le calcul de la constante additive pourrait être la source de cette distorsion. Le rôle de cette troisième dimension est d'accroître la distance entre les voyelles extrêmes [i], [a] et [u]. Ceci pourrait indiquer que ces distances ont été sous-estimées par l'utilisation des matrices de confusions ou par le calcul des distances. Cette sous-estimation semble se retrouver dans un certain nombre d'études multidimensionnelles (TEERBECK 1977, par ex.). De plus, toutes les études utilisant les matrices de confusions semblent présenter une dimension supplémentaire artefactuelle. L'absence de la matrice de confusions ne permet pas dans ce cas de vérifier l'hypothèse suivante : si le nombre de confusions entre [i] et [a], et [i] et [ɛ] est faible, les

distances calculées seront quasiment identiques. Ceci implique que [a] et [ɛ] sont très proches. Cette information est en conflit avec le nombre de confusions entre [a] et [ɛ], si celui-ci n'est pas très élevé. Ce conflit ne peut se résoudre qu'en plaçant ces trois voyelles sur une dimension supplémentaire qui "prend en charge" ce biais apparent. Cette hypothèse s'applique à toutes les triades de voyelles.

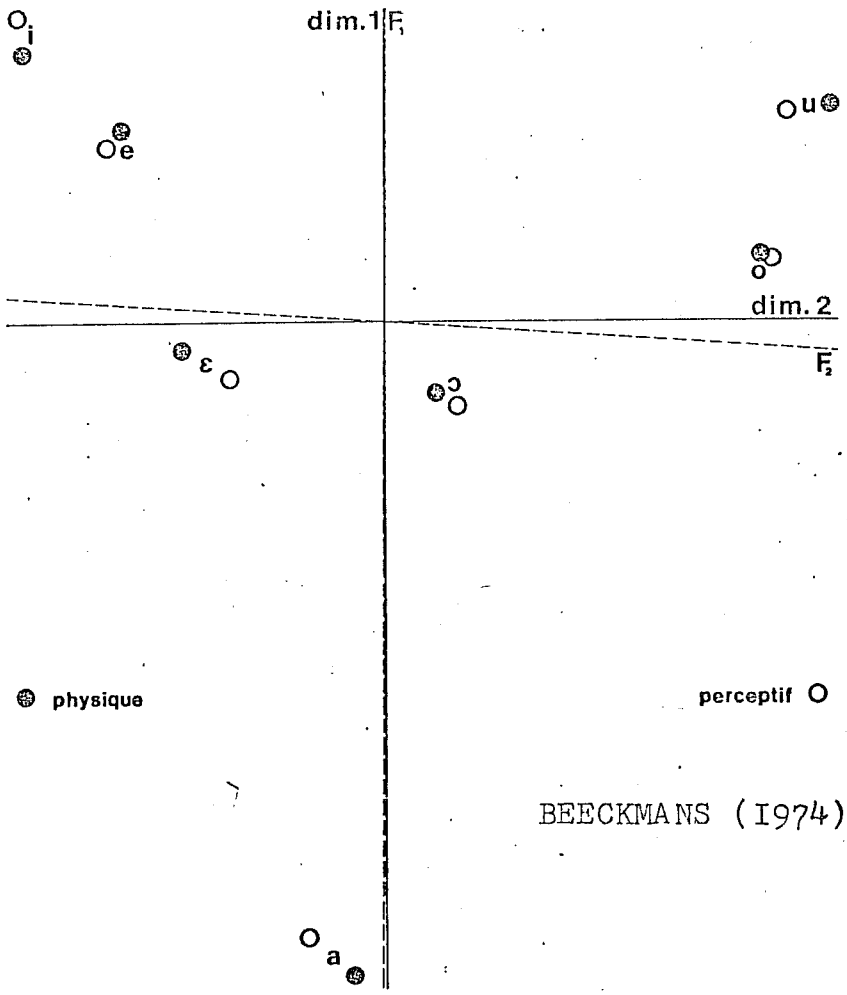
En conclusion, cette étude sur un échantillon incomplet de voyelles françaises confirme la présence de deux dimensions perceptives qui sont des transformations simples des variables physiques. Des doutes sérieux pèsent sur l'existence d'une troisième dimension.

0/ BEECKMANS (1977)

Dans son introduction, l'auteur insiste, à juste titre, sur le fait que ".. si la configuration perceptive à 2 dimensions peut être mise en rapport avec le plan formantique F1-F2,..., on ne peut affirmer l'hypothèse formantique selon laquelle ces valeurs seraient extraites... et utilisées comme paramètres de décision. La même configuration globale serait obtenue par n'importe quel processus perceptif utilisant une quelconque combinaison linéaire des deux premiers formants" (p. 58, op. cit.). Mais si cela est vrai, on ne voit pas comment on peut tester l'hypothèse d'une combinaison linéaire, du moins avec une méthode d'analyse multidimensionnelle. Une simple notion de simplicité nous conduit dans ce cas à proposer F1 et F2 seuls. De plus, comme nous le verrons au chapitre suivant, un modèle de perception prévoyant l'extraction de F1 et F2 (ou F'2) est compatible avec une large gamme de résultats provenant d'expériences diverses. Les scrupules de BEECKMANS, s'ils sont fondés en théorie, ne nous semblent pas importants en pratique. En revanche, la remarque que nous avons faite précédemment est primordiale : la non-indépendance de F2, F3 et F4 ne permet pas de déterminer la base physique de la deuxième dimension.

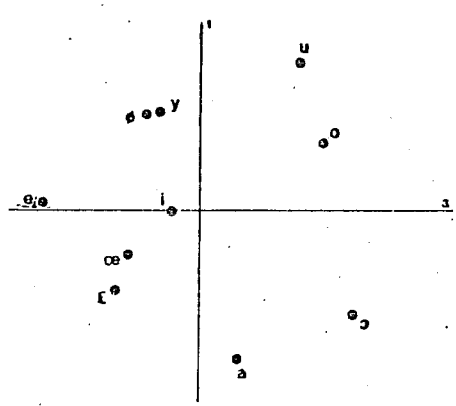
Le corpus vocalique utilisé dans cette expérience comprenait 10 voyelles françaises [i, e, ɛ, a, ɔ, o, u, y, ø, œ] prononcées par 10 locuteurs à une fréquence fondamentale contrôlée, mais différente, pour chacun . Après digitalisation sur ordinateur, 10 sujets ont identifié les sons vocaliques produits par 20 répétitions d'une même période. La méthode est très proche de celle employée par VAN DER KAMP et POLS (1971). Les mêmes critiques peuvent être portées : utilisation de matrices de confusions, non-représentativité de la voyelle après sélection d'une période unique, en raison des variations de la source vocale à court terme. La procédure de segmentation a, d'après l'auteur lui-même, un effet perturbateur sur la voyelle [i]. Le taux global de reconnaissance varie de 10 à 80 %. La longueur du segment, qui dépend de F_0, n'a pas d'effet sur le taux de reconnaissance. La matrice de confusions globale a été partitionnée, de 5 manières différentes, en 10 matrices correspondant à chaque locuteur, à chaque sujet, à chaque contexte objectif (voyelles précédentes), à chaque contexte subjectif (voyelles précédentes perçues effectivement par le sujet) et à l'ordre de passation. L'ensemble de ces partitions a été analysé par la méthode INDSCAL. Aucune indication n'étant donnée sur la transformation des indices en distances, on peut supposer que la technique présentée précédemment (BEECKMANS 1974) a été employée. L'influence des trois dernières conditions (contexte objectif, contexte subjectif et ordre de passation) ne se manifeste pas de façon spécifique pour chaque dimension extraite.

Pour la variable "sujet", l'analyse fournirait 3 dimensions pertinentes. La variance expliquée ne s'élève qu'à 58 %. Aucune discussion des variations individuelles n'est présentée. Les dimensions de rang supérieur n'apportent qu'une augmentation non-significative de la variance. L'auteur qualifie ce résultat de "médiocre". Il en attribue la responsabilité au nombre élevé de stimuli, à la difficulté de la tâche et à l'hétérogénéité des voyelles due à l'emploi de plusieurs locuteurs. La fig. 24 - III , qui reproduit la fig. 3a de l'article original présente la position des voyelles dans un espace



BEECKMANS (1974)

Fig.9



BEECKMANS 1977

Fig. 24 - III

bidimensionnel DI - DIII. La structure spatiale correspond grossièrement au plan formantique F1 - F2. La position de la voyelle [i] est aberrante. Comme l'écrit l'auteur "... il semble que les deux premiers formants soient les dimensions pertinentes pour l'ensemble du processus d'identification". La proximité de [ϕ, y] et de [ɛ, œ] pourrait s'expliquer par l'utilisation d'une période unique. Nous retrouvons là une difficulté que présente également l'étude de VAN DER KAMP et POLS (1971). BEECKMANS ne mentionne jamais la dimension II. Nous en déduisons que celle-ci est ininterprétable. Le pourcentage de variance expliquée associé à la représentation bidimensionnelle est sans doute largement inférieur à 58 %. Ce chiffre nous semble constituer une condamnation de la procédure expérimentale choisie par BEECKMANS. La fin de l'étude est une discussion de l'analyse INDSCAL en fonction des locuteurs. Les 10 matrices de confusions analysées contiennent les distances intervocaliques pour chaque locuteur séparément. On constate que F1 et F2 constituent également les dimensions perceptives pour cette condition. La structure de l'espace est très proche de celle obtenue dans la condition "sujet". Nous avouons être incapables de tirer une conclusion de ce résultat et de saisir le fil de la discussion de BEECKMANS. Nous ne voyons pas en particulier quel peut être le rapport de ce résultat avec l'hypothèse de normalisation discutée par BEECKMANS : ".. les sujets sont capables de s'adapter à un locuteur en ce sens qu'ils attribuent des poids différents à chacune des deux informations concernant F1 et F2 (p. 63)".

Il faudrait montrer que ces variations sont bien une adaptation, et non les phénomènes aléatoires que provoque toute expérience de perception. Or BEECKMANS ne présente pas de résultats détaillés concernant l'évolution des poids en fonction des locuteurs. De plus, il n'est pas certain qu'il soit possible de mettre une adaptation en évidence par une analyse multidimensionnelle. Si l'adaptation est constituée, en première approximation, par une multiplication de la fréquence formantique perçue par un facteur constant (inférieur

à l'unité pour les voix féminines par exemple), celle-ci ne changera pas le poids des dimensions perceptives.

En conclusion, nous ne pensons pas qu'il soit possible de tirer une quelconque conclusion de cette expérience en ce qui concerne le processus de normalisation. Plus généralement, cette étude confirme que les seules dimensions interprétables sont liées à F1 et F2 et que l'utilisation des matrices de confusions ne donne pas de résultats satisfaisants : le taux de variance expliquée est faible; une dimension artefactuelle apparaît sans doute.

RECAPITULATION DES PRINCIPAUX RESULTATS

L'étude critique et les nouvelles analyses auxquelles nous nous sommes livrés sur les 15 analyses multidimensionnelles publiées permettent de tirer dès à présent quelques conclusions que nous tenterons de vérifier dans une étude de l'espace perceptif des voyelles françaises.

La plus importante est qu'il n'existe pas d'indices sérieux de l'existence d'une troisième dimension, pour les systèmes vocaliques ne possédant pas de voyelles rétroflexes. L'espace perceptif des voyelles non-rétroflexes ne semble donc comporter que deux dimensions. Cette conclusion découle des constatations suivantes :

- la présence d'une dimension supplémentaire n'augmente pas de façon significative le taux de variance expliquée.
- la configuration des voyelles sur la troisième dimension n'est pas stable. Elle dépend de l'ordre choisi arbitrairement par l'expérimentateur pour l'initialisation de l'algorithme.
- il n'est pas possible d'exclure la possibilité d'un artefact provenant de la méthode d'estimation des distances subjectives.
- enfin, l'ordre des voyelles sur la troisième dimension ne correspond à aucun paramètre physique.

Le travail de SHEPARD indique que presque toute l'information contenue dans une matrice de confusions peut être décrite par la position des voyelles dans un espace à trois dimensions, dont l'une ne rend compte que de l'opposition rétroflexe - non-rétroflexe. Il est frappant de voir que toutes les études s'accordent pour conclure à l'existence de deux dimensions corrélant fortement avec F1 et F2. Peu d'auteurs soulignent que l'interprétation de la deuxième dimension comme corrélat perceptif de F2 n'est que l'une des possibilités que l'on peut envisager.

Un argument décisif pour notre interprétation serait obtenu si l'on montrait que l'analyse multidimensionnelle pouvait rendre compte de la quasi totalité de la variance avec deux dimensions seulement. Les résultats que nous allons présenter atteignent presque cet objectif.

V - ANALYSE MULTIDIMENSIONNELLE DE 10 VOYELLES FRANÇAISES

La préparation de cette expérience a tenu compte, dans la mesure du possible, des enseignements méthodologiques fournis par l'étude des travaux antérieurs. Le problème capital nous semble être la détermination du nombre de dimensions. L'interprétation des dimensions est un problème encore plus complexe auquel nous réserverons le chapitre IV. Cette interprétation ne peut s'effectuer que si l'on prend en compte des informations provenant d'expériences utilisant des techniques différentes.

A/ Choix des stimuli

10 voyelles françaises isolées ont été retenues : [i e ɛ a ɔ o u y ø œ]. Plusieurs séries de voyelles ont été enregistrées en studio par un phonéticien professionnel. Le choix du meilleur échantillon a été fait par deux phonéticiens sur la base du timbre perçu. Le locuteur devait maintenir un fondamental fixe à 120 Hz à l'aide d'un son de référence. Les stimuli retenus ont été analysés au spectrographe.

Les fréquences formantiques et les durées sont présentées dans la table ci-dessous et à la fig. 25-III . Le point d'interrogation signifie une valeur douteuse, le tiret indique un formant non-visible.

	F1	F2	F3	F4 (Hz)	durée (msec)
i	250	2220	2930	3430	320
e	340	2050	2650	3450	310
ɛ	590	1780	2350?	——	360
a	885	1150	2410	3770	280
ɔ	640	890	2650	3650	280
o	410	650	2650	3300?	300
u	300	780	2500	3500	350
y	250	1780	2170	3180	320
ø	350	1300	2350	3250	300
œ	640	1410	2475	——	320
ə	610	1360	2375	3650	320 - 410

Aucune manipulation acoustique n'a été effectuée. L'intensité subjective en particulier n'a pas été égalisée. Mais on a pris soin de ne pas faire varier l'intensité des voyelles lors de la création de la bande test.

B/ Technique de présentation

Tous les couples de voyelles ont été présentés, en ordre aléatoire, à l'aide d'un magnétophone Revox A 77 et d'enceintes AR 2, en champ libre dans une salle non-traitée. L'intensité a été réglée de façon à être considérée comme satisfaisante par tous les sujets. Chaque couple était séparé du couple précédent par une paire [ə - ə] dans le but d'éviter tout effet d'ordre dû à la mémoire auditive. Les fréquences formantiques (identiques) et les durées de ces deux voyelles ont été présentées au bas de la table ci-dessus.
Les 10 échantillons ont été présentés à plusieurs reprises avant l'expérience pour familiarisation. La voix du locuteur était connue des sujets. Une pause très brève intervenait après la présentation de 11 couples.

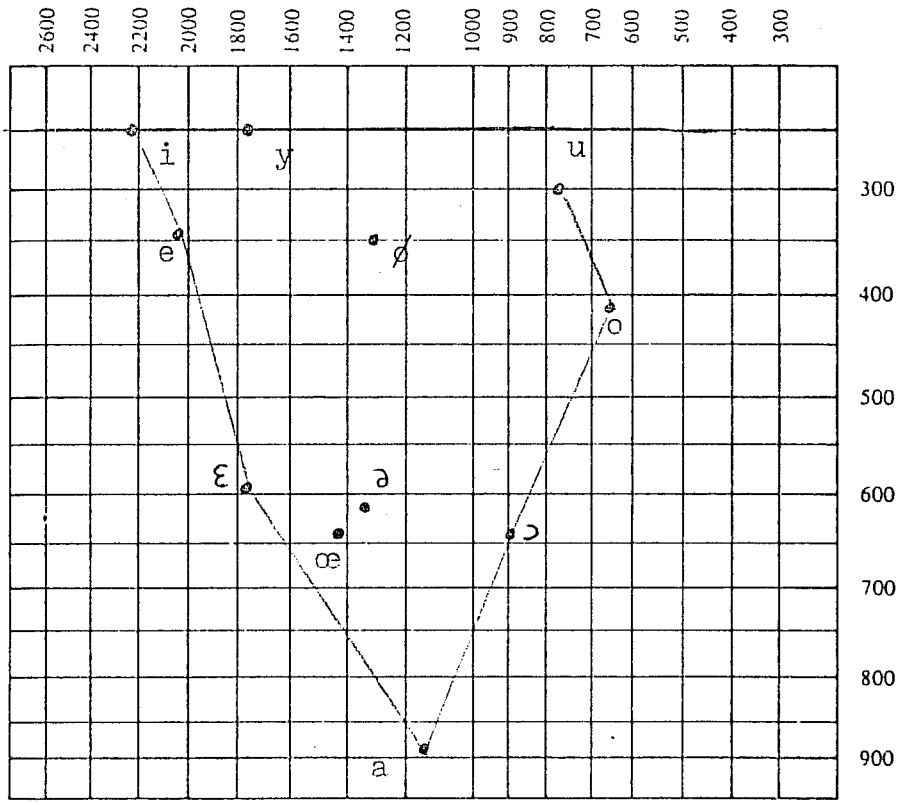

2600
2400
2200
2000
1800
1600
1400
1200
1000
900
800
700
600
500
400
300
300
400
500
600
700
800
900
i
y
u
e
ø
o
ɛ
ə
ɔ
œ
a

Fig. 25 - III

C/ Sujets

40 sujets, étudiant(e)s du certificat de phonétique générale ont été testés en deux groupes de taille égale. Ayant suivi un enseignement de transcription phonétique, ils étaient relativement entraînés à la reconnaissance des timbres vocaliques. L'influence éventuelle des connaissances de phonétique sera discutée plus bas. 4 sujets sur 40 n'étaient pas de langue maternelle française, tout en ayant une bonne pratique de celle-ci. Après examen des résultats et devant l'absence de caractéristiques particulières, les réponses de ces sujets ont été maintenues pour le calcul des moyennes de groupe.

D/ Estimation des dissimilarités

Dans les 5 secondes suivant la présentation de chaque couple, les sujets devaient cocher l'une des 7 cases d'une ligne de la feuille de réponses :

La case la plus à gauche devait être utilisée pour indiquer les stimuli "extrêmement semblables", la case la plus à droite étant réservée aux stimuli jugés "extrêmement différents". Les cases médianes étaient utilisées pour des couples aux caractéristiques subjectives moins extrêmes. Cette information était présentée sur un tableau visible pendant tout le temps de l'expérience. Les sujets étaient familiers de cette technique d'échelle* concernant des contours intonatifs (CARTON et LONCHAMP 1977). Les instructions ont donc été peu nombreuses ; elles concernaient l'utilisation de la feuille de réponses et la nécessité de répondre, dans un bref laps de temps, d'une façon spontanée. L'expérimentateur a souligné que l'expérience portait exclusivement sur le timbre, à l'exclusion de la durée ou de la hauteur des voyelles (concepts connus des sujets). De nombreuses questions ont été posées avant l'expérience, ce que nous considérons comme un signe de la motivation des sujets.

* d'intervalles, qu'ils avaient déjà utilisée lors de la collecte de données

L'expérimentateur a dû, en particulier, expliquer la consigne de "réponse obligatoire dans tous les cas, même au hasard". Les sujets n'étaient pas informés que les quinze premières paires étaient destinées à stabiliser les réponses et ne devaient pas être analysées.

E/ Traitement des données initiales

Le tableau des dissimilarités a été calculé en faisant la moyenne, pour chaque sujet, des couples présentant les mêmes voyelles dans un ordre différent. Pour le calcul des moyennes de groupe et l'estimation des constantes additives, les détails seront présentés dans la section traitant des résultats. Le programme INDSCAL utilisé est identique à celui que nous avons publié en 1974 (LONCHAMP 1974). Les résultats individuels sont présentés en annexe de ce chapitre.

F/ Résultats

a)- ANALYSE INDIVIDUELLE (sujet par sujet) -

Espace bidimensionnel.

Pour un espace à deux dimensions, sans estimation d'une constante additive, les coordonnées des voyelles sont les suivantes :

	Dim. I	Dim. II
i	-.62	.30
e	-.52	.01
ɛ	-1.17	-1.22
a	.90	-1.58
ɔ	1.44	-1.00
o	1.79	.51
u	1.30	.84
y	-.67	1.71
ø	-.12	.70
œ	-.32	-.50

Ces résultats sont présentés graphiquement à la fig. 26 -III . Le tableau I-III présente la valeur des poids pour chaque sujet, ainsi que le pourcentage de variance expliquée.

Le pourcentage moyen de variance expliquée est décevant (37,2 %). La moitié des sujets (20/40) obtiennent un pourcentage supérieur à 40 %. 3 sujets sont en-dessous du pourcentage minimum théorique (22 %). Pour ces derniers, une cause possible est une inversion des réponses "semblable" et "différent" sur la feuille de réponse. Décevante également est la configuration globale (fig. 26-III). Si la position des voyelles postérieures et de [ɛ] est relativement conforme à leur position sur le plan physique F1 - F2, on ne peut en dire autant pour [i] et [e] par rapport à [y], [o] et [ɛ]. Ces résultats indiquent une quantité considérable de "bruit" dans les réponses individuelles. Rappelons que le pourcentage de variance expliquée n'est pas le seul critère d'un bon sujet. Le sujet 40 possède le pourcentage le plus élevé : 67 %. Mais ce résultat est obtenu parce que le sujet estime les dissimilarités sur la seule dimension 'ouvert - fermé'. Sur les 67 %, seuls 1,7 % sont expliqués par la deuxième dimension. La simplification de la tâche explique le bon résultat. Les réponses du sujet s'inscrivent en fait dans un espace monodimensionnel. Le taux de 50 % obtenu par les sujets 13 et 14, avec des poids identiques sur les deux dimensions est bien préférable.

Sans vouloir discuter dès maintenant l'interprétation de cette configuration des points, on peut noter que la position de [y] est la plus éloignée de ce que le plan F1 - F2 permet de prévoir.

Deux raisons pourraient expliquer ces résultats décevants :

- l'absence d'une correction de constante additive ;
- un nombre de dimensions trop faible.

Nous avons testé successivement ces deux hypothèses.

	poids Dim. I	poids Dim. II	% var. expl.
Suj.I	.46	.56	.53
2	.05	.45	.21
3	.23	.47	.27
4	.48	.30	.32
5	.22	.45	.26
6	.15	.32	.12
7	.28	.41	.25
8	.46	.39	.36
9	.30	.47	.31
I0	.38	.30	.23
II	.14	.47	.24
I2	.39	.37	.29
I3	.51	.49	.50
I4	.53	.47	.50
I5	.30	.48	.32
I6	.37	.52	.41
I7	.28	.42	.26
I8	.30	.41	.26
I9	.36	.52	.40
20	.40	.50	.41
2I	.49	.47	.46

	poids Dim. I	poids Dim. II	% var. expl.
22	.25	.46	.28
23	.30	.62	.48
24	.29	.29	.16
25	.46	.45	.42
26	.39	.51	.41
27	.14	.52	.30
28	.47	.44	.41
29	.62	.39	.54
30	.53	.42	.45
3I	.50	.49	.49
32	.30	.65	.51
33	.39	.50	.41
34	.43	.36	.31
35	.27	.41	.24
36	.63	.29	.49
37	.38	.50	.40
38	.28	.76	.65
39	.27	.57	.39
40	.13	.81	.67

Table I - III

b)- ANALYSE INDIVIDUELLE - Estimation d'une constante additive - Espace bidimensionnel.

Les constantes additives individuelles sont présentées à l'annexe A de ce chapitre, avec les résultats individuels. Les coordonnées des voyelles sur les deux dimensions sont les suivantes :

	Dim. I	Dim. II
i	-.60	.28
e	-.47	.01
ɛ	-1.14	-1.22
a	.90	-1.57
ɔ	1.45	-1.02
o	1.77	.52
u	1.32	.85
y	-.73	1.70
ø	-.19	.71
œ	-.37	-.49

Ces résultats sont pratiquement identiques à ceux obtenus sans constante additive. A l'échelle de la fig. 26-III les deux configurations ne peuvent être distinguées. Les variations ne dépassent jamais 2 %. On constate donc que l'introduction d'une constante additive n'a aucun effet sur la forme de la configuration. On note une légère augmentation (1,1 %) de la variance expliquée moyenne, qui atteint 38,3 %. La moitié des sujets ont maintenant un pourcentage égal ou supérieur à 41 %. (table II-III). Deux sujets ont toujours un score inférieur au minimum théorique. La transformation des indices en distances ne semble jouer aucun rôle dans les résultats de l'analyse individuelle.

La première conclusion que l'on puisse tirer est que la configuration des points ne reflète que grossièrement le plan F1 - F2. Il apparaît néanmoins que ces deux paramètres physiques jouent un rôle dans l'établissement de l'espace perceptif.

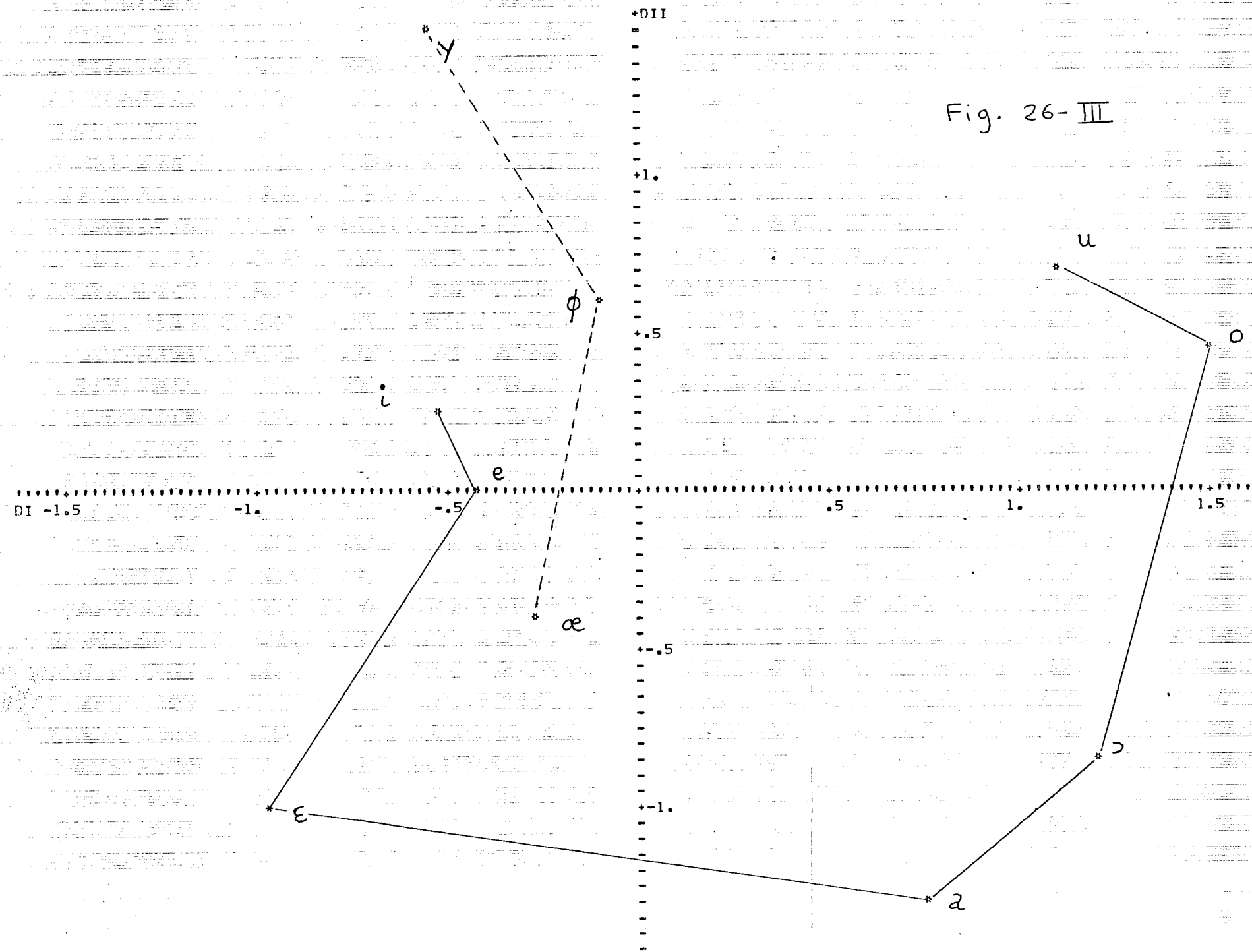


Fig. 26- III

	Poids DIM. I	Poids DIM 2	% var. expl.		Poids DIM I	Poids DIM 2	% var. expl.
SI	.47	.55	.53	S22	.31	.49	.33
S2	.08	.46	.22	S23	.32	.60	.47
S3	.27	.47	.30	S24	.31	.32	.20
S4	.48	.32	.34	S25	.47	.45	.42
S5	.24	.46	.27	S26	.39	.50	.41
S6	.21	.36	.17	S27	.15	.52	.30
S7	.28	.41	.24	S28	.47	.45	.42
S8	.46	.38	.36	S29	.62	.39	.53
S9	.30	.47	.31	S30	.53	.43	.46
SI0	.38	.32	.25	S3I	.50	.49	.49
SII	.21	.51	.30	S32	.32	.63	.50
SI2	.42	.38	.32	S33	.40	.51	.41
SI3	.51	.48	.50	S34	.43	.38	.33
SI4	.52	.47	.49	S35	.28	.42	.25
SI5	.32	.48	.34	S36	.63	.29	.48
SI6	.39	.55	.45	S37	.40	.50	.42
SI7	.27	.42	.25	S38	.30	.74	.64
SI8	.30	.41	.26	S39	.27	.58	.41
SI9	.37	.50	.39	S40	.15	.81	.67
S20	.41	.50	.42				
S2I	.50	.47	.47				

Table II - III

Pour vérifier l'hypothèse d'un nombre insuffisant de dimensions, nous avons procédé à une nouvelle analyse dans un espace à trois dimensions.

c)- ANALYSE INDIVIDUELLE - Espace tridimensionnel avec constante additive.

Les coordonnées des voyelles dans un espace tridimensionnel, après correction de constante additive, s'établissent ainsi :

	DIM. I	DIM. II	DIM. III
i	-.06	-.02	.00
e	.19	.06	.02
ɛ	-.68	-.50	1.78
a	.78	-1.70	.67
ɔ	1.23	-1.25	.51
o	.97	-.40	-1.64
u	1.42	.21	-1.39
y	.65	2.18	.08
ø	-1.47	.41	-.91
œ	-1.45	-.35	.62

Ces résultats sont présentés graphiquement à la fig. 27-III . La table III-III , qui présente les poids et le pourcentage de variance expliquée indique clairement que l'introduction d'une troisième dimension ne diminue en aucune manière la variance résiduelle. Au contraire, le taux de variance expliquée moyen diminue (35,25 %), pour se retrouver proche du minimum théorique (33 %). Le nombre de dimensions est donc supérieur au nombre de dimensions mathématiquement nécessaire. De ce fait, la répartition des points n'est plus unique dans l'espace ; elle dépend, de façon aléatoire, du "bruit" affectant les données initiales. L'examen de la configuration obtenu confirme cette analyse. La correspondance entre les deux premières dimensions et le plan F1 - F2 est encore plus mauvaise que dans l'analyse précédente. Les voyelles [y, ø, œ] sont placées largement à l'extérieur de la figure

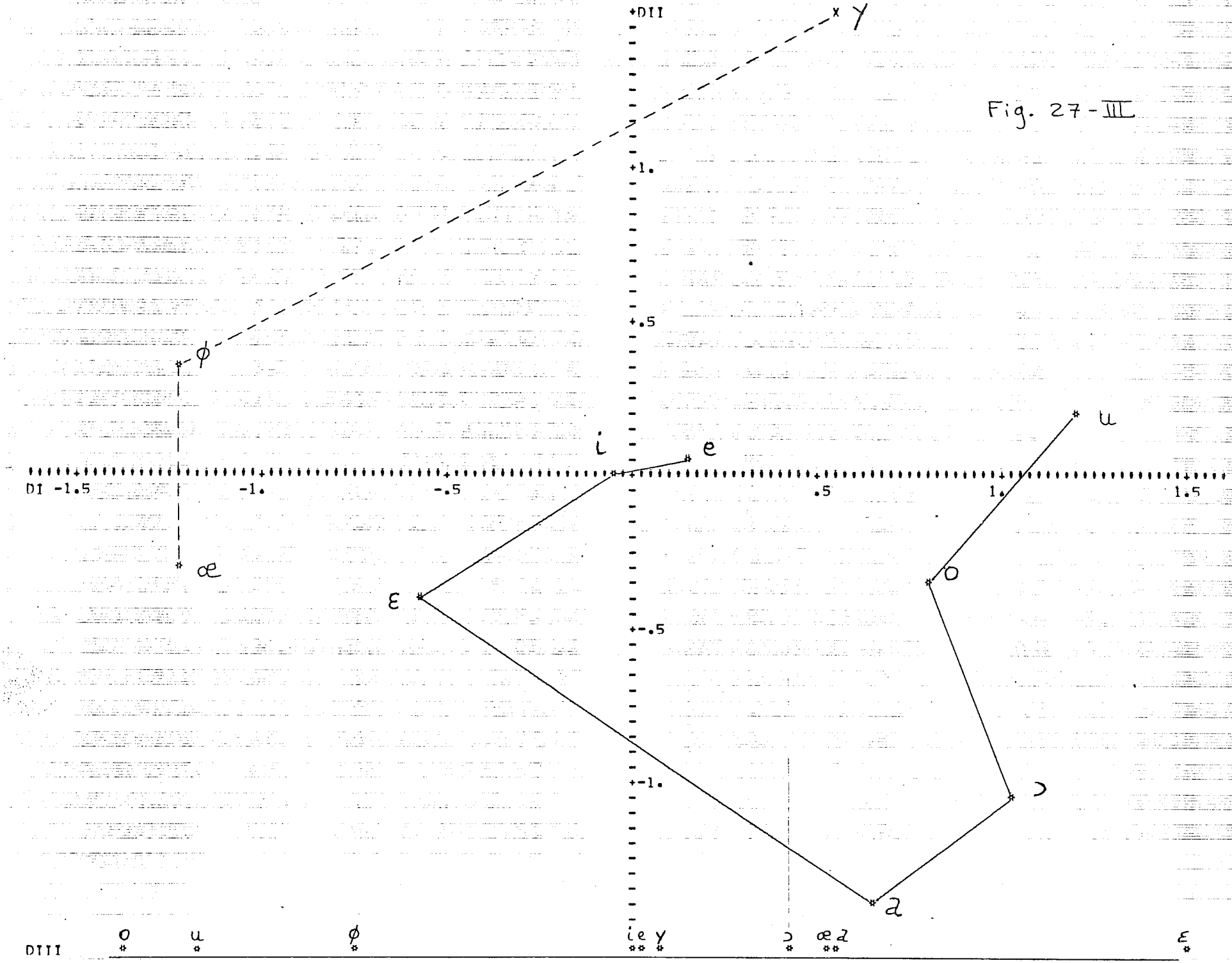


Fig. 27-III
+DII
+1.
+.5
+-.5
+-1.
DI -1.5
-1.
-.5
.5
1.
1.5
DIII
y
ø
i
e
u
œ
ɛ
o
ɔ
a
o
u
ø
ie y
ɔ
œ a
ɛ

Poids	DI	D2	D3	% var.
SI	.24	.47	.35	.40
S2	.05	.35	.35	.25
S3	.15	.36	.40	.31
S4	.17	.37	.22	.21
S5	.25	.34	.30	.26
S6	.42	.23	.23	.28
S7	.26	.38	.30	.31
S8	.32	.35	.37	.36
S9	.26	.38	.33	.32
SI0	.18	.23	.32	.19
SII	.30	.36	.37	.35
SI2	.41	.36	.21	.35
SI3	.20	.41	.41	.38
SI4	.25	.44	.43	.44
SI5	.45	.34	.37	.45
SI6	.32	.45	.31	.40
SI7	.12	.29	.32	.20
SI8	.34	.28	.38	.34
SI9	.27	.37	.45	.42
S20	.41	.45	.30	.46
S2I	.28	.52	.27	.42

	DI	D2	D3	% var.
S22	.23	.41	.24	.28
S23	.11	.54	.24	.37
S24	.53	.20	.24	.38
S25	.23	.37	.32	.30
S26	.25	.37	.37	.34
S27	.09	.40	.30	.26
S28	.22	.42	.34	.34
S29	.43	.40	.28	.42
S30	.23	.42	.27	.31
S3I	.33	.44	.30	.39
S32	.20	.48	.42	.44
S33	.14	.42	.41	.36
S34	.20	.37	.33	.29
S35	.15	.19	.59	.40
S36	.25	.36	.33	.31
S37	.28	.43	.30	.35
S38	.16	.52	.47	.52
S39	.15	.35	.45	.35
S40	.05	.49	.59	.59

Table III - III

formée par les 7 voyelles restantes. Quant à la répartition des voyelles sur la troisième dimension, elle ne semble pas interprétable. Il est clair que la structure spatiale nécessaire pour représenter les données initiales ne comporte pas plus de deux dimensions.

Si, comme nous le pensons, le faible taux de variance expliquée est dû au "bruit" et non à l'utilisation par différents sujets de dimensions spécifiques, ou de dimensions particulières à un très petit nombre de stimuli, on devrait observer une augmentation importante de ce taux pour l'analyse des indices moyens de groupe.

d)- ANALYSE DES MOYENNES DE GROUPE - Espace bidimensionnel.

Il s'agit en fait de réaliser une analyse INDSCAL sur les moyennes des distances issues des indices de dissimilarité. Le calcul de cette moyenne n'est pas une opération élémentaire, car il faut tenir compte de la différence de moyenne et d'écart type des distributions individuelles. Nous avons tout d'abord normalisé les distributions de chaque sujet en les ramenant à une distribution de moyenne nulle et de variance unité. Puis nous avons calculé l'indice de dissimilarité moyen en faisant la moyenne des estimations pour les 40 sujets. Chaque indice de dissimilarité est donc la moyenne de 80 mesures, alors que, pour l'analyse individuelle, les indices individuels n'étaient calculés qu'à partir de deux mesures. Nous avons enfin corrigé ces nouveaux indices (de moyenne nulle) en ajoutant une constante égale à la moyenne globale des indices. Ces opérations ont pour but d'éliminer le biais personnel des sujets et d'égaliser la dispersion des indices. La procédure la moins satisfaisante concerne la normalisation des variances. Nous verrons plus bas une procédure plus adéquate. Nous avons ensuite vérifié que ces indices vérifiaient les propriétés des distances (inégalité triangulaire). Aucune constante additive ne s'est avérée nécessaire.

On sait que l'analyse d'un sujet unique ne garantit pas l'orientation correcte des axes. Aussi avons-nous conservé l'orientation obtenue lors des analyses individuelles. On constatera plus loin que la nouvelle configuration est directement interprétable sans rotation.

L'analyse de ces nouvelles données moyennes dans un espace à deux dimensions donne les résultats suivants :

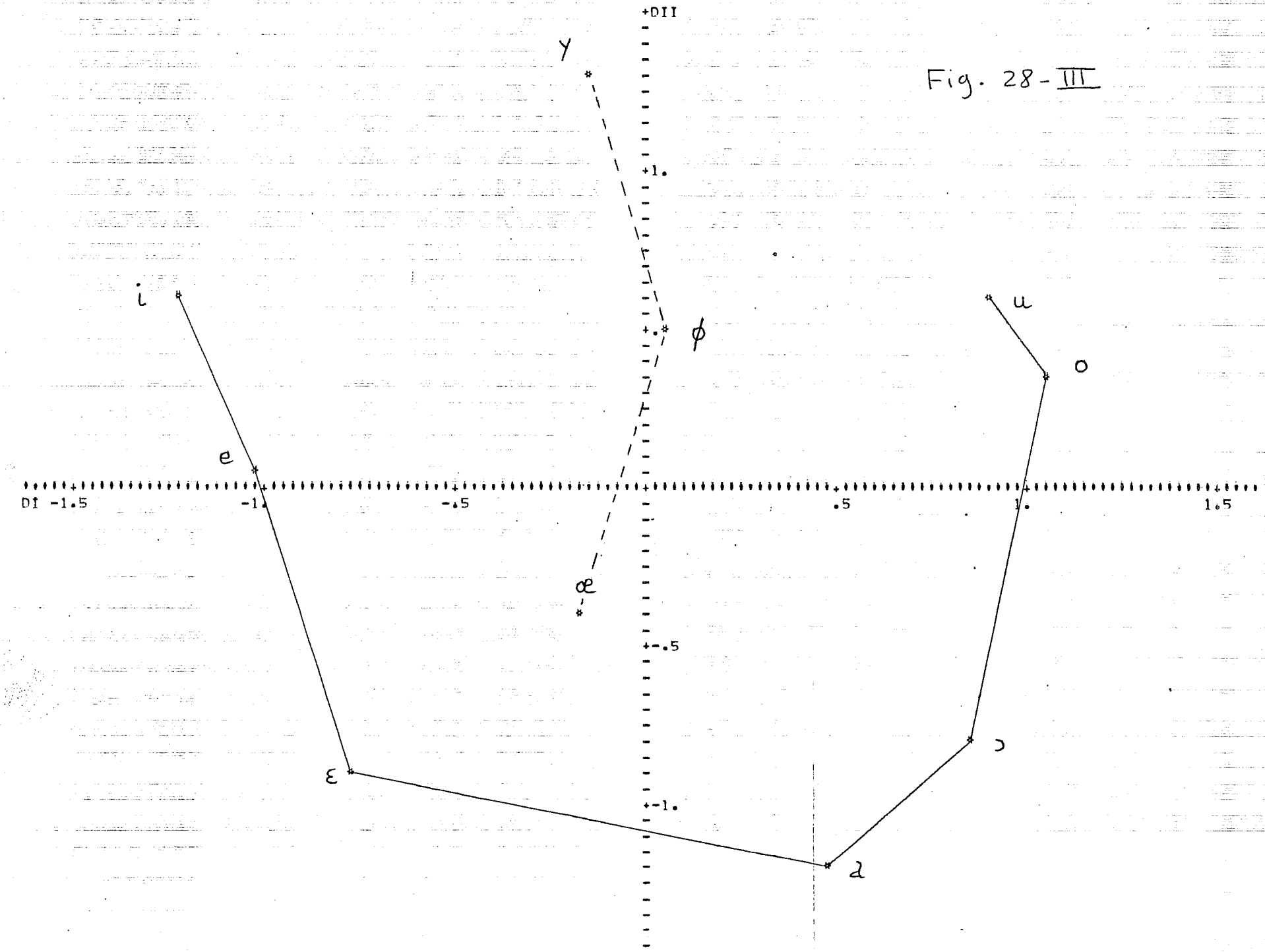
e	ɛ	a	ɔ	o	u	y	ø	œ	
3.62	4.4.7	5.63	5.39	5.34	4.91	4.39	4.55	4.71	i
	3.81	5.07	5.10	4.81	4.91	4.97	4.29	4.38	e
		4.36	4.29	5.28	5.22	5.37	4.62	3.74	ɛ
			4.01	4.91	4.95	5.65	4.92	4.62	a
				4.02	4.03	5.38	4.62	3.97	ɔ
					3.18	4.85	3.97	4.68	o
						4.20	4.02	4.71	u
							3.89	4.83	y
								3.36	ø

	Dim. I	Dim. II
i	-1.57	.81
e	-1.32	.04
ɛ	-.98	-1.20
a	.61	-1.58
ɔ	1.08	-1.07
o	1.36	.43
u	1.16	.79
y	-.19	1.66
ø	.07	.65
œ	-.23	-.54

La configuration est présentée graphiquement à la fig. 28-III
Les poids et le pourcentage de variance expliquée pour ce sujet "moyen" sont les suivants :

poids Dim. I	poids Dim. II	% variance expl.
.61	.59	72 %

Fig. 28-III



On observe l'augmentation considérable du taux de variance expliquée, qui double pratiquement par rapport à l'analyse individuelle (38 %). Ceci confirme dans une large mesure que le "bruit" était la seule cause de la faiblesse du taux de variance expliquée. Nous avons obtenu de meilleurs résultats à l'aide de la procédure suivante. L'examen du tableau des distances normalisées moyennes (table ci-dessus) révèle que toutes les distances sont relativement voisines en valeur absolue. Ceci est dû, pour une large part, au choix d'un écart type arbitraire pour la distribution globale. Ces indices vérifient l'inégalité triangulaire et sont donc techniquement des distances. Mais on remarque qu'il existe des écarts importants entre la représentation graphique bidimensionnelle et les valeurs numériques. Comme [i, e, ε] sont pratiquement alignés, la distance entre [i] et [ε] doit être égale à la somme des distances entre [i - e] et [e - ε]. Or le tableau révèle que d(i, e) + d(e, ε) est égal à 7,43, alors que la distance d(i, ε) ne vaut que 4.47. Rappelons que la valeur numérique des distances n'est que relative et dépend de la valeur numérique accordée à la case de l'échelle d'intervalles par l'expérimentateur. Nous avons arbitrairement accordé à la case "extrêmement semblable" la valeur numérique 1 et la case "extrêmement différent" la valeur 7, les cases intermédiaires étant effectuées d'une valeur comprise entre ces 2 extrêmes. On constate que pour l'exemple précédent, l'application d'une correction négative de -3 à chaque distance conduit à d(i, e) + d(e, ε) = 0,62 + 0,81 = 1,43, ce qui est pratiquement la distance entre [i] et [ε] : 1,47. Le coefficient négatif maximal que l'on puisse apporter au tableau est la valeur qui rend nulle la plus petite distance ; c'est-à-dire cette distance elle-même, affectée d'un signe négatif. Par définition, aucune distance entre deux stimuli différents ne peut être négative ou nulle. Nous avons tout d'abord appliqué la correction négative maximale (-3,18), puis nous avons transformé ces indices en distances en calculant la constante additive (0,76). On obtient le tableau ci-après :

e	ɛ	a	ɔ	o	u	y	ø	œ	
1.20	2.05	3.22	2.98	2.92	2.49	1.98	2.14	2.30	i
	1.40	2.66	2.69	2.40	2.49	2.56	1.87	1.96	e
		1.95	1.88	2.87	2.81	2.95	2.21	1.33	ɛ
			1.60	2.50	2.54	3.24	2.50	2.21	a
				1.60	1.62	2.96	2.21	1.56	ɔ
					.76	2.44	1.56	2.26	o
						1.78	1.61	2.30	u
							1.48	2.42	y
								.94	ø

On vérifie que l'écart entre la somme des distances d(i, e) + d(e, ɛ) = 2.60 est nettement réduit par rapport à la distance d(i, ɛ) = 2.05).

L'analyse INDSCAL de ce nouveau tableau donne les résultats suivants :

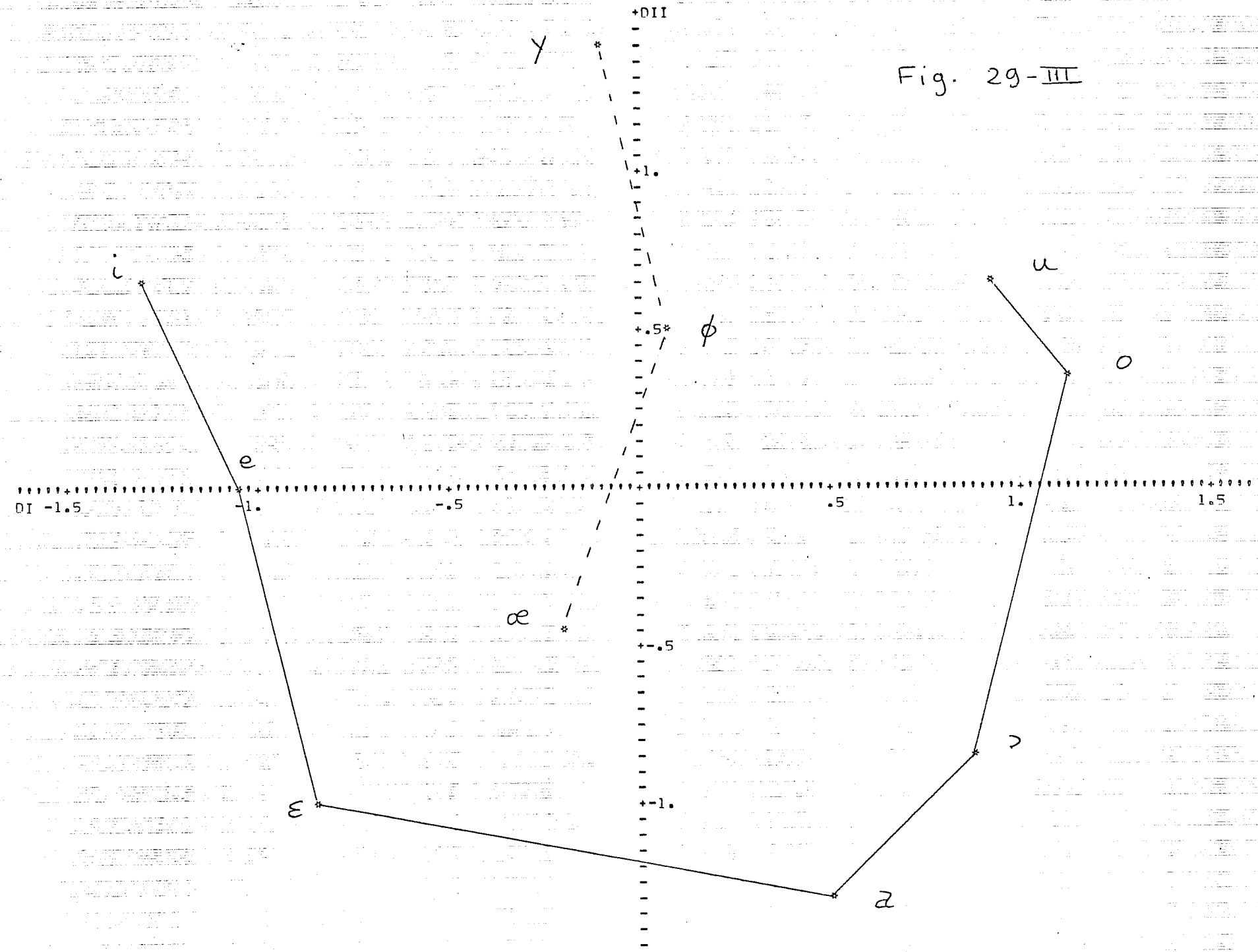
	Dim. I	Dim. II
i	-1.58	.80
e	-1.29	.00
ɛ	-1.04	-1.22
a	.62	-1.58
ɔ	1.07	-1.02
o	1.36	.42
u	1.14	.80
y	-.13	1.69
ø	.09	.62
œ	-.24	-.52

La configuration spatiale est présentée à la fig. 29-III Elle est pratiquement identique à celle obtenue sans correction négative. Les poids et le pourcentage de variance expliquée s'établissent aux valeurs suivantes :

poids Dim. I	poids Dim. II	% variance expl.
.67	.67	90 %

Ce taux de variance expliquée est exceptionnellement élevé. Il signifie que 10 % seulement de la variance n'est pas

Fig. 29-III



pris en compte dans la configuration spatiale bidimensionnelle. S'il est inférieur au chiffre obtenu par SHEPARD (1972), ce taux est nettement supérieur à tous les taux de variance obtenus par la méthode INDSCAL. Afin de vérifier ce chiffre, nous avons calculé le coefficient de corrélation entre les distances moyennes et les distances calculées.

Le tableau IV-III donne, pour chaque couple, la distance subjective, la distance calculée dans l'espace et cette même valeur normalisée. La distance calculée n'est en effet connue qu'à un facteur multiplicatif près, et il est légitime de la multiplifier par le facteur constant qui assure la meilleure correspondance avec la distance subjective. Nous avons utilisé la moyenne du rapport distance subjective/distance calculée comme facteur de normalisation.

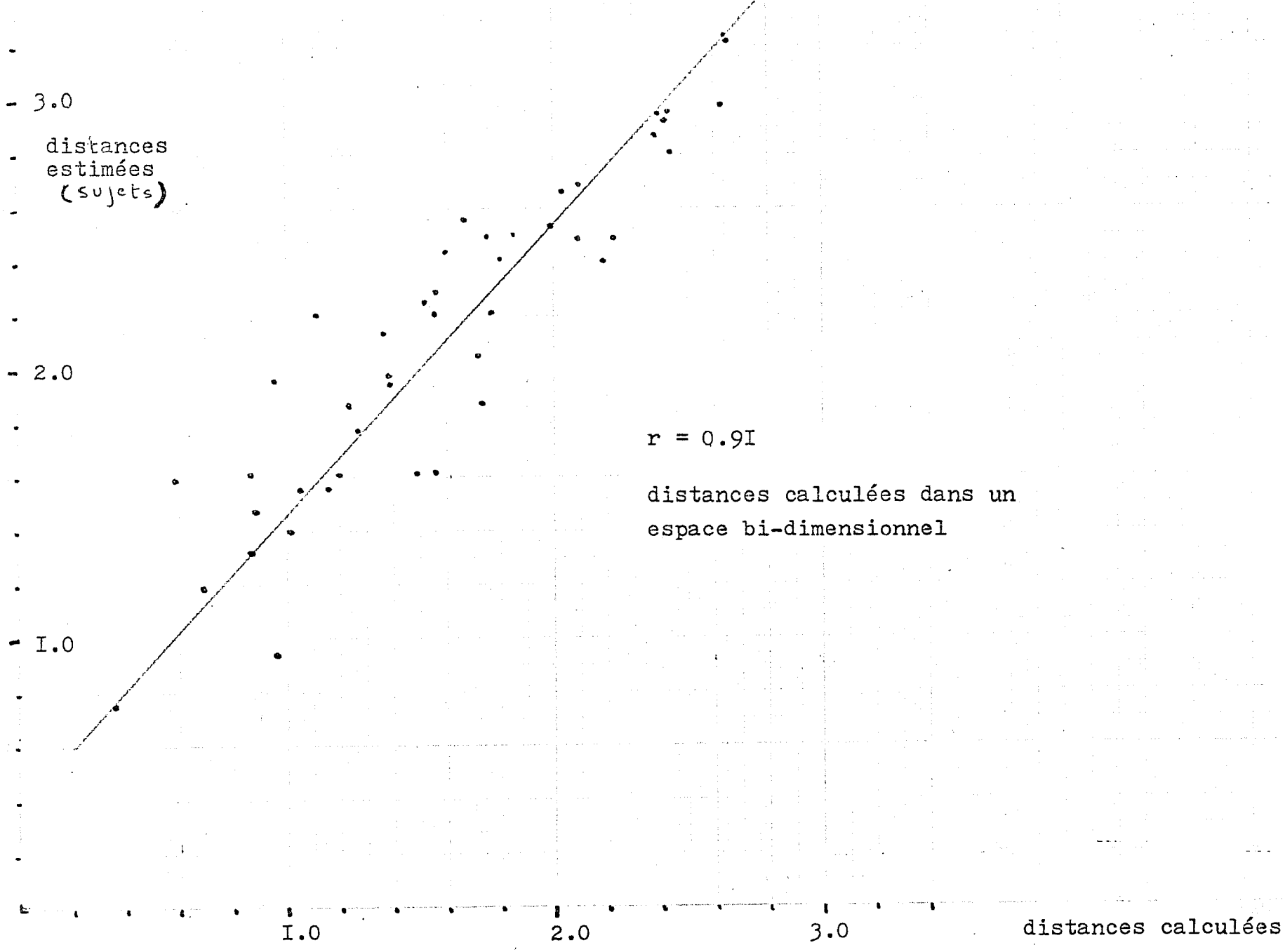
Le coefficient de corrélation s'élève à 0,91. Si l'on admet que la racine carrée de ce coefficient est approximativement égale à la variance expliquée, on obtient un taux de 95 % pour cette dernière. Ce taux est particulièrement élevé, et approche le pourcentage obtenu par SHEPARD (1972).

Nous avons ensuite procédé à une analyse en trois, puis en quatre dimensions. Les taux de variance expliquée tombent à 83 et 59 % respectivement. La position des points sur la troisième dimension n'est pas stable, et dépend de l'ordre initial choisi par l'expérimentateur. De plus, le coefficient de corrélation entre les distances subjectives et les distances calculées dans l'espace à trois dimensions est identique à celui obtenu précédemment (r = 0.91). L'introduction d'une dimension supplémentaire n'améliore pas l'accord entre les distances subjectives et les distances de l'espace multidimensionnel. Pour notre expérience, il est donc démontré qu'un espace bidimensionnel est suffisant pour décrire de façon adéquate les indices de dissimilarité fournis par les sujets.

Couples	distances initiales	distances calculées
i - e	I.20	o.93
i - ɛ	2.05	2.29
i - a	3.22	3.55
i - ɔ	2.98	3.52
i - o	2.92	3.25
i - u	2.49	2.98
i - y	I.98	I.86
i - ø	2.I4	I.84
i - œ	2.30	2.06
e - ɛ	I.40	I.36
e - a	2.66	2.7I
e - ɔ	2.69	2.8I
e - o	2.40	2.94
e - u	2.49	2.80
e - y	2.56	2.24
e - ø	I.87	I.66
e - œ	I.96	I.28
ɛ - a	I.95	I.86
ɛ - ɔ	I.88	2.32
ɛ - o	2.87	3.I8
ɛ - u	2.8I	3.25
ɛ - y	2.95	3.34
ɛ - ø	2.2I	2.36
ɛ - œ	I.33	I.I6
a - ɔ	I.60	0.79
a - o	2.50	2.33
a - u	2.54	2.67
a - y	3.24	3.67
a - ø	2.50	2.48
a - œ	2.2I	I.49

couples	distances initiales	distances calculées
ɔ - o	I.60	I.6I
ɔ - u	I.62	I.99
ɔ - y	2.96	3.24
ɔ - ø	2.2I	2.09
ɔ - œ	I.56	I.53
o - u	0.76	0.48
o - y	2.44	2.I4
o - ø	I.56	I.4I
o - œ	2.26	2.03
u - y	I.78	I.70
u - ø	I.6I	I.I7
u - œ	2.30	2.09
y - ø	I.48	I.20
y - œ	2.42	2.42
ø - œ	0.94	I.30

Table IV - III


3.0
distances
estimées
(sujets)
2.0
I.0
r = 0.9I
distances calculées dans un
espace bi-dimensionnel
I.0
2.0
3.0
distances calculées

G/ Discussion et interprétation des résultats

a) Corrélation entre les dimensions physiques et perceptives

Les distances fournies par les sujets ne comportent pas de trace de l'existence d'une troisième dimension. L'importance du taux de variance expliquée, et sa stagnation lorsque l'on introduit une troisième dimension, indiquent que la variance résiduelle est due au "bruit" qui accompagne les estimations subjectives. Pour placer l'interprétation des deux dimensions sur une base quantitative, nous avons calculé plusieurs coefficients de corrélation de rang (ou indices de SPEARMAN) entre les caractéristiques fréquentielles des stimuli et les coordonnées dans l'espace. Le coefficient de corrélation de rang (ρ) est donné par la formule

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2-1)}$$

où d est la différence de rang pour un même stimulus, et n le nombre de stimuli.

Pour n = 10, les seuils de signification sont les suivants :

P	0.10	0.05	0.02	0.01
ρ	0.55	0.63	0.72	0.76

Le tableau suivant présente les rangs des différents paramètres pertinents :

Dim. I	F2	F2 + F3/2	Dim. II	F1
1	1	1	2.5	1.5
2	2	2	6	4
3	3.5	3	9	7
7	7	7	10	10
8	8	8	8	8.5
10	10	9	5	6
9	9	10	2.5	3
5	3.5	4	1	1.5
6	6	6	4	5
4	5	5	7	8.5

Le coefficient de corrélation de rang entre F1 et la dimension II s'établit à 0.915, ce qui est hautement significatif ($p. \ll 0.01$). En revanche, le coefficient entre F2 et la deuxième dimension est nul ($\rho = 0.01$)*. Pour la première dimension et F2, celui-ci atteint 0.98. Il n'est que de 0.26 entre F1 et la première dimension, ce qui n'est pas significatif. Mais comme nous le rappelions dans la partie de ce chapitre consacrée aux critères d'identification des dimensions, on ne peut conclure que la dimension I soit le corrélat perceptif de F2. A titre d'illustration, nous avons calculé le coefficient de corrélation entre la première dimension et la quantité (F2 + F3 /2). Celui-ci atteint 0.98, c'est-à-dire une valeur égale à la corrélation entre DI et F2. On peut donc tout aussi bien considérer que (F2 + F3/2) est la base physique de la première dimension. Plus généralement nous considérerons la première dimension comme le corrélat perceptif d'une combinaison linéaire des formants F2, F3 et éventuellement F4. Cette conclusion est dictée par un souci de cohérence avec les résultats obtenus par la synthèse vocalique à 2 formants. De nombreuses expériences (DELATTRE et al. 1952, CARLSON et al. 1975 par exemple) ont montré que la synthèse à 2 formants n'était possible, pour les voyelles antérieures, qu'à condition que le deuxième formant soit plus élevé que celui des voyelles naturelles. Ce deuxième formant "synthétique" joue le rôle de F2 et F3. Le chapitre IV tout entier sera consacré à l'étude de ce "deuxième formant équivalent" (FANT 1959) ou F'2.

Il n'existe pas de méthode permettant de déduire exactement, à partir de la description physique du stimulus et les résultats de l'analyse multidimensionnelle, les rapports existant entre les dimensions physiques et les dimensions perceptives. Il faut utiliser toutes les sources d'information, en veillant à la cohérence des résultats. La possibilité de

---

* Sur un plan pratique, notons que l'ordre d'extraction des dimensions est arbitraire. Dans cette analyse, il se trouve que la deuxième dimension corrèle avec F1. Lorsque nous parlerons ailleurs de deuxième dimension, il s'agira de la dimension qui corrèle avec F2.

synthétiser une voyelle à l'aide de deux formants seulement suggère une liaison entre les formants de la synthèse (F1 et F'2) et les dimensions de l'espace. Cette interprétation a l'avantage d'unifier les résultats obtenus par deux méthodes d'investigation.

b) Un artefact majeur ?

La discussion précédente n'épuise cependant pas le problème de l'interprétation des dimensions. Les sujets ayant suivi un enseignement de phonétique, on peut se demander si les résultats ne sont pas dus à l'utilisation de l'information spatiale contenue dans le triangle vocalique traditionnel. Cet artefact majeur ne nous semble pas vraisemblable, bien qu'on ne puisse pas l'exclure à priori. Les positions des voyelles [y] et [ϕ] dans l'espace perceptif ne sont pas conformes à ce que prédirait l'hypothèse d'une utilisation du triangle vocalique. Notons également que nos résultats ne sont pas essentiellement différents de ceux obtenus par des sujets "naïfs". En revanche, l'utilisation de sujets connaissant la transcription phonétique assure qu'ils possèdent une certaine expérience du jugement du timbre des voyelles. Ceci devrait contribuer à améliorer les performances en réduisant le "bruit". Enfin, on peut penser que l'utilisation d'informations dérivées du triangle vocalique aurait facilité l'apparition d'une troisième dimension perceptive, en améliorant la précision des estimations sur les deux premières dimensions. Or on constate que les estimations individuelles des sujets sont entachées d'un "bruit" très important ; on peut penser que tel n'aurait pas été le cas si les sujets s'étaient servis d'une image mentale du triangle vocalique. Mais il est certain qu'une enquête post-expérimentale aurait été utile.

c) Structure circulaire ou effet de plafonnement ?

L'examen de la configuration des points montre que la position de [y] sur la dimension II n'est pas conforme à la valeur fréquentielle du premier formant. Cette distorsion explique la (relative) faiblesse du coefficient de corrélation

de rang (ρ = 0.915). La configuration des points indique que les voyelles tendent à se placer sur la circonférence d'un cercle dont le centre se trouve à l'origine des axes. Quelques études (SHEPARD 1962, CHIPMAN et CAREY 1975) ont déjà souligné la présence de structures circulaires dans différents domaines perceptifs. Mais, pour notre expérience, nous montrerons qu'il s'agit vraisemblablement d'un effet dû à la méthode d'estimation des distances subjectives.

L'étude des 8 indices de dissimilarité les plus importants montre que 6 voyelles sur 8 s'inscrivent quasiment à égale distance de deux ou trois voyelles. Ces indices sont compris entre 3.24 et 2.81 (fig. 30-III)

Les 9 indices les plus faibles (.076 - 1.60) relient toutes les voyelles adjacentes sur la deuxième dimension (F1) à l'exception de [ɛ - ɑ] et de [ɑ - ɔ]. Sur la première dimension, ils relient également [ɛ - œ, ø - o, œ - ɔ]. Il est clair qu'une structure p ossédant ces deux propriétés aura une forme très proche d'un cercle. Le problème est de savoir si la forme correspond effectivement à une réalité perceptive, ou si elle est en partie "gauchie" par une particularité des indices de dissimilarité; la forme réelle serait alors encore plus proche de la position des voyelles dans le plan F1 - F'2.

Nous pensons que cette forme quasi-circulaire est due à un effet de "plafonnement" des réponses "différentes" ou "très différentes" ("ceiling effect"). Les voyelles les plus proches sur le trapèze vocalique seraient jugées "assez" ou "modérément" proches. En revanche, les paires de voyelles plus "lointaines" seraient jugées systématiquement "très différentes".

Les indices de dissimilarité ne seraient plus une fonction relativement linéaire de la distance acoustique, comme le montre la figure 31-III . Un tel comportement peut s'observer dans la table donnant les distances intervocaliques. L'accroissement entre la distance [a - o] et [a - u] est très faible (2.50 - 2.54). Tout se passe comme si les sujets avaient du mal à classer entre elles les distances les plus

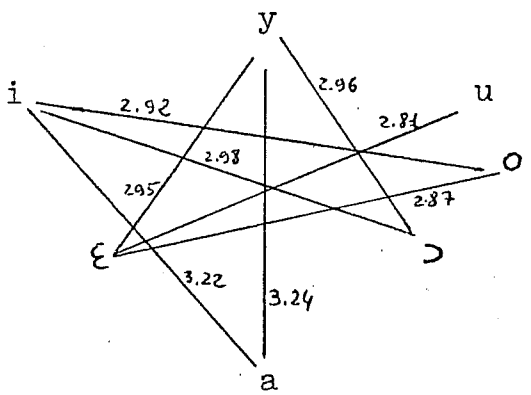


Fig. 30 - III a

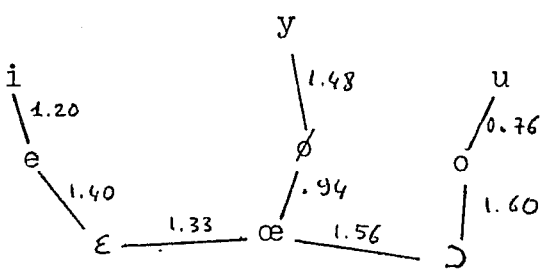


Fig 30 - III b

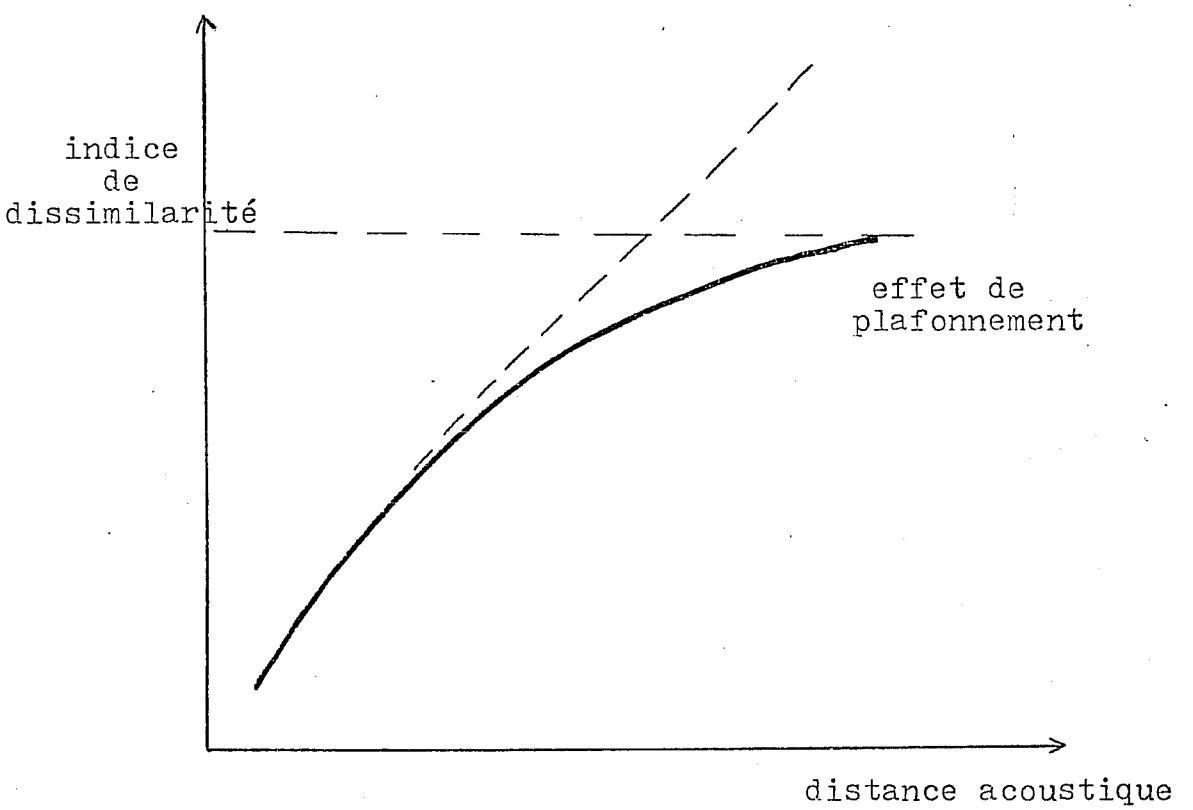

indice
de
dissimilarité
effet de
plafonnement
distance acoustique

Fig. 3I - III

longues. Pour obtenir une vue globale, nous avons comparé la répartition des distances subjectives et des distances acoustiques calculées dans l'espace physique log F1 - Log F2 (fig. 32-III et fig. 33-III). On note la distribution nettement dissymétrique de la répartition des indices de dissimilarité. Les réponses "longues" sont plus nombreuses que les réponses "courtes". Or cette dissymétrie ne se retrouve pas dans la répartition des distances acoustiques, qui accuserait plutôt un déséquilibre inverse : les distances "courtes" sont plus nombreuses que les distances "longues". La distribution des distances subjectives peut se déduire de celle des distances acoustiques si l'on fixe un plafond que les distances subjectives ne peuvent franchir. L'existence d'un tel plafonnement est une indication méthodologique importante pour l'utilisation des échelles d'intervalles. Il semble que toutes les méthodes d'estimation des distances se heurtent à un même problème : un biais systématique qui sous-estime (ou plafonne) les distances les plus longues. Il faudrait, comme le suggère TERBEECK, appliquer une correction en sens inverse. Le problème est d'estimer la taille de la correction à effectuer pour ne pas tomber dans le défaut inverse. Cet effet de plafonnement n'est pourtant pas assez important pour avoir forcé le programme à extraire une dimension supplémentaire rendant compte de cette non-linéarité. Nous ne pensons pas que cet effet de plafonnement infirme les conclusions essentielles que nous avons tirées de cette analyse de l'espace perceptif des voyelles françaises non-nasales.

- L'espace perceptif nécessaire pour rendre compte des dissimilarités entre les 10 voyelles françaises non-nasales suivantes [i e ɛ a ɔ o u y ø œ] ne possède que deux dimensions. Le taux de variance expliquée, s'il est faible dans l'analyse individuelle, est très important pour des moyennes de groupe, il dépasse légèrement 90 %. L'indépendance des dimensions physiques F1 et F2 et la forte corrélation entre la deuxième dimension et F1, suggèrent que celle-ci est le corrélat perceptif de la fréquence du premier formant. Nous proposons enfin d'interpréter la première dimension comme le corrélat perceptif

Fig. 32 - III

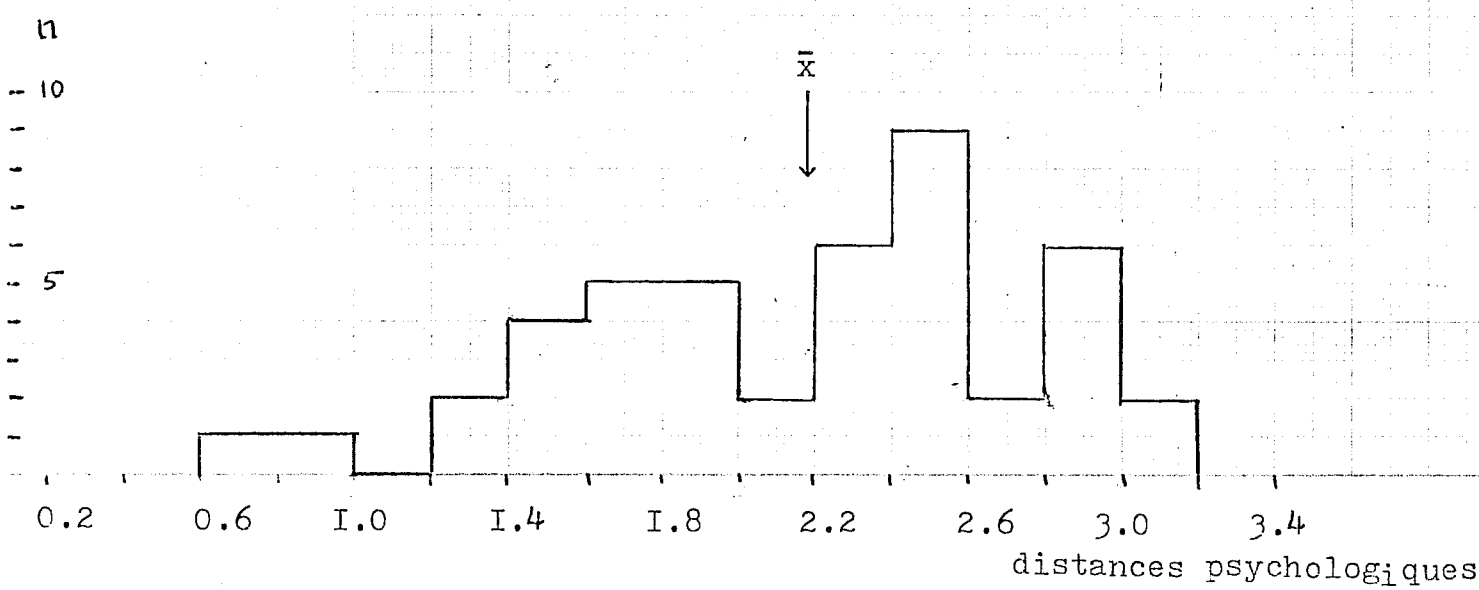
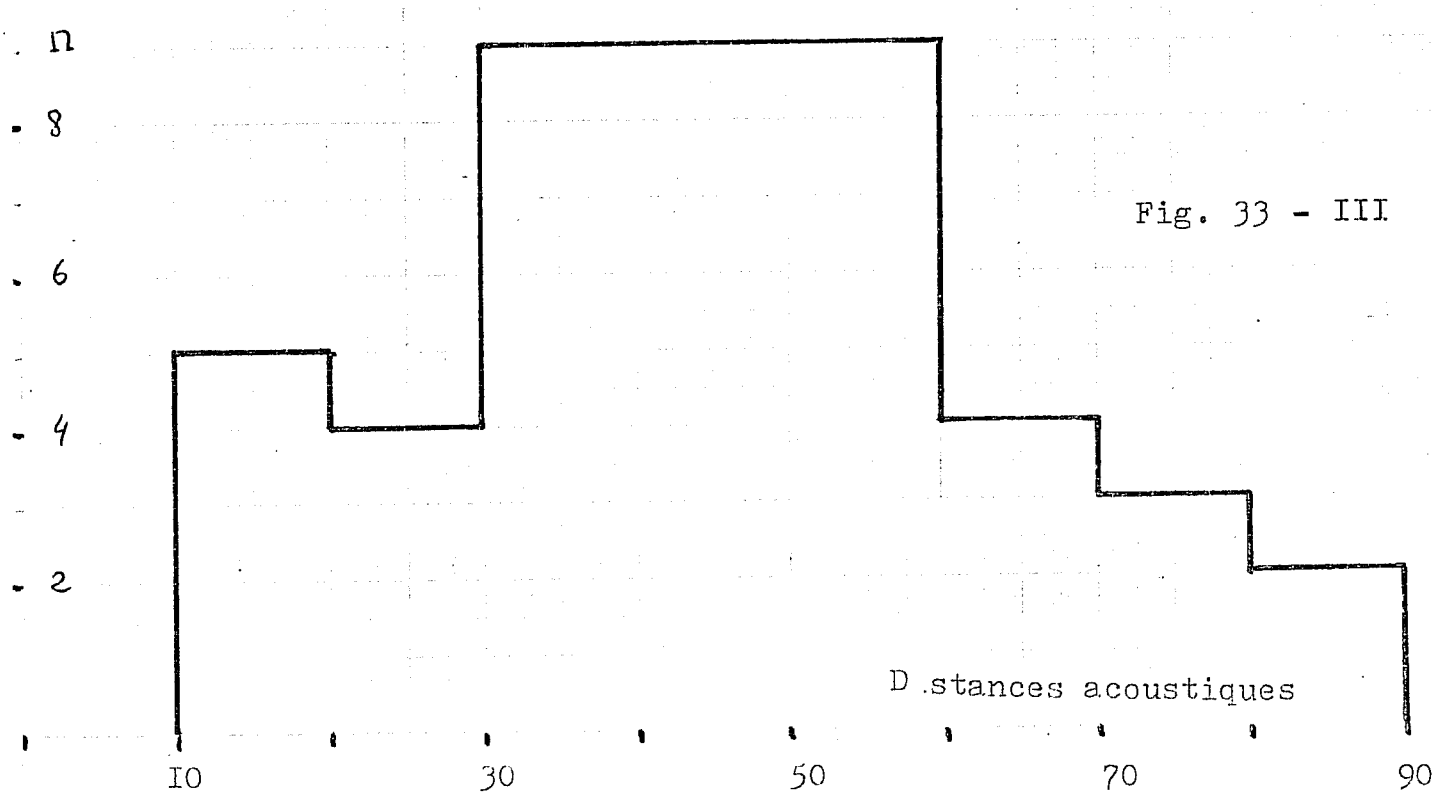


Fig. 33 - III



du "deuxième formant équivalent" mis en évidence dans les expériences de synthèse à deux formants. La première dimension serait donc le corrélat perceptif de la zone fréquentielle incluant les deuxième, troisième et éventuellement quatrième formants.

VI - CONCLUSIONS GENERALES : LA BIDIMENSIONNALITE DE L'ESPACE PERCEPTIF

La première partie de ce chapitre est consacrée à une étude détaillée des problèmes méthodologiques posés par la détermination du nombre des dimensions et par l'interprétation des résultats de l'analyse multidimensionnelle :

- la quantité de variance expliquée doit être vérifiée par le calcul du coefficient de corrélation entre les indices subjectifs et les distances dans l'espace perceptif.

- les méthodes d'estimation des distances subjectives sont le "maillon faible" de la technique.

Il est établi que toutes sous-estiment systématiquement les distances les plus longues. L'utilisation des matrices de confusions se heurte en outre à l'incertitude créée par la non-symétrie des confusions et au biais causé par la présence de "cases vides".

- le corrélat physique qui sous-tend une dimension perceptive ne peut être identifié avec certitude. Dans la plupart des cas, une dimension peut être une combinaison de plusieurs paramètres physiques.

La deuxième partie du chapitre est une revue critique des 15 analyses publiées jusqu'en 1977. Lorsque les conditions le permettaient, nous avons procédé à une nouvelle analyse. Toutes les expériences mettent en évidence deux dimensions corrélant avec les fréquences du premier et du deuxième formant. Il n'existe pas d'indices permettant de penser que les sujets utilisent une troisième dimension perceptive. On constate en effet :

- l'absence d'augmentation significative de la quantité de variance expliquée pour cet axe supplémentaire.

- l'instabilité de l'ordre des voyelles, qui dépend de celui imposé initialement par l'expérimentateur.

- la possibilité d'une interprétation par un artefact, ou par un biais systématique des estimations subjectives.

La seule exception est la troisième dimension qui sépare la voyelle rétroflexe [ɹ] des voyelles non-rétroflexes. On retrouve cette dimension dans toutes les analyses incluant cette voyelle.
Une preuve quasi-décisive serait obtenue si l'on montrait qu'il était possible de rendre compte de toute la variance à l'aide de deux dimensions. Nos résultats, qui portent sur 10 voyelles françaises non-nasales, sont très proches de cet objectif et peuvent se résumer ainsi :

- l'analyse individuelle révèle une quantité considérable de "bruit" dans les réponses issues d'indices de dissimilarité sur des échelles d'intervalles. Le pourcentage de variance expliquée pour deux dimensions ne dépasse pas 38 %. L'introduction d'une troisième dimension n'améliore aucunement la quantité d'information extraite.

- en revanche, l'analyse des moyennes de groupe permet de rendre compte de 90 % de la variance. Ce chiffre très élevé a été confirmé par le calcul du coefficient de corrélation entre les dissimilarités et les distances dans l'espace ($r = 0.91$). L'introduction d'une troisième dimension n'améliore pas ce coefficient, et l'ordre des voyelles devient instable. Cette expérience montre sans ambiguïté que l'espace perceptif de ces 10 voyelles est BIDIMENSIONNEL.

L'une des deux dimensions corrèle fortement avec la fréquence du permier formant. L'indépendance statistique de F1 par rapport à F2 indique vraisemblablement que cette dimension est le corrélat perceptif de F1 seul. Nous faisons l'hypothèse que la deuxième est la représentation perceptive de l'ensemble des formants supérieurs, F2, F3 et éventuellement F4. Cette

interprétation est compatible avec les résultats de la synthèse vocalique à deux formants. Pour les voyelles antérieures, ceux-ci indiquent la nécessité d'un deuxième formant (F'2 ou "deuxième formant équivalent") plus élevé que celui mesuré sur les voyelles naturelles.

La forme circulaire qui tend à prendre le "nuage" des voyelles dans l'espace bidimensionnel est sans doute liée à un effet de "plafonnement" qui biaise les estimations subjectives des couples les plus dissemblables. Au-delà d'un certain seuil, les sujets ne sont plus en mesure de donner une estimation "fine" des dissimilarités et utilisent la seule réponse "très différent". Cet artefact, s'il ne s'agit pas d'un comportement spécifique des sujets, est commun à toutes les méthodes d'estimation des distances. Mais ce biais systématique ne nous semble pas devoir remettre en question le résultat essentiel que nous avons obtenu : l'espace perceptif des voyelles non-rétroflexes est bidimensionnel.

Sujet 1

3.00	4.00	6.00	5.50	6.00	4.00	4.00	3.00	4.50	3.00	6.50	5.50	4.50	5.50	6.50
3.50	3.50	3.00	5.00	5.00	5.00	6.00	3.50	3.00	3.00	5.50	5.50	6.50	4.50	2.50
3.00	2.00	5.50	3.00	3.50	1.00	6.00	3.50	5.00	4.50	3.50	4.50	3.00	5.00	2.50
.00	3.00	4.00	6.00	5.50	6.00	4.00	4.00	3.00	4.50					
3.00	.00	3.00	6.50	5.50	4.50	5.50	6.50	3.50	3.50					
4.00	3.00	.00	3.00	5.00	5.00	5.00	6.00	3.50	3.00					
6.00	6.50	3.00	.00	3.00	5.50	5.50	6.50	4.50	2.50					
5.50	5.50	5.00	3.00	.00	3.00	2.00	5.50	3.00	3.50					
6.00	4.50	5.00	5.50	3.00	.00	1.00	6.00	3.50	5.00					
4.00	5.50	5.00	5.50	2.00	1.00	.00	4.50	3.50	4.50					
4.00	6.50	6.00	6.50	5.50	6.00	4.50	.00	3.00	5.00					
3.00	3.50	3.50	4.50	3.00	3.50	3.50	3.00	.00	2.50					
4.50	3.50	3.00	2.50	3.50	5.00	4.50	5.00	2.50	.00					

CST ADD. = 1.00000

Données non corrigées par cst. add. dans l'ordre exigé par le programme INDSCAL

tableau des indices de dissimilarité (sujet

(diagonale nulle)

ordre des voyelles

[i e ɛ a ɔ o u y ø œ]

S2

2.50	4.50	4.00	5.00	4.50	3.50	3.00	4.50	4.00	4.50	4.00	3.00	3.00	4.50	3.00
3.00	5.50	3.00	4.00	4.50	5.00	5.00	6.00	5.50	3.50	3.50	4.50	4.00	3.00	4.00
6.50	5.50	5.50	4.00	4.00	5.50	6.00	4.00	6.50	3.00	4.00	5.00	3.00	5.00	3.50
.00	2.50	4.50	4.00	5.00	4.50	3.50	3.00	4.50	4.00					
2.50	.00	4.50	4.00	3.00	3.00	4.50	3.00	3.00	5.50					
4.50	4.50	.00	3.00	4.00	4.50	5.00	5.00	6.00	5.50					
4.00	4.00	3.00	.00	3.50	3.50	4.50	4.00	3.00	4.00					
5.00	3.00	4.00	3.50	.00	6.50	5.50	5.50	4.00	4.00					
4.50	3.00	4.50	3.50	6.50	.00	5.50	6.00	4.00	6.50					
3.50	4.50	5.00	4.50	5.50	5.50	.00	3.00	4.00	5.00					
3.00	3.00	5.00	4.00	5.50	6.00	3.00	.00	3.00	5.00					
4.50	3.00	6.00	3.00	4.00	4.00	4.00	3.00	.00	3.50					
4.00	5.50	5.50	4.00	4.00	6.50	5.00	5.00	3.50	.00					

CST ADD. = .50000

S3

2.50	3.50	5.50	4.00	4.50	4.50	4.50	3.00	4.50	5.50	6.50	6.00	6.00	5.00	6.50
5.50	5.50	3.50	4.00	6.50	3.00	5.50	6.00	5.00	2.00	7.00	5.00	6.50	6.00	5.50
2.00	3.00	3.50	2.00	2.50	1.00	6.00	4.00	4.00	1.50	2.00	2.50	5.50	4.50	2.50
.00	2.50	3.50	5.50	4.00	4.50	4.50	4.50	3.00	4.50					
2.50	.00	5.50	6.50	6.00	6.00	5.00	6.50	5.50	5.50					
3.50	5.50	.00	3.50	4.00	6.50	3.00	5.50	6.00	5.00					
5.50	6.50	3.50	.00	2.00	7.00	5.00	6.50	6.00	5.50					
4.00	6.00	4.00	2.00	.00	2.00	3.00	3.50	2.00	2.50					
4.50	6.00	6.50	7.00	2.00	.00	1.00	6.00	4.00	4.00					
4.50	5.00	3.00	5.00	3.00	1.00	.00	1.50	2.00	2.50					
4.50	6.50	5.50	6.50	3.50	6.00	1.50	.00	5.50	4.50					
3.00	5.50	6.00	6.00	2.00	4.00	2.00	5.50	.00	2.50					
4.50	5.50	5.00	5.50	2.50	4.00	2.50	4.50	2.50	.00					

CST ADD. = 3.50000

S4

1.50	6.00	6.50	7.00	6.50	7.00	1.00	5.50	3.50	4.00	5.50	5.50	4.00	6.00	4.50
3.00	4.50	4.00	4.00	5.50	4.00	5.00	3.50	4.50	3.00	3.00	3.50	6.00	4.00	4.00
3.50	2.00	6.00	2.00	3.00	1.50	5.50	3.00	3.00	6.00	2.00	5.00	3.50	5.00	2.00
.00	1.50	6.00	6.50	7.00	6.50	7.00	1.00	5.50	3.50					
1.50	.00	4.00	5.50	5.50	4.00	6.00	4.50	3.00	4.50					
6.00	4.00	.00	4.00	4.00	5.50	4.00	5.00	3.50	4.50					
6.50	5.50	4.00	.00	3.00	3.00	3.50	6.00	4.00	4.00					
7.00	5.50	4.00	3.00	.00	3.50	2.00	6.00	2.00	3.00					
6.50	4.00	5.50	3.00	3.50	.00	1.50	5.50	3.00	3.00					
7.00	6.00	4.00	3.50	2.00	1.50	.00	6.00	2.00	5.00					
1.00	4.50	5.00	6.00	6.00	5.50	6.00	.00	3.50	5.00					
5.50	3.00	3.50	4.00	2.00	3.00	2.00	3.50	.00	2.00					
3.50	4.50	4.50	4.00	3.00	3.00	5.00	5.00	2.00	.00					

CST ADD. = 2.00000

S_5

6.50	4.50	7.00	6.00	7.00	4.00	7.00	6.00	4.00	4.00	6.50	6.00	7.00	4.50	6.00
2.50	6.00	4.00	2.00	4.00	6.00	4.00	2.50	2.50	6.00	7.00	4.00	7.00	4.50	4.50
5.00	4.00	6.00	2.00	4.50	4.00	4.00	4.00	6.00	2.50	3.00	6.00	4.00	6.00	2.00

.00	6.50	4.50	7.00	6.00	7.00	4.00	7.00	6.00	4.00
6.50	.00	4.00	6.50	6.00	7.00	4.50	6.00	2.50	6.00
4.50	4.00	.00	4.00	2.00	4.00	6.00	4.00	2.50	2.50
7.00	6.50	4.00	.00	6.00	7.00	4.00	7.00	4.50	4.50
6.00	6.00	2.00	6.00	.00	5.00	4.00	6.00	2.00	4.50
7.00	7.00	4.00	7.00	5.00	.00	4.00	4.00	4.00	6.00
4.00	4.50	6.00	4.00	4.00	4.00	.00	2.50	3.00	6.00
7.00	6.00	4.00	7.00	6.00	4.00	2.50	.00	4.00	6.00
6.00	2.50	2.50	4.50	2.00	4.00	3.00	4.00	.00	2.00
4.00	6.00	2.50	4.50	4.50	6.00	6.00	6.00	2.00	.00

CST ADD. = 1.50000

S_6

2.00	4.00	5.50	5.00	6.00	2.00	5.50	4.50	3.00	2.50	4.00	5.50	2.00	2.00	4.50
4.00	3.00	5.00	2.50	2.00	6.00	6.50	2.50	4.00	3.00	4.50	2.00	2.50	3.00	6.00
3.00	3.50	4.50	6.00	2.00	1.00	3.00	2.00	4.50	2.00	2.50	6.00	2.50	7.00	1.00

.00	2.00	4.00	5.50	5.00	6.00	2.00	5.50	4.50	3.00
2.00	.00	2.50	4.00	5.50	2.00	2.00	4.50	4.00	3.00
4.00	2.50	.00	5.00	2.50	2.00	6.00	6.50	2.50	4.00
5.50	4.00	5.00	.00	3.00	4.50	2.00	2.50	3.00	6.00
5.00	5.50	2.50	3.00	.00	3.00	3.50	4.50	6.00	2.00
6.00	2.00	2.00	4.50	3.00	.00	1.00	3.00	2.00	4.50
2.00	2.00	6.00	2.00	3.50	1.00	.00	2.00	2.50	6.00
5.50	4.50	6.50	2.50	4.50	3.00	2.00	.00	2.50	7.00
4.50	4.00	2.50	3.00	6.00	2.00	2.50	2.50	.00	1.00
3.00	3.00	4.00	6.00	2.00	4.50	6.00	7.00	1.00	.00

CST ADD. = 3.50000

S_7

5.50	5.00	7.00	5.50	7.00	7.00	7.00	5.50	7.00	3.50	6.50	6.50	6.50	5.50	5.50
4.50	5.00	6.50	5.00	6.50	5.50	6.00	4.00	2.50	6.00	7.00	7.00	7.00	5.50	5.50
6.50	6.50	7.00	6.00	4.50	6.50	7.00	4.50	5.50	5.00	5.50	6.00	6.50	5.50	2.50

.00	5.50	5.00	7.00	5.50	7.00	7.00	7.00	5.50	7.00
5.50	.00	3.50	6.50	6.50	6.50	5.50	5.50	4.50	5.00
5.00	3.50	.00	6.50	5.00	6.50	5.50	6.00	4.00	2.50
7.00	6.50	6.50	.00	6.00	7.00	7.00	7.00	5.50	5.50
5.50	6.50	5.00	6.00	.00	6.50	6.50	7.00	6.00	4.50
7.00	6.50	6.50	7.00	6.50	.00	6.50	7.00	4.50	5.50
7.00	5.50	5.50	7.00	6.50	6.50	.00	5.00	5.50	6.00
7.00	5.50	6.00	7.00	7.00	7.00	5.00	.00	6.50	5.50
5.50	4.50	4.00	5.50	6.00	4.50	5.50	6.50	.00	2.50
7.00	5.00	2.50	5.50	4.50	5.50	6.00	5.50	2.50	.00

CST ADD. = .00000

S_8

5.50	7.00	7.00	7.00	6.50	6.50	6.50	5.50	6.00	4.00	6.00	7.00	6.00	6.00	6.50
4.50	4.50	6.00	6.00	6.50	7.00	6.50	5.00	2.50	6.50	6.00	4.50	7.00	6.00	6.00
5.50	6.50	6.50	6.00	4.50	5.00	7.00	5.00	5.00	6.50	4.50	5.50	5.50	6.00	3.00

.00	5.50	7.00	7.00	7.00	6.50	6.50	6.50	5.50	6.00
5.50	.00	4.00	6.00	7.00	6.00	6.00	6.50	4.50	4.50
7.00	4.00	.00	6.00	6.00	6.50	7.00	6.50	5.00	2.50
7.00	6.00	6.00	.00	6.50	6.00	4.50	7.00	6.00	6.00
7.00	7.00	6.00	6.50	.00	5.50	6.50	6.50	6.00	4.50
6.50	6.00	6.50	6.00	5.50	.00	5.00	7.00	5.00	5.00
6.50	6.00	7.00	4.50	6.50	5.00	.00	6.50	4.50	5.50
6.50	6.50	6.50	7.00	6.50	7.00	6.50	.00	5.50	6.00
5.50	4.50	5.00	6.00	6.00	5.00	4.50	5.50	.00	3.00
6.00	4.50	2.50	6.00	4.50	5.00	5.50	6.00	3.00	.00

I.U.C.A. NANC

CST ADD. = .00000

S_9

4.00	5.50	6.50	7.00	7.00	6.50	5.50	6.00	6.00	6.50	5.00	5.00	7.00	5.00	6.50
6.50	6.50	4.00	5.50	6.50	6.50	6.00	6.50	6.00	4.00	4.00	6.00	6.50	5.50	6.00
7.00	6.00	7.00	6.00	6.50	4.00	4.00	7.00	5.00	6.50	6.00	6.50	5.00	6.50	6.00
.00	4.00	5.50	6.50	7.00	7.00	6.50	5.50	6.00	6.00					
4.00	.00	6.50	5.00	5.00	7.00	5.00	6.50	6.50	6.50					
5.50	6.50	.00	4.00	5.50	6.50	6.50	6.00	6.50	6.00					
6.50	5.00	4.00	.00	4.00	4.00	6.00	6.50	5.50	6.00					
7.00	5.00	5.50	4.00	.00	7.00	6.00	7.00	6.00	6.50					
7.00	7.00	6.50	4.00	7.00	.00	4.00	4.00	7.00	5.00					
6.50	5.00	6.50	6.00	6.00	4.00	.00	6.50	6.00	6.50					
5.50	6.50	6.00	6.50	7.00	4.00	6.50	.00	5.00	6.50					
6.00	6.50	6.50	5.50	6.00	7.00	6.00	5.00	.00	6.00					
6.00	6.50	6.00	6.00	6.50	5.00	6.50	6.50	6.00	.00					

CST ADD. = .00000

S_{10}

2.50	2.50	6.50	4.00	6.50	6.00	3.50	6.50	6.00	1.00	6.50	5.50	6.50	5.50	5.50
6.00	2.50	4.00	5.00	6.00	5.50	5.50	5.50	2.50	5.00	2.50	4.00	4.50	4.50	5.00
1.50	4.00	5.50	3.00	4.00	3.00	2.50	1.50	3.00	3.00	2.00	5.00	2.50	4.00	2.00
.00	2.50	2.50	6.50	4.00	6.50	6.00	3.50	6.50	6.00					
2.50	.00	1.00	6.50	5.50	6.50	5.50	5.50	6.00	2.50					
2.50	1.00	.00	4.00	5.00	6.00	5.50	5.50	5.50	2.50					
6.50	6.50	4.00	.00	5.00	2.50	4.00	4.50	4.50	5.00					
4.00	5.50	5.00	5.00	.00	1.50	4.00	5.50	3.00	4.00					
6.50	6.50	6.00	2.50	1.50	.00	3.00	2.50	1.50	3.00					
6.00	5.50	5.50	4.00	4.00	3.00	.00	3.00	2.00	5.00					
3.50	5.50	5.50	4.50	5.50	2.50	3.00	.00	2.50	4.00					
6.50	6.00	5.50	4.50	3.00	1.50	2.00	2.50	.00	2.00					
6.00	2.50	2.50	5.00	4.00	3.00	5.00	4.00	2.00	.00					

CST ADD. = 1.50000

S_{11}

3.50	5.00	4.00	6.00	1.50	1.00	2.50	3.50	6.00	7.00	1.00	3.50	1.00	2.00	1.50
3.00	4.50	5.00	7.00	4.50	7.00	6.00	5.50	4.50	4.00	4.00	3.50	7.00	7.00	7.00
7.00	4.50	4.00	6.00	7.00	1.00	3.00	1.50	4.50	1.50	5.00	6.00	2.00	3.00	7.00
.00	3.50	5.00	4.00	6.00	1.50	1.00	2.50	3.50	6.00					
3.50	.00	7.00	1.00	3.50	1.00	2.00	1.50	3.00	4.50					
5.00	7.00	.00	5.00	7.00	4.50	7.00	6.00	5.50	4.50					
4.00	1.00	5.00	.00	4.00	4.00	3.50	7.00	7.00	7.00					
6.00	3.50	7.00	4.00	.00	7.00	4.50	4.00	6.00	7.00					
1.50	1.00	4.50	4.00	7.00	.00	1.00	3.00	1.50	4.50					
1.00	2.00	7.00	3.50	4.50	1.00	.00	1.50	5.00	6.00					
2.50	1.50	6.00	7.00	4.00	3.00	1.50	.00	2.00	3.00					
3.50	3.00	5.50	7.00	6.00	1.50	5.00	2.00	.00	7.00					
6.00	4.50	4.50	7.00	7.00	4.50	6.00	3.00	7.00	.00					

CST ADD. = 4.50000

S_{12}

5.50	6.00	5.50	5.50	5.50	4.50	4.00	5.50	4.00	3.00	4.00	5.00	6.00	3.50	5.00
6.00	4.50	4.00	4.00	4.50	4.00	4.50	6.00	2.00	2.50	5.50	3.50	5.50	6.00	5.00
1.50	1.50	5.00	4.00	2.50	1.50	4.50	7.00	5.50	3.00	5.00	3.50	4.50	3.00	2.50
.00	5.50	6.00	5.50	5.50	5.50	4.50	4.00	5.50	4.00					
5.50	.00	3.00	4.00	5.00	6.00	3.50	5.00	6.00	4.50					
6.00	3.00	.00	4.00	4.00	4.50	4.00	4.50	6.00	2.00					
5.50	4.00	4.00	.00	2.50	5.50	3.50	5.50	6.00	5.00					
5.50	5.00	4.00	2.50	.00	1.50	1.50	5.00	4.00	2.50					
5.50	6.00	4.50	5.50	1.50	.00	1.50	4.50	7.00	5.50					
4.50	3.50	4.00	3.50	1.50	1.50	.00	3.00	5.00	3.50					
4.00	5.00	4.50	5.50	5.00	4.50	3.00	.00	4.50	3.00					
5.50	6.00	6.00	6.00	4.00	7.00	5.00	4.50	.00	2.50					

4.00	4.50	2.00	5.00	2.50	5.50	3.50	3.00	2.50	.00

CST ADD. = 1.50000

S_{13}

4.00	3.00	6.00	6.00	5.50	7.00	5.00	5.00	4.50	2.50	5.00	4.50	6.00	5.50	5.50
2.50	3.50	4.50	6.00	6.50	5.50	6.50	3.50	3.50	4.00	7.00	6.50	7.00	5.00	4.50
3.00	3.50	6.00	4.50	3.00	1.50	6.00	4.00	4.00	6.50	3.00	5.50	3.50	4.50	3.00

.00	4.00	3.00	6.00	6.00	5.50	7.00	5.00	5.00	4.50
4.00	.00	2.50	5.00	4.50	6.00	5.50	5.50	2.50	3.50
3.00	2.50	.00	4.50	6.00	6.50	5.50	6.50	3.50	3.50
6.00	5.00	4.50	.00	4.00	7.00	6.50	7.00	5.00	4.50
6.00	4.50	6.00	4.00	.00	3.00	3.50	6.00	4.50	3.00
5.50	6.00	6.50	7.00	3.00	.00	1.50	6.00	4.00	4.00
7.00	5.50	5.50	6.50	3.50	1.50	.00	6.50	3.00	5.50
5.00	5.50	6.50	7.00	6.00	6.00	6.50	.00	3.50	4.50
5.00	2.50	3.50	5.00	4.50	4.00	3.00	3.50	.00	3.00
4.50	3.50	3.50	4.50	3.00	4.00	5.50	4.50	3.00	.00

CST ADD. = .00000

S_{14}

1.00	4.00	5.00	5.50	2.50	4.50	2.00	4.00	2.50	2.50	6.00	6.00	6.50	5.50	5.00
4.00	4.50	6.00	4.00	6.00	6.00	4.50	6.00	3.50	3.50	5.00	6.50	6.50	5.50	2.50
3.50	2.50	6.00	5.00	2.00	1.50	6.50	5.00	4.50	4.00	4.00	5.00	4.00	4.00	2.50

.00	1.00	4.00	5.00	5.50	2.50	4.50	2.00	4.00	2.50
1.00	.00	2.50	6.00	6.00	6.50	5.50	5.00	4.00	4.50
4.00	2.50	.00	6.00	4.00	6.00	6.00	4.50	6.00	3.50
5.00	6.00	6.00	.00	3.50	5.00	6.50	6.50	5.50	2.50
5.50	6.00	4.00	3.50	.00	3.50	2.50	6.00	5.00	2.00
2.50	6.50	6.00	5.00	3.50	.00	1.50	6.50	5.00	4.50
4.50	5.50	6.00	6.50	2.50	1.50	.00	4.00	4.00	5.00
2.00	5.00	4.50	6.50	6.00	6.50	4.00	.00	4.00	4.00
4.00	4.00	6.00	5.50	5.00	5.00	4.00	4.00	.00	2.50
2.50	4.50	3.50	2.50	2.00	4.50	5.00	4.00	2.50	.00

CST ADD. = 3.00000

S_{15}

4.00	5.00	5.50	3.50	6.00	4.50	6.50	3.50	4.50	4.50	5.50	6.00	6.00	6.00	6.50
2.00	5.00	4.00	2.50	5.00	6.00	6.50	4.00	2.00	3.50	4.50	5.50	5.00	4.00	6.50
6.00	4.50	6.50	5.50	6.00	1.00	5.50	5.50	6.00	3.00	4.00	6.50	4.00	5.50	2.50

.00	4.00	5.00	5.50	3.50	6.00	4.50	6.50	3.50	4.50
4.00	.00	4.50	5.50	6.00	6.00	6.00	6.50	2.00	5.00
5.00	4.50	.00	4.00	2.50	5.00	6.00	6.50	4.00	2.00
5.50	5.50	4.00	.00	3.50	4.50	5.50	5.00	4.00	6.50
3.50	6.00	2.50	3.50	.00	6.00	4.50	6.50	5.50	6.00
6.00	6.00	5.00	4.50	6.00	.00	1.00	5.50	5.50	6.00
4.50	6.00	6.00	5.50	4.50	1.00	.00	3.00	4.00	6.50
6.50	6.50	6.50	5.00	6.50	5.50	3.00	.00	4.00	5.50
3.50	2.00	4.00	4.00	5.50	5.50	4.00	4.00	.00	2.50
4.50	5.00	2.00	6.50	6.00	6.00	6.50	5.50	2.50	.00

CST ADD. = 1.50000

S_{16}

2.00	5.50	7.00	7.00	4.50	5.00	2.00	3.50	5.50	2.00	4.00	6.00	2.50	4.50	4.00
3.00	2.50	2.50	1.00	5.00	5.50	4.00	2.50	2.00	1.50	4.00	4.00	6.00	5.00	2.50
1.50	4.50	4.00	5.00	2.50	3.00	4.50	5.00	3.50	3.00	4.00	5.50	2.50	5.00	3.00

.00	2.00	5.50	7.00	7.00	4.50	5.00	2.00	3.50	5.50
2.00	.00	2.00	4.00	6.00	2.50	4.50	4.00	3.00	2.50
5.50	2.00	.00	2.50	1.00	5.00	5.50	4.00	2.50	2.00
7.00	4.00	2.50	.00	1.50	4.00	4.00	6.00	5.00	2.50
7.00	6.00	1.00	1.50	.00	1.50	4.50	4.00	5.00	2.50
4.50	2.50	5.00	4.00	1.50	.00	3.00	4.50	5.00	3.50
5.00	4.50	5.50	4.00	4.50	3.00	.00	3.00	4.00	5.50
2.00	4.00	4.00	6.00	4.00	4.50	3.00	.00	2.50	5.00

I.U.C.A. NANC

3.50	3.00	2.50	5.00	5.00	5.00	4.00	2.50	.00	3.00	
5.50	2.50	2.00	2.50	2.50	3.50	5.50	5.00	3.00	.00	

CST ADD. = 3.00000

S_{17}

3.00	2.00	7.00	6.00	7.00	7.00	6.00	6.00	6.00	2.50	7.00	6.00	6.00	6.50	6.50
5.50	6.00	6.50	4.00	6.00	6.50	7.00	6.00	6.00	6.50	6.50	7.00	6.50	6.50	6.00
4.50	5.00	6.00	5.00	4.00	2.50	4.50	3.00	2.50	4.50	5.00	4.00	3.50	4.00	2.00
.00	3.00	2.00	7.00	6.00	7.00	7.00	6.00	6.00	6.00					
3.00	.00	2.50	7.00	6.00	6.00	6.50	6.50	5.50	6.00					
2.00	2.50	.00	6.50	4.00	6.00	6.50	7.00	6.00	6.00					
7.00	7.00	6.50	.00	6.50	6.50	7.00	6.50	6.50	6.00					
6.00	6.00	4.00	6.50	.00	4.50	5.00	6.00	5.00	4.00					
7.00	6.00	6.00	6.50	4.50	.00	2.50	4.50	3.00	2.50					
7.00	6.50	6.50	7.00	5.00	2.50	.00	4.50	5.00	4.00					
6.00	6.50	7.00	6.50	6.00	4.50	4.50	.00	3.50	4.00					
6.00	5.50	6.00	6.50	5.00	3.00	5.00	3.50	.00	2.00					
6.00	6.00	6.00	6.00	4.00	2.50	4.00	4.00	2.00	.00					

CST ADD. = .00000

S_{18}

4.50	6.50	6.50	6.50	7.00	4.50	6.50	4.50	6.00	6.50	6.50	6.50	7.00	7.00	6.50
5.50	6.50	5.00	6.00	6.50	6.50	6.50	6.00	5.50	6.00	7.00	7.00	6.50	6.50	6.50
6.50	5.00	6.50	6.50	6.00	6.00	6.00	4.00	7.00	7.00	7.00	7.00	6.50	7.00	5.50
.00	4.50	6.50	6.50	6.50	7.00	4.50	6.50	4.50	6.00					
4.50	.00	6.50	6.50	6.50	7.00	7.00	6.50	5.50	6.50					
6.50	6.50	.00	5.00	6.00	6.50	6.50	6.50	6.00	5.50					
6.50	6.50	5.00	.00	6.00	7.00	7.00	6.50	6.50	6.50					
6.50	6.50	6.00	6.00	.00	6.50	5.00	6.50	6.50	6.00					
7.00	7.00	6.50	7.00	6.50	.00	6.00	6.00	4.00	7.00					
4.50	7.00	6.50	7.00	5.00	6.00	.00	7.00	7.00	7.00					
6.50	6.50	6.50	6.50	6.50	6.00	7.00	.00	6.50	7.00					
4.50	5.50	6.00	6.50	6.50	4.00	7.00	6.50	.00	5.50					
6.00	6.50	5.50	6.50	6.00	7.00	7.00	7.00	5.50	.00					

CST ADD. = .00000

S_{19}

5.50	5.50	6.50	7.00	7.00	3.00	6.50	4.50	6.00	2.00	6.00	6.00	4.50	4.50	6.50
5.00	4.50	6.50	4.50	6.00	6.50	7.00	4.50	2.50	5.50	7.00	6.50	7.00	6.50	4.50
4.00	4.50	6.50	5.50	4.50	1.50	6.00	3.00	5.50	6.50	4.00	4.50	5.50	6.50	4.00
.00	5.50	5.50	6.50	7.00	7.00	3.00	6.50	4.50	6.00					
5.50	.00	2.00	6.00	6.00	4.50	4.50	6.50	5.00	4.50					
5.50	2.00	.00	6.50	4.50	6.00	6.50	7.00	4.50	2.50					
6.50	6.00	6.50	.00	5.50	7.00	6.50	7.00	6.50	4.50					
7.00	6.00	4.50	5.50	.00	4.00	4.50	6.50	5.50	4.50					
7.00	4.50	6.00	7.00	4.00	.00	1.50	6.00	3.00	5.50					
3.00	4.50	6.50	6.50	4.50	1.50	.00	6.50	4.00	4.50					
6.50	6.50	7.00	7.00	6.50	6.00	6.50	.00	5.50	6.50					
4.50	5.00	4.50	6.50	5.50	3.00	4.00	5.50	.00	4.00					
6.00	4.50	2.50	4.50	4.50	5.50	4.50	6.50	4.00	.00					

CST ADD. = 2.50000

S_{20}

3.00	4.00	6.00	6.00	4.00	4.50	5.00	3.50	5.50	4.00	5.00	3.50	4.50	4.00	6.00
5.50	3.00	5.00	2.50	5.00	5.50	5.00	4.00	3.00	3.00	4.00	4.50	6.00	5.00	4.50
3.00	3.50	5.00	5.00	4.50	2.00	5.50	4.50	5.50	3.00	4.00	3.50	2.50	5.50	1.00
.00	3.00	4.00	6.00	6.00	4.00	4.50	5.00	3.50	5.50					
3.00	.00	4.00	5.00	3.50	4.50	4.00	6.00	5.50	3.00					
4.00	4.00	.00	5.00	2.50	5.00	5.50	5.00	4.00	3.00					
6.00	5.00	5.00	.00	3.00	4.00	4.50	6.00	5.00	4.50					
6.00	3.50	2.50	3.00	.00	3.00	3.50	5.00	5.00	4.50					
4.00	4.50	5.00	4.00	3.00	.00	2.00	5.50	4.50	5.50					
4.50	4.00	5.50	4.50	3.50	2.00	.00	3.00	4.00	3.50					

5.00	6.00	5.00	6.00	5.00	5.50	3.00	.00	2.50	5.50
3.50	5.50	4.00	5.00	5.00	4.50	4.00	2.50	.00	1.00
5.50	3.00	3.00	4.50	4.50	5.50	3.50	5.50	1.00	.00

CST ADD. = 2.00000

S_{21}

2.00	3.50	4.50	5.00	4.00	5.50	4.00	5.50	4.50	5.00	6.50	4.50	4.00	4.00	4.00
3.50	4.50	5.00	6.50	5.00	5.50	7.00	5.00	4.50	2.50	4.00	3.50	6.50	5.50	4.50
3.50	3.00	6.50	5.00	3.50	1.00	6.00	2.50	4.50	5.50	3.00	5.00	3.00	4.50	2.00
.00	2.00	3.50	4.50	5.00	4.00	5.50	4.00	5.50	4.50					
2.00	.00	5.00	6.50	4.50	4.00	4.00	4.00	3.50	4.50					
3.50	5.00	.00	5.00	6.50	5.00	5.50	7.00	5.00	4.50					
4.50	6.50	5.00	.00	2.50	4.00	3.50	6.50	5.50	4.50					
5.00	4.50	6.50	2.50	.00	3.50	3.00	6.50	5.00	3.50					
4.00	4.00	5.00	4.00	3.50	.00	1.00	6.00	2.50	4.50					
5.50	4.00	5.50	3.50	3.00	1.00	.00	5.50	3.00	5.00					
4.00	4.00	7.00	6.50	6.50	6.00	5.50	.00	3.00	4.50					
5.50	3.50	5.00	5.50	5.00	2.50	3.00	3.00	.00	2.00					
4.50	4.50	4.50	4.50	3.50	4.50	5.00	4.50	2.00	.00					

CST ADD. = .50000

S_{22}

1.50	1.50	4.50	5.00	6.00	6.00	2.00	6.00	5.00	2.00	2.50	6.00	2.50	6.00	6.50
6.50	3.50	3.00	2.50	4.00	5.00	4.50	7.00	5.00	2.00	4.50	5.00	6.00	5.50	5.50
1.50	2.00	7.00	4.50	4.50	4.00	1.50	4.00	5.50	2.00	4.00	4.50	1.00	1.50	1.50
.00	1.50	1.50	4.50	5.00	6.00	6.00	2.00	6.00	5.00					
1.50	.00	2.00	2.50	6.00	2.50	6.00	6.50	6.50	3.50					
1.50	2.00	.00	3.00	2.50	4.00	5.00	4.50	7.00	5.00					
4.50	2.50	3.00	.00	2.00	4.50	5.00	6.00	5.50	5.50					
5.00	6.00	2.50	2.00	.00	1.50	2.00	7.00	4.50	4.50					
6.00	2.50	4.00	4.50	1.50	.00	4.00	1.50	4.00	5.50					
6.00	6.00	5.00	5.00	2.00	4.00	.00	2.00	4.00	4.50					
2.00	6.50	4.50	6.00	7.00	1.50	2.00	.00	1.00	1.50					
6.00	6.50	7.00	5.50	4.50	4.00	4.00	1.00	.00	1.50					
5.00	3.50	5.00	5.50	4.50	5.50	4.50	1.50	1.50	.00					

CST ADD. = 4.00000

S_{23}

3.00	5.00	6.50	5.00	5.00	4.50	3.50	4.50	2.50	2.50	5.00	3.50	4.00	4.00	5.00
3.00	3.00	3.00	5.00	4.50	3.50	5.00	3.00	3.00	3.50	4.50	5.00	7.00	3.50	2.00
3.00	3.00	5.00	2.50	1.50	3.00	3.00	3.00	3.00	4.00	2.00	3.00	3.00	4.00	1.50
.00	3.00	5.00	6.50	5.00	5.00	4.50	3.50	4.50	2.50					
3.00	.00	2.50	5.00	3.50	4.00	4.00	5.00	3.00	3.00					
5.00	2.50	.00	3.00	5.00	4.50	3.50	5.00	3.00	3.00					
6.50	5.00	3.00	.00	3.50	4.50	5.00	7.00	3.50	2.00					
5.00	3.50	5.00	3.50	.00	3.00	3.00	5.00	2.50	1.50					
5.00	4.00	4.50	4.50	3.00	.00	3.00	3.00	3.00	3.00					
4.50	4.00	3.50	5.00	3.00	3.00	.00	4.00	2.00	3.00					
3.50	5.00	5.00	7.00	5.00	3.00	4.00	.00	3.00	4.00					
4.50	3.00	3.00	3.50	2.50	3.00	2.00	3.00	.00	1.50					
2.50	3.00	3.00	2.00	1.50	3.00	3.00	4.00	1.50	.00					

CST ADD. = 2.00000

S_{24}

2.50	3.00	6.00	4.00	4.00	6.00	4.50	4.50	4.00	1.00	2.50	4.50	4.50	4.50	3.00
4.50	2.00	2.50	2.00	3.00	6.50	3.00	4.50	2.00	3.00	3.00	7.00	3.50	4.50	6.00
3.00	3.00	3.50	6.00	5.00	4.00	3.50	4.50	4.50	2.00	6.50	4.50	5.00	5.00	1.50
.00	2.50	3.00	6.00	4.00	4.00	6.00	4.50	4.50	4.00					
2.50	.00	1.00	2.50	4.50	4.50	4.50	3.00	4.50	2.00					
3.00	1.00	.00	2.50	2.00	3.00	6.50	3.00	4.50	2.00					
6.00	2.50	2.50	.00	3.00	3.00	7.00	3.50	4.50	6.00					
4.00	4.50	2.00	3.00	.00	3.00	3.00	3.50	6.00	5.00					
4.00	4.50	3.00	3.00	3.00	.00	4.00	3.50	4.50	4.50					

6.00	4.50	6.50	7.00	3.00	4.00	.00	2.00	6.50	4.50
4.50	3.00	3.00	3.50	3.50	3.50	2.00	.00	5.00	5.00
4.50	4.50	4.50	4.50	6.00	4.50	6.50	5.00	.00	1.50
4.00	2.00	2.00	6.00	5.00	4.50	4.50	5.00	1.50	.00

CST ADD. = 1.50000

S_{25}

1.50	3.00	6.00	6.50	6.50	6.50	4.50	5.50	4.50	2.50	5.50	6.50	7.00	6.00	7.00
3.00	3.00	3.00	4.00	7.00	6.00	7.00	3.00	3.50	6.00	7.00	6.50	7.00	6.00	6.50
2.00	2.00	6.50	4.00	2.50	4.00	7.00	4.50	5.00	5.00	4.50	4.50	3.50	4.50	2.50
.00	1.50	3.00	6.00	6.50	6.50	6.50	4.50	5.50	4.50					
1.50	.00	2.50	5.50	6.50	7.00	6.00	7.00	3.00	3.00					
3.00	2.50	.00	3.00	4.00	7.00	6.00	7.00	3.00	3.50					
6.00	5.50	3.00	.00	6.00	7.00	6.50	7.00	6.00	6.50					
6.50	6.50	4.00	6.00	.00	2.00	2.00	6.50	4.00	2.50					
6.50	7.00	7.00	7.00	2.00	.00	4.00	7.00	4.50	5.00					
6.50	6.00	6.00	6.50	2.00	4.00	.00	5.00	4.50	4.50					
4.50	7.00	7.00	7.00	6.50	7.00	5.00	.00	3.50	4.50					
5.50	3.00	3.00	6.00	4.00	4.50	4.50	3.50	.00	2.50					
4.50	3.00	3.50	6.50	2.50	5.00	4.50	4.50	2.50	.00					

CST ADD. = 1.00000

S_{26}

5.00	3.00	7.00	6.00	6.00	5.00	6.00	3.50	5.50	3.00	6.50	5.50	5.00	5.50	6.00
4.50	4.00	6.50	3.50	5.50	6.00	6.00	3.50	3.00	5.50	7.00	6.00	6.50	6.50	5.50
3.50	1.50	6.00	2.50	3.50	1.00	4.50	4.00	4.00	4.00	2.00	4.00	3.50	5.50	2.50
.00	5.00	3.00	7.00	6.00	6.00	5.00	6.00	3.50	5.50					
5.00	.00	3.00	6.50	5.50	5.00	5.50	6.00	4.50	4.00					
3.00	3.00	.00	6.50	3.50	5.50	6.00	6.00	3.50	3.00					
7.00	6.50	6.50	.00	5.50	7.00	6.00	6.50	6.50	5.50					
6.00	5.50	3.50	5.50	.00	3.50	1.50	6.00	2.50	3.50					
6.00	5.00	5.50	7.00	3.50	.00	1.00	4.50	4.00	4.00					
5.00	5.50	6.00	6.00	1.50	1.00	.00	4.00	2.00	4.00					
6.00	6.00	6.00	6.50	6.00	4.50	4.00	.00	3.50	5.50					
3.50	4.50	3.50	6.50	2.50	4.00	2.00	3.50	.00	2.50					
5.50	4.00	3.00	5.50	3.50	4.00	4.00	5.50	2.50	.00					

CST ADD. = 1.00000

S_{27}

2.00	7.00	6.00	6.50	3.50	6.50	3.00	4.00	5.50	6.00	6.50	6.50	3.50	7.00	3.00
4.50	6.50	3.50	3.00	6.00	3.50	7.00	6.00	3.00	4.50	6.50	5.00	6.00	5.50	5.00
6.00	3.00	7.00	6.50	3.50	4.50	6.50	3.00	6.00	4.50	5.00	4.00	3.50	5.50	4.50
.00	2.00	7.00	6.00	6.50	3.50	6.50	3.00	4.00	5.50					
2.00	.00	6.00	6.50	6.50	3.50	7.00	3.00	4.50	6.50					
7.00	6.00	.00	3.50	3.00	6.00	3.50	7.00	6.00	3.00					
6.00	6.50	3.50	.00	4.50	6.50	5.00	6.00	5.50	5.00					
6.50	6.50	3.00	4.50	.00	6.00	3.00	7.00	6.50	3.50					
3.50	3.50	6.00	6.50	6.00	.00	4.50	6.50	3.00	6.00					
6.50	7.00	3.50	5.00	3.00	4.50	.00	4.50	5.00	4.00					
3.00	3.00	7.00	6.00	7.00	6.50	4.50	.00	3.50	5.50					
4.00	4.50	6.00	5.50	6.50	3.00	5.00	3.50	.00	4.50					
5.50	6.50	3.00	5.00	3.50	6.00	4.00	5.50	4.50	.00					

CST ADD. = .00000

S_{28}

4.00	5.00	6.50	6.50	7.00	5.00	6.50	6.00	4.50	4.00	7.00	4.50	7.00	5.50	6.50
6.00	4.50	6.00	4.00	5.50	6.00	5.50	5.50	2.50	2.50	6.50	5.00	6.50	5.00	2.00
2.50	2.50	6.00	4.00	2.50	1.00	6.50	6.00	4.00	4.00	3.50	4.00	3.50	4.50	2.50
.00	4.00	5.00	6.50	6.50	7.00	5.00	6.50	6.00	4.50					
4.00	.00	4.00	7.00	4.50	7.00	5.50	6.50	6.00	4.50					
5.00	4.00	.00	6.00	4.00	5.50	6.00	5.50	5.50	2.50					
6.50	7.00	6.00	.00	2.50	6.50	5.00	6.50	5.00	2.00					
6.50	4.50	4.00	2.50	.00	2.50	2.50	6.00	4.00	2.50					

7.00	7.00	5.50	6.50	2.50	.00	1.00	6.50	6.00	4.00
5.00	5.50	6.00	5.00	2.50	1.00	.00	4.00	3.50	4.00
6.50	6.50	5.50	6.50	6.00	6.50	4.00	.00	3.50	4.50
6.00	6.00	5.50	5.00	4.00	6.00	3.50	3.50	.00	2.50
4.50	4.50	2.50	2.00	2.50	4.00	4.00	4.50	2.50	.00

CST ADD. = 1.50000

S_{29}

5.50	5.50	7.00	6.50	7.00	6.00	6.00	5.00	5.00	2.00	5.50	6.00	6.00	5.50	5.50
4.00	2.50	5.00	6.00	7.00	6.00	7.00	2.50	2.50	3.00	5.00	5.00	6.50	5.00	7.00
1.00	4.50	7.00	4.50	4.00	1.00	6.00	4.50	5.00	5.50	5.00	5.50	3.00	3.00	1.00
.00	5.50	5.50	7.00	6.50	7.00	6.00	6.00	5.00	5.00					
5.50	.00	2.00	5.50	6.00	6.00	5.50	5.50	4.00	2.50					
5.50	2.00	.00	5.00	6.00	7.00	6.00	7.00	2.50	2.50					
7.00	5.50	5.00	.00	3.00	5.00	5.00	6.50	5.00	7.00					
6.50	6.00	6.00	3.00	.00	1.00	4.50	7.00	4.50	4.00					
7.00	6.00	7.00	5.00	1.00	.00	1.00	6.00	4.50	5.00					
6.00	5.50	6.00	5.00	4.50	1.00	.00	5.50	5.00	5.50					
6.00	5.50	7.00	6.50	7.00	6.00	5.50	.00	3.00	3.00					
5.00	4.00	2.50	5.00	4.50	4.50	5.00	3.00	.00	1.00					
5.00	2.50	2.50	7.00	4.00	5.00	5.50	3.00	1.00	.00					

CST ADD. = 2.50000

S_{30}

2.00	3.50	6.50	6.00	5.50	5.00	3.50	3.50	4.00	3.00	3.00	5.00	4.00	6.00	4.00
4.50	3.00	5.50	6.00	6.50	5.00	6.00	3.00	3.00	2.00	5.00	3.50	6.50	4.00	4.00
2.00	2.00	5.00	4.00	2.00	2.00	4.00	2.00	3.00	3.50	2.50	2.00	1.50	3.00	2.00
.00	2.00	3.50	6.50	6.00	5.50	5.00	3.50	3.50	4.00					
2.00	.00	3.00	3.00	5.00	4.00	6.00	4.00	4.50	3.00					
3.50	3.00	.00	5.50	6.00	6.50	5.00	6.00	3.00	3.00					
6.50	3.00	5.50	.00	2.00	5.00	3.50	6.50	4.00	4.00					
6.00	5.00	6.00	2.00	.00	2.00	2.00	5.00	4.00	2.00					
5.50	4.00	6.50	5.00	2.00	.00	2.00	4.00	2.00	3.00					
5.00	6.00	5.00	3.50	2.00	2.00	.00	3.50	2.50	2.00					
3.50	4.00	6.00	6.50	5.00	4.00	3.50	.00	1.50	3.00					
3.50	4.50	3.00	4.00	4.00	2.00	2.50	1.50	.00	2.00					
4.00	3.00	3.00	4.00	2.00	3.00	2.00	3.00	2.00	.00					

CST ADD. = 1.50000

S_{31}

5.00	4.50	7.00	6.50	7.00	5.50	3.00	5.50	5.50	2.00	5.50	6.00	7.00	6.50	6.00
3.00	2.00	6.50	5.00	7.00	4.00	6.00	2.50	2.00	2.00	7.00	6.50	7.00	5.50	5.00
2.00	6.00	6.00	5.50	5.00	2.00	6.00	5.00	6.00	2.50	4.00	5.00	2.50	4.00	2.50
.00	5.00	4.50	7.00	6.50	7.00	5.50	3.00	5.50	5.50					
5.00	.00	2.00	5.50	6.00	7.00	6.50	6.00	3.00	2.00					
4.50	2.00	.00	6.50	5.00	7.00	4.00	6.00	2.50	2.00					
7.00	5.50	6.50	.00	2.00	7.00	6.50	7.00	5.50	5.00					
6.50	6.00	5.00	2.00	.00	2.00	6.00	6.00	5.50	5.00					
7.00	7.00	7.00	7.00	2.00	.00	2.00	6.00	5.00	6.00					
5.50	6.50	4.00	6.50	6.00	2.00	.00	2.50	4.00	5.00					
3.00	6.00	6.00	7.00	6.00	6.00	2.50	.00	2.50	4.00					
5.50	3.00	2.50	5.50	5.50	5.00	4.00	2.50	.00	2.50					
5.50	2.00	2.00	5.00	5.00	6.00	5.00	4.00	2.50	.00					

CST ADD. = 3.00000

S_{32}

3.50	5.00	6.00	6.00	5.50	5.50	4.00	3.50	6.00	2.00	5.50	5.00	3.00	4.50	4.50
4.00	3.50	4.50	3.00	6.50	3.50	6.00	5.50	2.00	3.50	5.50	7.00	6.50	4.00	4.00
2.50	4.00	6.00	5.00	3.00	2.50	4.50	2.50	4.50	4.00	4.50	4.50	3.50	5.50	2.00
.00	3.50	5.00	6.00	6.00	5.50	5.50	4.00	3.50	6.00					
3.50	.00	2.00	5.50	5.00	3.00	4.50	4.50	4.00	3.50					
5.00	2.00	.00	4.50	3.00	6.50	3.50	6.00	5.50	2.00					
6.00	5.50	4.50	.00	3.50	5.50	7.00	6.50	4.00	4.00					

6.00	5.00	3.00	3.50	.00	2.50	4.00	6.00	5.00	3.00
5.50	3.00	6.50	5.50	2.50	.00	2.50	4.50	2.50	4.50
5.50	4.50	3.50	7.00	4.00	2.50	.00	4.00	4.50	4.50
4.00	4.50	6.00	6.50	6.00	4.50	4.00	.00	3.50	5.50
3.50	4.00	5.50	4.00	5.00	2.50	4.50	3.50	.00	2.00
6.00	3.50	2.00	4.00	3.00	4.50	4.50	5.50	2.00	.00

CST ADD. = 1.50000

S_{33}

3.00	4.00	7.00	7.00	7.00	5.50	4.50	6.00	5.00	4.00	5.50	7.00	4.50	5.00	5.00
5.50	4.50	3.00	3.00	5.50	6.00	5.50	5.50	4.00	4.50	5.00	6.50	7.00	5.00	3.50
3.00	3.00	5.00	3.50	4.00	2.00	6.50	3.50	5.50	3.50	3.50	4.00	2.50	4.00	2.00
.00	3.00	4.00	7.00	7.00	7.00	5.50	4.50	6.00	5.00					
3.00	.00	4.00	5.50	7.00	4.50	5.00	5.00	5.50	4.50					
4.00	4.00	.00	3.00	3.00	5.50	6.00	5.50	5.50	4.00					
7.00	5.50	3.00	.00	4.50	5.00	6.50	7.00	5.00	3.50					
7.00	7.00	3.00	4.50	.00	3.00	3.00	5.00	3.50	4.00					
7.00	4.50	5.50	5.00	3.00	.00	2.00	6.50	3.50	5.50					
5.50	5.00	6.00	6.50	3.00	2.00	.00	3.50	3.50	4.00					
4.50	5.00	5.50	7.00	5.00	6.50	3.50	.00	2.50	4.00					
6.00	5.50	5.50	5.00	3.50	3.50	3.50	2.50	.00	2.00					
5.00	4.50	4.00	3.50	4.00	5.50	4.00	4.00	2.00	.00					

CST ADD. = 1.00000

S_{34}

3.50	5.50	7.00	6.00	7.00	5.50	3.50	3.50	4.00	2.50	2.50	4.50	3.00	5.00	5.50
4.00	2.50	2.00	3.00	5.50	4.50	3.00	3.00	2.00	2.00	3.50	3.50	5.50	3.00	2.50
3.50	5.00	5.00	4.00	2.50	3.50	6.50	3.00	4.00	4.50	4.00	4.50	4.00	5.50	2.00
.00	3.50	5.50	7.00	6.00	7.00	5.50	3.50	3.50	4.00					
3.50	.00	2.50	2.50	4.50	3.00	5.00	5.50	4.00	2.50					
5.50	2.50	.00	2.00	3.00	5.50	4.50	3.00	3.00	2.00					
7.00	2.50	2.00	.00	2.00	3.50	3.50	5.50	3.00	2.50					
6.00	4.50	3.00	2.00	.00	3.50	5.00	5.00	4.00	2.50					
7.00	3.00	5.50	3.50	3.50	.00	3.50	6.50	3.00	4.00					
5.50	5.00	4.50	3.50	5.00	3.50	.00	4.50	4.00	4.50					
3.50	5.50	3.00	5.50	5.00	6.50	4.50	.00	4.00	5.50					
3.50	4.00	3.00	3.00	4.00	3.00	4.00	4.00	.00	2.00					
4.00	2.50	2.00	2.50	2.50	4.00	4.50	5.50	2.00	.00					

CST ADD. = 1.00000

S_{35}

2.00	2.50	4.50	3.00	5.50	4.00	3.00	4.50	2.50	4.50	6.00	4.50	3.00	4.00	3.50
4.50	4.00	4.50	3.50	6.50	6.00	6.00	6.50	4.00	6.00	3.00	3.00	4.50	5.50	4.00
4.50	5.50	4.00	6.50	4.00	2.00	4.00	3.50	5.00	4.00	3.50	4.00	4.50	4.00	5.50
.00	2.00	2.50	4.50	3.00	5.50	4.00	3.00	4.50	2.50					
2.00	.00	4.50	6.00	4.50	3.00	4.00	3.50	4.50	4.00					
2.50	4.50	.00	4.50	3.50	6.50	6.00	6.00	6.50	4.00					
4.50	6.00	4.50	.00	6.00	3.00	3.00	4.50	5.50	4.00					
3.00	4.50	3.50	6.00	.00	4.50	5.50	4.00	6.50	4.00					
5.50	3.00	6.50	3.00	4.50	.00	2.00	4.00	3.50	5.00					
4.00	4.00	6.00	3.00	5.50	2.00	.00	4.00	3.50	4.00					
3.00	3.50	6.00	4.50	4.00	4.00	4.00	.00	4.50	4.00					
4.50	4.50	6.50	5.50	6.50	3.50	3.50	4.50	.00	5.50					
2.50	4.00	4.00	4.00	4.00	5.00	4.00	4.00	5.50	.00					

CST ADD. = .50000

S_{36}

1.50	2.00	5.50	5.00	6.00	4.50	4.00	3.50	3.00	3.00	7.00	6.00	5.50	5.00	5.00
3.00	4.00	6.50	6.00	6.00	6.50	6.00	5.00	4.50	4.00	4.50	5.50	6.50	5.50	6.00
3.00	2.00	5.50	4.00	3.00	2.50	6.50	3.50	5.00	5.50	2.50	4.50	5.00	5.00	4.00
.00	1.50	2.00	5.50	5.00	6.00	4.50	4.00	3.50	3.00					
1.50	.00	3.00	7.00	6.00	5.50	5.00	5.00	3.00	4.00					
2.00	3.00	.00	6.50	6.00	6.00	6.50	6.00	5.00	4.50					

I.U.C.A. NANC

5.50	7.00	6.50	.00	4.00	4.50	5.50	6.50	5.50	6.00
5.00	6.00	6.00	4.00	.00	3.00	2.00	5.50	4.00	3.00
6.00	5.50	6.00	4.50	3.00	.00	2.50	6.50	3.50	5.00
4.50	5.00	6.50	5.50	2.00	2.50	.00	5.50	2.50	4.50
4.00	5.00	6.00	6.50	5.50	6.50	5.50	.00	5.00	5.00
3.50	3.00	5.00	5.50	4.00	3.50	2.50	5.00	.00	4.00
3.00	4.00	4.50	6.00	3.00	5.00	4.50	5.00	4.00	.00

CST ADD. = .00000

S_{37}

5.50	4.00	7.00	7.00	7.00	7.00	4.00	7.00	7.00	2.50	4.00	5.00	4.00	5.50	6.50
3.00	6.00	2.50	5.50	5.00	6.50	7.00	3.50	3.50	3.00	5.00	4.00	7.00	3.00	2.00
2.50	3.00	7.00	5.00	5.50	1.00	5.00	2.00	3.00	6.00	2.50	6.00	4.00	7.00	1.50

.00	5.50	4.00	7.00	7.00	7.00	7.00	4.00	7.00	7.00
5.50	.00	2.50	4.00	5.00	4.00	5.50	6.50	3.00	6.00
4.00	2.50	.00	2.50	5.50	5.00	6.50	7.00	3.50	3.50
7.00	4.00	2.50	.00	3.00	5.00	4.00	7.00	3.00	2.00
7.00	5.00	5.50	3.00	.00	2.50	3.00	7.00	5.00	5.50
7.00	4.00	5.00	5.00	2.50	.00	1.00	5.00	2.00	3.00
7.00	5.50	6.50	4.00	3.00	1.00	.00	6.00	2.50	6.00
4.00	6.50	7.00	7.00	7.00	5.00	6.00	.00	4.00	7.00
7.00	3.00	3.50	3.00	5.00	2.00	2.50	4.00	.00	1.50
7.00	6.00	3.50	2.00	5.50	3.00	6.00	7.00	1.50	.00

CST ADD. = 2.00000

S_{38}

1.00	5.50	6.50	6.50	4.00	3.50	2.00	2.00	4.00	6.00	4.50	5.00	7.00	4.50	2.50
4.50	5.50	2.00	5.00	7.00	6.50	6.50	6.00	3.00	1.00	1.50	6.00	7.00	7.00	5.00
4.00	5.50	7.00	5.00	4.00	2.00	1.00	1.50	4.50	1.00	2.00	5.50	1.50	6.50	6.00

.00	1.00	5.50	6.50	6.50	4.00	3.50	2.00	2.00	4.00
1.00	.00	6.00	4.50	5.00	7.00	4.50	2.50	4.50	5.50
5.50	6.00	.00	2.00	5.00	7.00	6.50	6.50	6.00	3.00
6.50	4.50	2.00	.00	1.00	1.50	6.00	7.00	7.00	5.00
6.50	5.00	5.00	1.00	.00	4.00	5.50	7.00	5.00	4.00
4.00	7.00	7.00	1.50	4.00	.00	2.00	1.00	1.50	4.50
3.50	4.50	6.50	6.00	5.50	2.00	.00	1.00	2.00	5.50
2.00	2.50	6.50	7.00	7.00	1.00	1.00	.00	1.50	6.50
2.00	4.50	6.00	7.00	5.00	1.50	2.00	1.50	.00	6.00
4.00	5.50	3.00	5.00	4.00	4.50	5.50	6.50	6.00	.00

CST ADD. = 4.50000

S_{39}

2.00	4.50	6.50	6.00	6.00	6.50	2.50	4.00	5.50	1.50	6.50	6.00	6.00	6.00	4.50
4.00	6.50	3.50	5.00	6.50	5.50	6.00	6.50	4.00	5.00	6.00	4.50	6.00	5.50	4.00
5.00	3.00	5.50	5.00	2.00	1.00	1.50	2.00	4.00	1.50	2.00	4.00	1.50	6.00	2.50

.00	2.00	4.50	6.50	6.00	6.00	6.50	2.50	4.00	5.50
2.00	.00	1.50	6.50	6.00	6.00	6.00	4.50	4.00	6.50
4.50	1.50	.00	3.50	5.00	6.50	5.50	6.00	6.50	4.00
6.50	6.50	3.50	.00	5.00	6.00	4.50	6.00	5.50	4.00
6.00	6.00	5.00	5.00	.00	5.00	3.00	5.50	5.00	2.00
6.00	6.00	6.50	6.00	5.00	.00	1.00	1.50	2.00	4.00
6.50	6.00	5.50	4.50	3.00	1.00	.00	1.50	2.00	4.00
2.50	4.50	6.00	6.00	5.50	1.50	1.50	.00	1.50	6.00
4.00	4.00	6.50	5.50	5.00	2.00	2.00	1.50	.00	2.50
5.50	6.50	4.00	4.00	2.00	4.00	4.00	6.00	2.50	.00

CST ADD. = 2.50000

S_{40}

1.00	6.00	7.00	6.50	4.00	2.50	2.00	2.00	6.50	5.00	7.00	5.50	3.50	4.00	3.50
2.00	6.50	1.00	1.00	7.00	7.00	6.50	6.50	2.50	1.00	5.00	6.50	7.00	7.00	5.00
6.50	6.00	7.00	7.00	2.00	1.50	2.50	1.50	6.00	1.50	2.00	5.50	1.00	6.00	6.00

.00	1.00	6.00	7.00	6.50	4.00	2.50	2.00	2.00	6.50
1.00	.00	5.00	7.00	5.50	3.50	4.00	3.50	2.00	6.50

```
6.00   5.00    .00   1.00   1.00   7.00   7.00   6.50   6.50   2.50
7.00   7.00   1.00    .00   1.00   5.00   6.50   7.00   7.00   5.00
6.50   5.50   1.00   1.00    .00   6.50   6.00   7.00   7.00   2.00
4.00   3.50   7.00   5.00   6.50    .00   1.50   2.50   1.50   6.00
2.50   4.00   7.00   6.50   6.00   1.50    .00   1.50   2.00   5.50
2.00   3.50   6.50   7.00   7.00   2.50   1.50    .00   1.00   6.00
2.00   2.00   6.50   7.00   7.00   1.50   2.00   1.00    .00   6.00
6.50   6.50   2.50   5.00   2.00   6.00   5.50   6.00   6.00    .00
          CST ADD. =  2.00000

*STOP* TERMINE 3
JOB STEP 03 TERMINATED AT  11*01*47* AFTER 0000.23 MIN
.       CORE USED  0032     DISC USED  0014   WAIT TIME 0000.00
*************************************************************************
     TIME        TIME*CORE      CORE-USE       TIME*DISC     DISC-USE
     00.42         18.87            81%          17.15          15%

      DS          DS            SHR          SHR          CR          CP          LP
   I/O-BYTES I/O-CALLS     I/O-BYTES I/O-CALLS    CARDS       CARDS       PAGES
     811718       475          00          00      222          00          13
*ETTIJOB TERMINATED LONCHAMP-AB  IUCA  11*01*47*1
```

CHAPITRE IV

F'2, CORRELAT PHYSIQUE DE LA DEUXIEME DIMENSION PERCEPTIVE

I - INTRODUCTION

Le chapitre précédent a montré que, selon toute vraisemblance, l'espace perceptif vocalique ne comporte que deux dimensions. L'une de ces dimensions est le corrélat perceptif d'une caractéristique physique unique : la fréquence du premier formant. Par contre, pour des raisons que nous avons développées, et qui tiennent essentiellement au fait que la deuxième dimension peut être considérée comme le corrélat perceptif d'une combinaison linéaire de plusieurs formants, la méthode d'analyse multidimensionnelle ne permet pas de préciser le rapport exact entre cette dimension et les caractéristiques physiques des stimuli. L'étude de ces rapports passe par l'utilisation de méthodes différentes.

L'existence d'un espace bidimensionnel est parfaitement compatible avec la possibilité de synthétiser l'ensemble des voyelles à l'aide de deux formants seulement. Mais on sait, depuis de nombreuses années, que cette synthèse exige que le deuxième formant des voyelles antérieures soit placé à une fréquence plus élevée que celui des voyelles naturelles. Ce résultat a été obtenu bien avant les premières analyses multidimensionnelles. On a donné au deuxième formant utilisé en synthèse le nom de F'2, pour le différencier du deuxième formant des voyelles naturelles.

Ce chapitre est donc consacré à l'étude de F'2. Après avoir présenté les études antérieures sur le rôle perceptif des formants supérieurs à F1, nous étudierons les raisons de la difficulté d'une étude directe de F'2. On montrera en particulier que les estimations de F'2 sont peu fiables pour les voyelles antérieures. Après avoir simulé un modèle simple de la détection de F'2 qui ne donne pas les résultats escomptés, nous présenterons une méthode originale d'estimation de F'2. Nous terminerons ce chapitre par la description d'un modèle de perception dans un espace bidimensionnel. Ce modèle sera testé sur des données expérimentales provenant d'une expérience de perception de voyelles synthétiques. Ce chapitre, s'il ne constitue pas une "preuve" de l'existence d'un espace perceptif

bidimensionnel, montre néanmoins que cette hypothèse est compatible avec un large ensemble de résultats expérimentaux.

II - F'2 : LE DEUXIEME FORMANT "EQUIVALENT"

FANT (1959, p. 80) définit F'2 comme le "deuxième formant efficace", c'est-à-dire la fréquence formantique (supérieure) qui permet la synthèse vocalique à deux formants. Nous préférons le terme "équivalent", qui indique que F'2 équivaut perceptivement aux formants F2, F3 et F4.

Les chercheurs des laboratoires HASKINS (DELATTRE et al. 1952) ont montré que la synthèse vocalique à deux formants permettait d'obtenir un taux d'identification voisin de 80 % pour 10 voyelles françaises. Un taux d'identification de 100 % peut être obtenu pour [i] et [e] à condition de placer le deuxième formant équivalent à 2900 et 2400 Hz, c'est-à-dire à des fréquences nettement supérieures à celles observées pour des voyelles naturelles. Ces résultats ont été confirmés par MILLER (1953). F'2 est égal à F2 pour toutes les voyelles postérieures et les antérieures labialisées. Les résultats qualitatifs de FANT (1959, p. 80) indiquent que F'2 doit être à une fréquence équivalente à F3 pour [e] et proche de F4 pour [i]. La différence entre ces résultats et ceux de DELATTRE et al. est attribuée à une technique de synthèse différente. FANT (op. cit., p. 80) propose une formule de calcul de F'2 à partir des fréquences de F2 et F3, tout en indiquant immédiatement que d'autres formules "qui augmenteraient le poids de F3 pourraient être préférables" :

$$F'2 = F2 + 1/2\ (F3 - F2)\ ((F2 - F1)\ /\ (F3 - F1))$$

La proximité de F2 et F1 pour les voyelles ouvertes et postérieures diminue le coefficient (F2 - F1) / (F3 - F1), et minore de façon appréciable la pondération 1/2 (F3 - F2). Pour F2 proche de F3, (i , e) , F'2 s'établit à égale distance de F2 et F3.

En 1970, CARLSON et al. ont présenté une importante étude expérimentale de F'2, faisant suite aux travaux préliminaire

de FANT et RISBERG (1963). Cette étude a conduit (CARLSON, FANT et GRANSTROM 1975) à une nouvelle formule de calcul de F'2 :

$$F'2 = (F2 + C(F3 \cdot F4)^{1/2}) / (1 + C) \; ; C = (F1/500)^2 \cdot ((F2 - F1)/(F4 - F3))^4 \cdot ((F3 - F2)/(F3 - F1))^2$$

Cette formule ne nous paraît pas satisfaisante pour les voyelles antérieures fermées, car certains résultats expérimentaux de CARLSON et al. (1970) sont discutables.

L'importance de la notion de F'2 nous a conduit à étudier la validité des procédures utilisées pour la détermination de F'2.

III - ECHEC DE NOTRE METHODE EXPERIMENTALE

Apparemment, la mesure de F'2 ne pose pas de problèmes. Deux méthodes peuvent être envisagées. CARLSON et al. (1970) demandaient à des sujets de comparer des voyelles synthétiques à 4 et 2 formants. Pour les voyelles à 2 formants, ceux-ci devaient ajuster la fréquence de F'2 de manière à produire un timbre identique à celui de la voyelle à 4 formants.

Les résultats montrent une dispersion importante de l'estimation de F'2 pour la voyelle [i]. Les résultats extrêmes diffèrent de près de 1000 Hz (2700 - 3700 Hz ; moyenne 3210 Hz). Au contraire, pour la voyelle [y], les variations ne dépassent pas 250 Hz (1900 - 2150 ; moyenne 2010 Hz). Des écarts importants sont également observés pour [e] et [æ]. Cette variabilité semble interdire toute étude détaillée de F'2. Les problèmes essentiels se posent pour les voyelles [i] et [y]. F'2 est égal à F2 pour les voyelles postérieures.

Désireux d'étudier à notre tour la fréquence de F'2 pour les voyelles antérieures [i] et [y], et ne pouvant, pour des raisons techniques, utiliser la même procédure que CARLSON et al., nous avions envisagé une deuxième méthode. Lorsque F'2 varie de 2000 à 3000 Hz et que F1 est fixé à 250 ou 300 Hz, la voyelle perçue passe de [y] à [i]. Nous

pensions faire identifier une série de voyelles synthétiques pour F'2 variant de 2000 à 3000 Hz, et choisir comme estimation de F'2, la valeur minimale donnant un pourcentage d'identification supérieur ou égal à 95 %. Un deuxième test, portant sur des voyelles à 4 formants, aurait permis d'estimer les combinaisons de F2, F3 et F4 donnant au minimum 95 % de réponses [i].

Ces deux séries de voyelles ont été réalisées sur le synthétiseur à formants que nous avons simulé sur le mini-ordinateur NOVA 2 du Laboratoire d'Electricité et d'Automatique de l'Université de Nancy I*. Ce synthétiseur est décrit en annexe de cette thèse. Précisons simplement que le choix des paramètres a été effectué de manière à obtenir la meilleure qualité de synthèse possible. L'onde glottale est la fonction jugée la plus naturelle par ROSENBERG (1971). La fréquence fondamentale et l'intensité sont affectées de micro-modulations aléatoires qui augmentent dans une large mesure le "naturel". Tous les paramètres sont variables à chaque période, ce qui permet une attaque naturelle. Le pas de quantification des paramètres est très fin : 1 % pour F_o, 1 dB pour l'intensité, quelques Hz pour les fréquences formantiques. Les voyelles sont filtrées par un filtre passe-bande 70 - 4700 Hz à - 36 dB par octave. Ce filtrage élimine tout bruit de synthèse ainsi que le bruit d'échantillonnage. 4 formants sont utilisés, car des tests informels ont montré que la présence d'un cinquième formant n'a pas d'effet sur la qualité perçue.

A/ Synthèse de voyelles à 4 formants

Pour la voyelle [i], de nombreuses sessions d'écoute de voyelles synthétiques (à 4 formants) ont montré que la synthèse la plus naturelle était obtenue pour les valeurs formantiques suivantes ;

---

* Je remercie Monsieur M. LAMOTTE d'avoir mis à notre disposition le matériel du groupe de recherche qu'il dirige.

F1 300 Hz
F2 2300
F3 2900
F4 3400

Ces valeurs n'ont pas été vérifiées par un groupe de sujets. Mais les avis recueillis auprès de phonéticiens et de membres du laboratoire ont montré que cette voyelle était décrite sans hésitation comme un [i].

Pour obtenir une voyelle acceptable, les fréquences formantiques doivent être supérieures aux limites (approximatives) suivantes :

F2 2100 Hz
F3 2700 Hz
F4 3400 Hz

Les remarques suivantes ont été faites pendant ces séances d'audition :

- Une fréquence de F1 inférieure ou égale à 250 Hz donne une qualité "sourde" ou "étouffée" à la voyelle [i]. Pour F1 = 300 Hz, aucune confusion avec [e] n'est possible et la qualité de [i] est bien meilleure. La voyelle synthétique ne "vire" à [e] que pour une fréquence de F1 proche de 400 Hz.

- La plage de variation acceptable pour un formant supérieur dépend, dans une large mesure, des fréquences des formants voisins. On peut encore obtenir un timbre [i] acceptable pour F2 = 2100 Hz, à condition que F3 soit proche de F4 (3000 - 3400 Hz respectivement). Une autre combinaison acceptable est F2 = 2350 Hz, F3 - 2500 Hz, F4 = 3400 Hz. Il est encourageant de constater que des valeurs proches des moyennes formantiques "naturelles" donnent un timbre de bonne qualité pour [i](cf. infra). Ces variations importantes résultent de la sensibilité de l'intensité des formants à la distance inter-formantique. FANT (1956) a montré que l'intensité de deux formants augmente de 6 dB lorsque la distance qui les sépare est réduite de moitié. Pour F2 et F3 proches, on constate que l'amplitude du groupe F2 - F3 est plus importante que l'amplitud

de F4. Pour F3 proche de F4 la situation est inversée. Il est donc possible de synthétiser une voyelle de timbre [i] avec des valeurs formantiques "paradoxales". Il y a là une difficulté supplémentaire pour l'évaluation de F'2.

B/ Synthèse de voyelles à 2 formants

Nous avons synthétisé une série de voyelles dont le deuxième formant (F'2) variait de 2000 à 3000 Hz (F1 = 300 Hz). Pour F'2 inférieur à 2500 Hz, la voyelle perçue est une voyelle relativement sourde, à laquelle il est peut-être possible d'attacher le label [y]. Mais, au-delà de cette fréquence, la voyelle se présente comme un "accord musical" dans lequel on distingue une "basse continue" (F1) et un son aigrelet de fréquence élevée (F'2). Par déduction logique, la seule voyelle compatible avec ce timbre est [i]. Mais ce son n'a qu'un rapport lointain avec le timbre de la voyelle naturelle. FANT (1975) indique d'ailleurs que la voyelle [i] à 2 formants possède un timbre un peu "maigre" ("thin").

Il ne nous a pas semblé souhaitable (ou honnête) de poursuivre cette expérience d'identification du timbre d'une voyelle à 2 formants. Nous sommes nous-mêmes dans l'impossibilité de déterminer la zone de passage entre le timbre de [y] et [i]. L'écoute de ces stimuli synthétiques donne la clé des écarts constatés par CARLSON et al. (1970). Il est très difficile d'associer le timbre d'une voyelle à 2 formants à celui d'une voyelle à 4 formants dans cette zone de fréquence.

Nous concluons donc à l'échec de notre procédure expérimentale. De plus, nous sommes convaincus que les résultats de CARLSON et al. pour la voyelle [i] doivent être utilisés avec une extrême prudence.

Le paragraphe suivant est une revue des résultats qui mettent en doute la nouvelle formule de calcul de F'2 (CARLSON et al. 1975) fondée sur les résultats expérimentaux obtenus en 1970.

IV - ESTIMATION DE F'2 : LES DIFFICULTES

Il paraît donc difficile, pour ne pas dire impossible, d'estimer directement la fréquence de F'2. La comparaison directe entre le timbre d'une voyelle à 4 et à 2 formants (CARLSON et al.) donne lieu à des écarts très importants pour la voyelle [i]. Il est également impossible d'associer avec précision un label phonétique à une voyelle à 2 formants quand la fréquence de F'2 dépasse 2500 Hz environ.

Avant d'étudier un certain nombre de résultats qui mettent en doute la précision des estimations de CARLSON et al., il convient d'étudier un autre aspect de la méthode : Est-il possible de choisir une voyelle à 4 formants qui soit véritablement représentative de l'ensemble des voyelles [i] observées ?

A/ La variabilité des fréquences formantiques pour les voyelles naturelles

L'établissement d'une formule de calcul de F'2 à partir des 4 premiers formants n'est possible que si leurs fréquences sont relativement stables. C'est l'hypothèse que font les auteurs suédois quand ils n'utilisent pour [i] qu'une seule combinaison de fréquences formantiques (225, 2065, 2960 et 3400 Hz).

Pour vérifier ce point, nous avons rassemblé les résultats expérimentaux portant sur un nombre élevé de locuteurs dans diverses langues. La comparaison suppose que les conditions expérimentales n'ont qu'une faible influence sur les résultats, et que la voyelle [i] est phonétiquement identique dans toutes les langues considérées.

Langue		F1	F2	F3	F4 (Hz)	Contexte	Référence
anglais	américain	270	2290	3010	—	h - d	1
"	"	280	2320	2950	—	h - d	2
"	"	300	2340	2920	—	isolée	2
"	"	300	2260	2910	—	C - C	2
suédois		256	2066	2960	3400	isolée	3
"		255	2190	3150	3730	isolée	4

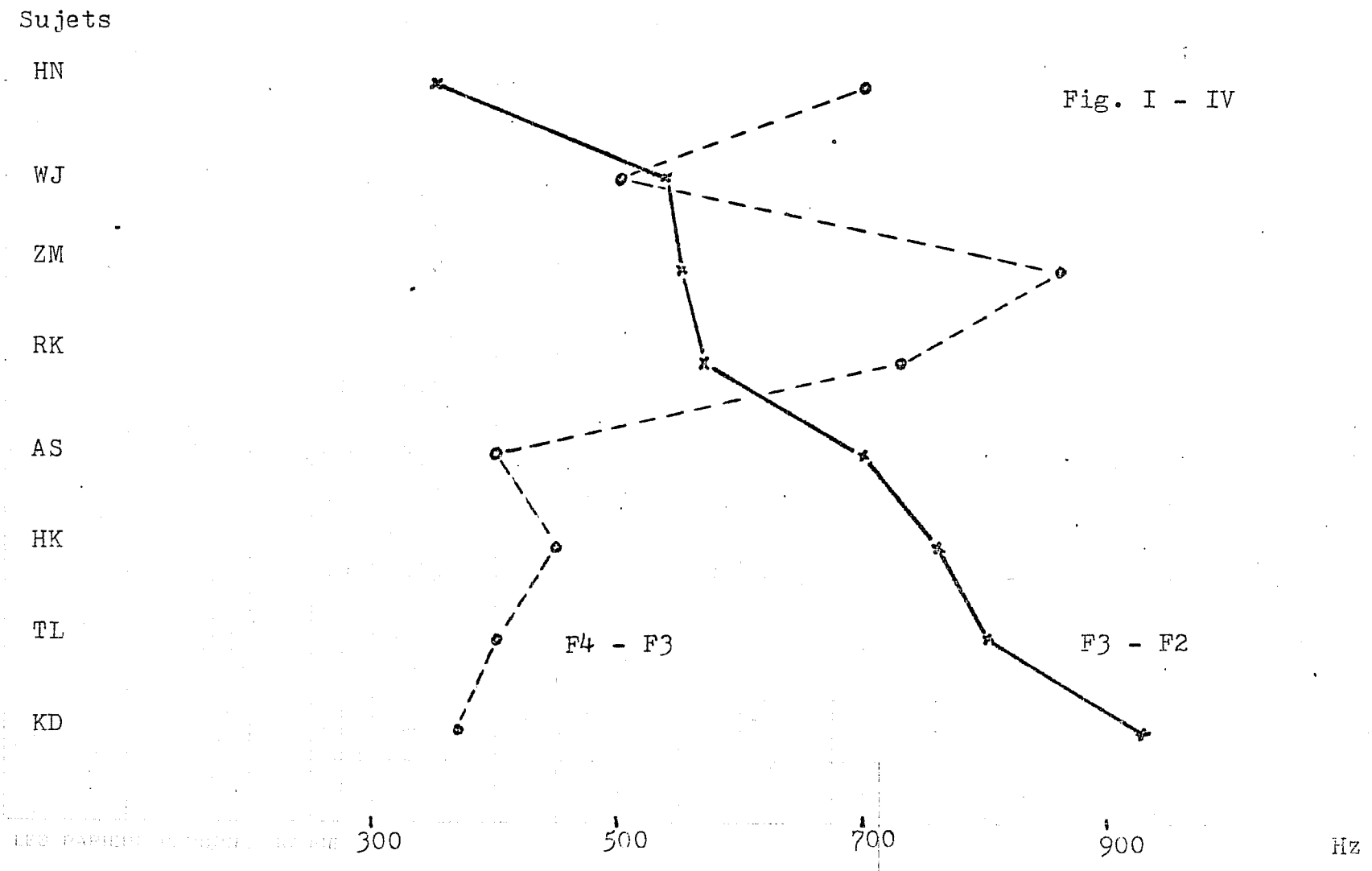


Fig. I - IV

Langue	F1	F2	F3	F4	Contexte	Référence
néerlandais	294	2208	2766	—	h - t	5
polonais	213	2310	2960	3500	isolée	6
français	281	2093	2973	—	C - C	7

1 : PETERSON et BARNEY : hommes (1952)
2 : STEVENS et HOUSE (1963)
3 : FANT (1959)
4 : FANT, HENNINGSSON et STÅLHAMMAR (1969)
5 : POLS, TROMP et PLOMP (1973)
6 : JASSEM (1968)
7 : DEBROCK et FORREZ (1976).

Le faible écart, inférieur à 10 % en moyenne, est encourageant, d'autant que les conditions expérimentales sont variées. On note cependant quelques valeurs "aberrantes" : F2, FANT 1959; F3, POLS et al. 1973; F1, JASSEM 1968. Encourageante également est la comparaison avec les valeurs optimales que nous avons obtenues par la synthèse à 4 formants (300, 2300, 2900, 3400 Hz). La reproduction des valeurs moyennes donne donc une synthèse acceptable. Cette table révèle en outre que la valeur de F2 utilisée par CARLSON et al. (1970) est trop faible.

Mais l'accord apparent des valeurs moyennes cache en fait des variations individuelles très importantes. L'analyse des données individuelles publiées par JASSEM (1968) révèle un phénomène intéressant, qui ne semble pas avoir été noté à ce jour. La fig. 1-IV, basée sur les données de 8 locuteurs masculins, (table I, p. 24-25, op. cit.) montre une corrélation négative ($r = -0,875$) entre les différences F3 - F2, et F4 - F3. Ces différences sont les moyennes de 5 mesures pour chaque locuteur. Cette corrélation implique que la séparation entre F2 et F3 est d'autant plus importante que la distance F3 - F4 est faible. En pourcentage, les distances F3 - F2 représentent de 33 à 71,5 % de la distance F4 - F2. En valeur absolue les distances entre F2 et F3 s'étagent de 350 à 930 Hz ; les différences F4 - F3 varient entre 370 et 860 Hz. Nous avons choisi

d'étudier les différences, et non les valeurs absolues, car cette procédure minimise les variations imputables aux différences anatomiques. Cette corrélation inverse confirme une des conclusions que nous avons tirées de notre expérience limitée de synthèse à 4 formants. Un timbre [i] acceptable peut être obtenu par des combinaisons relativement différentes de F2, F3 et F4. Pour la détermination de F'2, une difficulté supplémentaire apparaît donc : la formule de calcul de F'2 devra être compatible, non seulement avec une combinaison particulière de fréquences formantiques, mais également avec toute la gamme des variations observées dans la parole naturelle.

B/ Inadéquation de la nouvelle formule de calcul de F'2 (CARLSON et al. 1974).

Cette nouvelle formule de calcul de F'2 a été présentée au paragraphe II. Son application conduit à des résultats contradictoires avec un certain nombre de données expérimentales.

a) Dans l'article même de CARLSON et al. (1970), les résultats détaillés d'une expérience complémentaire, dans laquelle l'intensité et la fréquence de F'2 étaient variables, montrent que celui-ci ne dépasse jamais 3050 Hz, alors qu'il atteint une moyenne de 3210 Hz

dans l'expérience principale. Pour la voyelle [e] en revanche, les résultats sont identiques, ce qui indique que les deux expériences peuvent être comparées (op. cit., fig. II-A-3)

b) Dans ce même travail, une autre expérience complémentaire (fig. II-A-9) compare l'estimation auditive de F'2 à des voyelles dont le 3ème formant varie de 2000 à 2800 Hz. Nous avons reproduit les résultats moyens à la fig. 2-IV, ainsi que les valeurs obtenues par la formule. Celle-ci sous-estime F'2. D'autre part, comme le reconnaissent les auteurs, les estimations de F'2 sont presque dichotomiques : Pour F3 variant de 2000 à 2650 Hz, F'2 varie de 2250 à 2650 Hz. Pour F3 supérieur à 2700 Hz, F'2 saute brutalement à 3000 Hz et plus. On peut faire l'hypothèse d'une sorte de "frontière phonématique"

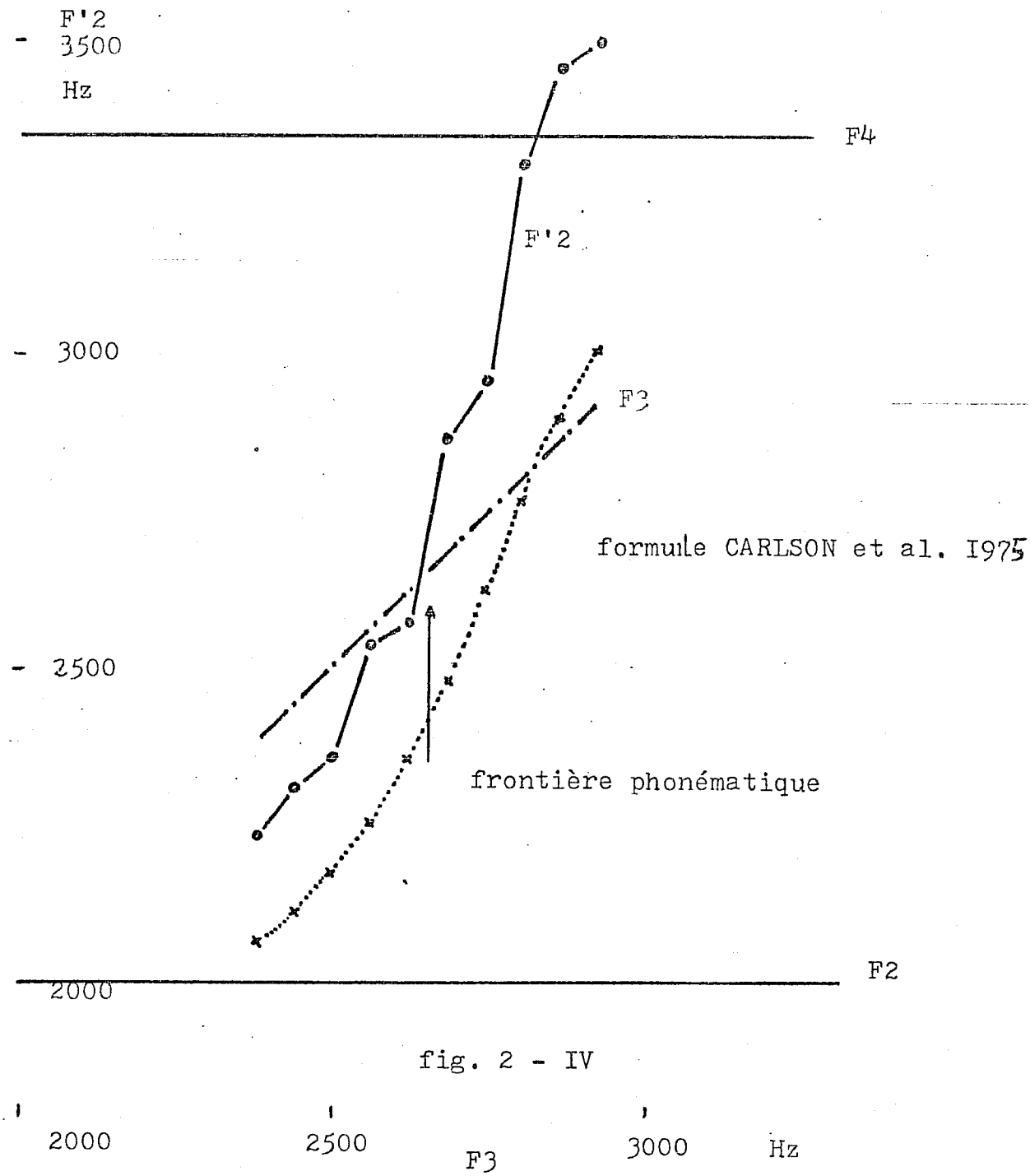

F'2
3500
Hz
F4
F'2
3000
F3
formule CARLSON et al. 1975
2500
frontière phonématique
F2
2000
2000
2500
F3
3000
Hz

fig. 2 - IV

vers 2650 - 2700 Hz. Cette frontière se place exactement à mi-distance de F3 et F4. Nous donnerons plus loin une tentative d'explication du comportement de F'2.

c) Un travail récent (FANT, CARLSON et GRANSTRÖM 1974) sur l'ambiguïté potentielle des réalisations de [e] et [ϕ] pour des locuteurs masculins et féminins révèle que la formule ne peut rendre compte des résultats expérimentaux : Si F1, F2 et F3 peuvent être identiques pour un [e] masculin et un [ϕ] féminin, la valeur de F4, qui dépend principalement de la longueur du conduit, est différente. La plus grande élévation de F4 provoque un abaissement de F'2, ce qui est contraire aux prédictions de la formule.

d) enfin, comme nous le verrons plus loin, les prédictions ne sont pas compatibles avec les résultats de FUJIMURA (1967).

Il convient sans doute d'améliorer les prédictions en utilisant les données expérimentales ci-dessus. Nous proposerons plus loin une nouvelle technique d'estimation de F'2. Mais auparavant, il convient d'expliquer les causes des difficultés rencontrées dans l'estimation auditive directe de F'2.

V - LA REALITE PERCEPTIVE DE F'2 : SPECTRES DE SONIE ET EXPERIENCES DE MASQUAGE

Ce paragraphe rassemble quelques conclusions d'expériences assez disparates qui permettent de comprendre les difficultés d'une étude expérimentale directe de F'2. Ces conclusions suggèrent également que le "deuxième formant équivalent" a bien une réalité perceptive.

Pour la voyelle [i] la zone fréquentielle contenant F2, F3 et F4 est perçue comme un "bloc" unique, mais largement étalé en fréquence. Nous émettons l'hypothèse que cette large répartition est la cause des variations observées dans l'estimation de F'2.

BRANDT et THOMAS (1967 ; non publié; cités dans BRANDT 1976) ont mesuré le seuil de détection de sinusoïdes masquées

par un bruit blanc ayant la même répartition spectrale que la voyelle [i] (fig. 3-IV). On remarquera que le groupe F2 - F3 du spectre provoque l'apparition d'un pic assez large dans le tracé du seuil de masquage. La différence de niveau global est sans doute attribuable à la diffraction de la tête du sujet et au conduit de l'oreille externe. Le décalage que l'on observe pour F1 n'est pas explicable. Les résultats pour [æ] indiquent que F2 et F3 provoquent une élévation du seuil entre F2 et F3.

HOUTGAST (1974) présente également une expérience de masquage de sinusoïdes par un spectre comprenant des formants. Il fait appel à un nouveau paradigme utilisant le seuil de pulsations. Les spécialistes de psychophysique s'accordent pour considérer ce paradigme comme nettement supérieur aux méthodes antérieures. Il est compatible avec les phénomènes d'inhibition latérale révélés par l'électrophysiologie. La fig. 4-IV présente les résultats pour 3 spectres correspondant approximativement à [i], [ɛ] et [ɑ]. La ligne supérieure (2) est le spectre mesuré par un filtre 1/3 d'octave. La ligne (1) est le spectre acoustique ; la ligne (3) présente la moyenne des estimations du seuil de masquage. Pour la voyelle [i], le seuil présente un "plateau" entre 2000 et 3000 Hz. La "vallée" acoustique entre les formants 2 et 3 est comblée. On comprend qu'il soit difficile de placer un formant unique de manière à provoquer la même impression auditive que celle produite par un "bloc" fréquentiel s'étendant sur près de 1000 Hz. Ce "plateau" est mis en valeur par une "vallée" placée un peu en-dessous de 2000 Hz, et qui pourrait être une trace d'inhibition latérale. Les seuils de masquage pour les voyelles [ɛ] et [ɑ] montrent, au contraire, 2 pics dans la zone des 2ème et 3ème formants. Mais le pic correspondant à F3 est moins élevé.

Des résultats similaires sont obtenus par la mesure des spectres de sonie. Cette technique est jugée "utile" par FANT (1975). L'annexe B présente les concepts et les résultats fondamentaux conduisant au calcul du spectre de sonie.

Les auteurs suivants ont publié des spectres de sonie pour les voyelles [i] et [y] :

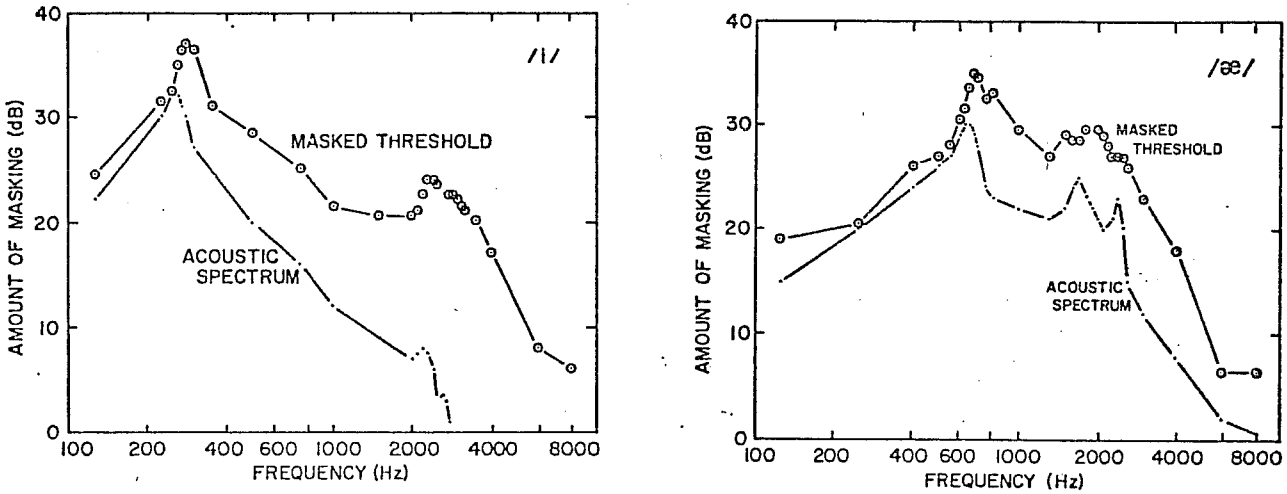


Fig. 3 - IV

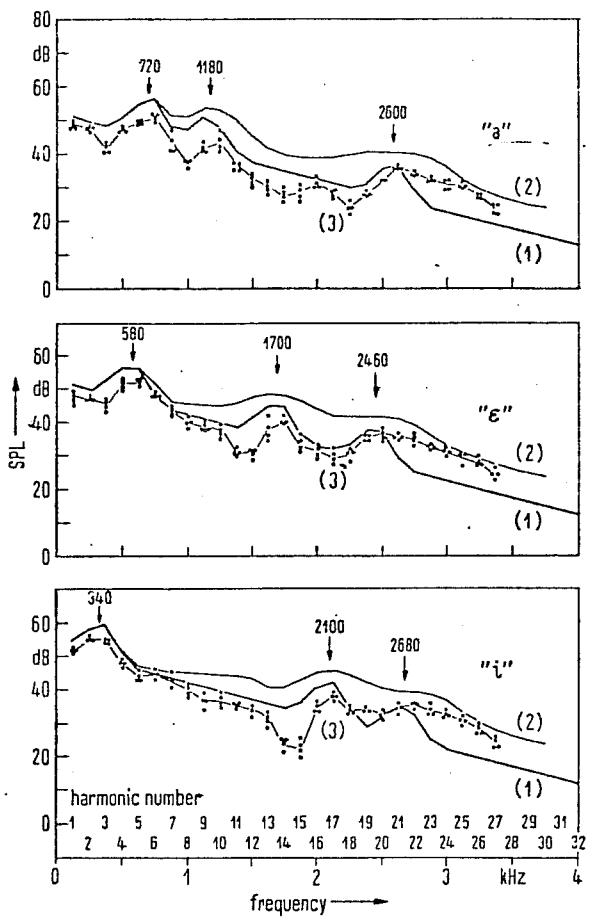
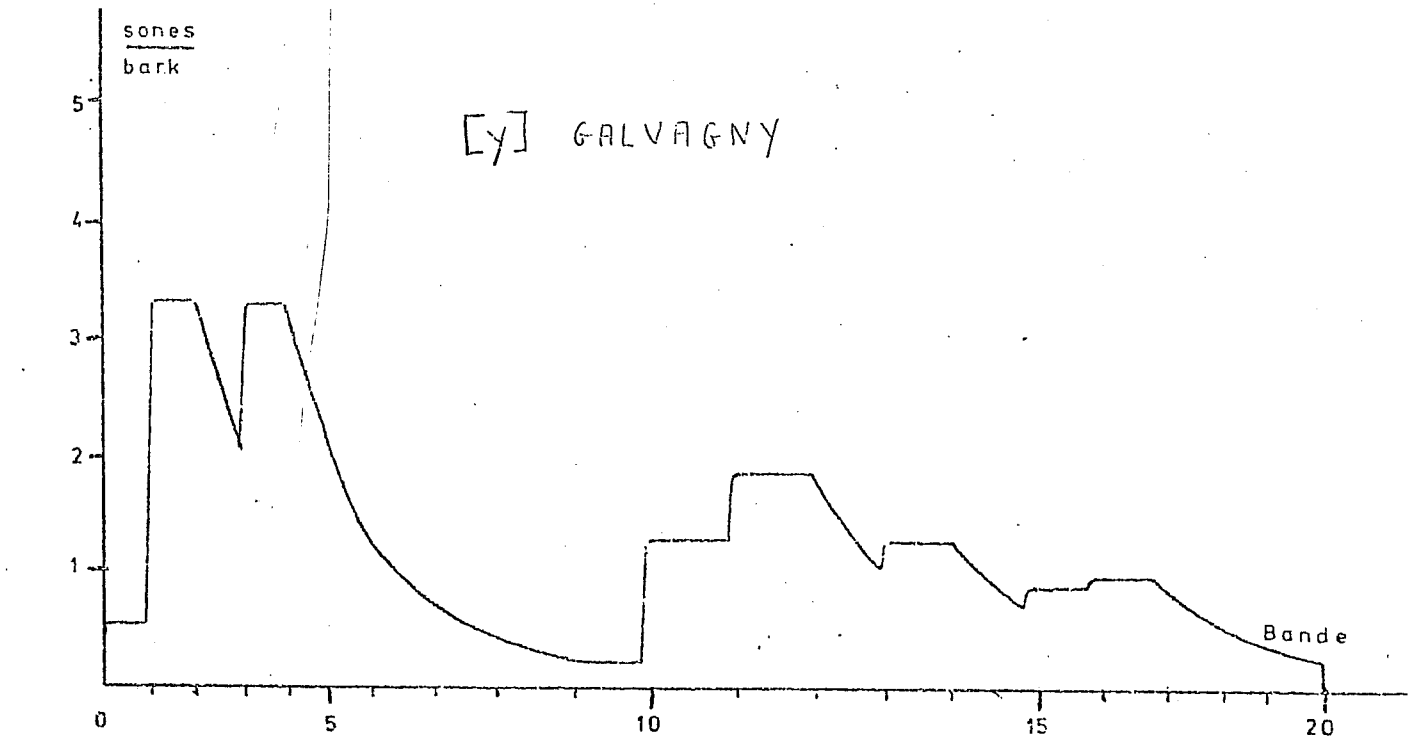
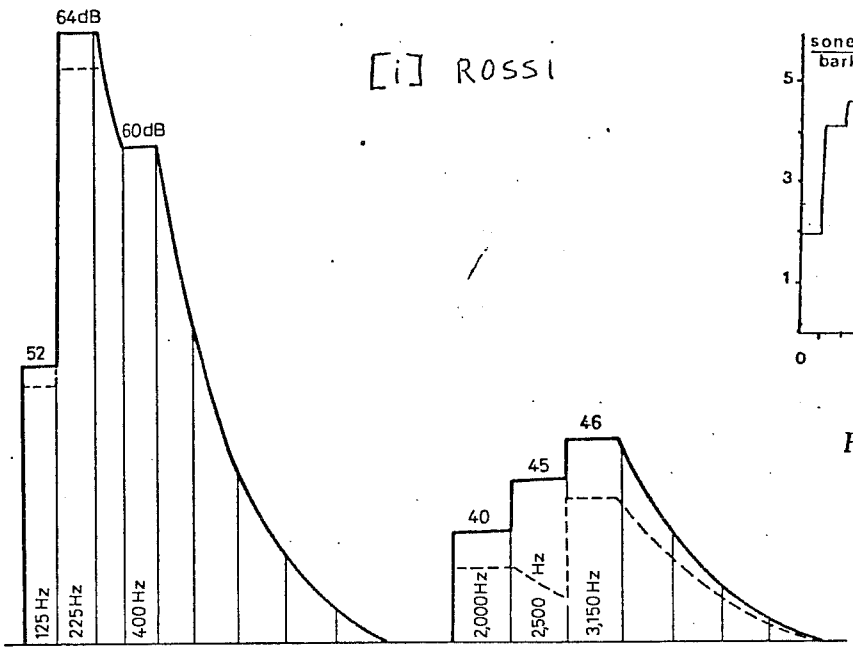


FIg. 4 - IV





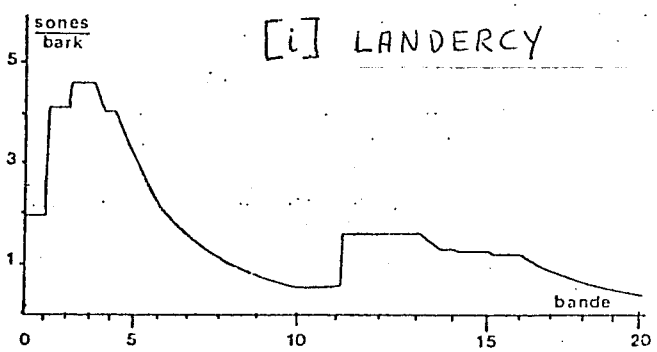
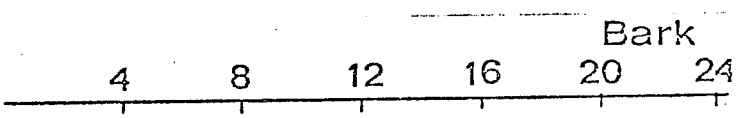
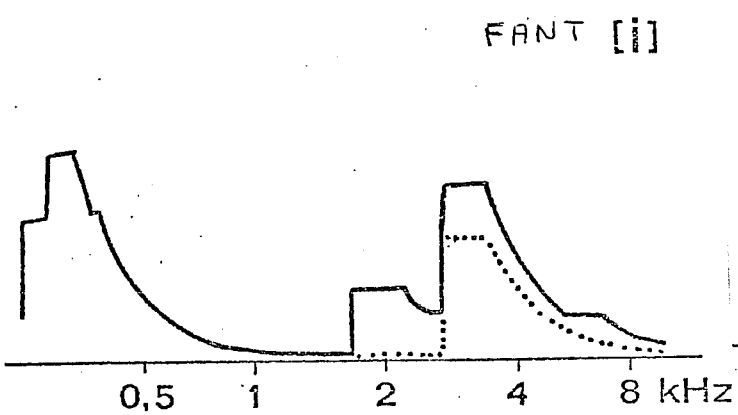


Fig. 11. — Sprectres de sonie du même [i].





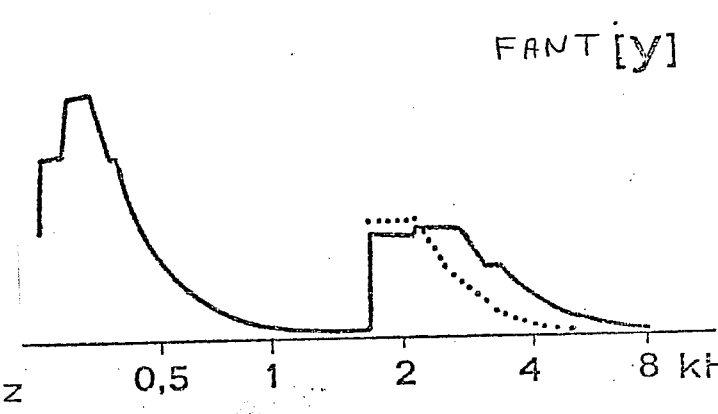


Fig. 5- IV

ROSSI (1971) fig. 6a [i] ; FANT (1975) fig. 10 [i - y] ; LANDERCY (1973) fig. 11, p. 17 [i] ; GALVAGNY (1973) [y]. La fig. 5-IV présente ces différents résultats. Pour FANT (1975) le tracé en pointillé se rapporte à une voyelle synthétique à 2 formants.

On remarque immédiatement que la zone F2 - F3 (et éventuellement F4) est marquée par un plateau s'étendant sur plus d'une bande critique. Pour ROSSI (1971), les fréquences de F2 et F3 étaient à 2160 et 3480 Hz. Pour LANDERCY (1973), de 2100 et 2800 Hz. Les résultats de FANT (en trait plein) se rapportent à une voyelle synthétique à 4 formants, et ils sont différents de ceux obtenus pour les voyelles naturelles. Ceci indique que les résultats de CARLSON et al. (1970), qui ont utilisé cette voyelle, pourraient être incorrects . Le spectre de sonie montre un pic élevé vers F3 et F4, au lieu de F2 - F3 pour les voyelles naturelles. Cette différence mériterait une étude approfondie.

L'existence d'un corrélat perceptif unique des formants F2, F3 et F4 est donc hautement probable pour la voyelle [i]. Mais, à la différence de ce que l'on observe pour les autres voyelles, celui-ci couvre une large zone qui s'étend de 2000 à 3000 Hz. Plus large qu'une bande critique, son effet auditif est différent de celui obtenu par un formant unique, qui ne présente de l'énergie que dans une bande critique unique. Ce formant unique sera placé aléatoirement dans toute la zone fréquentielle occupée par ce "bloc" perceptif ; ceci rend compte des variations importantes observées.

Nous proposons dans les paragraphes suivants deux méthodes d'estimation de F'2, qui ne dépendent pas de l'évaluation directe de la similitude entre des voyelles à 4 et à 2 formants, ou d'un test d'identification.

VI - EXTRACTION DIRECTE DE F'2 PAR UNE METHODE FREQUENTIELLE

CARLSON et al. (1970) envisagent plusieurs méthodes permettant d'extraire expérimentalement F'2 de l'onde sonore. (cf. également CARLSON et al. 1975). Ils rejettent le modèle simple de filtrage passe-haut et de comptage des passages par

zéro, en raison des mauvais résultats obtenus pour [i] et [y].

	F'2 auditif	F'2 passage/zéro	Diff.	F1	F2	F3	F4
i	3 210	2860	+ 350	255	2065	2960	3400
y	2 010	2380	- 370	255	1930	2420	3300
I/e	2 370	2380	- 10	375	2060	2560	3400
æ	1 960	1950	+ 10	605	1550	2450	3400
ø	1 270	1890	- 170	360	1690	2200	3390
ʉ	1 730	1720	+ 10	280	1630	2140	3310

On sait que cette méthode équivaut, pour des bandes de bruit filtré, à calculer le centre de gravité des composantes. Certains des résultats présentés au tableau ci-dessus nous paraissent surprenants : on obtiendrait le même nombre de passages par zéro à partir de fréquences formantiques aussi différentes que celle de [I/e] et [y]. D'autre part, il est possible que la valeur mesurée de F'2 soit trop élevée pour [i].

Nous avons reproduit l'expérience de CARLSON et al. (1970) à l'aide du dispositif présenté à la fig. 6-IV.

Ce montage permet la synthèse d'une voyelle à 4 formants. Celle-ci est soumise à un filtrage passe-bande (1000 - 4700 Hz) très raide (- 36 dB/oct.), afin d'éliminer le premier formant et le bruit de synthèse. En "bouclant" la sortie des échantillons, on obtient une voyelle continue. CARLSON et al. (1970) ne précisant pas les largeurs de bande utilisées, nous avons choisi arbitrairement les valeurs suivantes : 60, 80, 90 et 120 Hz.

La mesure du nombre de passages par zéro, dont l'inverse donne la fréquence cherchée, se heurte à un problème pratique. Pour le fréquencemètre utilisé, un passage n'est compté que lorsque le signal dépasse un seuil légèrement plus élevé que le zéro. (cf. fig. 7-IV a et b). Pour de très faibles oscillations, ce seuil peut ne pas être atteint. Pour l'expérimentateur, la densité moyenne semble dépendre du gain de l'amplificateur. Afin d'éviter cet artefact, nous avons visualisé sur oscilloscope une période de la voyelle, et nous avons éliminé les faibles oscillations en faisant légèrement varier la fréquence fondamentale.

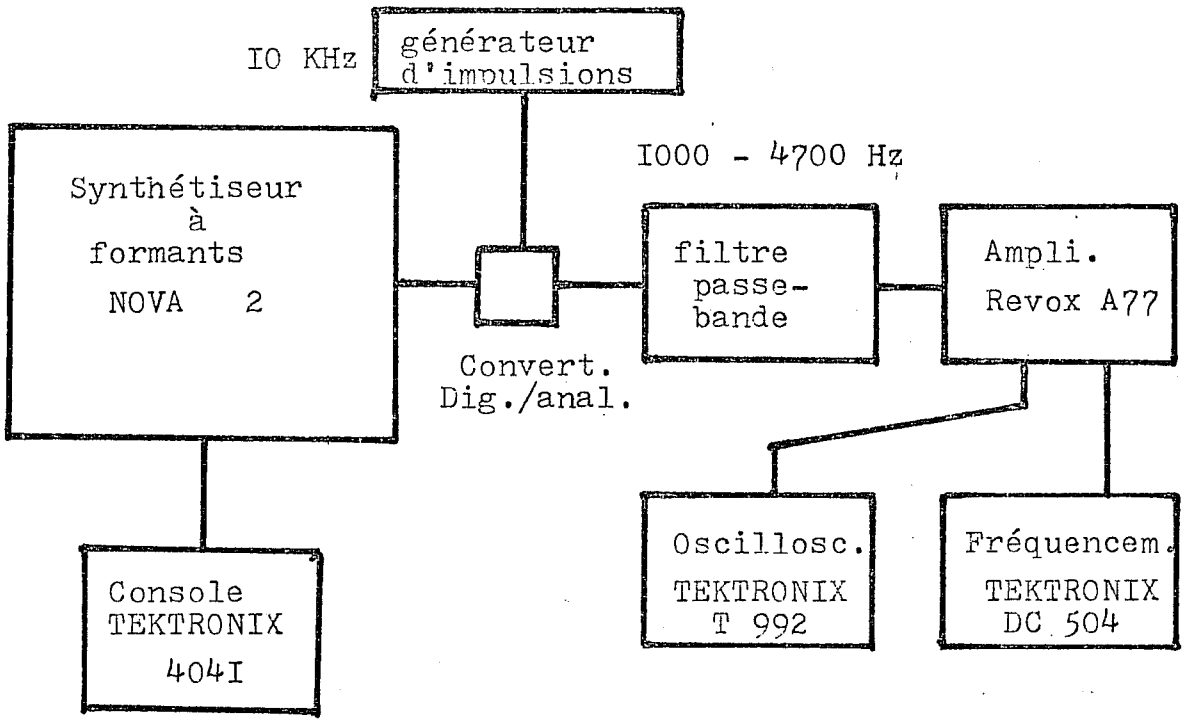


Fig. 6 - IV

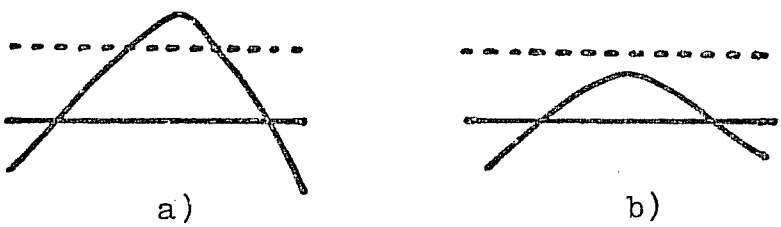


Fig. 7 - IV

Pour des voyelles synthétiques identiques à celles employées par CARLSON et al., nous avons obtenu les résultats suivants :

	F'2 auditif	F'2 (CARLSON et al.) 1970	F'2 nos résultats
i	3210	2860	2600
y	2370	2380	2300
I/e	2010	2380	2100
æ	1960	1950	1700
ø	1720	1890	1800
ɯ	1730	1720	1700

Nos résultats sont sensiblement différents de ceux de CARLSON et al. (1970) pour les voyelles [i, y, æ] . Pour [i] et [æ] , nos valeurs semblent un peu basses. Pour [y] l'accord entre la fréquence de F'2 auditif et notre mesure est bon. L'écart entre les mesures de [e] et [y] est de 200 Hz environ, ce qui est acceptable. De nombreux facteurs peuvent expliquer les différences entre les deux expériences. Parmi les plus probables, on peut citer les différences de pente spectrale et de largeurs de bande. L'hypothèse d'une détection de F'2 par comptage des passages par zéro ne peut être rejetée. Un travail récent de SCOTT (1976) a montré que des facteurs temporels jouent un rôle important dans la perception du premier formant ; ils expliquent en particulier l'interaction entre la fréquence fondamentale et la fréquence formantique réellement perçue. Mais les différences constatées entre les deux expériences ne permettent pas d'affirmer que ce modèle très simple rend compte de la perception du "deuxième formant équivalent". CARLSON et al. (1970) ont montré que la position de F'2 auditif ne dépendait guère de l'intensité des formants. Or, c'est à ce facteur qu'il faut attribuer les différences notées plus haut.

Une deuxième méthode étudiée par CARLSON et al. semble donner de meilleurs résultats. Cette méthode incorpore pourtant un modèle de la membrane basilaire (FLANAGAN 1972) dont on sait aujourd'hui qu'il est incorrect. Ceci montre à l'évidence qu'un accord entre les prédictions d'un modèle et les résultats

expérimentaux n'est pas une preuve absolue de la validité du modèle.

Les résultats ambigus que nous avons obtenus nous ont conduit à rechercher une méthode différente d'estimation de F'2.

VII - ESTIMATION DE F'2 : UNE TECHNIQUE ORIGINALE

Nous proposons une méthode d'estimation indirecte de F'2 à partir des scores d'identification de voyelles à 4 formants. Cette méthode nous a été suggérée par une expérience de FUJIMURA (1967), dont nous utiliserons les résultats dans une première évaluation. Notre méthode repose sur l'hypothèse suivante : le score d'identification de la voyelle [i] par rapport à [y] est proportionnel à F'2. Le score d'identification de [i] sera d'autant plus élevé que F'2 est élevé. Les résultats de FUJIMURA indiquent que les stimuli dont les formants supérieurs s'établissent à 2200, 2850, 3300 Hz et à 2400 2620, 3300 Hz ont le même score d'identification (= 50 % [i]). Par définition, nous considérerons que la fréquence de F'2 est identique pour ces deux stimuli. La forme de la fonction liant F'2 et les pourcentages d'identification n'est pas précisée. Nous exigeons seulement une relation ordinale entre ces deux séries de valeurs. Cette méthode suppose la synthèse d'un nombre important de voyelles à 4 formants pour lesquelles F2, F3 et F4 couvrent une large gamme de fréquences. Après l'obtention des scores d'identification, on recherche empiriquement une formule de calcul de F'2 telle que l'ordre des stimuli sur F'2 soit conforme à celui des scores d'identification. L'expérience doit être répétée pour la paire [e - ø]. On peut ensuite tenter d'établir une formule générale couvrant l'ensemble des voyelles. Mais la partie cruciale concerne la prédiction de F'2 dans la zone fréquentielle correspondant à [i , e , y et ø] . Nous avons utilisé une partie des résultats de FUJIMURA (1967) pour tester notre hypothèse. La table I-IV présente les fréquences formantiques de F2 et F3 pour les 14 stimuli retenus. Les stimuli de cette liste possèdent tous un deuxième formant supérieur à 2000 Hz et ont été perçus en majorité comme [i] et [y].

Stimulus	F2	F3	% identification [i]
I	2770	3200	I00
2	2800	3II0	I00
3	2400	3II0	I00
4	2470	3020	I00
5	2540	2930	I00
6	26I0	2850	8I
7	2200	2850	46
8	2330	2700	4I
9	2400	2630	48
I0	2260	2770	I2
II	2020	2620	7
I2	2080	2550	0
I3	2200	2400	5
I4	2I40	2470	0

Table I - IV

Stimulus	F'2	% d'identification [i]	formule utilisée
I	3250	I00	2
2	3205	I00	2
3	3205	I00	2
4	3I60	I00	2
5	3II5	I00	2
6	2730	8I	I
7	2525	46	I
8	25I5	4I	I
9	25I0	48	I
I0	25I5	I2	I
II	2320	7	I
I2	23I5	0	I
I3	2300	5	I
I4	2305	0	I

Table II - IV

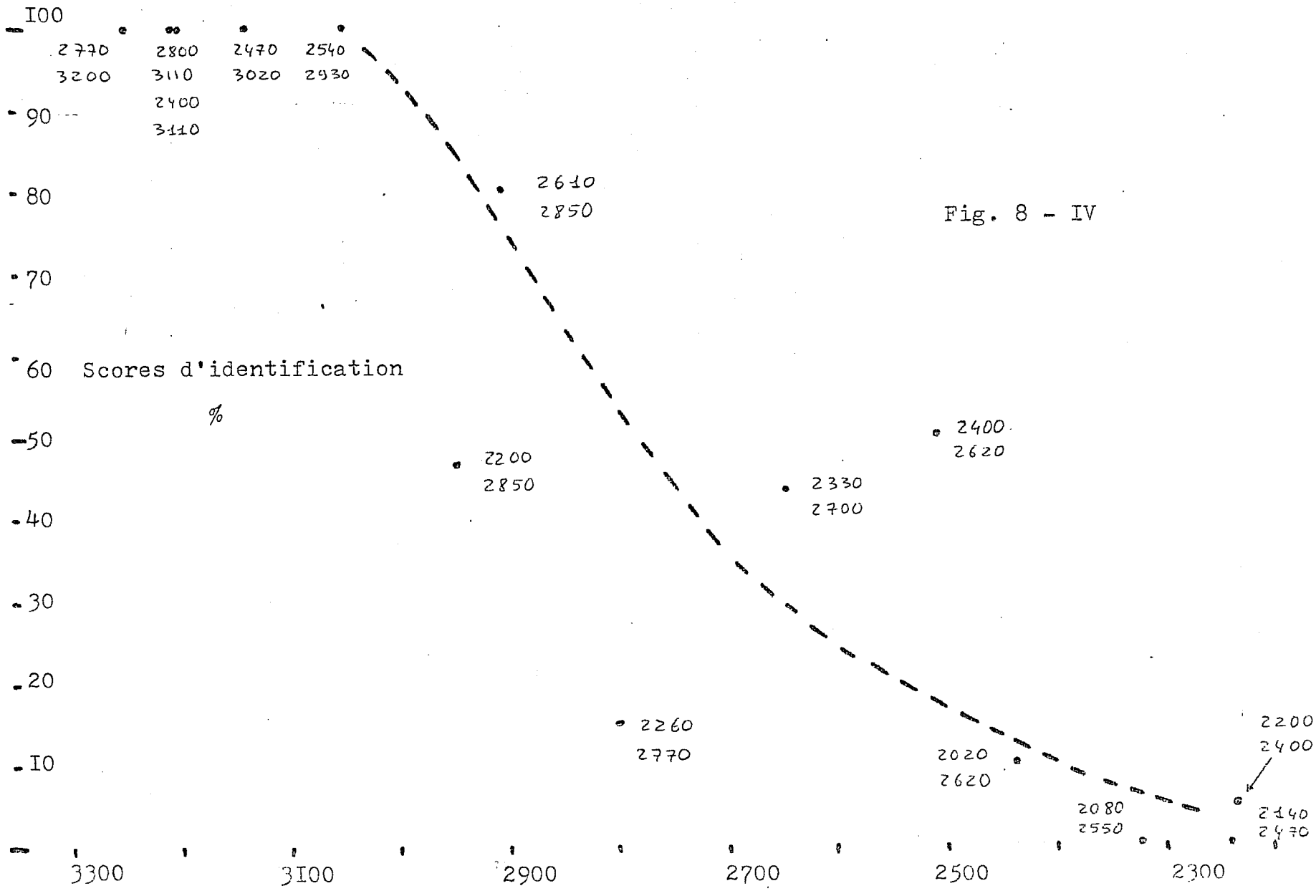

100
90
80
70
60
50
40
30
20
10
Scores d'identification
%
2770
3200
2800
3110
2400
3110
2470
3020
2540
2930
2610
2850
2200
2850
2330
2700
2400
2620
2260
2770
2020
2620
2080
2550
2200
2400
2140
2470
3300
3100
2900
2700
2500
2300
F'2: formule CARLSON et al. 1975

Fig. 8 - IV

F1 et F4 étaient fixés à 270 et 3300 Hz. Les pourcentages d'identification ont été estimés d'après les résultats graphiques. FUJIMURA conclut son travail en affirmant "qu'aucune combinaison linéaire de F2 et F3 déterminée par des coefficients positifs (éventuellement dépendants de la fréquence), ou même qu'aucune fonction croissante quelconque de ces deux variables ne permet la prédiction du deuxième formant équivalent" (p. 188). Ceci est exact d'un point de vue formel. Mais plusieurs chercheurs ont critiqué cette conclusion. CHISTOVICH et KOZHEVNIKOV (1970, cités par CARLSON et al. 1975) et LONCHAMP (1971) ont indépendamment remarqué que cette conclusion n'est dictée que par un tout petit nombre de réponses à des stimuli ayant F2 et F3 très proches (le stimulus(1850, 2020 Hz) est identifié comme [y], alors que le stimulus (1850, 2400 Hz) est perçu majoritairement comme [ʉ]; or F3 est plus élevé dans ce dernier cas). CHISTOVICH et KOZHEVNIKOV proposent de prendre en compte la notion de "bande critique". CARLSON et al. (1975) suggèrent un déplacement important du centre de gravité des formants supérieurs en raison des variations d'amplitude pour les formants proches.

La fig. 8-IV présente les scores d'identification en fonction de la valeur de F'2 calculée par la nouvelle formule de FANT. La dispersion des points est importante. Un même score d'identification est obtenu pour des fréquences de F'2 allant de 2500 à 2950 Hz.

Nous proposons une méthode très simple, proche de celle de FANT (1959).

si $F4 - F3 > 400$ Hz $\quad F'2 = F2 + \frac{(F3 - F2)}{2}$ (1)

si $F4 - F3 < 400$ Hz $\quad F'2 = F3 + \frac{(F4 - F3)}{2}$ (2)

La valeur de 400 Hz est approximativement celle de la largeur d'une bande critique pour cette zone de fréquence (SCHARF 1970 p. 162 ; f = 2900 Hz ; bande critique = 450 Hz).

On a représenté graphiquement (cf. fig. 9-IV) l'évolution de F'2 prévue par notre formule. F4 et F2 sont fixes ;

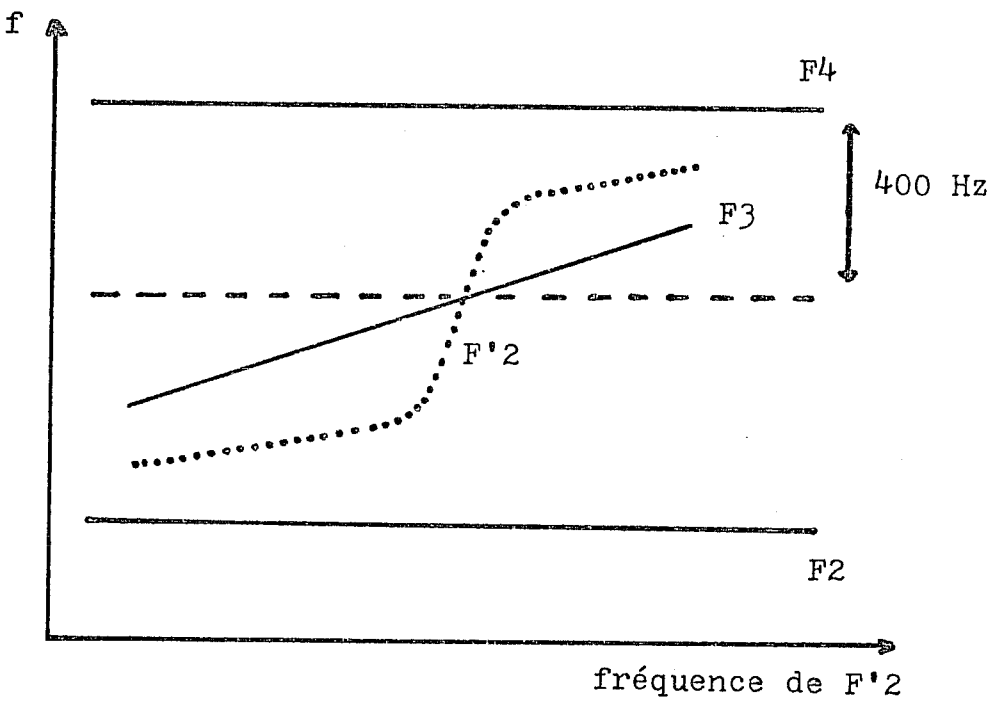


fig. 9 - IV

F3 est la variable principale. Lorsque F3 est à plus de 400 Hz de F4, F'2 est à mi-distance entre F2 et F3. Mais lorsque F3 et F4 se placent dans une même bande critique, F'2 passe brutalement à une fréquence égale à (F3 + F4)/2.

La fig. 2-IV présentée plus haut n'est pas incompatible avec une telle interprétation. Si l'on représente les scores d'identification en fonction de la fréquence de F'2 calculée par notre formule (fig. 10-IV), on constate une relation relativement régulière. La position d'un point (2260 - 2770 Hz) est aberrante. La table II-IV présente les résultats numériques et indique la relation utilisée. Le score d'identification de la voyelle [i] est bien proportionnel à la fréquence de F'2 calculée.

Notre formule est également compatible avec les résultats de CARLSON et al. (1970). F3 et F4, pour la voyelle utilisée, n'étaient séparés que de 440 Hz, ce qui justifie l'emploi de la relation (2). Celle-ci donne 3180 Hz alors que la moyenne observée est de 3210 Hz. Pour [y] , la relation(1) prévoit une fréquence de 2175 Hz, contre 2010 Hz pour la valeur observée. Par contre, notre formule échoue pour les données expérimentales de CARLSON et al. (1975, fig. 5, p. 80). Une variation de 25 Hz pour F'2 entraînerait une augmentation de 34 % du score d'identification, ce qui paraît excessif. Dans cette expérience, une variation de F3 de 100 Hz peut provoquer une variation du score d'identification de 66 %. Ceci confirme que F'2 fait un "bond" important lorsque F3 s'approche de F4. Pour cette dernière expérience la "zone critique" possède une largeur de 500 Hz environ. Le score d'identification, pour F3 placé à 2900 Hz (450 Hz d'écart avec F4) est de 85 %. Il tombe à 19 % pour F3 placé à 2800 Hz (550 Hz d'écart).

Il serait très intéressant de calculer le spectre de sonie pour ces stimuli, afin d'observer le comportement du "bloc" auditif correspondant aux formants supérieurs lorsque F3 varie. Notre proposition nécessite, de toute évidence, une vérification expérimentale plus étoffée. Les premiers indices sont néanmoins encourageants. Il semble bien que la notion de bande critique, dont l'importance est dégagée par la mesure du

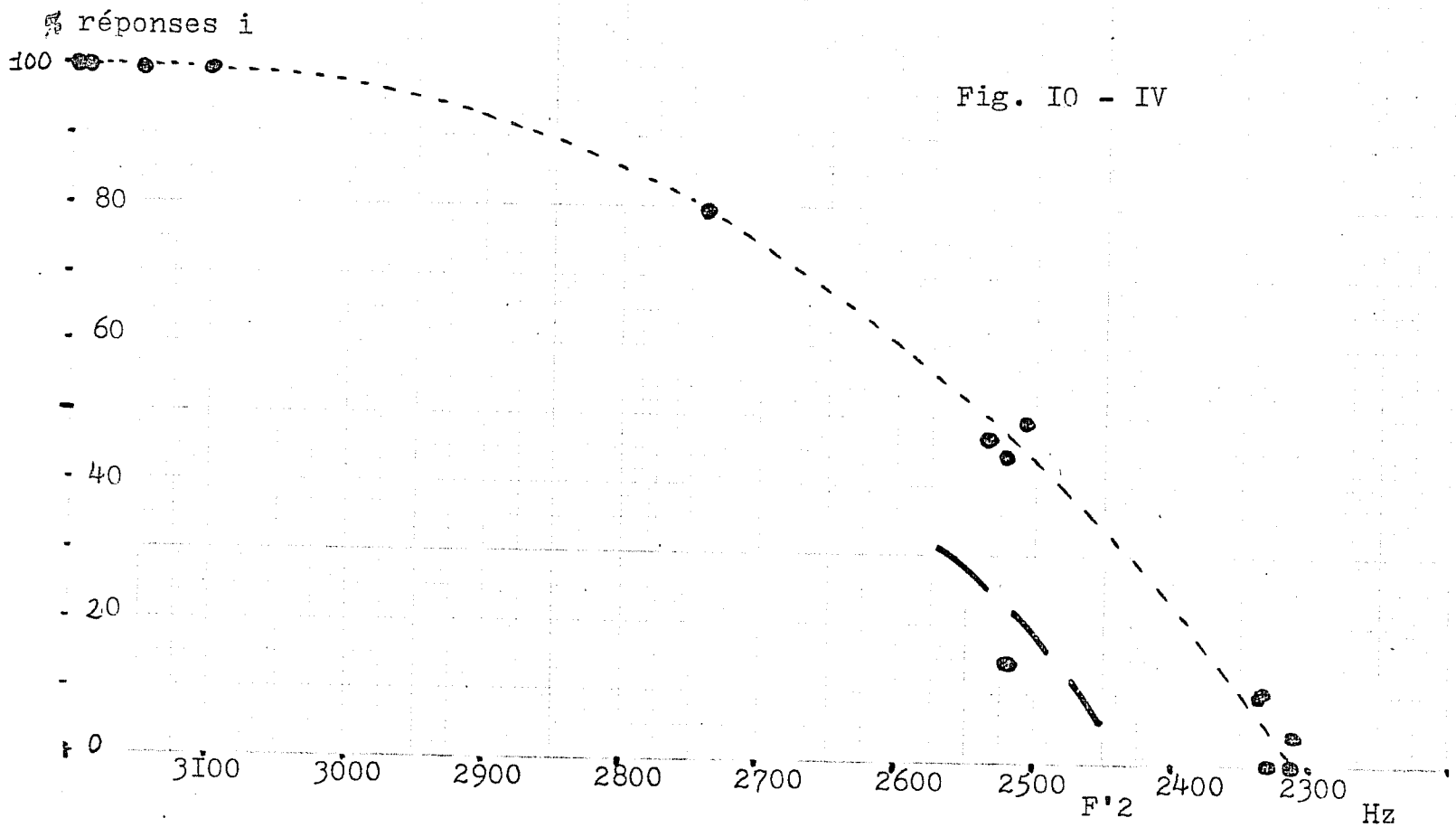

% réponses i
100
80
60
40
20
0
3100
3000
2900
2800
2700
2600
2500
2400
2300
F'2
Hz

Fig. IO - IV

spectre de sonie, permet d'élucider les problèmes liés à la perception des voyelles antérieures.

Pour terminer ce chapitre, nous présentons un modèle d'identification des voyelles dans un espace bidimensionnel. Les dimensions de cet espace sont les corrélats perceptifs de F1 et F'2.

VIII - VERS UN MODELE PROBABILISTE D'IDENTIFICATION DES VOYELLES DANS UN ESPACE BIDIMENSIONNEL

A/ Présentation du modèle

Ce modèle d'identification nous semble une preuve indirecte importante de l'existence de deux dimensions perceptives. On doit pourtant remarquer que les résultats expérimentaux ne concernent qu'une partie de l'espace, et qu'il n'est pas permis, en théorie, de généraliser. Nous avons présenté antérieurement (LONCHAMP 1973) l'esquisse d'un modèle probabiliste d'identification. A la fin de ce travail, nous indiquions que certaines modifications devaient être apportées au modèle, et qu'une vérification expérimentale plus complète devait être effectuée. C'est cette nouvelle étape que nous présentons ici.

Le modèle est bâti autour d'une constatation fondamentale que LINDBLOM (1974) exprime en ces termes : "... Pour un couple de voyelles, le risque de confusion est, en première approximation, une fonction inverse de la distance acoustique des éléments du couple". Nous avions exprimé cette constatation en des termes voisins (LONCHAMP 1971) : "... On constate.... que plus la distance entre deux voyelles est courte, plus le pourcentage d'erreurs augmente" (page 15).
Notre modèle a pour but de prévoir la probabilité de catégorisation phonologique d'un stimulus vocalique en fonction de ses caractéristiques physiques (fréquences formantiques). Il incorpore un certain nombre d'hypothèses complémentaires qui demanderaient une vérification expérimentale. Mais l'accord entre les prédictions et les résultats expérimentaux suggère que ces hypothèses sont, à tout le moins, compatibles avec les résultats expérimentaux.

Les hypothèses fondamentales de notre modèle sont les suivantes :

a) le timbre d'une voyelle peut être représenté comme un point dans un espace à deux dimensions. Ces deux dimensions sont indépendantes. Elles correspondent à une transformation (non-linéaire) des fréquences F1 et F'2. Les fréquences sont converties en unités de hauteur subjective (ou mels ; cf. FANT 1959, p. 75). ".. l'échelle "mel" a été constituée à partir des évaluations subjectives des intervalles résultant du doublement ou de la division par deux de la fréquence d'un son pur"). Jusqu'à 3000 Hz, l'échelle mel peut se déduire de l'échelle des fréquences par la formule (FANT 1959, p. 76).

$$Y = 1000 \operatorname{Log}_2 (1 + F/1000) = (1000/\operatorname{Log}_{10} 2) \cdot \operatorname{Log}_{10} (1 + F/1000$$

F = fréquence en Hz ; Y = hauteur subjective en mels.

b) chaque individu possède un "étalon" perceptif interne de chacun des phonèmes (et allophones majeurs). La position de cet étalon est fixe pour une population linguistiquement homogène. Pour un individu, la position de cet étalon fluctue de façon aléatoire autour de cette position moyenne. Suivant une suggestion de M. le Professeur CARTON, nous avons appelé ces étalons des "voyelles non-ambiguës" (VNA). Par définition, une voyelle placée dans l'espace des timbres à la même position que celle occupée par une VNA, sera reconnue sans ambiguïté comme la réalisation d'une unité phonologique. L'hypothèse d'un étalon interne a souvent été proposée par G. FANT et ses collaborateurs. Il n'existe pas de données expérimentales corroborant cette hypothèse. Tout au plus pouvons-nous faire appel à une impression subjective. Il est possible d'avoir une "représentation mentale" d'une voyelle sans émettre le moindre son. Cette représentation mentale pourrait être liée à la position articulatoire. Il semble impossible de faire naître cette "image mentale" lorsque la position des articulateurs n'est pas compatible avec cette voyelle. C'est le cas de [i] lorsque les mâchoires sont largement écartées. Nous ne pensons pas à une théorie motrice stricte. Celle-ci est exclue par la réorganisation motrice qu'effectuent les fumeurs de pipe

(MacNEILAGE 1970). Mais nous pensons à une intégration beaucoup plus complexe des phénomènes moteurs et perceptifs qui permet au sujet d'avoir une représentation mentale du son à partir d'une position articulatoire quelconque,et, inversement, de déterminer les ajustements articulatoires nécessaires pour produire un son quelconque. Des expériences préliminaires de LINDBLOM et SUNDBERG (1971) indiquent qu'un sujet reproduit des voyelles perçues auditivement avec la même précision lorsque son articulation est libre, que lorsqu'un paramètre articulatoire (écartement maxillaire) est fixé. La méthode employée exclut tout "feedback" auditif. Cette expérience semble impliquer un centre intégrateur articulo-perceptif qui peut tenir compte de la position particulière d'un paramètre articulatoire en plaçant, le cas échéant, les autres articulateurs dans des positions "paradoxales". Dans ce cadre, la VNA serait la voyelle "extrême" compatible avec les données anatomiques d'un sujet. Pour [i] par exemple, la VNA serait placée en fonction du premier formant le plus bas et du F'2 le plus élevé qu'il soit possible d'émettre .. Cette hypothèse n'est pas incompatible avec la stabilité de la VNA dans la population, car il est nécessaire de prévoir une procédure de normalisation pour rendre compte de la perception des sons émis par des locuteurs ayant des conduits vocaux de longueurs différentes.

c) Le sujet est capable d'estimer la distance entre un point de l'espace représentant un timbre particulier, et les VNA les plus proches. L'identification de la voyelle s'opère en comparant les distances entre le point et les VNA proches, et en sélectionnant la catégorie correspondant à la VNA la plus proche. Pour un grand nombre de sujets tirés d'une population homogène, ou pour de nombreuses répétitions par un sujet unique, on fait l'hypothèse que les mesures des distances sont affectées d'une erreur aléatoire. Cette hypothèse est compatible avec les résultats de MUSHNIKOV et CHISTOVICH (1974) sur les fluctuations à moyen terme des "frontières" phonématiques.

Pour le calcul des distances, nous avons retenu une métrique euclidienne dans l'espace "mel". Les distances

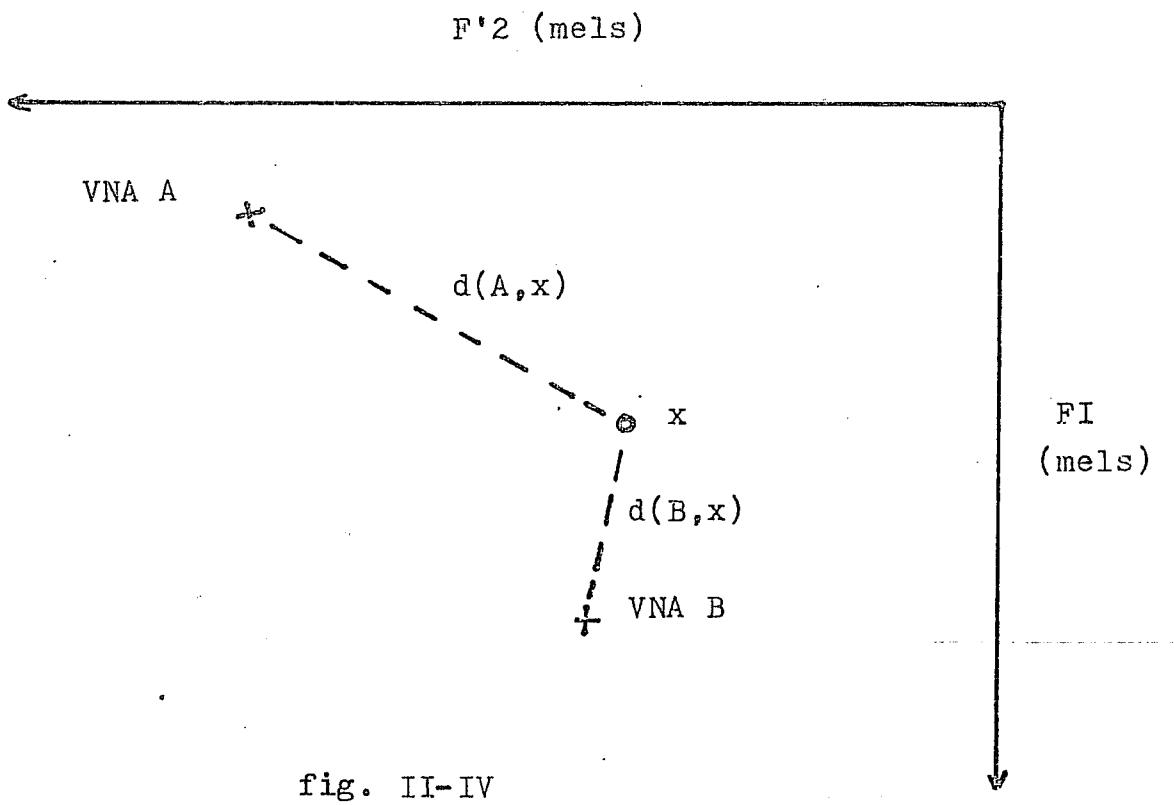


fig. II-IV

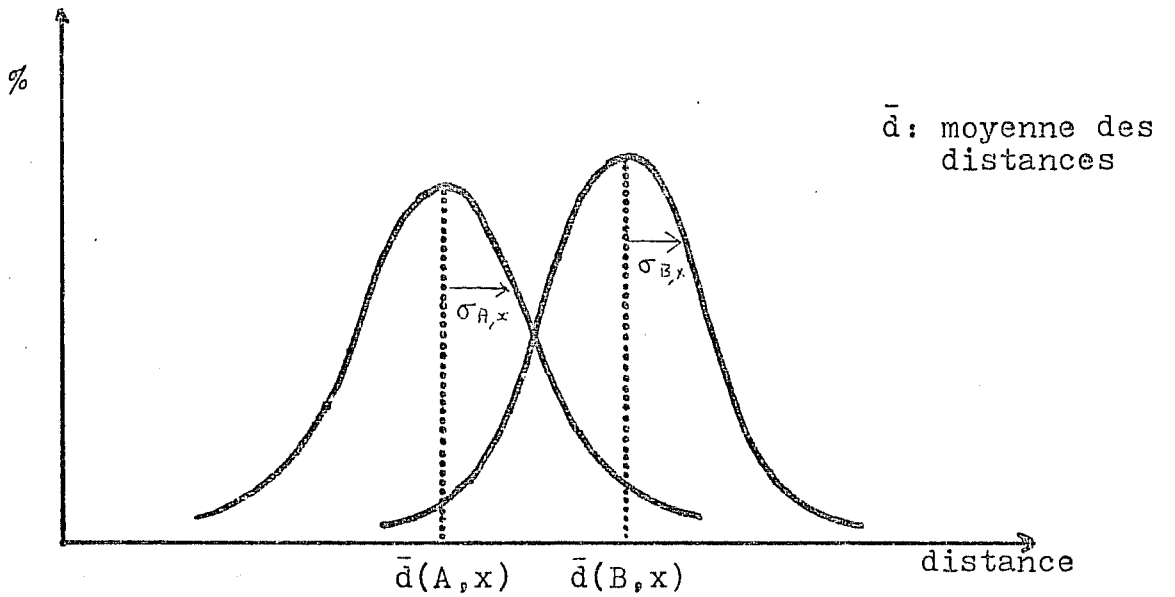


fig.I2-IV

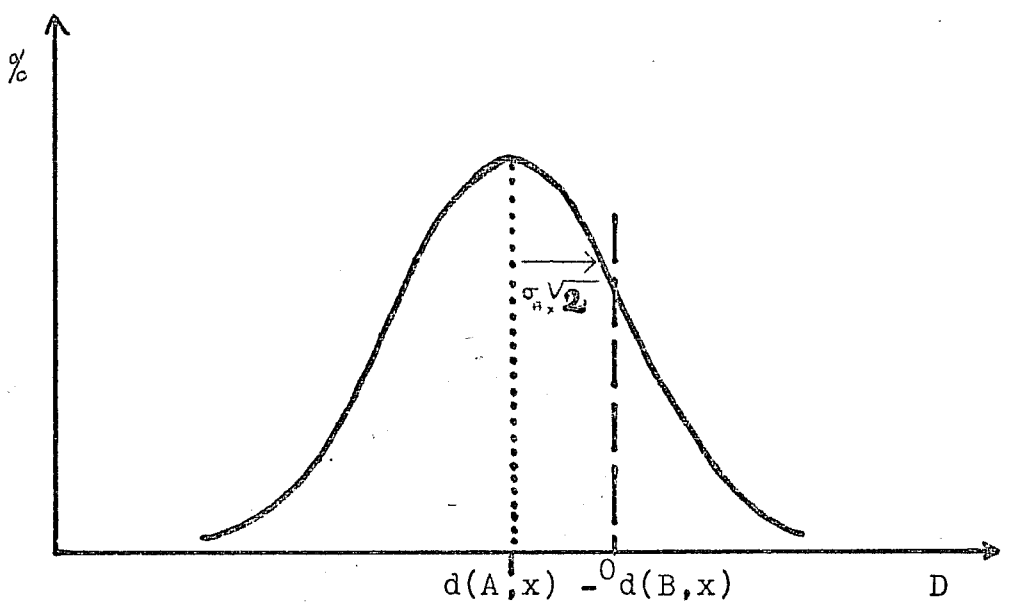


fig. 13 - IV

(indépendantes) sont donc considérées comme des variables aléatoires de moyenne d_i et d'écart type σ_i .

- Si l'on ne considère que la "compétition" entre deux VNA, il est possible de dériver une expression simple de la probabilité de classification. Dans le cas où plus de deux VNA sont en compétition, il n'existe plus de forme analytique et il faut recourir à une simulation (par une méthode de Monte-Carlo, par exemple). Nous n'examinerons que le cas d'une classification dépendant de deux VNA. En d'autres termes, on suppose nulle la probabilité qu'une troisième distance soit plus courte que les deux distances examinées. Les notations employées sont présentées à la fig. 11-IV. Dans ce cas, la règle d'identification est simplement : la voyelle est placée dans la catégorie A si $d(A, x)$ est inférieur à $d(B, x)$, ou de façon équivalente, si $d(A, x) - d(B, x) < 0$.

Nous supposons que les distributions des erreurs sont normales et indépendantes. L'hypothèse de non-corrélation des résultats implique que la distance $d(A, x) - d(B,x)$, qui est la différence entre deux distributions normales, est elle-même une distribution normale de moyenne $d(A, x) - d(B, x)$ et d'écart type égal à $\sqrt{\sigma^2_{A,x} + \sigma^2_{B,x}}$. Pour simplifier, on peut considérer que les écarts type $\sigma_{A,x}$ et $\sigma_{B,x}$ sont identiques. Dans ce cas l'écart type de la différence est égal à $\sqrt{2}\,\sigma_{A,x}$ (fig. 12-IV). Notre modèle de catégorisation est identique à la "loi des jugements comparatifs" (TORGERSON 1958 ch. 9). L'aire sous la courbe (fig. 13-IV) indique la probabilité que $d(A, x) - d(B, x)$ soit inférieur à 0, c'est-à-dire que le stimulus soit classé dans la catégorie A. On peut démontrer que la probabilité de classification est une fonction linéaire de la différence des distances $d(A, x) - d(B, x)$, si l'échelle des pourcentages est proportionnelle à la fonction de répartition d'une loi normale (méthode de la droite de Henry).

B/ Résultats expérimentaux

a) Présentation des résultats

Les résultats expérimentaux qui nous serviront à tester notre modèle sont tirés d'une expérience d'identification de voyelles synthétiques à 4 formants. Les stimuli synthétiques nous ont été fournis par le laboratoire de Traitement du Signal de la C.G.E. La méthode de synthèse et les résultats complets sont présentés en annexe de ce chapitre. Nous avons sélectionné pour la vérification du modèle, une partie de l'espace formantique où les réponses [e] et [ɛ] sont largement majoritaires. Ce choix répondait à deux critères :

- tester le modèle dans le cas où seules deux VNA étaient en concurrence ;
- disposer d'un nombre suffisant de points expérimentaux.

La probabilité de catégorisation (ou d'identification) du stimulus x dans la catégorie A(p (x/A)) s'exprime par

$$\frac{n(x/\ )}{n(x/A) + n(x/B)}$$

où n(x/A) et n(x/B) sont le nombre de réponses "A" et "B". L'élimination des réponses autres que [e] et [ɛ] est conforme à la règle du "rapport constant" ("constant ratio rule" : CLARKE 1957) qui indique que le rapport des réponses ne change pas si l'on élimine les réponses à une troisième catégorie.

Le tableau II-IV présente les pourcentages d'identification et les caractéristiques acoustiques des 10 stimuli retenus. Les résultats numériques "absolus" peuvent être trouvés en annexe de ce chapitre. On a indiqué le numéro d'ordre du stimulus dans le coin inférieur droit de chaque case.

F1/F2	F3 : 2750 2350	F3 : 2500 2200	F3 : 2500 2050	F3 : 2500 2500
400	e 95 % ε 5 % 9	e 100 ε 0 15		e 65 ε 35 27
Hz 500	e 88 ε 12 10	e 57 ε 43 16	e 38 ε 62 22	e 9 ε 91 28
600	e 40 ε 60 11	e 13 ε 87 17	e 6 ε 23 23	e 6 ε 94 29

Table III- IV

b) Choix des VNA

Le stimulus F1 = 400 Hz, F2 = 2200 Hz satisfait au critère de la non-ambiguïté. La position de la VNA [e] sera donc définie par la transformation de ces valeurs dans l'espace perceptif. Aucun stimulus ne correspond exactement à la position* suivante : F1 = 570, F2 = 1970 Hz. Mais nous ne pouvons affirmer qu'il s'agit de la position optimale de la VNA .

c) Mesure des distances et résultats

Nous avons calculé les distances euclidiennes (dans un espace "mel") entre les stimuli et les VNA pour deux conditions :

- F'2 = F2 ; c'est-à-dire en considérant F2 comme une approximation suffisante de F'2 pour cette zone de l'espace formantique.
- F'2 = F2 + (F3 - F2) ((F2 - F1)/(F3 - F1) (FANT 1915).

Les résultats du calcul des distances sont présentés à la table IV-IV. L'indice A renvoit à la condition F'2 = F2 et l'indice B à la formule de FANT. M'2 est la valeur du stimulus sur la deuxième dimension, exprimée en mels ; M1 est la valeur correspondante pour F1. d(x/e) et d(x/ε) sont les distances en mels entre les stimuli et les VNA. On a porté sous diff. la valeur absolue de la différence. Le pourcentage

*de la VNA [ε] . Après de nombreux essais sur ordinateur, nous avons choisi la position

stimulus	FI	F'2 A	F'2 B	MI	M'2 A	M'2 B
9	400	2350	25I5	485	I744	I8I4
I0	500	2350	25I5	585	I744	I8I4
II	600	2350	25I5	678	I744	I8I4
I6	500	2200	2327	585	I678	I734
I7	600	2200	2327	678	I678	I734
22	500	2050	2225	585	I609	I689
23	600	2050	2225	678	I609	I689
27	400	I900	2II0	485	I536	I637
28	500	I900	2II0	585	I536	I637
29	600	I900	2II0	678	I536	I637

stimulus	A d(x/e)	A d(x/ε)	A diff.	A %	B d(x/e)	B d(x/ε)	B diff.	B %
9	66	240	I74	95	80	226	I4I	95
I0	I20	I86	66	88	I28	I68	40	88
II	204	I76	28	60	209	I56	53	60
I6	I00	I27	27	57	I00	99	I	57
I7	I93	III	82	87	I93	79	II4	87
22	I2I	77	44	62	II0	72	38	62
23	205	47	I58	94	I98	40	I58	94
27	I42	I69	27	65	97	I68	7I	65
28	I78	74	I00	9I	I39	70	69	9I
29	240	43	I97	94	2I6	35	I8I	94

	FI	F'2A	F'2 B	MI	M'2 A	M'2 B
VNA e	400	2200	2330	485	I678	I735
VNA ε	570	I970	2I60	650	I570	I660

Table IV -IV

d'identification (%) correspond soit à [e], soit à [ɛ] selon que la différence est positive ou négative. La symétrie du modèle permet de ne considérer que les pourcentages supérieurs à 50 %, ce qui améliore la précision graphique.

La fig. 14-IV présente les résultats pour les conditions A et B sur une échelle de probabilité linéaire. Les fig. 15-IV et 16-IV présentent ces mêmes résultats sur un système d'axes permettant de tracer la droite de HENRY. Rappelons que selon notre modèle, les probabilités doivent être, dans ce cas, une fonction linéaire des distances. Pour la condition A, où F'2 est égal à F2, il est impossible de tracer une droite de régression (fig. 15-IV). Les points sont relativement bien alignés. Pour la condition B, où nous avons utilisé une formule de calcul de F'2 à partir de F1, F2 et F3, il est possible de tracer une droite de régression comme le prévoit notre modèle, à condition de considérer que les stimuli 500 - 2350 et 500 - 1900 Hz ont un comportement aberrant. L'erreur moyenne absolue est de 4,5 % pour les 8 stimuli alignés et de 7,8 % pour les 10 stimuli. Ces erreurs nous semblent voisines de la précision des mesures, si l'on considère le petit nombre de réponses pour chaque stimulus. Ces premiers résultats montrent que les données expérimentales calculées par la formule d'estimation de F'2 due à FANT (1959) ne sont pas incompatibles avec les prédictions du modèle. Mais ces premiers résultats devront être corroborés par l'étude d'autres paires dans l'espace perceptif. Il ne convient pas de surestimer l'importance théorique de ces résultats. Il est probable qu'une large classe de modèles est compatible avec ces résultats expérimentaux. Mais notre modèle permet néanmoins une importante "compression" de l'information contenue dans une matrice d'identification. La description acoustique des stimuli, la position des VNA et le coefficient de pente de la droite de régression permettent de recalculer les probabilités d'identification des stimuli avec une bonne précision. Notre méthode est un outil intéressant pour l'étude synthétique des matrices d'identification. En reconnaissance automatique de la parole, cette méthode pourrait remplacer avantageusement le

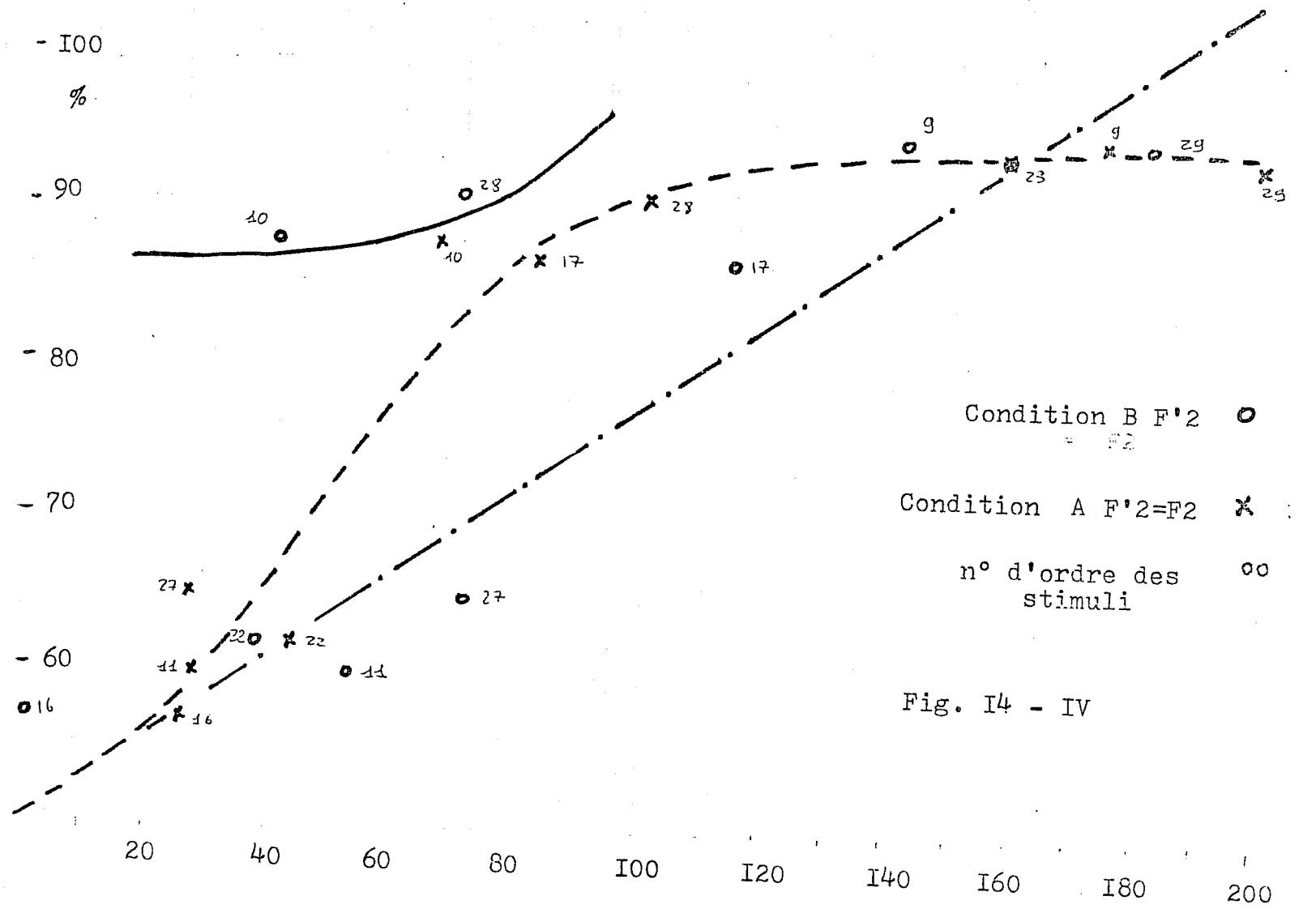

100
%
90
80
70
60
20
40
60
80
100
120
140
160
180
200
Condition B F'2 = F2
Condition A F'2=F2
n° d'ordre des stimuli

Fig. 14 - IV

Fig. 16-IV

CONDITION A

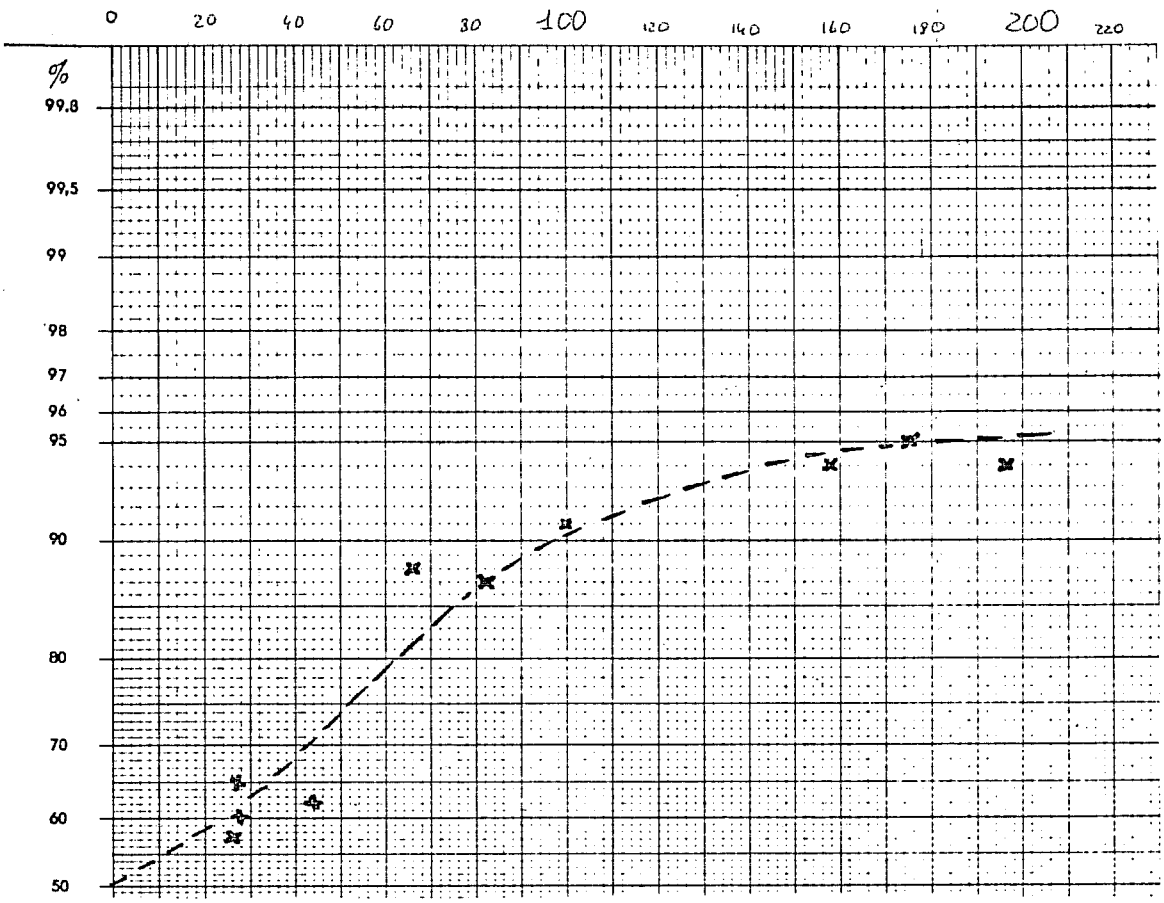
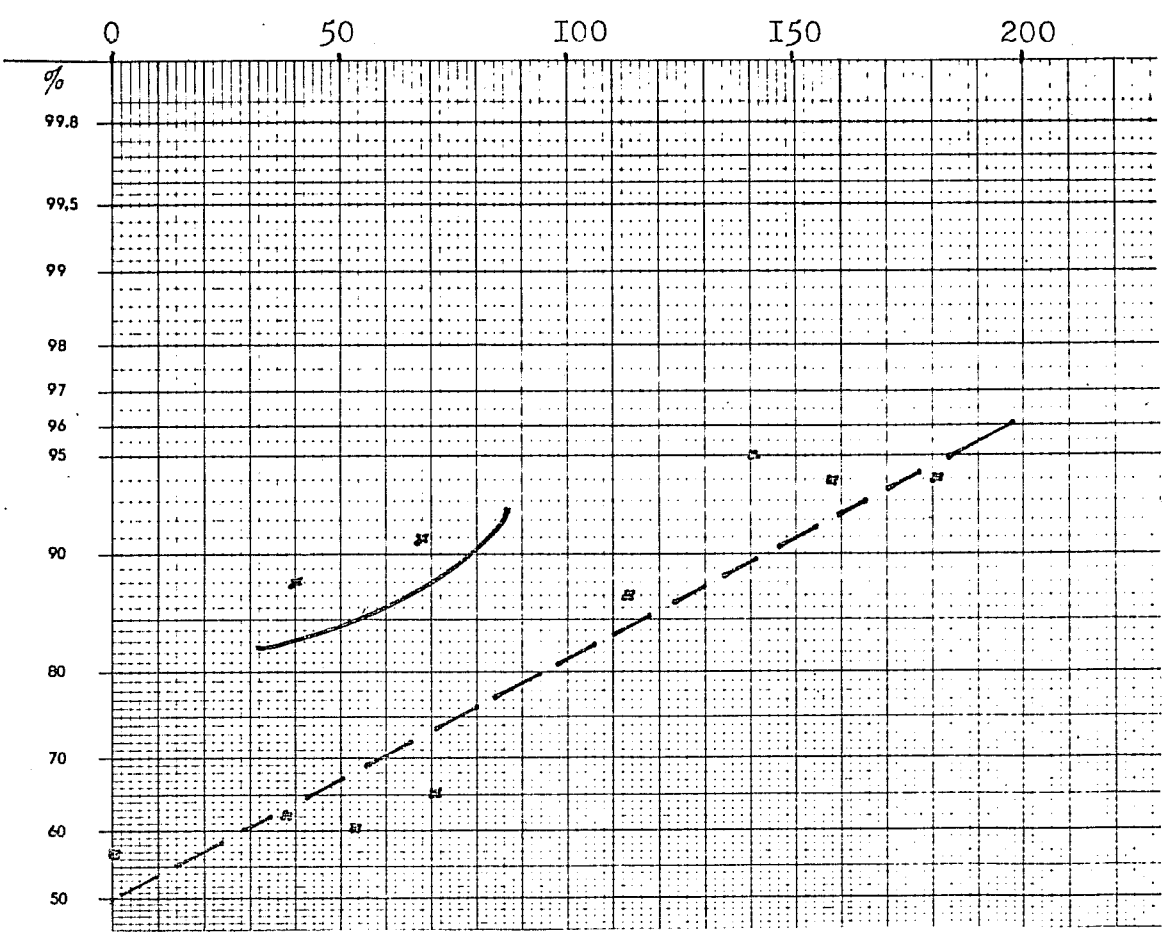


Fig. I5 - IV

CONDITION B



calcul des fonctions discriminantes. L'obtention de probabilités fiables devrait permettre leur classement et la réduction du nombre de voyelles "possibles". La position fixe des VNA permettrait peut-être d'améliorer le taux de reconnaissance pour des locuteurs multiples.

IX - CONCLUSIONS

Ce chapitre a été consacré à l'étude de la deuxième dimension perceptive. Au chapitre précédent, nous avons fait l'hypothèse que cette deuxième dimension est le corrélat perceptif du "deuxième formant équivalent" mis en évidence par les expériences de synthèse de la parole. Cette identification ne peut pas être "prouvée", au sens strict du terme, mais elle apporte une cohérence certaine au domaine de la perception vocalique, en unifiant les résultats et les hypothèses de deux domaines de recherche jusqu'ici distincts.

Une étude directe de F'2 par comparaison de voyelles synthétiques à 4 et à 2 formants révèle une très grande variabilité des estimations. Notre propre expérience de la synthèse montre qu'il est très difficile, sinon impossible, de créer une voyelle [i] satisfaisante avec 2 formants. Cette comparaison directe néglige d'ailleurs la corrélation inverse que l'on observe, pour des voyelles naturelles, entre les distances F2 - F3 et F3 - F4. La nouvelle formule de calcul de F'2 (CARLSON et al. 1975) n'est pas compatible avec certains résultats expérimentaux. La difficulté d'une comparaison directe entre des voyelles synthétiques à 4 et à 2 formants, ainsi que la variabilité observée, doivent être attribuées au fait que F2, F3 et F4 forment un "bloc" perceptif s'étendant sur plus d'une bande critique. Ce résultat est dicté par des expériences de détermination du seuil de masquage et par la mesure des spectres de sonie.

Les résultats d'une simulation de l'extraction de F'2 par comptage des passages par zéro sont ambigus. Ils semblent dépendre exagérément de l'amplitude des formants. Une technique indirecte d'estimation de F'2, à partir des scores d'identification

de voyelles à 4 formants, donne de meilleurs résultats. Cette technique utilise la notion de bande critique pour prévoir la forte non-linéarité de F'2 lorsque F3 est proche de F4. On observe une relation non-linéaire, mais régulière, entre les scores d'identification et F'2 calculé par notre méthode ; ceci n'est pas le cas lorsqu'on utilise la formule de CARLSON et al. Nous utilisons l'espace bidimensionnel corrélant avec F1 et F'2 dans un modèle de perception des voyelles. La probabilité de catégorisation est calculée à partir des distances entre le point représentant le timbre perçu et une position fixe assurant une catégorisation sans ambiguïté (VNA). L'emploi de F'2 semble donner de meilleurs résultats que celui de F2.

En conclusion, il apparaît que l'étude de la deuxième dimension perceptive est bien avancée, bien que des zones d'ombre subsistent. L'utilisation des spectres auditifs permettra sans doute des progrès rapides à brève échéance.

ANNEXE A

SCORE D'IDENTIFICATION DE VOYELLES SYNTHETIQUES ISOLEES

A - SYNTHETISEUR

Les voyelles synthétiques ont été mises à notre disposition par le département "Recherches Physiques de Base-Traitement du signal" (M. G. ROGER) des Laboratoires de MARCOUSSIS de la C.G.E.

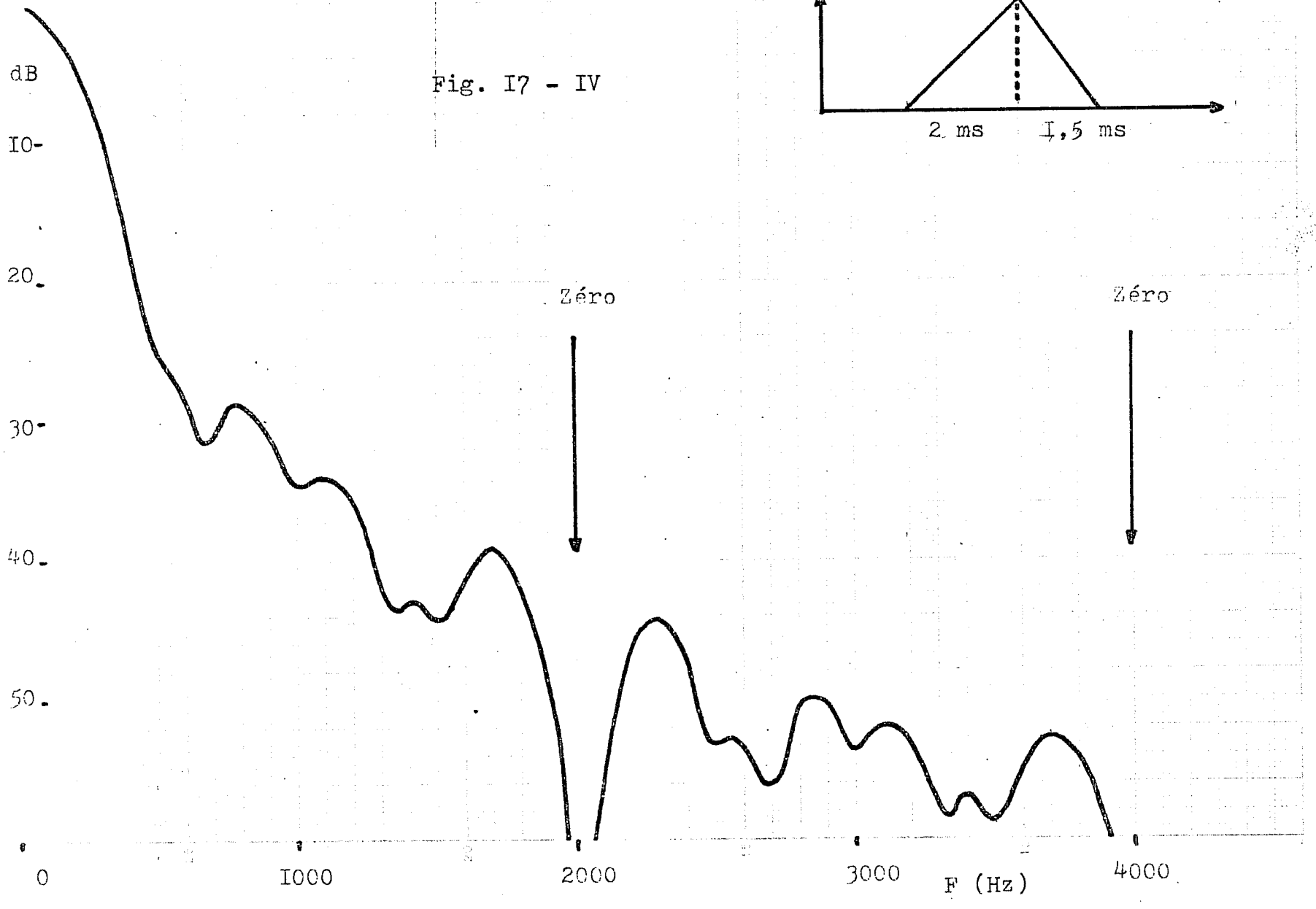
Le synthétiseur a été décrit par GREVIN, BERTIN et SAP (1975) et GREVIN (1974). Ce synthétiseur à 4 formants en série utilise un filtre récursif câblé, piloté par un ordinateur. Les paramètres (fortement quantifiés) sont actualisés toutes les 8 ms. La vitesse d'échantillonnage est de 10 KHz et la sortie s'effectue par un convertisseur 10 bits. L'algorithme câblé permet de s'approcher d'un fonctionnement en temps réel.

B - STIMULI SYNTHETIQUES

107 stimuli ont été synthétisés. Les paramètres fixes étaient : F_0 : 130 Hz ; F4 : 3500 Hz ; largeur de bande du 4ème formant (B4) : 200 Hz ; front de montée (attaque) : durée 48 ms, pente de + 6dB tous les 8 ms ; front de descente : durée 48 ms, - 6 dB tous les 8 ms ; durée totale : 248 ms, dont 152 ms à l'amplitude maximale ; B1 : 70 Hz ; B2 : 100 Hz ; B3 : 120 Hz.

Les paramètres variables (F1 et F2) sont présentés dans les tableaux donnant les scores d'identification. F3 a été fixé à 2800 Hz pour F2 = 2500 Hz, et à 2750 Hz pour F2 = 2350 Hz Tous les autres stimuli ont été synthétisés avec F3 = 2500 Hz. Les paramètres de l'onde glottale triangulaire (forme imposée par le synthétiseur) sont présentés à la fig. 17-IV avec le spectre correspondant. Cette onde présente malheureusement deux zéros à 2000 et 4000 Hz.

Fig. I7 - IV



C - QUALITE SUBJECTIVE

Celle-ci est très moyenne. Les raisons en sont certaines caractéristiques du synthétiseur ainsi qu'un choix de paramètres peu judicieux dans certains domaines. On peut noter :

- la présence d'un "bruit de synthèse" important à la fréquence d'échantillonnage.

- un choix défectueux de la pente d'attaque des stimuli. Sa raideur provoque un claquement audible au début et, dans une moindre mesure, et à la fin du stimulus.

- la présence d'un zéro à 2000 Hz qui "assourdit" les stimuli ayant leur deuxième formant dans cette zone.

- le choix d'une valeur de F1 trop faible (200 Hz) pour une partie des stimuli. Notre propre expérience ultérieure de la synthèse nous a montré qu'il ne fallait pas descendre en-dessous de 250 Hz si l'on veut éviter un caractère exagérément "assourdi".

D - METHODE DE TEST

Les stimuli ont été présentés, en ordre aléatoire, par blocs de 40, à des groupes de sujets différents. Les sujets devaient compter à haute voix de 1 à 10 pour minimiser l'effet d'ordre dû au timbre du stimulus précédent. La présentation était faite à partir d'un magnétophone REVOX A77 en champ libre dans une salle non-traitée. La présentation de stimuli de qualité moyenne exige une concentration intense de la part des sujets. Pour des raisons matérielles, nous n'avons pu présenter l'ensemble des stimuli à tous les groupes, ce qui explique le nombre de réponses différent selon les stimuli. Les sujets, étudiants du certificat phonétique générale, avaient suivi un entraînement à la transcription phonétique. Ils avaient pour consigne de transcrire chaque stimulus à l'aide du symbole A.P.I. de leur choix. Ils ont été encouragés à utiliser le symbole de nasalisation, et à ne pas répondre s'ils jugeaient que le stimulus n'était pas une voyelle française.

E - RESULTATS

Les réponses "nasales" seront examinées au chapitre X. Les réponses orales sont présentées aux tables V-IV (A et B). Le chiffre au bas de chaque case indique le nombre de réponses après élimination des non-réponses. Certains sujets n'ayant utilisé qu'un seul symbole pour la voyelle "a", nous avons regroupé toutes les réponses [ɑ] ou [a] sous le même sigle "ɑ". Les cases sans réponse indiquent une majorité de non-réponses. Chaque case du tableau présente la réponse la plus fréquente, ainsi que les réponses phonétiquement proches. Nous avons mis en valeur par un grand symbole les cases qui ont obtenu les scores les plus élevés.

Peu de stimuli ont été identifiés unanimement :
[e] 400 - 2200 - 2500 "ɑ" 875 - 1200 - 2500 Hz.

On obtient des scores d'identification élevés par les stimuli suivants :

i	300 - 2500 - 2800	:	84 %
ɛ	600 - 1900 - 2500	:	94 %
u	200 - 900 - 2500	:	92 %
œ	600 - 1300 - 2500	:	85 %
ɔ	600 - 1000 - 2500	:	80 %
o	400 - 800 - 2500	:	75 %

Par contre, le score d'identification de [y] ne dépasse pas 28 % pour 300 - 1800 - 2500 Hz et celui de [ø] plafonne à 35 % : 400 - 1400 - 2500 Hz. La raison est que F3 est trop élevé.

Notons enfin que l'on obtient 95 % de réponses [ɯ] pour le stimulus 200 - 1700 - 2500 Hz. Ce résultat est confirmé par les scores des stimuli voisins. A l'audition, le timbre est très assourdi. La fréquence de F1, inférieure à 250 Hz, en est la cause probable.

Aucune expérience comparable n'ayant été réalisée pour le français, il n'est pas possible de juger de la représentativité de ces résultats. La fig. 18-IV compare les valeurs "optimales" de synthèse et les moyennes de DEBROCK et FORREZ (1976).

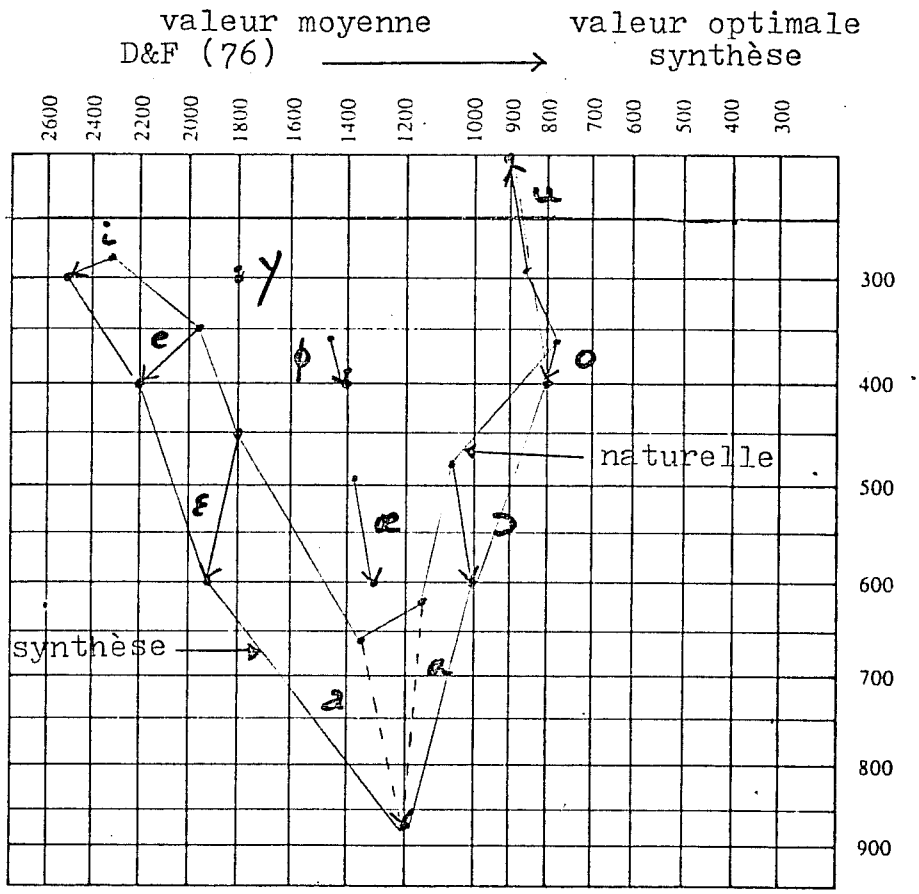

valeur moyenne
D&F (76)
valeur optimale
synthèse
2600
2400
2200
2000
1800
1600
1400
1200
1000
900
800
700
600
500
400
300
300
400
500
600
700
800
900
naturelle
synthèse

Fig. I8 - IV

F1 Hz \ F2 Hz	2500	2350	2200	2050	I900	I800	I700	I600
200	i I5 20	i 5 u 6 20	e 7 u 5 i 4 y 3 20	u 5 e 4 y 4 I5	u I0 y 2 I6	u I2 I6	u I9 (u) 20	u I2 y 4 20
300	i 3I i 37	i I2 e 7 20	e 6 i 6 u 3 y 3 20	e 9 I6	u 7 e 4 y 3 I9	u I4 y I0 e 5 y 36	20	u II ø 3 I6
400	e 2I 26	e I8 ɛ I 26	e 20 e 20	u 7 ɛ 2 e 2 20	e I3 ɛ 3 20	œ 9 ø 4 I6	ø 5 œ 5 I6	ø 4 œ 5 I6
500	e I5 20	e I4 ɛ 2 I6	e 20 ɛ I5 36	ɛ I0 e 6 I6	ɛ 30 e 3 37	œ 9 ɛ 6 e I I6	œ I4 ø 3 20	œ 8 ø 4 20
600	20	e 4 ɛ 6 I6	ɛ I4 e 2 20	ɛ 30 e 2 36	ɛ I6 ɛ I7	ɛ I0 œ 3 e I I6	œ 9 ɛ 8 20	œ 7 ɛ 4 ø 3 I9
700	i I0 20	ɛ 5 ɛ̃ 4 e 3 I6	ɛ II e 2 20	ɛ I0 e I I5	ɛ 6 ɑ I 26	ɛ I3 œ 5 20	ɛ 9 I5	œ I5 ɛ 2 ɑ I 26
800						ɛ I0 ɑ I 20	ɛ 7 œ̃ 5 ɛ̃ 4 ɑ I ɛ̃ œ̃ 20	ɑ I3 œ 5 ɛ 5 ɛ̃ 4 œ̃ 4 35
875							ɑ 6 ɛ 4 ɛ̃ 3 I6	ɑ I4 ɛ 3 20

Table V -IV A

F1 \ F2	I500	I400	I300	I200	II00	I000	900	800 Hz
200 Hz	u I4 œ 2 I6						u 24 u 26	u I6 20
300	u I2 y 6 26	u 8 I6			o 7 20	u II I6	u I2 o 4 20	u 8 o 4 20
400	u I0 œ I0 ø 8 36	ø 6 o 3 œ 3 ø I7	o 6 ø 4 œ 3 I6	ø 5 o 4 œ 4 I5	o I0 20	o 8 ø 3 I6	o II u 3 20	o I5 ɔ 3 o 20
500	œ 30 ø 5 37	œ I6 ø 2 20	œ I4 I7	œ I3 20	œ I0 ɔ̃ 4 ɔ̃ 20	ɔ I0 o 4 I6	ɔ II 20	ɔ 6 o 4 œ 3 I6
600	œ I6 ø 2 20	œ I5 ø 3 20	œ I7 œ 20	œ 30 37	ɔ 9 I5	ɔ I6 ɔ 20	ɑ I6 I7	ɔ 5 ɛ 6 I5
700	œ I9 26	œ I4 20	œ 7 ɑ 5 20	ɑ I3 20	ɑ 2I 26	ɑ I4 20	ɑ 9 ɑ̃ 8 20	ɔ 8 œ 4 ɛ 6 ɑ 3 20
800	ɑ 8 ɛ̃ 4 I6	ɑ I3 20	ɑ 23 26	ɑ I5 I7	ɑ I5 I7	ɑ II I6	ɑ̃ 9 ɑ 6 ɑ̃ I6	
875	ɑ I6 ɛ I 20	ɑ I6 ɛ I I7	ɑ 34 37	ɑ 20 ɑ 20	ɑ I8 20	ɑ I5 I7		

Table V - IV B

- On notera la nette centralisation des voyelles naturelles [i, e, ɛ, ə, a, ɔ]

- Les voyelles naturelles ont un premier formant moins élevé, sauf [u] La centralisation est d'autant plus forte que la fréquence de F1 est plus élevée.

- Le deuxième formant de la voyelle naturelle est moins élevé qu'en synthèse pour [i, e, ɛ] . La différence est plus faible pour les voyelles postérieures.

- L'écart entre les voyelles synthétiques et les voyelles naturelles est beaucoup plus faible pour les voyelles labialisées [y, ø, o, u]

- La voyelle synthétique correspondant à un "a" moyen est bien intermédiaire entre [a] et [ɑ] naturel.

ANNEXE B

INTENSITE SONORE : TERMINOLOGIE ET METHODES

Cette annexe a pour but de rappeler les principes de la mesure de l'intensité subjective des sons. On trouvera une bonne synthèse dans LANDERCY (1973).

I - SONS PURS OU BANDES DE BRUIT ETROITES

A - Niveau de pression sonore

Le niveau de pression sonore (anglais "sound pressure level" SPL) correspond à une intensité physique, mesurée en décibels (dB). Il est défini par le niveau de pression par rapport à une pression de référence p_o égale à 2×10^{-4} dynes/cm^2 (un dyne est la force qu'il faut appliquer à une masse d'un gramme pour lui communiquer une accélération de 1 cm^2/s^2)

$$SPL_{dB} = 20 \text{ Log}_{10} \; P/P_o$$

L'emploi d'une transformation Log ramène le rapport entre 0 et 150 environ. De plus, les valeurs exprimées en dB sont plus semblables, sans être identiques, aux rapports d'intensité perçus par l'oreille.

B - Intensité sonore subjective

La mesure de l'intensité sonore subjective (anglais "loudness") suppose que l'on définisse d'abord le niveau d'égale intensité subjective en fonction de la fréquence. L'étalon (arbitraire) choisi est un son à la fréquence de 1000 Hz. Le niveau d'égale intensité subjective, exprimé en PHONES, est défini comme le nombre qui mesure le niveau de pression sonore d'un son subjectivement équivalent à 1000 Hz. Par définition, l'échelle "phone" et "dB" coïncident donc pour la fréquence de 1000 Hz. Pour les fréquences différentes, on utilise les courbes expérimentales d'isophonie, dont la première a été publiée par FLETCHER et MUNSON (1937). Ces courbes d'isophonie précisent la sensibilité de l'oreille en fonction de la fréquence.

Pour pouvoir comparer les intensités subjectives entre elles, on utilise une échelle de SONIE. Celle-ci permet d'exprimer qu'un son est par exemple deux fois plus intense qu'un autre. Par définition, un sonie est l'intensité subjective d'un son de 40 dB à 1000 Hz, c'est-à-dire de 40 phones. Les résultats expérimentaux montrent que, pour les sons supérieurs ou égaux à 40 phones, la valeur de la sonie(s) est égale à

$$S = 2^{(1 - 40)/10}$$ 1 : intensité subjective en phones.

L'exemple suivant récapitule les notions précédentes :

Soit un son p créant à proximité de l'oreille une pression de 0,1 dynes/cm^2. Le niveau de pression sonore est, par définition, de 20 Log$_{10}$ (0,1/2 x 10^{-4}) = 54 dB. Si la fréquence de ce son est de 300 Hz, les courbes d'isophonie indiquent que 54 dB à 300 Hz correspondent à 49 phones. Ce son a donc la même intensité subjective qu'un son de 49 dB à 1000 Hz. La sonie correspondante est de $2^{(49 - 40)/10} = 2^{0,9} = 1,87$ sones Ce son est presque deux fois plus intense qu'un son de 40 dB à 1000 Hz.

II - SONS PRESENTANT DE L'ENERGIE DANS UNE LARGE BANDE DE FREQUENCE

Les calculs précédents ne sont valables que pour des sons purs ou des bandes de bruit étroites car ils supposent que la fréquence du son peut être définie. On constate expérimentalement que l'intensité subjective varie en fonction de la largeur de la bande de fréquence contenant de l'énergie.

A - Notion de "Bande Critique"

On trouvera un excellent exposé dans SCHARF (1970). Une bande critique est définie expérimentalement comme une bande fréquentielle (de part et d'autre d'une fréquence centrale) à l'intérieur de laquelle la largeur d'une bande de bruit peut varier, sans provoquer de changement dans l'intensité subjective. Si la zone fréquentielle couverte par la bande de bruit est plus large que la bande critique, l'intensité subjective sera différente de celle calculée par la méthode présentée plus

haut. La table suivante présente quelques valeurs de bande critique en fonction de la fréquence centrale.

fréquence centrale (Hz)	largeur de la bande critique (Hz)
100	100
250	100
570	120
1000	160
2150	320
3400	550

La précision de l'estimation de la largeur de la bande critique est de l'ordre de 15 % seulement. Les fréquences centrales, placées de manière que les bandes critiques soient adjacentes (sans recouvrement ni hiatus), correspondent approximativement aux fréquences des filtres "tiers d'octave", utilisés en acoustique pour mesurer le spectre instantané. Les centres de ces bandes définissent des unités de fréquence appelées "Bark" (en l'honneur de l'acousticien allemand BARKHAUSEN). Le phonéticien remarquera que la largeur d'un formant est largement inférieure à une bande critique. Par contre, le fondamental et la première harmonique appartiennent à des bandes critiques différentes. Dans l'analyse, on réunit souvent des filtres de basse fréquence pour éviter le plus possible cette séparation. On utilise alors une batterie de filtres dont les fréquences centrales s'établissent à :
63, 125, 224, 315, 400, 500, 630, 800, 1000, 1250, 1600, 2000, 2500, 3150, 4000, 5000 Hz.

B - Effet de Masque

La présence de deux sons séparés de plus d'une bande critique provoque une augmentation de l'intensité subjective par rapport au niveau dû au plus intense d'entre eux, lorsqu'il est présenté seul. Cet accroissement dépend de la séparation fréquentielle. Si les sons sont très éloignés en fréquence, les intensités subjectives de chacun des sons s'ajoutent. Mais si la fréquence entre les deux diminue, cette règle d'addition

n'est plus valide. Le son de fréquence inférieure tend à masquer le son de fréquence plus élevée. ZWICKER a publié, après de très longues expérimentations, des diagrammes permettant de connaître l'effet de masque en fonction de la fréquence et de la sonie des composantes du son. La fig. 19-IV présente la sonie d'un son à trois composantes (a, b et c), en fonction de la fréquence. La ligne pointillée descendante, partant du sommet de chaque composante, indique, en fonction de la fréquence, le niveau qu'une autre composante doit atteindre pour contribuer à la sonie globale. La composante b, par exemple est entièrement masquée par la composante a. ceci implique que la présence (ou l'absence) de b ne change pas l'intensité subjective totale. La composante c est partiellement masquée par a. Mais c contribue partiellement à l'intensité subjective totale. L'intensité subjective totale est obtenue en faisant la moyenne des sonies de chaque bande fréquentielle. Cette méthode a été cependant rejetée par ROSSI (1971) dans une étude de l'intensité spécifique des voyelles. Il propose une méthode plus précise, dérivée de STEVENS (1956), qui fait intervenir la notion de largeur de bande "utile". Le phonéticien s'intéresse plus souvent aux caractéristiques qui peuvent être masquées, qu'à l'intensité subjective totale. La fig. 20-IV présente le spectre de sonie, ou densité fréquentielle de l'intensité subjective. Si les trois composantes précédentes sont les formants d'une voyelle, le phonéticien pourra conclure que le deuxième formant b ne peut jouer aucun rôle, car il est masqué par le premier formant ; le troisième formant, par contre, peut être utilisé perceptivement et contribue sans doute au timbre perçu. La société HEWLETT-PACKARD commercialise un instrument (8051 A "Loudness Analyser") qui effectue automatiquement le calcul du spectre de sonie selon le principe des nomogrammes de ZWICKER.

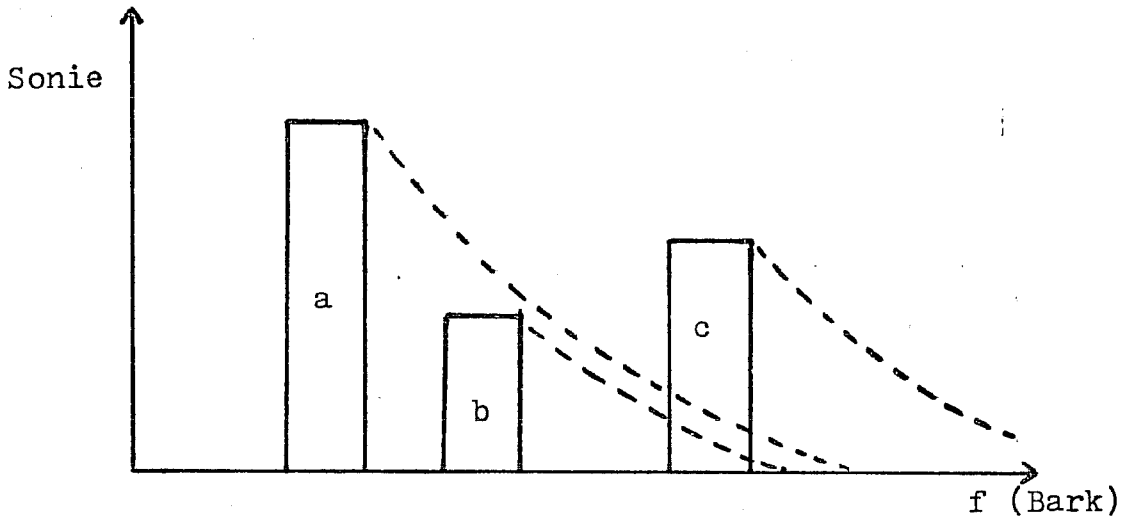


fig. I9-IV

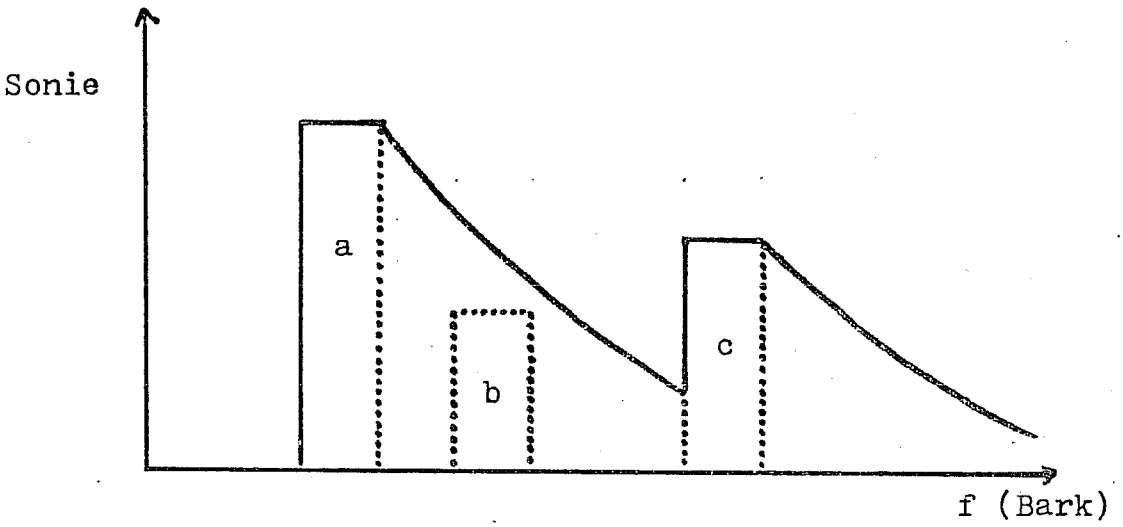


fig. 20 - IV

C H A P I T R E V

ESPACE PERCEPTIF ET SYSTEMES VOCALIQUES : ECHEC D'UNE EXPLICATION PUREMENT PERCEPTIVE

I - INTRODUCTION

Ce chapitre a pour but d'examiner, à la lumière des travaux de LILJENCRANTS et LINDBLOM (1972), l'hypothèse selon laquelle la structure des systèmes vocaliques serait gouvernée par le principe de différenciation maxima dans l'espace des timbres. Nous tirons des conclusions opposées à celles des auteurs, ce qui justifie l'analyse détaillée de leurs résultats. Les chapitres VIII et IX tenteront de cerner quelques paramètres ou contraintes articulatoires qui jouent un rôle important dans la structuration des systèmes vocaliques. Ce chapitre apporte néanmoins une preuve supplémentaire de l'importance de l'espace perceptif bidimensionnel. LILJENCRANTS et LINDBLOM posent également le problème d'une phonologie fondée sur la substance, par le biais de "méta-principes" perceptifs. Ce dernier point sera examiné au chapitre XI.

II - LE PRINCIPE DE "DIFFERENCIATION MAXIMA"

L'école de Prague, qui définit les phonèmes au moyen des oppositions inter-systémiques, considère la notion de "système" comme une donnée fondamentale et quasi-axiomatique de la phonologie. LILJENCRANTS et LINDBLOM (1972) tentent de dériver la notion de "système" et d'"opposition phonologique" d'un principe général qui serait une donnée empirique. Les forces agissant sur le système linguistique, et en particulier sur le système vocalique, ne proviendraient pas du système lui-même, comme l'affirme avec force MARTINET (1955, p. 80-81) dans sa théorie des "cases vides". Mais elles seraient le résultat de l'action d'un principe général de contraste ou de "différenciation maxima", pour reprendre la propre expression de MARTINET. Ce principe s'appliquerait dans l'espace perceptif des timbres. La différence entre la position de MARTINET (op. cit. p. 150-152, 255-256) et celle de LILJENCRANTS et LINDBLOM est que le linguiste français ne se prononce pas sur la nature de ce principe. Au contraire, les phonéticiens suédois le placent exclusivement au niveau perceptif, et excluent ainsi toute influence des données

articulatoires. Si pour MARTINET ou HAUDRICOURT la notion de différenciation maxima joue un rôle important en diachronie, LILJENCRANTS et LINDBLOM lui attribuent un rôle synchronique. Le principe est maintenant susceptible d'être testé empiriquement. JAKOBSON (1941), comme nous le rappelions dans le chapitre II, avait déja tenté une explication psychophysique du système ternaire /i - a - u/, sur une base analogue. Pour LILJENCRANTS et LINDBLOM, il est possible de prédire synchroniquement la structure phonétique des systèmes vocaliques (en fonction du nombre de voyelles) à partir de ce principe unique. L'utilisation des termes "phonétique" et "phonologique" dans l'article cité n'est pas claire : "... on cherche (dans quelle mesure) le principe de différenciation maxima peut être utilisé en phonologie pour expliciter la structure phonétique des systèmes vocaliques" (op. cit., p. 839, résumé). Nous croyons que ce principe, si tant qu'il soit valide, ne peut s'appliquer qu'aux données phonétiques.

III - STRUCTURES PHONETIQUES OU STRUCTURES PHONOLOGIQUES ?

Notre position nécessiterait un développement qui examinerait le rapport entre la phonologie et la phonétique. Nous nous contenterons de quelques exemples. L'étude du système vocalique du polonais a provoqué de nombreuses controverses (on trouvera un bon résumé dans ŁOBACZ 1973). Le problème est de savoir si les voyelles [i] et [ɨ] sont des variantes ou des phonèmes distincts. Le débat porte sur un nombre infime de formes et nous ne le trancherons pas, parce que nous croyons que ce débat doit être évité. Nous pensons qu'il peut être évité si l'on reconnaît entre le phonème et la variante contextuelle, la notion d'"allophone majeur".

Soit une langue ne connaissant (à une certaine époque) qu'un phonème /i/ avec une variante unique [i]. Par suite d'une évolution phonétique naturelle, cette langue développe deux variantes [i] et [ɨ] au contact des consonnes A, B, C et D, E, F respectivement. Supposons enfin qu'à une époque ultérieure, et par un processus phonétique indépendant du précédent, F vienne à se confondre avec B, et ceci

dans un petit nombre de cas. Ce processus créera un nouveau phonème car les variantes [i] et [ɨ] ont maintenant un contexte commun. Ce processus de phonologisation peut être linguistiquement important, comme la création des antérieures labialisées en haut-allemand ; mais que dire s'il ne touche qu'une ou deux paires ? C'est le cas en français pour les voyelles [ø] et [œ] (WALTER 1976). Est-il raisonnable de faire dépendre le choix phonologique le plus important, c'est-à-dire la détermination des phonèmes d'une langue, de la survie de mots comme "jeûne" ou "veule" ?

Une condition essentielle pour la reconnaissance d'un allophone majeur dans une langue est que cet allophone ait statut phonologique dans une autre langue. La présence d'un signe orthographique distinct est un indice non négligeable. L'intérêt de notre proposition pour l'étude des systèmes vocaliques est évident. Il permet de ne pas interpréter le système du polonais comme étant identique à celui de l'espagnol.... Il nous semble également, que d'un point de vue perceptif, il soit indispensable de considérer l'arabe dialectal de Constantine (Algérie) comme un système à 5 éléments, dont deux allophones majeurs, (ɛ, o) et non comme un système à 3 voyelles.

Ce point de vue permettra de prévoir que les deux allophones majeurs de l'arabe dialectal sont [ɛ] et [o], et non [y] ou [ɤ] . Ils sont structurellement déterminés par les mêmes règles qui régissent la structure de l'espagnol, par exemple. En résumé, le principe de différenciation maxima ne doit pas prévoir la structure phonologique, ce qu'il ne peut faire d'ailleurs dans certain cas, comme nous le verrons. Mais il doit prévoir la structure des éléments phonétiques jusqu'au niveau des allophones majeurs.

IV - FORMALISATION DU CRITERE DE DIFFERENCIATION MAXIMA

Ce critère, sous sa forme mathématique, consiste à placer les voyelles dans un espace de telle manière que les distances entre les voyelles soient les plus grandes possibles.

Si r_{ij} est la distance euclidienne entre les voyelles i et j (dont les coordonnées dans l'espace sont a_{ik} et a_{jk} pour la k ième dimension), on peut écrire

$$r_{ij} = \sqrt{\sum_{k=1}^{n} (a_{ik} - a_{jk})^2}$$

Le critère revient alors à minimiser la quantité $\sum_{1}^{m} 1/r_{ij}^2$. L'algorithme de minimisation suggéré est très simple, mais il est peu efficace quand le nombre de voyelles devient important.

V - L'ESPACE PERCEPTIF DES TIMBRES VOCALIQUES

Celui-ci est défini par LILJENCRANTS et LINDBLOM par la transformation d'un espace physique à 3 dimensions en un espace perceptif à deux dimensions. Les dimensions physiques sont les fréquences des 3 premiers formants. Les valeurs extrêmes sont 250 - 2250 - 3000, 250 - 750 - 2300 et 800 - 1200 - 2500 Hz, correspondant aux voyelles [i], [y] et [ɑ]. Ces fréquences sont en fait exprimées en mels (FANT 1959, p. 76). Une fréquence en Hz est convertie en unités mels par la formule.

$$y \text{ (mels)} = (1000/\text{Log } 2).\text{Log } (1 + F/1000)$$

La transformation en un espace perceptif bidimensionnel s'effectue en combinant les dimensions correspondant à F2 et F3 en une dimension unique. Cette transformation utilise la notion de "deuxième formant équivalent" ou F'2. "La représentation de l'espace correspond à celle du plan M1-M2 (F1 et F2 exprimés en mels), corrigé en fonction de F3 (p. 841)". Si l'on utilise l'ancienne formule de FANT (1959, p. 80).

$$F'2 = F2 + 0{,}5\ (F3 - F2)\ ((F2-F1/F3-F1))$$

pour F1 = 250, F2 = 2250 et F3 = 3000 Hz, c'est-à-dire pour le point limite de l'espace correspondant à [i], F'2 est égal à 2523 Hz (1817 mels).

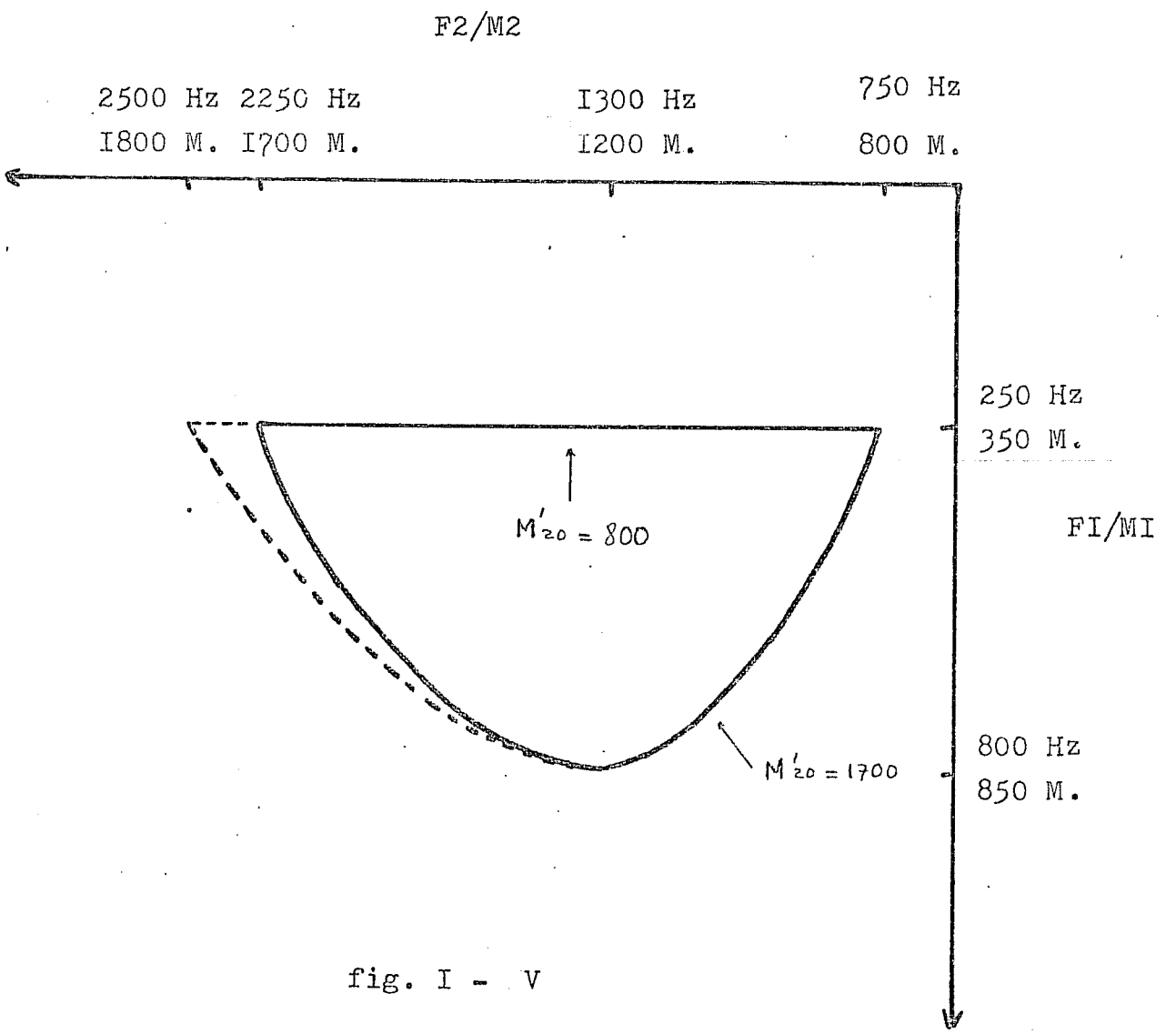

F2/M2
2500 Hz
I800 M.
2250 Hz
I700 M.
I300 Hz
I200 M.
750 Hz
800 M.
250 Hz
350 M.
FI/MI
800 Hz
850 M.
M'20 = 800
M'20 = 1700

fig. I - V

Cette valeur est bien inférieure à celle obtenue par la nouvelle formule (CARLSON, FANT et GRANSTROM 1974). F'2 s'élève alors à 3068 Hz (2025 mels). Mais elle n'a été publiée que postérieurement à l'article étudié.

Dans l'espace perceptif M1-M2, les voyelles s'inscrivent dans une figure définie par l'équation paramétrée suivante (fig. 1-VI), où M'2 représente M2 corrigé en fonction de F3.

$$M'2 = 1150 + (M'_{20} - 1150) \sqrt{(850 - M1)/(850-350)}$$

M'_{20} = 1700 définit la courbe allant de [u] à [a], et de [a] à [i]

M1 peut varier de 350 à 850 mels (250-800 Hz)

M'_{20} = 800 définit la droite allant de [i] à [u].

Un calcul simple montre que pour M1 = 350 mels, M'2 varie entre 800 et 1700 Hz.(750 - 2250 Hz). Pour M1 = 350 mels et = 1700 mels, M'2 vaut

$$1150 + (1700 - 1150) \quad \frac{850 - 350}{850 - 350} = 1700 \text{ mels } (2250 \text{ Hz}).$$

CE RESULTAT INDIQUE QUE, CONTRAIREMENT AUX AFFIRMATIONS DE LILJENCRANTS ET LINDBLOM, M'2 N'A SUBI AUCUNE CORRECTION PAR RAPPORT A M2.

M'2 devrait atteindre, comme nous l'avons calculé plus haut, 1800 mels environ (2500 Hz) pour satisfaire au calcul de F'2 selon FANT (1959). La distance entre [i] et [u] a été sous-estimée, même par rapport au calcul le plus conservateur de cette distance. Si l'on considère d'autre part que "l'erreur principale du modèle est de ne pas prédire l'apparition de voyelles du type [ø] et de prédire un nombre trop élevé de voyelles fermées centrales telles que [ɨ] et [ɨ] " (p. 854), c'est-à-dire sur la ligne joignant [i] à [u], on ne peut s'empêcher de se demander ce que les prédictions auraient été si M'2 avait été effectivement corrigé ! Nous avons porté en ligne pointillée la forme approximative de l'espace pour M'2 corrigé (fig. 1-V). Il est probable que

les prédictions auraient été plus mauvaises encore. Cette "erreur", que nous espérons fortuite, enlève beaucoup de leur intérêt aux résultats et constitue la raison essentielle de notre désaccord quant aux conclusions.

VI - LES DONNEES EMPIRIQUES : STRUCTURES VOCALIQUES ATTESTEES DANS LES LANGUES DU MONDE

LILJENCRANTS et LINDBLOM ont réalisé un travail utile en rassemblant les résultats de 3 études sur les systèmes vocaliques, dont la plus récente est inaccessible en France : TRUBETZKOY (1949), HOCKETT (1955), SEDLAK (1969). On peut y ajouter le travail de LINDAU (1975). D'autres typologies, comme celle de REID (1971) mériteraient une analyse approfondie. Dans notre discussion, les lettres T, H ou S indiquent la provenance de l'information. LILJENCRANTS et LINDBLOM comparent les résultats de la simulation en fonction du nombre de voyelles avec les systèmes attestés. Ils attribuent à chaque voyelle de l'espace perceptif un symbole, en fonction de sa position dans l'espace perceptif. A cette première approximation s'ajoutent les différences notables entre les trois sources de données. 12 symboles différents par exemple sont employés pour les voyelles fermées. Dans de nombreux cas, il y a désaccord sur le nombre des voyelles d'une langue. Les analyses sont plus phonologiques que phonétiques. Elles ne prennent pas en compte les "allophones majeurs". Nous présenterons donc les voyelles entre deux barres inclinées. Il convient de montrer une grande prudence dans l'interprétation des systèmes attestés.

On peut faire quelques remarques sur les trois voyelles que JAKOBSON définissait comme fondamentales.

- la voyelle /i/ apparaît dans tous les systèmes, à l'exception des systèmes dits "linéaires" (cf. infra), de certaines langues du Caucase (ABKHAZ, ADYGE, KABARDIAN...). Dans ces systèmes, elle n'apparaît que comme allophone majeur (HALLE 1970, TRUBETZKOY 1949, p. 102).

- La voyelle /a/ apparaît dans tous les systèmes attestés sauf l'ARAPAHO(S). Cette exception mériterait confirmation.

- La voyelle /u/, au contraire, est absente dans un nombre non-négligeable de systèmes. Mais si on élimine à nouveau les systèmes linéaires, on constate que toutes les langues ont soit /u/, soit /o/, soit /ɯ/. Il serait intéressant de connaître la qualité de la voyelle notée /o/ dans de nombreux systèmes amérindiens. Tous les systèmes de plus de 5 voyelles possèdent un /u/. Ces remarques seront amplifiées au chapitre IX.

Nous ne croyons pas qu'il soit intéressant pour l'étude des systèmes de considérer les systèmes linéaires :

. ABKHAZ, ADYGE (T)

e ə
a

Les travaux plus récents de DUMEZIL (1954) ou de HALLE (1970) confirment l'existence de ces systèmes phonologiques. Mais ces systèmes possèdent de nombreuses variantes en fonction de l'environnement consonantique. Ces variantes se disposent exactement comme les voyelles des systèmes plus riches en voyelles. Il semble donc que ces variantes obéissent aux mêmes règles que les voyelles au statut phonologique certain. TRUBETZKOY (1949, p. 102) indique que le système phonétique de l'ABKHAZ comprend 6 voyelles :

i u
ə
e o
a

Ce système est très proche du système bulgare par exemple.

A/ Les systèmes de 3 à 5 voyelles

Nous appelerons système "théorique" le résultat de la simulation de LILJENCRANTS et LINDBLOM. Les systèmes théoriques ne comportent pas de voyelles "centrales", c'est-à-dire de voyelles autres qu'antérieures non-arrondies, postérieures arrondies ou /a/ :

i u a (3)	i u ɛ ɑ (4)	i u e o a (5)

Si les systèmes observés confirment cette règle pour 3 voyelles, on note quelques exceptions pour 4 voyelles :

i ɨ u a	(S) BINOKID BONTOC ICOCANO IVATAN	(H) AMAHUACA
i u ə a	(S) SQUAMISH PASHTO	(H) ILOCANO DIBABAON
i ĩ o e	(S) ARAPAHO	

Le nombre de ces systèmes est trop élevé pour que l'on puisse suggérer dans tous les cas une erreur d'analyse.

Les systèmes à 5 voyelles apportent d'autres exemples :

i y/ɨ u e a	(T) TABASSARAN KYURI (H) HUICHOL (avec ɨ)	(H) POTAWATOMI i e ə o a

Deux langues possèdent même 2 voyelles "centrales" sur 5 :

(S) CHACOBO i y ɨ u a	(S) MANUBO i ɨ ə o a

Le modèle de LILJENCRANTS et LINDBLOM ne rend pas compte de la minorité de langues qui possèdent une voyelle "centrale" pour un nombre total ne dépassant pas 5. On pourrait songer à incriminer la trop faible distance entre [i] et [u]. Une augmentation de cette distance intercalerait peut-être une voyelle "centrale", mais nous ne rendrions plus

compte alors de la majorité des systèmes qui ne possèdent pas de voyelles centrales. Il ne semble pas y avoir de solutions dans le cadre du modèle.

B/ Systèmes comportant 6 ou 7 voyelles

Le système théorique à 6 voyelles comprend une voyelle que les auteurs notent par [ɨ] . Cette identification nous paraît contestable en raison de l'extrême rareté de cette voyelle. Nous considérons comme conforme aux prédictions tout système comportant une voyelle centrale.

Or il existe un nombre important de systèmes, dont certains sont bien connus, qui ne comportent pas cette voyelle centrale. Citons seulement

i u	(H) UKRAINIEN
e o	(H; S) CHIPEWYAN
ɛ	
a	(T, S) UZBEK (avec ɔ pour ɛ)

Le chinois mandarin semble posséder 3 voyelles centrales :

i y ɨ u	(H) chinois mandarin
ə	
a	

Le système théorique à 7 voyelles comporte deux voyelles centrales :

i y ɨ u

ɛ ɔ

a

Aucun système attesté (sur plus de 30), à l'exception peut-être d'un dialecte chinois (T), ne comporte 2 voyelles centrales fermées sur 7. Par contre un nombre non-négligeable de langues ne possèdent pas de voyelle centrale. C'est, de l'aveu même des auteurs, le système le plus fréquent dans les sources consultées (p. 849). Citons seulement :

i	u	(T, S, H)	ITALIEN
e	o	(S, H)	PORTUGAIS
ɛ	ɔ		TUNICA
a		(H)	LATIN VULGAIRE

A nouveau, le principe de différenciation maxima ne rend pas compte de la diversité des systèmes observés. Les erreurs commises sur la dimension /i - u/ font que le nombre de degrés d'aperture est également incorrect.

C/ Systèmes à plus de 7 voyelles

Si pour 8 voyelles, la simulation prévoit 4 voyelles fermées, ce chiffre passe à 5 pour les systèmes à 9 voyelles : Or on ne connaît aucune langue possédant plus de 4 voyelles fermées : Suédois : /i y ʉ u/ . On peut se demander si ce chiffre n'aurait pas été dépassé si la deuxième dimension perceptive avait été corrigée en fonction de F3. De plus, tous les systèmes à partir de 8 voyelles connaissent une voyelle centrale d'aperture moyenne : /ɸ, ə, ʌ etc.../ Il faut attendre les systèmes théoriques à 11 voyelles pour voir une voyelle centrale non-fermée. Le système théorique à 12 voyelles prévoit trois voyelles ouvertes /æ ɑ ɑ/ , ce qui paraît extrêmement douteux. Toutes ces erreurs sont liées entre elles.

VII - DISCUSSION ET CONCLUSION

Cette simulation de la structure des systèmes vocaliques par l'introduction du concept de différenciation maxima dans un espace perceptif se solde par un échec. Celui-ci se traduit par

- la prédiction d'un nombre trop important de voyelles fermées centrales, c'est-à-dire comprises entre [i] et [u] Il semble que ce résultat n'ait été obtenu qu'en diminuant, contrairement aux conclusions des études antérieures sur F'2, la distance entre les voyelles [i] et [u]. La structure du

sous-système des voyelles "centrales" sera étudiée en détail au chapitre IX.

- l'absence de voyelles centrales d'aperture intermédiaire.

- Par voie de conséquence, des prédictions incorrectes sur le nombre des degrés d'aperture pour un nombre de voyelles donné.

Un phonéticien français ne peut s'empêcher de rapprocher ces faits des problèmes de phonétique diachronique concernant l'apparition de la voyelle "centrale" [y] en vieux-français. Nous pensons qu'une meilleure connaissance des mécanismes expliquant la mutation [u > y] permettrait de faire progresser largement l'étude de la typologie des systèmes vocaliques. Le chapitre IX présentera une ébauche de solution originale à ce problème.

Les raisons de cet échec sont plus difficiles à cerner. Trois hypothèses peuvent être émises. La première a été formulée par LINDBLOM (1974) lui-même : "... Si on fait l'hypothèse que les voyelles tendent à s'opposer non seulement entre elles, mais également avec les consonnes, la théorie... prédit alors des systèmes qui sont beaucoup plus proches des systèmes observés que ce que l'on obtenait avec la théorie antérieure" (résumé). On pense tout de suite à [j], [ɰ] ou [ɥ]. L'argument ne paraît pas très convaincant par rapport aux autres hypothèses, mais il mérite une étude détaillée.

La deuxième hypothèse met en cause la métrique utilisée dans l'espace perceptif. L'utilisation des unités mels en particulier ne conviendrait pas. Cet argument se heurte à une objection de taille : la diversité des systèmes observés est telle que l'on ne peut prédire à la fois un système de 4 voyelles contenant un [ɨ], et un système de 7 voyelles sans voyelle centrale. Nous ne pensons pas que l'analyse multidimensionnelle soit d'un grand secours pour estimer la forme de l'espace perceptif, et en particulier la distance entre [i] et [u]. L'estimation des distances inter-stimuli

les plus longues sont les plus contestables, comme l'a démontré le chapitre III.

On peut également attribuer l'échec de cette tentative à l'absence de contraintes articulatoires dans le modèle. Cette dernière hypothèse a notre préférence. Cette absence s'explique aisément par le faible niveau de nos connaissances dans ce domaine, mais également par les difficultés de formalisation. Nous étudierons, dans un prochain chapitre, deux contraintes articulatoires. Mais il nous faut auparavant présenter les deux "outils" qui nous seront nécessaires. Ces deux "outils" nous permettront de passer d'une représentation fréquentielle à une représentation articulatoire, et vice versa. Des travaux importants ont été menés dans ces deux domaines depuis quelques années. L'étude de ZERLING (1974) sur le calcul des fréquences formantiques à partir de la forme du conduit vocal a été l'une des premieres du genre en France. Nous avons poursuivi ces recherches et présentons des corrections qui améliorent dans une large mesure les prédictions. Nous présenterons également au chapitre VII la méthode d'estimation de la forme du conduit par prédiction linéaire sur l'onde de pression.

L'échec de la tentative de LILJENCRANTS et LINDBLOM ne nous permet cependant pas de rejeter le principe de différenciation maxima. Au contraire, nous pensons que ce principe est parfaitement valide et nous en fournirons une preuve au chapitre IX. Ce principe joue un rôle certain dans les évolutions diachroniques. Mais il ne peut, à lui seul, rendre compte de la structuration des systèmes vocaliques. L'échec de la tentative de LILJENCRANTS et LINDBLOM devrait inciter les phonéticiens à accentuer l'effort de recherche dans le domaine, trop négligé, de la structuration des systèmes vocaliques. C'est un domaine où la contribution de la phonétique à la phonologie, contrairement à l'opinion de TRUBETZKOY, peut être déterminante.

C H A P I T R E VI

CALCUL DES FREQUENCES FORMANTIQUES A PARTIR DE LA FORME DU CONDUIT VOCAL.

I - INTRODUCTION

Ce chapitre a pour but d'étudier les problèmes associés au calcul des fréquences de résonance (formants) à partir de la forme du conduit vocal. Ces résultats seront utilisés dans les chapitres suivants qui traiteront des contraintes articulatoires sur la structure des systèmes vocaliques. Ces résultats seront en outre utilisés dans le chapitre traitant du problème inverse : peut-on, avec une bonne approximation, calculer la forme du conduit vocal à partir de la seule connaissance des fréquences des 4 ou 5 premiers formants ? On comprend l'intérêt fondamental, pour le phonéticien, de cette dernière question quand on connaît les difficultés que présentent les mesures directes (cinéradiographie latérale couplée à la tomographie frontale).

C'est un chercheur du Laboratoire de Phonétique de Nancy (ZERLING 1974) qui, le premier en France, a utilisé une méthode de calcul sur ordinateur mise au point aux U.S.A. (COCKER 1968, FLANAGAN 1972, p. 272-273). Cette méthode est fondée sur une équation de Webster discrétisée. MRAYATI et GUERIN, du laboratoire de la communication parlée à l'ENSERG, ont développé un modèle différent (MRAYATI et GUERIN 1976) qui considère le conduit vocal comme une série de quadripôles électriques. Ce modèle est beaucoup plus complet car il permet le calcul des largeurs de bande en plus du calcul des fréquences formantiques, en raison de l'introduction des différentes pertes acoustiques. Mais la longueur et la lourdeur des calculs est en proportion de la précision obtenue. Il y a, nous semble-t-il, un terrain d'application pour un algorithme plus simple et ne nécessitant pas de trop longs calculs. Nous pensons en particulier aux applications en temps réel telles que l'on peut les concevoir dans un système de rééducation d'enfants sourds. De plus, nous allons montrer que des procédures de corrections peuvent être apportées à la méthode pour lui donner une précision égale à celle de la méthode de MRAYATI et GUERIN. Cette dernière est néanmoins la seule qui permette le

calcul des largeurs de bande. Cette information spectrale nous paraît pourtant présenter peu d'intérêt pratique.

II - EQUATIONS DE WEBSTER

Depuis 1970, l'étude de la propagation du son dans un tube de section variable a progressé de façon notable. Si les équations fondamentales (équations de Webster) sont connues depuis 1912 (WEBSTER 1919), elles ne s'appliquent rigoureusement qu'à une propagation unidimensionnelle du son dans un tuyau aux parois rigides, en négligeant l'ensemble des pertes. L'équation de Webster pour la pression en fonction du temps et de la position dans le tube, p(t, x), est l'équation différentielle suivante :

$$(1) \quad \frac{\partial^2 p}{\partial x^2} + \frac{1}{A}\frac{\partial p}{\partial x}\frac{\partial A}{\partial x} - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0$$

FLANAGAN (1972) p. 272

où c représente la vitesse du son (353 m/s), A (x) la fonction décrivant l'aire de la section du tube en fonction de la distance par rapport à une des extrémités (appelée improprement, mais d'un terme consacré par l'usage, fonction d'aire), p la pression et t le temps.

Une dérivation particulièrement claire se trouve dans MRAYATI (1976). Celle-ci est trop complexe pour être reproduite ici. Une résolution directe n'étant pas possible, divers chercheurs ont proposé des méthodes approximatives ou indirectes de résolution (CHIBA et KAJIYAMA 1941, p. 229-232, UNGEHEUER 1962, p. 107-112). COKER (1968) utilise une représentation discrétisée (c'est-à-dire définie numériquement en un grand nombre de points discontinus du temps et de l'espace) des fonctions donnant la pression en fonction du temps et de l'espace). Si le nombre de points est suffisamment grand, on peut approximer les dérivées premières et secondes de p(x) par

$$\left.\frac{d^2 p}{dx^2}\right|_{x=x_i} \cong \frac{p_{i+1} - 2p_i + p_{i-1}}{\Delta x^2}$$

Δx étant l'intervalle de longueur entre deux valeurs de p.

$$\left.\frac{dp}{dx}\right|_{x=x_i} \cong \frac{p_i - p_{i-1}}{\Delta x}$$

Après introduction de ces relations dans (1), on montre qu'il est possible de former une équation récursive liant p_{i+1} à p_i et p_{i-1}

$$p_{i+1} = \left[p_i \left(1 - \frac{\omega^2 \Delta^2 x}{c^2} + \frac{A_{i-1}}{A_i}\right) - \left(\frac{A_{i-1}}{A_i}\right) p_{i-1} \right] \qquad \omega = 2\pi F \text{ F en Hz}$$

(2)

(Flanagan 1972, p. 273).

La pression au point p_{i+1} est parfaitement définie par la connaissance des pressions aux points i et $i-1$, des aires en ces mêmes points, de la longueur séparant deux points successifs, de la vitesse du son et de la fréquence angulaire ω des variations de pression à une extrémité.

III - LES CONDITIONS D'UTILISATION - PROBLEMES D'INTERPOLATION

L'utilisation de l'équation de Webster suppose que l'on ne considère que la propagation plane unidimensionnelle de l'onde. On sait (cf. par ex. CHIBA et KAJIYAMA 1941, p. 147) que l'excitation d'un tuyau fermé à une seule extrémité (glotte) à une fréquence de résonance produit une onde de pression qui est nulle à l'embouchure (lèvres) et qui possède une valeur maximale à la glotte. Pour le calcul point par point de l'onde de pression, Zerling (1974) a montré qu'un algorithme correct devait partir des conditions initiales $p_0 = 0$, $p_1 = a$ (a : constante arbitraire), le calcul de p_2, p_3 ... se faisant en utilisant l'équation récursive (2). Le calcul se fait donc à partir des lèvres. L'algorithme de ZERLING détermine ω par approximations successives de manière à maximiser la pression à la glotte. Les valeurs de ω qui satisfont ce critère sont les fréquences des formants cherchés. ESPESSER(1977, p. 71-73) a montré qu'une telle approximation n'est valide que pour des fréquences inférieures à $\frac{1.84c}{2\ r_0}$ où c est la vitesse du son, et r_0 le rayon du plus grand cylindre de l'approximation par une série de cylindres du tube acoustique de section variable. Pour une aire maximale de 16 cm^2, (FANT 1960, p. 115), la fréquence limite est de 4600 Hz.

L'équation de Webster peut donc représenter le comportement acoustique du conduit pour les 5 premiers formants.

D'autre part, ZERLING (1974) a montré quantitativement que l'erreur due à la discrétisation devenait négligeable si l'on approximait le conduit vocal par un nombre de cylindres de l'ordre de 100. En réalité, on ne dispose que de 8 à 40 valeurs. Il convient de construire les sections intermédiaires par interpolation. Nous avons essayé divers types de lissage en vue de réduire le nombre de cylindres nécessaires, en diminuant le saut quantitatif entre deux sections. Pour tous les procédés de lissage envisagés, l'erreur est plus grande qu'en l'absence de lissage. Pour les interpolations nécessaires, nous avons simplement divisé chaque cylindre élémentaire en un nombre suffisant de cylindres ayant une section identique. Des travaux de simulation ont montré que le nombre optimum de sections dépend du formant considéré. Ce nombre est beaucoup plus faible pour F1 que pour F4. Nous avons adopté comme valeur d'interpolation la valeur qui minimise l'erreur pour F2. Ce résultat est obtenu pour une valeur d'interpolation qui divise le conduit vocal en 150 sections environ. Nous présentons ci-dessous les résultats pour un tube uniforme de 17 cm de long. Si la vitesse du son est prise égale à 340 m/s, les fréquences théoriques de résonance s'établissent à 500, 1500, 2500 et 3500 Hz.

Les résultats en fonction du nombre de sections sont :

Nombre de sections	F1	F2	F3	F4	(Hz)
60	504	1512	2520	3524	
80	503	1509	2515	2519	
100	503	1507	2512	3516	
120	501	1506	2510	3513	
140	501	1505	2509	3411	
160	500	1505	2508	3510	

Pour des valeurs supérieures, F1 décroît assez rapidement alors que l'erreur sur F2 ne diminue pas. Des résultats analogues sont obtenus pour d'autres tubes.

Deux ensembles de données permettent de vérifier la précision de notre méthode de calcul des fréquences formantiques. MERMELSTEIN (1967) présente un calcul des fréquences de résonance pour les 6 voyelles du russe à partir des fonctions d'aire mesurées par FANT (1960).

Voyelles	MERMELSTEIN(1967)			Equation de Webster			Différences		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
i	224	2256	3166	220	2245	3207	4	9	-41
e	421	1971	2908	420	1981	2932	1	-10	-24
a	669	1139	2487	657	1128	2518	-12	11	-31
o	523	911	2350	523	913	2362	0	- 2	-12
u	250	601	2308	246	598	2325	- 4	3	-17
ɨ	283	1773	2319	288	1755	2389	- 5	-22	-70

La concordance est très bonne pour les valeurs de F1, et bonne pour F2 et F3. On peut également comparer les valeurs calculées par notre méthode avec les valeurs de MRAYATI et CARRE (1976) pour 12 voyelles françaises, si l'on utilise la partie du tableau (op. cit., p. 287, tab. 1) ne tenant compte que des pertes par viscosité et par transfert de chaleur, qui n'influent pratiquement pas sur le résultat, comme nous le verrons plus bas. Les fonctions d'aire retenues sont des constructions relativement hypothétiques qui ont été choisies de manière à redonner des fréquences formantiques compatibles avec les valeurs mesurées pour des voyelles naturelles.

Voyelles	MRAYATI & CARRE(1976)			Equation de Webster			Différences		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
i	195	2248	3208	189	2246	3239	6	2	-31
e	396	2193	3053	392	2197	3099	4	4	-46
ɛ	444	1957	2757	439	1960	2785	5	3	-28
ɑ	690	1438	2644	689	1440	2673	1	-2	-29
a	684	1275	2258	682	1277	2270	2	-2	-12
ɔ	519	1149	2583	515	1148	2605	4	-1	-22
o	375	865	2355	372	862	2472	3	3	-117
u	247	769	2276	245	763	2272	2	6	- 1
y	231	1949	2271	225	1952	2287	6	-3	-16
ø	344	1742	2513	341	1746	2528	3	-4	-15
œ	522	1510	2358	519	1510	2374	3	0	-16
ə	515	1539	2551	512	1538	2564	3	1	-13

Pour F1 et F2, l'accord entre les deux méthodes de calcul est extrêmement bon, et ceci sans aucune manipulation des résultats par un facteur comme la vitesse du son (cf. ZERLING 1974). Pour F3, nos valeurs sont légèrement surévaluées, ce qui pourrait provenir de la discrétisation. L'erreur importante pour le troisième formant de [o] reste inexplicable. Peut-être s'agit-il d'une erreur typographique.
ESPESSER (1977) a utilisé le même jeu de données et un programme dérivé de celui de ZERLING (1974). Les résultats qu'il obtient sont nettement différents. L'analyse des résultats détaillés laisse supposer qu'il n'a pas procédé à l'interpolation nécessaire.

Les résultats précédents montrent que l'erreur due à la discrétisation peut être minimisée en augmentant par interpolation le nombre des sections. Il convient d'étudier maintenant l'incorporation de corrections à la méthode fondée sur l'équation de Webster, afin de la rendre numériquement aussi précise que la méthode de MRAYATI et GUERIN.

IV - LES CORRECTIONS

Le calcul des fréquences de résonance par l'équation de Webster néglige en particulier les pertes par viscosité et conduction de chaleur qui ont été dérivées par FLANAGAN (1972, p. 28-34), de façon rigoureuse.
Le travail de MRAYATI et GUERIN permet de vérifier que ces pertes n'influent que d'une façon négligeable sur les fréquences de résonance (formants). Par contre, elles jouent un rôle important dans l'établissement des largeurs de bande. Les résultats détaillés pour les voyelles [ə] et [ɑ] sont les suivants :

	Fréquences formantiques (Hz) incluant l'effet des pertes par viscosité et conduction thermique			Fréquences formantiques sans l'effet des pertes			Différences		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
ə	521	1463	2413	520	1463	2414	1	0	1
ɑ	676	1192	2208	676	1192	2207	0	0	1

Aucune correction n'est donc nécessaire pour les pertes par viscosité ou conduction thermique.

FANT (1972) et SONDHI (1974) ont étudié les pertes par vibrations des parois. Il apparaît que ces pertes peuvent être représentées en fonction de la fréquence de résonance du conduit par l'équation suivante :

(3) $F_P = \sqrt{F_T^2 + F_F^2}$

où F_P désigne la fréquence de résonance, compte tenu des pertes par vibrations des parois, F_T la fréquence (théorique) du tube rigide et F_F la fréquence de résonance du conduit lorsque celui-ci est fermé aux lèvres. On peut trouver une dérivation rigoureuse dans MRAYATI et GUERIN (1976). Pour un tube rigide, la fréquence de résonance doit être nulle pour un tube fermé aux deux extrémités. En fait, la non-rigidité des parois permet une fréquence de résonance non-nulle. Pour un locuteur masculin, la fréquence de résonance "lèvres fermées" est de l'ordre de 180-200 Hz. Une section au spectrographe dans la tenue d'une consonne occlusive sonore montre que le premier harmonique (à une fréquence Fo x 2) est relativement intense. On trouvera une étude détaillée de ce dernier point dans UMEDA et COKER (1975).

F_P est donc sensiblement égal à $\sqrt{F_T^2 + 180^2}$

La table suivante montre que cette correction est particulièrement importante pour le premier formant lorsque celui-ci est à une fréquence peu élevée. La correction atteint 35 % à 200Hz, mais tombe à 0.16 % pour une fréquence de l'ordre de 3000 Hz.

F_T (Hz)	F_P (Hz)	Correction (Hz)
200	269	69
250	308.5	58.5
300	350	50
400	438.5	38.5
500	531.5	31.5
750	771	21
1000	1016	16
1500	1510.5	10.5
2000	2000	8
3000	3005	5

Mais les travaux de MRAYATI et GUERIN montrent que la formule (3) n'est valable que pour "les voyelles ayant une configuration autre qu'avant et fermée" (op. cit., p. 27).

Nous avons donc cherché à estimer de façon plus précise la contribution des vibrations des parois à l'augmentation des fréquences de résonance. On peut estimer, à partir des graphiques publiés par MRAYATI (1976), l'effet dû à la radiation, troisième et dernière source des pertes négligées par l'équation de Webster. Il est donc possible par soustraction d'obtenir une estimation quantitative de l'effet des vibrations des parois pour chaque voyelle. On peut ensuite calculer F1 pour chacune des voyelles. On a en effet :

$F1 = F_T + \Delta F_{vT} + \Delta F_P + \Delta F_R$: F_T : fréquence théorique pour conduit rigide ; ΔF_{vT} : correction pour pertes de viscosité et conduction thermique (celles-ci sont, comme nous l'avons vu, négligeables) ; ΔF_P : correction pour pertes par vibrations des parois ; ΔF_R : correction pour pertes dues à la radiation.

D'où $F_1 \cong F_T + \Delta F_P + \Delta F_R$

et $\Delta F_P = F1 - F_T - \Delta F_R$

On peut calculer F_F pour chaque voyelle ; $F_P = \sqrt{F_T^2 + F_F^2}$

d'où $F_F = \sqrt{F_P^2 - F_T^2}$

La figure 1-VI présente F_F en fonction de F_T. Le tableau suivant présente les valeurs numériques essentielles pour ce calcul.

Voyelles	Fréquences théoriques MRAYATI & G. F	F1 (toutes pertes incl.)	Effet de la radiation MRAYATI (76)	$\Delta F_P = F1 - F_T - \Delta F_R$	F_P	F_F
i	195	244	- 6	55	250	156 (Hz)
e	396	418	-14	36	432	173
ɛ	444	458	-20	34	478	177
a	690	678	-51	39	729	235
ɑ	684	676	-49	41	725	238
ɔ	519	540	-20	41	560	210
o	375	402	-15	42	417	182
u	247	296	-13	62	309	186
y	231	280	- 8	57	288	172
ø	344	384	- 8	48	392	188
œ	522	525	-26	39	561	206
ə	515	521	-28	34	549	190

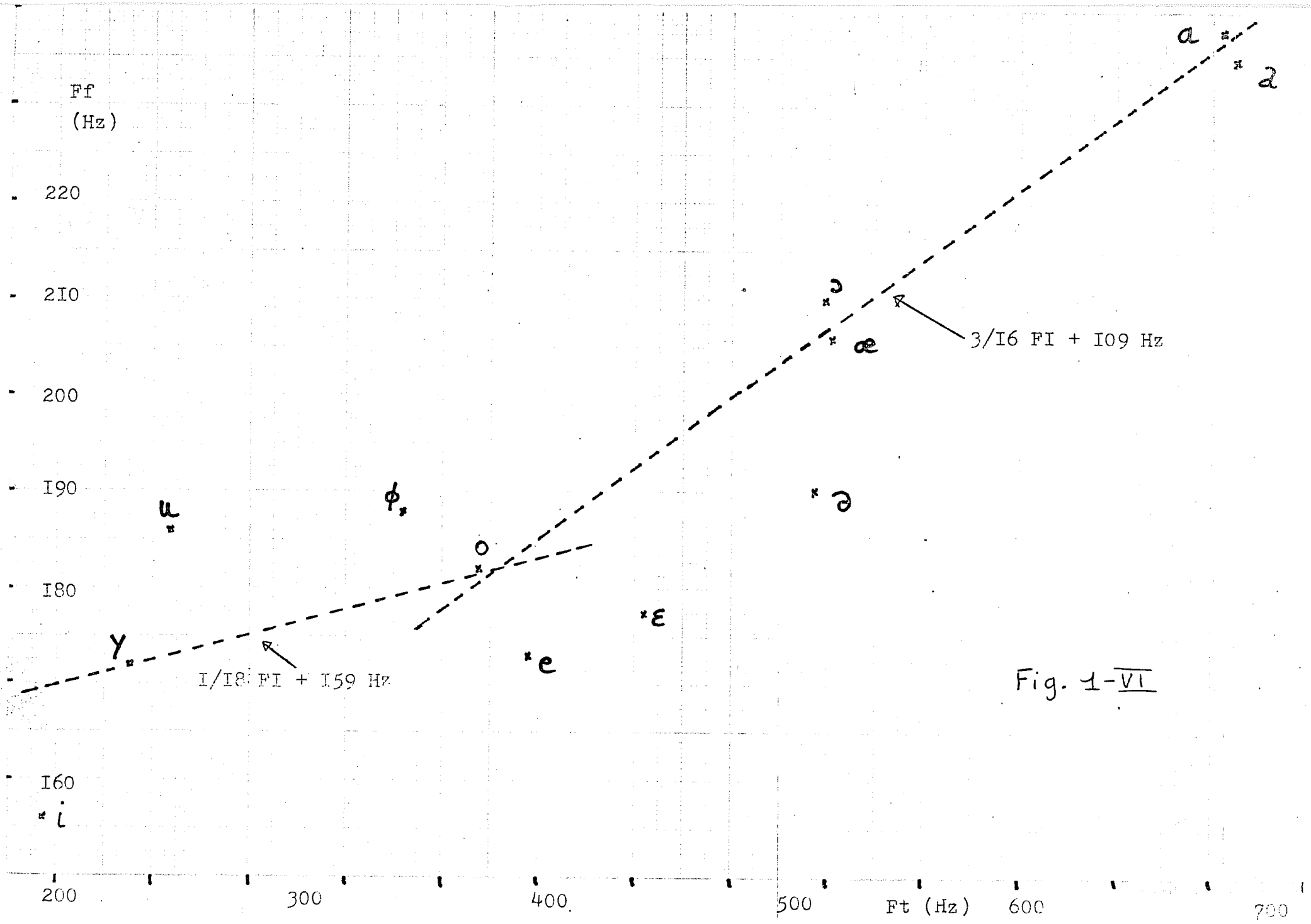

Ff
(Hz)
220
2I0
200
I90
I80
I60
200
300
400
500
Ft (Hz)
600
700
a
ɑ
ɔ
æ
ə
ɛ
e
o
ɸ
u
y
i
3/I6 FI + I09 Hz
I/I8 FI + I59 Hz

Fig. 1-VI

La figure 1-VI montre que F_F est une fonction croissante de F_T.
On peut approximer F_F par deux segments de droite selon que F_F est inférieur ou supérieur à 380 Hz.

Pour $190 < F_T < 380$ Hz $F_F = 1/18\ F1 + 158.9$ Hz (4)
Pour $380 < F < 750$ Hz $F_F = 3/16\ F1 + 108.75$ Hz (5)

Ces deux équations nous permettent de calculer de façon précise la contribution des vibrations des parois aux fréquences de résonance correspondant à F1. L'erreur résiduelle est faible comme le montre la comparaison entre la correction implicite effectuée par le modèle de MRAYATI et GUERIN et la correction $\Delta\rho$ calculée à l'aide des équations (3), (4) et (5). Ce tableau présente également l'effet d'une correction fixe (F_F = 190 Hz).

	$\Delta\rho$ calculé MRAYATI (76)	$\Delta\rho$ calculé à partir eq. 3,4 et 5	Différence	$\Delta\rho$ calculé à partir F fixe (190 Hz)	Différence
i	55	64	-9	77	-22
e	36	40	-4	43	- 7
ɛ	34	40	-6	39	- 5
ə	39	40	-1	26	+13
a	41	40	+1	26	+15
ɔ	41	40	+1	34	+ 7
o	42	41	+1	45	- 3
u	62	54	+8	65	- 3
y	57	57	0	68	-11
ø	48	43	+5	49	- 1
œ	39	39	0	34	+ 5
ə	34	39	-5	34	0

On voit que la correction variable permet de réduire l'erreur résiduelle au tiers de la valeur obtenue avec une correction fixe. En valeur absolue, l'erreur résiduelle ne dépasse pas 10 Hz.

Pour la dernière correction, qui concerne l'effet de la radiation aux lèvres (qui tend à faire baisser la fréquence

de résonance), le calcul est facilité par le fait qu'on peut établir une relation (FANT 1960, p. 293) entre la longueur du conduit et l'effet de la radiation : l'effet de la radiation est analogue à une augmentation Δl de la dernière section du conduit = $0.8\ (A/\pi)^{1/2}$ A : aire de la dernière section en cm2. C'est d'ailleurs cette correction qui est utilisée (en terme de circuit électrique par MRAYATI et GUERIN (1976). Quelques valeurs typiques sont données dans la table suivante :

A (cm2)	Δl (cm)
0.1	0.1427
0.2	0.2019
0.5	0.3192
1.0	0.4514
2.0	0.6383
5.0	1.009
10.0	1.4273

Nous avons incorporé cette correction dans le calcul des fréquences de résonance. Cette application sur les données initiales (fonction d'aire) conduit à une légère erreur que nous avons négligée. En effet le calcul de la correction due aux vibrations des parois sera effectué sur les fréquences de résonance déjà corrigées par l'effet de la radiation, et non par rapport à la fréquence théorique. Mais cette erreur est négligeable par rapport à la somme des erreurs provenant en particulier de la discrétisation.

Nous présentons en l'annexe A un programme Fortran qui incorpore l'ensemble des corrections. Ce programme remplace le programme écrit par ZERLING (1974). Pour les formes de conduit données par MRAYATI et GUERIN (1976), nous présentons ci-dessous un tableau comparatif des fréquences de résonance calculées par les deux méthodes.

	MRAYATI et GUERIN (1976)			Equation de Webster et corrections			Différences		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
i	244	2250	3082	254	2248	3095	10	-2	13
e	418	2111	2665	420	2119	2744	2	8	79
ɛ	458	1881	2584	460	1878	2598	2	-3	14
ɑ	678	1300	2583	683	1310	2604	5	10	21
a	676	1192	2208	678	1194	2215	2	2	7
ɔ	540	1007	2480	534	1014	2498	-6	7	18
o	402	797	2417	399	809	2436	-3	12	19
u	296	783	2286	289	783	2284	-7	0	-2
y	280	1846	2230	279	1856	2246	-1	10	16
ø	384	1592	2461	377	1591	2468	-7	-1	7
œ	535	1411	2269	530	1409	2277	-5	-2	8
ə	521	1463	2413	521	1467	2429	0	4	16

On voit que les différences sont négligeables pour F1 et F2 ; l'erreur ne dépasse pas 12 Hz. Quant à F3 l'erreur reste inférieure à 20 Hz, sauf pour [e] où elle ne dépasse pas 80 Hz. Les deux méthodes sont bien, comme nous l'affirmions dans l'introduction de ce chapitre, numériquement équivalentes.

V - CONCLUSIONS

On peut considérer que les principaux problèmes du calcul des fréquences de résonance à partir de la forme du conduit vocal sont aujourd'hui résolus.

Il reste néanmoins (FANT 1975) à préciser les interactions entre les vibrations des parois et les pertes à la glotte. Les phonéticiens ont à leur disposition des méthodes de calcul précises. Encore faut-il disposer des données de base, c'est-à-dire des fonctions d'aire pour les voyelles. C'est ce problème, lié à celui des limites de la méthode de prédiction linéaire, que nous aborderons dans le chapitre suivant, avant d'appliquer ces deux "outils" au problème des contraintes articulatoires dans la structuration des systèmes vocaliques.

A N N E X E

P R O G R A M M E FOR:

CALCUL DES FREQUENCES FORMANTIQUES A PARTIR DE LA FORME DU CONDUIT VOCAL

Ce programme a pour objet de calculer les 4 premières fréquences de résonance d'un tube acoustique de section variable. Deux corrections sont effectuées, afin de tenir compte de l'influence des pertes par vibration des parois et par effet de radiation. Ce programme utilise une équation de Webster discrétisée.

PARAMETRES: I - NA: nombre de conduits à analyser
NF: nombre de formants (max = 4)
(la numérotation renvoie à celle reproduite sur le listing du prog.)
2 - N : nombre de sections décrivant la fonction d'aire
XL: longueur d'une section (XL x K = longueur tot.)
V : vitesse du son (347 - 353 m/s)
NX: facteur d'interpolation (K x NX doit être ≅I50)

ENTREE: 3 - A(I): lecture d'une fonction d'aire

PROGRAMME: 4 - Interpolation
4'- Incrément de longueur simulant l'effet de radiation
5 - Calcul des fréquences par approximations successives
6 - Correction pour effet de paroi

SORTIE: 7 - Fonction d'aire après interpolation
8 - affichage des paramètres(2).
9 - Fréquences formantiques (correction radiation incluse mais sans correction de l'effet de paroi)
I0 - Fréquences formantiques (correction paroi incluse)

EXEMPLE NUMERIQUE (II): NA=I,NF = 4; fonction d'aire: voyelle e MRAYATI et GUERIN (I976) table II page 26; V = 347,5 ; NX = 8
K = I6, XL = I cm.
Résultats de MRAYATI et GUERIN: FI = 4I8, F2 = 2III, F3 = 2665 Hz

```
      DIMENSION A(200),P(200),F(4),AX(200)
      READ(105,70) NA,NF                                  I
   70 FORMAT(I2,1X,I1)
      IF(NF.NE.4) F(4)=0.
      DO71IL=1,NA
      READ(105,10)K,XL,V,NX                               2
   10 FORMAT(I2,1X,F4.3,1X,F5.0,1X,I2)
      READ(105,20)(A(I),I=1,K)                            3
   20 FORMAT(16(F4.2,1X))
      PI=3.1415926
      IC=0
      DO62I=1,K
      AX(I)=A(I)
   62 CONTINUE
      JJ=1
      DO60I=1,K                                           4
      DO61J=1,NX
      A(JJ)=AX(I)
      JJ=JJ+1
   61 CONTINUE
   60 CONTINUE
      K=K*NX
      XL=XL/NX
   63 CST=(2*PI*XL/V)**2
      YL=0.8*SQRT(AX(1)/PI)
      YLL=YL/XL
      IYLL=YLL
      IF((YLL-IYLL).GE.0.5)IYLL=IYLL+1
      DO80J=(K+IYLL),(IYLL+1),-1                          4'
      A(J)=A(J-IYLL)
   80 CONTINUE
      DO81J=1,(IYLL)
      A(J)=A(IYLL+I)
   81 CONTINUE
      K=K+IYLL
      P(1)=0.
      P(2)=2.
      FB=200
      I=1
    6 DP1=(-1.)**(I+1)
      DF=32
    7 DO8J=3,K+1
      P(J)=(1.-CST*FB**2+A(J-2)/A(J-1))*P(J-1)-(A(J-2)/A(J-1))*(P(J-2))
    8 CONTINUE
      DP2=P(K+1)-P(K)
      Q=DP1/DP2
      IF(Q.GE.0.)GOTO9                                    5
      DF=DF/(-2)
      IF(ABS(DF).GE.1.)GOTO9
      FB=FB+DF
      GOTO12
    9 FB=FB+DF
      IC=IC+1
      IF(IC.GT.200)STOP'BOUCLE'
      DP1=DP2
      GOTO7
   12 F(I)=FB
      IF(I.EQ.NF)GOTO11
      I=I+1
      FB=FB+300
      IC=0
      GOTO6
```

```
11 WRITE(108,40)(A(I),I=1,K)
40 FORMAT(10(3X,F7.4))
   WRITE(108,50)K,XL,V,NX
50 FORMAT(10X,I3,7X,F5.3,7X,F6.0,7X,I2)
   WRITE(108,30)(F(I),I=1,4)
30 FORMAT(5X,'F1 = ',F6.1,10X,'F2 = ',F6.1,10X,'F3 = ',F6.1,10X,'F4 =
  1',F6.1/)
   IF(F(1).GE.380.)GOTO82
   FW=F(1)/18.+159.
   GOTO83
82 FW=3*F(1)/16.+109.
83 CONTINUE
   F(1)=SQRT(F(1)**2+FW**2)
   FW=200.
   DO72I=2,NF
   F(I)=SQRT(F(I)**2+FW**2)
72 CONTINUE
   WRITE(108,73)(F(I),I=1,3)
73 FORMAT(5X,'CORRECTION VIBRATIONS PAROIS',5X,'F1 = ',F6.1,5X,
  1'F2 = ',F6.1,5X,'F3 = ',F6.1/)
71 CONTINUE
   STOP
```

II	5.5000	5.5000	5.5000	5.5000	5.5000	5.5000	5.5000	5.5000	5.5000	5.5000
	5.5000	5.5000	5.5000	5.5000	5.5000	5.5000	5.0000	5.0000	5.0000	5.0000
	5.0000	5.0000	5.0000	5.0000	3.3000	3.3000	3.3000	3.3000	3.3000	3.3000
	3.3000	3.3000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000
	1.7000	1.7000	1.7000	1.7000	1.7000	1.7000	1.7000	1.7000	3.4000	3.4000
	3.4000	3.4000	3.4000	3.4000	3.4000	3.4000	4.5000	4.5000	4.5000	4.5000
	4.5000	4.5000	4.5000	4.5000	5.5000	5.5000	5.5000	5.5000	5.5000	5.5000
	5.5000	5.5000	8.0000	8.0000	8.0000	8.0000	8.0000	8.0000	8.0000	8.0000
	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000
	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000	10.5000
	10.5000	10.5000	10.5000	10.5000	8.2000	8.2000	8.2000	8.2000	8.2000	8.2000
	8.2000	8.2000	7.1000	7.1000	7.1000	7.1000	7.1000	7.1000	7.1000	7.1000
	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000	2.3000	2.3000
	2.3000	2.3000	2.3000	2.3000	2.3000	2.3000				

```
        136        .125        34750.        8
 F1 =  379.5        F2 = 2109.0        F3 = 2736.5        F4 =3425.0

 CORRECTION VIBRATIONS PAROIS     F1 =  420.1     F2 = 2118.5     F3 = 2743.8
```

C H A P I T R E VII

APPLICATION DE LA METHODE DE PREDICTION LINEAIRE AU CALCUL DE LA FORME DU CONDUIT VOCAL A PARTIR DES FREQUENCES FORMANTIQUES

I - INTRODUCTION

La méthode d'analyse du signe vocal par prédiction linéaire (MARKEL et GRAY 1976) est une méthode récente qui présente de nombreux avantages par rapport aux méthodes classiques analogiques ou numériques, telles que l'analyse spectrographique, la transformée de Fourier rapide (FFT : BRIGHAM 1974, COOLEY et TUCKEY 1965), la méthode du cesptre (SCHAFER et RABINER 1970) ou la méthode de PINSON (PINSON 1963). Si le premier article qui lui a été consacré, en japonais, date de 1966 (SAITO et ITAKURA 1966), il a fallu attendre 1972 (MARKEL 1972) pour voir apparaître une analyse claire et détaillée des concepts et algorithmes nécessaires. Comme la méthode de PINSON (LONCHAMP 1974), la prédiction linéaire considère chaque période du signal vocal comme une somme de 3 à 6 sinusoïdes amorties. Ce modèle, moins général que le modèle issu du théorème de Fourier (qui considère une période comme le résultat d'une somme de 30 à 60 sinusoïdes harmoniques) permet, pour des sons vocaliques sans nasalité, une détection aisée et un calcul précis des fréquences de résonance ou formants. Notons dès à présent que le modèle de production postulé par l'analyse prédictive linéaire ne s'applique pas aux parties du signal vocal dont le spectre comprend, en plus des pôles (formants), un certain nombre de zéros. Ceci touche principalement les voyelles et consonnes nasales, et dans une moindre mesure, les consonnes sonantes, les fricatives et les occlusives sonores. Cet aspect sera repris dans le chapitre traitant des voyelles nasales. Un avantage de cette méthode est la simplicité et la rapidité d'exécution des algorithmes; on peut envisager un calcul des fréquences de formants en temps réel à l'aide d'un petit nombre de microprocesseurs (KNUDSEN 1975).

Le dernier avantage, et celui qui nous intéresse plus particulièrement dans ce travail, est la possibilité de calculer, de proche en proche, à partir des coefficients extraits par prédiction linéaire, l'aire d'un tube cylindrique de section variable qui possède la même fonction de transfert que le

conduit vocal (WAKITA 1973). Toutes ces caractéristiques expliquent l'extraordinaire floraison d'études sur, ou utilisant, la prédiction linéaire. Son emploi est quasi général dans les laboratoires s'occupant d'analyse, de synthèse ou de reconnaissance de la parole.

II - PREDICTION LINEAIRE : LES FONDEMENTS THEORIQUES

Malheureusement, il n'existe pas, à notre connaissance, d'introduction élémentaire satisfaisante à la théorie de la prédiction linéaire. En particulier, il existe un hiatus considérable entre l'exposé purement descriptif, sans dérivations mathématiques, (cf. par ex. WAKITA 1976) et les exposés très techniques (MARKEL et GRAY 1976). Nous essaierons de combler partiellement cette lacune en présentant une dérivation des équations de base. Certaines notions nécessaires sont présentées en annexe de ce chapitre.

L'observation d'une période d'une voyelle (fig. 1-VII) conduit à discerner une certaine "régularité" dans la forme. Notons en particulier l'absence de discontinuités brutales. Cette "régularité" permet, ayant observé le début de la période, de prédire la fin de celle-ci. Par exemple, si l'on considère la figure 2-VII, où un signal a été tracé jusqu'à l'instant t-1, on peut prédire, par continuité, que le signal à l'instant t, sera plus proche du point 10 ou 11 que du point 5 ou 15.

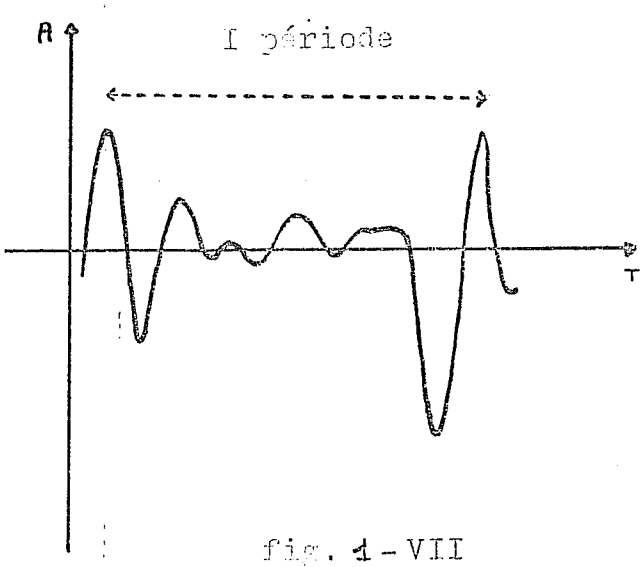


fig. 1-VII

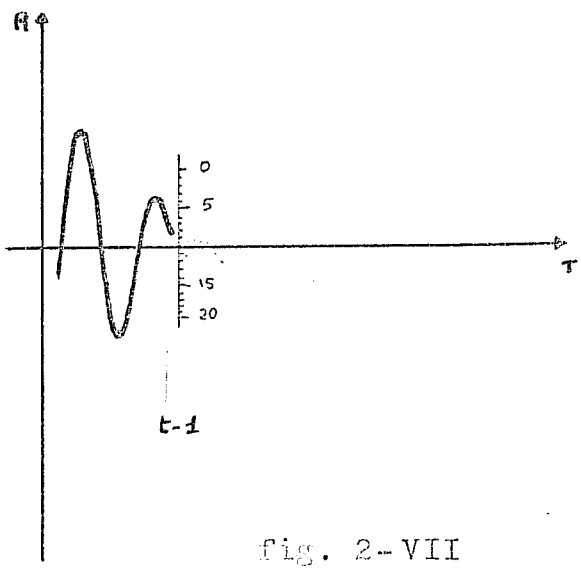


fig. 2-VII

Le chasseur, avant de tirer un oiseau en vol, fait une prédiction linéaire. L'oiseau, pendant le temps que mettent les plombs pour parvenir jusqu'à lui, se déplacera d'une certaine distance dans une certaine direction, ce qui oblige le chasseur à tirer devant l'oiseau. Cette prédiction ne sera correcte que si le vol est régulier, c'est à dire si l'oiseau ne modifie pas trop sa vitesse ou son angle de vol dans l'intervalle de temps considéré.

On exprime formellement ce qui précède en notant que la valeur d'un échantillon de parole, préalablement digitalisé, à l'instant n (s_n), peut être prédit avec une faible erreur par une somme pondérée de m valeurs antérieures du même signal aux instants n-1, n-2, ..., n-m. On note $\hat{s}_n$ le signal prédit.

$$\hat{s}_n = \sum_{i=1}^{m} a_i s_{n-i} = a_1 s_{n-1} + a_2 s_{n-2} + \ldots\ldots + a_m s_{n-m}$$

Cette équation n'est autre que l'équation d'une droite dans un espace à m + 1 dimensions passant par l'origine. Montrons-le par exemple pour un espace à 3 dimensions. Notons que $\hat{s}_n$ est une fonction linéaire de s_{n-1}

$$\hat{s}_n = \alpha_1 s_{n-1}$$

Si $\hat{s}_n$ est de la même façon une fonction linéaire de s_{n-2}, on peut écrire $\hat{s}_n = \alpha_2 s_{n-2}$. On a $2\hat{s}_n = \alpha_1 s_{n-1} + \alpha_2 s_{n-2}$ d'où $\hat{s}_n = \frac{\alpha_1}{2} s_{n-1} + \frac{\alpha_2}{2} s_{n-2}$ qui est bien de la forme z=ax+by+c (avec c=0), équation d'une droite dans un espace à 3 dimensions. On peut aisément généraliser. c = 0 car il est évident que si les m échantillons précédents sont nuls (absence de signal) $\hat{s}_n$ sera nul lui aussi.

L'erreur de prédiction e_n commise est $e_n = s_n - \hat{s}_n = s_n - \sum_{i=1}^{m} a_i s_{n-i}$ $= \sum_{i=0}^{m} a_i s_{n-i}$ avec $a_0 = 1$.

Nous verrons plus loin que les coefficients de pondération a_i permettent de calculer le spectre de la voyelle, ou plus précisément la fonction de transfert d'un filtre ayant la voyelle comme signal de sortie. Le problème est donc de calculer les coefficients a_i, i = 1,....., m de manière à rendre l'erreur de prédiction e_n la plus petite possible.

Pour améliorer la précision et pour des raisons de simplicité mathématique, on ne minimise pas un seul e_n, mais la somme des carrés des e_n (notée E) pour une séquence d'échantillons commençant à n_0 et se terminant à n_1. $E = \sum_{n=n_0}^{n_1} e^2_n$

E peut encore s'écrire : $\sum_{n=n_0}^{n_1} \left[\sum_{i=0}^{m} a_i \, s(n-i) \right]^2$

La quantité entre crochets est de la forme $\left[\sum_{i=0}^{m} \alpha_i \right]^2$ et cette expression vaut :

$$[\alpha_0 + \alpha_1 + \ldots + \alpha_m][\alpha_0 + \alpha_1 + \ldots + \alpha_m] =$$

$$\begin{array}{l} \alpha_0\alpha_0 + \alpha_0\alpha_1 + \alpha_0\alpha_2 + \ldots + \alpha_0\alpha_m + \\ \alpha_1\alpha_0 + \alpha_1\alpha_1 + \alpha_1\alpha_2 + \ldots + \alpha_1\alpha_m + \\ \vdots \\ \alpha_m\alpha_0 + \alpha_m\alpha_1 + \alpha_m\alpha_2 + \ldots + \alpha_m\alpha_m \end{array}$$

c'est-à-dire $\sum_{i=0}^{m} \sum_{j=0}^{m} \alpha_i \alpha_j$. $E = \sum_{n=n_0}^{n_1} \sum_{i=0}^{m} \sum_{j=0}^{m} a_i \, s(n-i) a_j \, s(n-j)$

que l'on peut réécrire $E = \sum_{i=0}^{m} \sum_{j=0}^{m} a_i a_j \left[\sum_{n=n_0}^{n_1} s(n-i) s(n-j) \right]$

L'expression entre parenthèses, que nous noterons Cij ne dépend que de la valeur des échantillons aux instants n-i et n-j.

$$E = \sum_{i=0}^{m} \sum_{j=0}^{m} a_i a_j C_{ij}$$

Pour que E soit minimum, il faut que pour chaque a_i, i = 0,...m la dérivée partielle correspondante de E soit égale à zéro.

Développons E : $E = a_0 a_0 C_{00} + a_0 a_1 C_{01} + \ldots + a_0 a_m C_{0m} +$

$$\begin{array}{l} a_1 a_0 C_{10} + a_1 a_1 C_{11} + \ldots + a_1 a_m C_{1m} + \\ \vdots \\ a_m a_0 C_{m0} + a_m a_1 C_{m1} + \ldots + a_m a_m C_{mm} \end{array}$$

Calculons par exemple $\frac{\partial E}{\partial a_0}$. Les dérivées des termes ne comportant pas a_0 sont nulles.

Ils se comportent en effet comme des constantes par rapport à a_0

Les seuls termes qui interviennent sont donc :

$$S = a_0 a_0 C_{00} + a_0 a_1 C_{01} + a_0 a_2 C_{02} + \ldots + a_0 a_m C_{0m} +$$
$$a_1 a_0 C_{10} + a_2 a_0 C_{20} + \ldots + a_m a_0 C_{m0}$$

Or Cij = Cij . En effet $C_{ij} = \sum_{n=n_0}^{n_1} s(n-i) s(n-j) = C_{ji} = \sum_{n=n_0}^{n_1} s(n-j) s(n-i)$

On a donc $a_i a_0 C_{i0} = a_i a_0 C_{0i}$

S peut donc s'écrire :

$$a_0^2 C_{00} + 2 a_0 a_1 C_{01} + 2 a_0 a_2 C_{02} + \ldots + 2 a_0 a_m C_{0m}$$

$$\frac{\partial E}{\partial a_0} = 2\, a_0 C_{00} + 2\, a_1 C_{01} + 2\, a_2 C_{02} + \dots\dots + 2\, a_m C_{0m}$$
$$= 2 \sum_{i=0}^{m} a_i C_{0i}$$

Pour un a_k quelconque, on a de même $\frac{\partial E}{\partial a_k} = 2 \sum_{i=0}^{m} a_i C_{ki} = 0$

On se souvient que a_0 a été pris égal à 1. On peut donc écrire

$$\frac{\partial E}{\partial a_k} = 0 = 2\, a_0 C_{0k} + 2 \sum_{i=1}^{m} a_i C_{ki}$$
$$= 2\, C_{0k} + 2 \sum_{i=1}^{m} a_i C_{ki} = 0$$

ce qui entraîne finalement $-C_{0k} = \sum_{i=1}^{m} a_i C_{ki}$ pour k = 1,....,m.
Les a_i sont donc les m inconnues de m équations linéaires simultanées.
Le terme Cok et les termes Cki sont directement calculables car ils ne dépendent que de la valeur des échantillons. Le problème du calcul des a_i qui minimisent l'erreur de prédiction est donc résolu.

De nombreuses méthodes de résolution sont disponibles et beaucoup d'efforts ont été déployés pour trouver des algorithmes rapides prenant en compte les caractéristiques des ai et Cij (qui ont la forme d'une matrice de T oeplitz) : méthode de Levinson (MARKEL 1972, méthode de DURBIN (MAKHOUL 1975), méthode de LE ROUX 1976). Deux méthodes peuvent être utilisées pour le calcul des Cij. Nous utiliserons exclusivement, pour des raisons de stabilité et de simplicité des algorithmes, la méthode dite "d'autocorrélation". On suppose $n_0 = -\infty$, $n_1 = \infty$, $s_n = 0$ pour $n < 0$, et $n \geqslant N$. Cij se simplifie alors et on le note R_L avec $L = |i - j|$

$$R_L = \sum_{n=0}^{N-1-L} s(n)\, s(n+L) \quad L \geqslant 0$$

(MARKEL et GRAY 1976, p. 14-15).

III - ANALYSE SPECTRALE

Nous allons montrer maintenant que les coefficients a_i permettent de calculer le spectre de la voyelle. La démonstration classique utilise la transformée en z du signal d'entrée et de sortie. On peut cependant, en remarquant que l'équation définissant $\hat{s}_n$ n'est pas récursive, donner une démonstration

plus simple utilisant la notion de "phaseur" (STEIGLITZ 1974, cf. annexe C de ce chapitre). Les coefficients a_i peuvent être considérés comme les coefficients d'un "filtre". Appliquons le phaseur $X(k) = e^{j\omega k}$ à l'entrée du filtre défini par

$$S(k) = a_1 s(k-1) + a_2 s(k-2) + \ldots\ldots + a_m s(k-m)$$

La valeur du phaseur à l'instant S(k) est $e^{j\omega k}$

A l'instant S(k-i) le phaseur vaut $e^{j\omega(k-i)}$. Or $e^{j\omega(k-i)}$ est égal à $e^{j\omega k - j\omega i} = e^{j\omega k} . e^{-j\omega i}$

comme on le démontre à l'annexe B.

S(k) peut donc s'écrire :

$$S(k) = a_1 e^{j\omega k} . e^{-j\omega} + a_2 e^{j\omega k} . e^{-2j\omega} + \ldots + a_m e^{j\omega k} . e^{-mj\omega}$$

$$= e^{j\omega k} \left(a_1 e^{-j\omega} + a_2 e^{-2j\omega} + \ldots\ldots + a_m e^{-mj\omega} \right)$$

$$= X(k) \left(a_1 e^{-j\omega} + a_2 e^{-2j\omega} + \ldots\ldots + a_m e^{-mj\omega} \right)$$

Nous obtenons le résultat fondamental suivant : Le signal de sortie n'est autre que le phaseur appliqué comme signal d'entrée modifié par la quantité

$$a_1 e^{-j\omega} + a_2 e^{-2j\omega} + \ldots + a_m e^{-mj\omega}$$

Précisons la signification de cette quantité. C'est une somme de nombres complexes, donc un nombre complexe $H(\omega)$, comme le montre la dérivation suivante :

$$H(\omega) = a_1 \cos\omega - a_1 j \sin\omega + a_2 \cos 2\omega - a_2 j \sin 2\omega + \ldots + a_m \cos m\omega - a_m j \sin m\omega$$

$$H(\omega) = \left(a_1 \cos\omega + a_2 \cos 2\omega + \ldots + a_m \cos m\omega\right) - j\left(a_1 \sin\omega + \ldots + a_m \sin m\omega\right)$$

$H(\omega)$ est appelé fonction de transfert du filtre. On voit que $H(\omega)$ est égal à S(k)/X(k) qui est le rapport signal de sortie/signal d'entrée.

On a en effet $S(k) = X(k) . H(\omega)$. La valeur de $H(\omega)$ dépend uniquement de ω. Notons $Re^{-j\theta}$ le nombre complexe correspondant à $H(\omega)$.

$$S(k) = e^{j\omega k} R e^{-j\theta} = R e^{j(\omega k - \theta)}$$

On voit que S(k), c'est-à-dire le signal de sortie, est un phaseur de même fréquence que X(k), le signal d'entrée, mais

dont l'amplitude est multipliée par la quantité réelle R et qui est déphasé d'une quantité θ par rapport au phaseur d'entrée.

Par définition, $R = \sqrt{A^2 + B^2}$ avec $A = a_1 \cos \omega + \dots + a_m \cos m\omega$

$B = a_1 \sin \omega + \dots + a_m \sin m\omega$

et θ = arctan (B/A).

En faisant varier ω, on peut calculer, pour chaque valeur de ω, la valeur du coefficient multiplicateur R. Pour une voyelle, R varie schématiquement en fonction de ω de la manière représentée à la figure 3-VII.

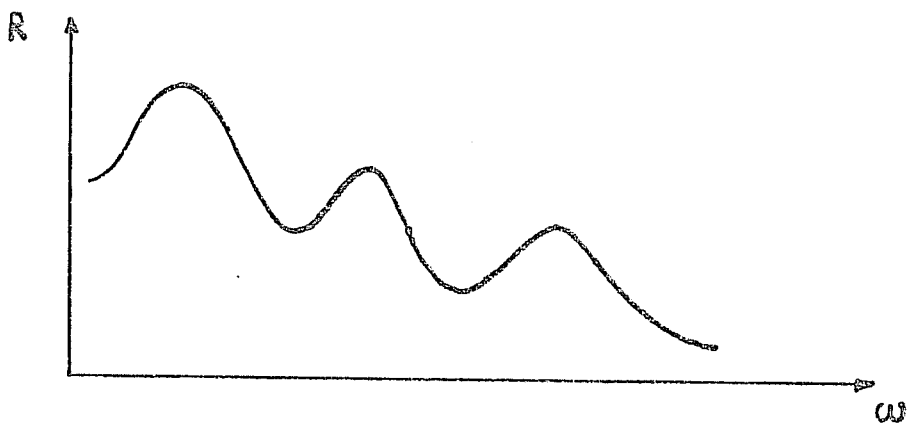


fig. 3- VII

Le phonéticien notera la présence de vallées et de pics dans lesquels il reconnaîtra les formants.

Le filtre défini par les coefficients a_i a donc pour effet d'atténuer de façon sélective, en fonction de la fréquence, l'amplitude d'une sinusoïde présentée à l'entrée.

On peut montrer qu'un tel filtre est linéaire, c'est-à-dire que l'application de la somme de plusieurs phaseurs (sinusoïdes) de fréquences différentes redonne bien une somme de phaseurs sans changement de fréquences. Les modifications des amplitudes et des phases sont les mêmes que si les phaseurs avaient été présentés de façon isolée.

IV - LE MODELE DE PRODUCTION IMPLICITE

Nous venons de montrer que la méthode de prédiction linéaire permet de calculer les coefficients d'un filtre d'ordre m. Mais il nous faut préciser la forme du modèle de production implicite dans cette formulation. Pour ce faire, nous serons obligés d'utiliser la transformée en z dont la définition et le théorème fondamental sont présentés à l'annexe D.

Présentons d'abord le modèle de production de la parole couramment utilisé, pour les sons vocaliques sans nasalité.
Il se compose de la série de filtres décrits à la figure 4-VII.

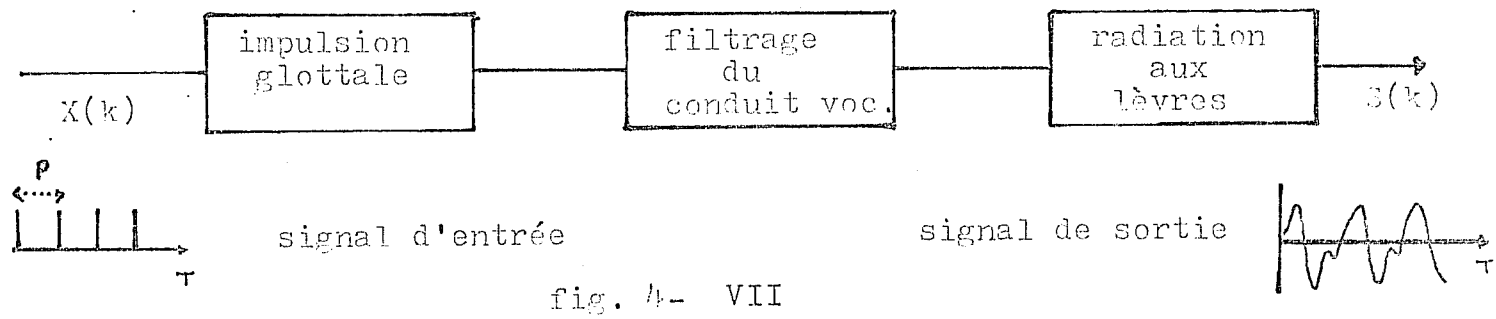


fig. 4- VII

Dans ce modèle, le signal d'entrée est constitué par une série d'impulsions espacées de $P = 1/F_o$ secondes. Le premier filtre possède une fonction de transfert qui donne aux composantes spectrales du signal d'entrée une pente moyenne de -12 dB/octave à partir de 100-300 Hz. La fonction des autres filtres est évidente. On peut envisager un modèle de production modifié, où le signal d'entrée est représenté directement par l'impulsion glottale. Ce nouveau modèle permet de supprimer le premier filtre (fig. 5-VII). C'est le modèle utilisé en synthèse de la parole (par ex. ESPESSER 1977). Mais la discussion qui suit exige que nous conservions le premier modèle.

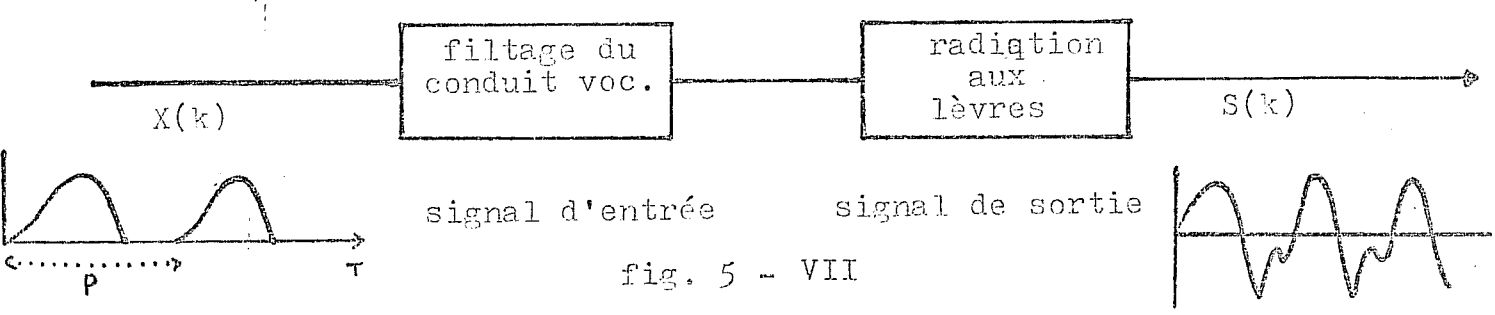


fig. 5 - VII

Le deuxième filtre représente l'effet du conduit vocal. Pour les sons sans nasalité, la fonction de transfert ne comporte que des pôles. Le troisième filtre représente l'effet de la radiation aux lèvres. En utilisant la transformée en z, on peut écrire que le signal de sortie (signal de parole) est le filtrage d'un signal d'entrée E(z) constitué par des impulsions espacées de $1/F_{os}$ par 3 filtres : G(z) décrivant la mise en forme de l'impulsion glottale, V(z) le filtrage dû à la cavité articulatoire et R(z) l'effet de la radiation :

$$S(z) = E(z) \cdot G(z) \cdot V(z) \cdot R(z)$$

Notons que E(z) n'est autre qu'une constante. En effet, sur une période, la suite des échantillons est A, 0, 0, 0,.....,0. la transformée en z d'une telle suite est, par définition, $E(z) = Az^{0} + 0z^{-1} + 0z^{-2} + \ldots\ldots = A$. G(z) peut être considéré comme un filtre passe-bas à deux pôles donnant une pente spectrale moyenne de -12 dB par octave (eq. 1.4 : MARWEL et GRAY 1976). L'effet de la radiation peut être représenté par une dérivation (discrète) en fonction du temps. En effet

$$\frac{d s_n}{dt} = \frac{s_n - s(n-1)}{\Delta t} = \frac{1}{\Delta t} s_n - \frac{1}{\Delta t} s(n-1)$$

La transformée en z est $S(z) = \frac{1}{\Delta t} S(z) - \frac{1}{\Delta t} z^{-1} S(z) = \frac{1}{\Delta t} S(z)(1 - z^{-1})$

La fonction de transfert est $\frac{1}{\Delta t}(1 - z^{-1})$. La constante $1/\Delta t$ ne dépend que de l'échelle de temps adoptée, qui est arbitraire. On peut donc prendre $1/\Delta t = 1$. Du point de vue spectral, l'effet de la radiation est de rehausser approximativement le spectre de + 6 dB par octave.

Le modèle de production implicite dans l'utilisation de la prédiction linéaire est le suivant. On sait que l'erreur de prédiction est

$$e_n = s_n - \sum_{i=1}^{m} a_i s_{n-i} = s_n - a_1 s_{n-1} - a_2 s_{n-2} - \ldots - a_m s_{n-m}$$

La transformée en z des deux premiers membres de l'égalité donne (on sait que la transformée en z de $s(n-i) = z^{-i} s(n)$)

$$\begin{aligned} E(z) &= S(z) - a_1 z^{-1} S(z) - a_2 z^{-2} S(z) - \ldots\ldots - a_m z^{-m} S(z) \\ &= S(z)\left[1 - a_1 z^{-1} - a_2 z^{-2} - \ldots\ldots - a_m z^{-m}\right] \\ &= S(z)\left[1 - \sum_{i=1}^{m} a_i z^{-i}\right] \end{aligned}$$

Notons par H(z) la partie entre crochets, que l'on peut réécrire $\sum_{i=0}^{m} a_i z^{-i}$, avec $a_0 = 1$. On a donc E(z) = S(z). H(z).

L'erreur de prédiction peut donc être considérée comme la sortie d'un filtre H(z) ayant pour entrée les échantillons du signal de parole (Fig. 6-VII).

fig. 6 - VII

H(z) est l'inverse du filtre décrivant la production des échantillons de parole, d'où le nom de filtrage inverse donné à la méthode par MARKEL. Le modèle implicite de production est donc (Fig. 7- VII) S(z) = E(z)/H(z).

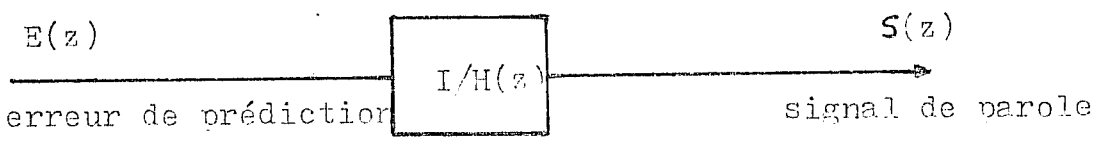


fig. 7 - VII

Si l'on identifie le signal d'entrée au signal d'entrée E(z)=A du modèle général de production présenté plus haut, il s'ensuit que : 1/H(z) = G(z). V(z). R(z)

$$H(z) = \sum_{i=0}^{m} a_i z^{-i}$$ ne comprend que des zéros.

Ceci est évident si l'on considère l'expansion suivante de H(z)

$$H(z) = a_0 z^0 + a_1 z^{-1} + \ldots\ldots + a_m z^{-m}$$

mettons z^{-m} en facteur :

$$H(z) = z^{-m}\left(a_0 z^m + \frac{a_1}{a_0} z^{m-1} + \ldots + \frac{a_m}{a_0}\right)$$

La partie entre parenthèses est une équation du $m^{\text{ième}}$ degré, qui a donc m racines, $z_1, z_2, \ldots\ldots\ldots z_m$.

H(z) peut donc s'écrire $H(z) = z^{-m}(z - z_1)(z - z_2)\ldots(z - z_m)$

La présence d'un pôle supposerait une racine infinie, ce qui

est impossible, étant donné que le système est physiquement réalisable et stable. La fonction de transfert de $1/H(z)$ ne comprend donc que des pôles. On en déduit donc le résultat fondamental suivant :

L'utilisation de la prédiction linéaire présuppose que la fonction de transfert de $G(z).V(z).R(z)$ ne comprend que m pôles à l'exclusion de tout zéro. On peut vérifier que ceci est approximativement vrai. On sait (FANT 1960) que $V(z)$ ne comprend que des pôles conjugués complexes pour des sons sans nasalité. $R(z)$ en première analyse est égal à $1-z^{-1}$.

$$G(z) = \frac{1}{(1-\alpha z^{-1})(1-\beta z^{-1})}$$

avec

$$\alpha \cong \beta = e^{-cT} \qquad c = 2\pi B$$

$B \cong$ 100 à 300Hz

T= fréquence d'échantillonnage

$$G(z).V(z).R(z) = \frac{1-z^{-1}}{(1-\alpha z^{-1})(1-\beta z^{-1})}$$

Comme $\alpha \cong 1$, le numérateur $1-z^{-1}$ annule approximativement $1-\alpha z^{-1}$. Il ne reste donc plus que des pôles.

On a en effet $G(z).V(z).R(z) = \dfrac{V(z)}{(1-\alpha z^{-1})} = 1/H(z)$.

Pour obtenir l'estimation de la fonction de transfert du seul conduit vocal $V(z)$ (c'est-à-dire pour pouvoir calculer les fréquences des formants), il faudra annuler l'effet du filtre passe-bas $(1-\alpha z^{-1})$ qui donne un effet spectral moyen de -6dB par octave. Cette annulation sera obtenue en procédant à une dérivation préalable ("pre-emphasis") utilisant une nouvelle fois la relation approximative $\dfrac{1-z^{-1}}{1-\alpha z^{-1}} \cong 1$.

Pour résumer le dernier point : si $S(z) = E(z)/H(z)$, alors

$$\begin{aligned} S(z)\,(1-z^{-1}) &= E(z)\,(1-z^{-1})\,/H(z) \\ &= E(z)\,(1-z^{-1})\;V(z)/(1-\alpha z^{-1}) \\ &= E(z).V(z) \quad \text{d'où } E(z) = \frac{S(z)\,(1-z^{-1})}{V(z)} \end{aligned}$$

L'ESTIMATION DE LA FONCTION DE TRANSFERT DU CONDUIT VOCAL SERA DONC OBTENUE EN OPERANT UN FILTRAGE INVERSE DE LA DERIVEE TEMPORELLE DU SIGNAL DE PAROLE. (fig. 8-VII)

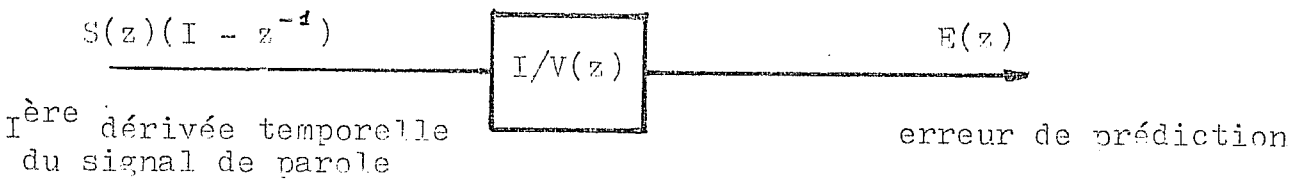


fig. 8 - VII

V - PREDICTION DE LA FORME DU CONDUIT VOCAL

La deuxième partie de ce chapitre traite du calcul de la forme du conduit vocal. La méthode de prédiction linéaire permet une estimation directe de la fonction d'aire à partir du signal de parole. Cette méthode est supérieure aux méthodes proposées antérieurement, et en particulier à celle utilisée par ZERLING (1974) (méthode de KADOKAWA et SUZUKI 1968). WAKITA (1973) a montré que les coefficients de corrélations partielles k_i, calculés comme étape intermédiaire du calcul des a_i, sont égaux aux coefficients de réflexion acoustique μ d'une onde de pression dans un tube non-uniforme. On trouvera une bonne illustration dans WAKITA (1976, p. 15-18). La complexité de la partie acoustique ne permet pas une démonstration détaillée.

Les coefficients de réflexion acoustique μ_m sont égaux à

$$\frac{S_m - S_{m+1}}{S_m + S_{m+1}} = \mu_m$$

où S_m et S_{m+1} représentent les aires de deux sections adjacentes. ATAL et HANAUER (1971) ont montré que les fréquences formantiques et les largeurs de bande sont suffisantes pour définir de façon unique la forme du conduit acoustique. Ils ont montré de plus qu'une fonction de transfert (spectre) à M pôles conjugués complexes peut, dans tous les cas, être

considérée comme la fonction de transfert d'un tube acoustique à M sections de longueurs égales. Une analyse acoustique permet la mesure de 4 ou 5 formants. Il est donc possible, à partir de ces 8 ou 10 pôles, de calculer une approximation de la forme du conduit vocal en 8 ou 10 sections de 2 cm environ de longueur. Ceci suppose pourtant que l'on fournisse une estimation des largeurs de bande. Il est intéressant de noter que les conditions théoriques sont les mêmes que celles qui régissent les équations de Webster : la propagation de l'onde est plane, les vibrations des parois et les autres sources de perte sont négligeables. L'influence de la radiation doit également être compensée* des pertes dans le chapitre précédent apparaît maintenant.

Pour permettre un calcul aussi précis que possible de la forme du conduit, il convient de "corriger" l'onde acoustique. Cette correction ne peut se faire directement au niveau de l'onde de pression captée par le microphone. Pour tourner cette difficulté, nous proposons la chaîne de traitement suivante: (fig. 9-VII).

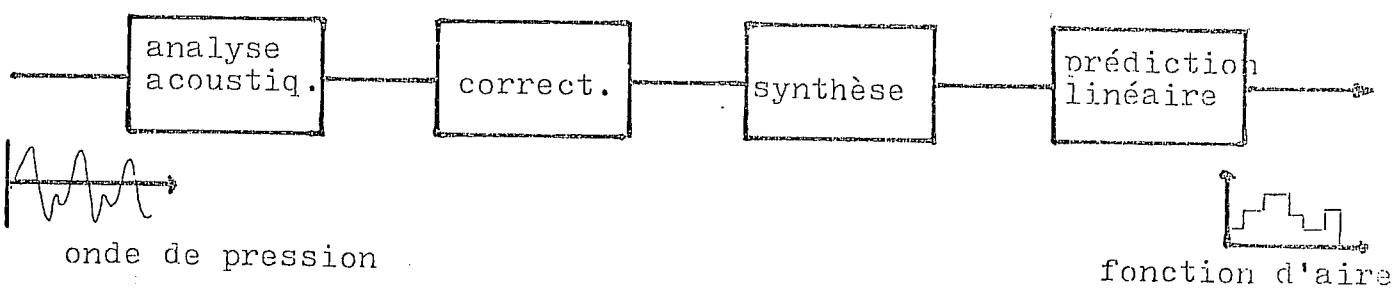


fig. 9 - VII

Dans un premier temps, on réalise une analyse acoustique du signal de parole. La méthode employée importe peu (analyse spectrographique, prédiction linéaire, F.F.T.). Cette analyse fournit les fréquences formantiques. Celles-ci sont corrigées en appliquant, en sens inverse, les équations présentées au chapitre précédant. On obtient, par ce biais, les caractéristiques fréquentielles d'un tube acoustique conforme aux conditions d'application de la méthode de prédiction linéaire. On

* par une pré-emphase spectrale. La raison de notre discussion détaillée

réalise ensuite une synthèse numérique de l'onde de pression, à partir de laquelle on calcule une estimation de la forme du conduit vocal. Le choix des largeurs de bande pose un problème délicat. La théorie acoustique du tube rigide (FLANAGAN 1972, p. 60) indique que les largeurs de bande théoriques sont très faibles (de 5 à 15 Hz). Une simulation nous a permis de constater que les fonctions d'aire obtenues avec des valeurs aussi faibles ne sont pas de bonne qualité. On peut envisager l'emploi de valeurs moyennes (MRAYATI et GUERIN 1976, FANT 1972) bien qu'il soit difficile de donner une justification théorique. Le problème n'est pas résolu par l'estimation des largeurs de bande à partir de l'onde de pression initiale, car on sait que celles-ci peuvent être incorrectes si l'analyse n'est pas faite sur la phase fermée de la voyelle (EL-MALLAWANY 1977, CHANDRA et LIN 1974). WAKITA (1977a) préconise également des corrections pour les largeurs de bande, sans autre précision. Signalons enfin qu'est paru, alors que ce chapitre était en cours de rédaction, un travail de WAKITA (1977a) qui développe une méthode de correction essentiellement identique à la nôtre. WAKITA fait appel au modèle de synthèse de ISHIZAKA et FLANAGAN (1972) qui incorpore les différentes sources de pertes acoustiques.

Notre procédure peut paraître lourde, mais elle est rendue obligatoire par une contrainte de la méthode. La forme du conduit n'est définie de façon unique que pour une longueur particulière du conduit vocal. WAKITA (1973) a démontré que la période d'échantillonnage T devait être choisie en fonction de la longueur du conduit de manière à satisfaire

$$l = c\,M\,T/2 = c\,M/2\,Fs$$

(l : longueur du conduit ; c : vitesse du son ; M : nombre de sections ; Fs : fréquence d'échantillonnage = 1/T). En règle générale, la longueur du conduit vocal n'est pas connue. Elle peut varier de 16 à 19 cm pour un sujet masculin (cf. chapitre VIII ; FANT 1960). L'utilisation d'une longueur fixe (17,5 cm par exemple), et par conséquent d'une fréquence d'échantillonnage unique, conduit à des formes de conduit qui ne correspondent pas à la réalité. Mais ces formes, que nous proposons

d'appeler "pseudo-formes", sont suffisamment proches des formes réelles pour certaines applications particulières, comme le calcul des paramètres articulatoires dans un système de rééducation des enfants sourds (CRICHTON et FALLSIDE 1974a, b). L'erreur est encore plus réduite si l'on compare ces "pseudo-formes" à des "pseudo-formes" typiques obtenues dans les mêmes conditions (avec des enfants à l'audition normale).

D'un point de vue pratique, il est impossible de faire varier la fréquence d'échantillonnage en cours d'analyse. La méthode de prédiction linéaire n'est donc pas entièrement adaptée au temps réel. EL-MALLAWANY (1975, 1976, 1977) a présenté une étude très complète des procédures permettant une utilisation optimale dans ces conditions. Rappelons également qu'il n'est pas possible de calculer la forme du conduit buccal pour les sons nasalisés.

VI - TEST DE LA PRECISION DE L'ESTIMATION DES FONCTIONS D'AIRE

Le problème de la précision des estimations est difficile à résoudre car nous devons disposer d'un ensemble de fonctions d'aire précises déterminées par une autre méthode. La seule source disponible (FANT 1960) présente des fonctions d'aire pour 6 voyelles russes : [i, e, a, o, u, ɨ]. FANT souligne d'ailleurs la difficulté de la mesure des fonctions d'aire par radiographie latérale et tomographie. Il ne se prononce pas sur la précision de ses mesures. MARKEL et GRAY (1976, p. 82-83, d'après WAKITA 1974 ; non publié) présentent une étude comparative des fonctions d'aire mesurées et calculées pour les 5 voyelles non-centrales de FANT.

La procédure adoptée est complexe. On recherche d'abord par essai et erreur sur une simulation d'un conduit vocal avec pertes (ISHIZAKA et FLANAGAN 1972) une fonction à 8 sections dont les fréquences de résonance soient égales aux fréquences mesurées par FANT. Ces fonctions d'aire "théoriques" sont les meilleures estimations possibles en fonction de nos connaissances actuelles. On calcule ensuite l'écart normalisé

moyen ε entre ces formes théoriques et les fonctions d'aire mesurées par FANT. Cet écart est donné par

$$\varepsilon = \sum_{i=1}^{m} \sqrt{\frac{(Aim - Aic)^2}{m}}$$

(m : nombre de sections ; Aim : aire de la $i^{ème}$ section mesurée ; Aic : aire de la $i^{ème}$ section calculée ; Acmx : aire (calculée) de la plus grande section). L'écart normalisé varie entre 10 % et 3 %, pour la voyelle [e] et [u] respectivement. Ceci indique que, toute source d'erreur confondue, les mesures de FANT sont affectées d'une erreur ne dépassant pas 10 % dans le cas le plus défavorable. On peut enfin, à partir des voyelles synthétisées sur l'analogue du conduit vocal, estimer la précision de la méthode de prédiction linéaire. L'erreur normalisée entre les fonctions mesurées et le résultat de l'analyse varie entre 8 et 13 % (voyelle [a]). On voit que la méthode de prédiction linéaire n'est guère plus imprécise que la mesure directe sur des tracés radiologiques. Nous avions procédé à une estimation personnelle de la précision de la méthode avant de prendre connaissance des résultats de WAKITA. Notre méthode, moins sophistiquée, a consisté à calculer les fréquences de résonance des fonctions d'aire de FANT par une équation de Webster (cf. chapitre VI). Ces estimations n'ont pas été corrigées, afin de respecter les conditions d'application du modèle. Ces fréquences formantiques ont servi à synthétiser une onde de pression acoustique sur laquelle a été effectuée la prédiction linéaire. Les largeurs de bande retenues (50, 70, 80, 100 et 150 Hz) sont des valeurs moyennes. Cette méthode est sans doute légèrement moins précise que la précédente car elle suppose exactes les fonctions d'aire mesurées par FANT. Par contre nous supprimons l'influence des pertes acoustiques. Une combinaison des deux méthodes devrait être effectuée, ainsi qu'une évaluation de l'influence des largeurs de bande.

Nous présenterons les résultats pour la voyelle [e]. Les autres résultats sont comparables. On a d'abord calculé une approximation, en 8 sections, de la fonction d'aire en 33 sections donnée par FANT.

	section n°	Aire (cm2)
lèvres	1	6.25
	2	2.70
	3	3.70
	4	5.80
	5	9.15
	6	9.50
	7	6.70
glotte	8	2.35

La longueur du conduit vocal a été réduite à 15,5 cm (cf. ch. VII) car il nous semble que FANT a surestimé la longueur du conduit vocal.

Le calcul des fréquences de résonance donne les résultats suivants :

	F1	F2	F3	F4
[e]	436	2056	3040	4000 (Hz)

Le léger décalage avec les valeurs présentées au chapitre précédent s'explique par un choix différent de la vitesse du son (353 m/s, au lieu de 347,5 précédemment).

L'estimation des aires par la méthode de WAKITA donne les résultats présentés ci-dessous. Rappelons que les fonctions d'aire calculées ne sont connues qu'à une constante multiplicative près. Elles ont été normalisées de telle manière que l'aire calculée la plus grande soit égale à la plus forte aire mesurée (section n° 6). Ces résultats sont présentés graphiquement à la fig. 10-VII.

	section n°	FANT	Calculé
lèvres	1	6,25	5.18 [cm²]
	2	2,70	2.92
	3	3,70	2.60
	4	5,80	4,68
	5	9,15	7.29
	6	9.50	9.50
	7	6,70	9.05
glotte	8	2,35	5.38

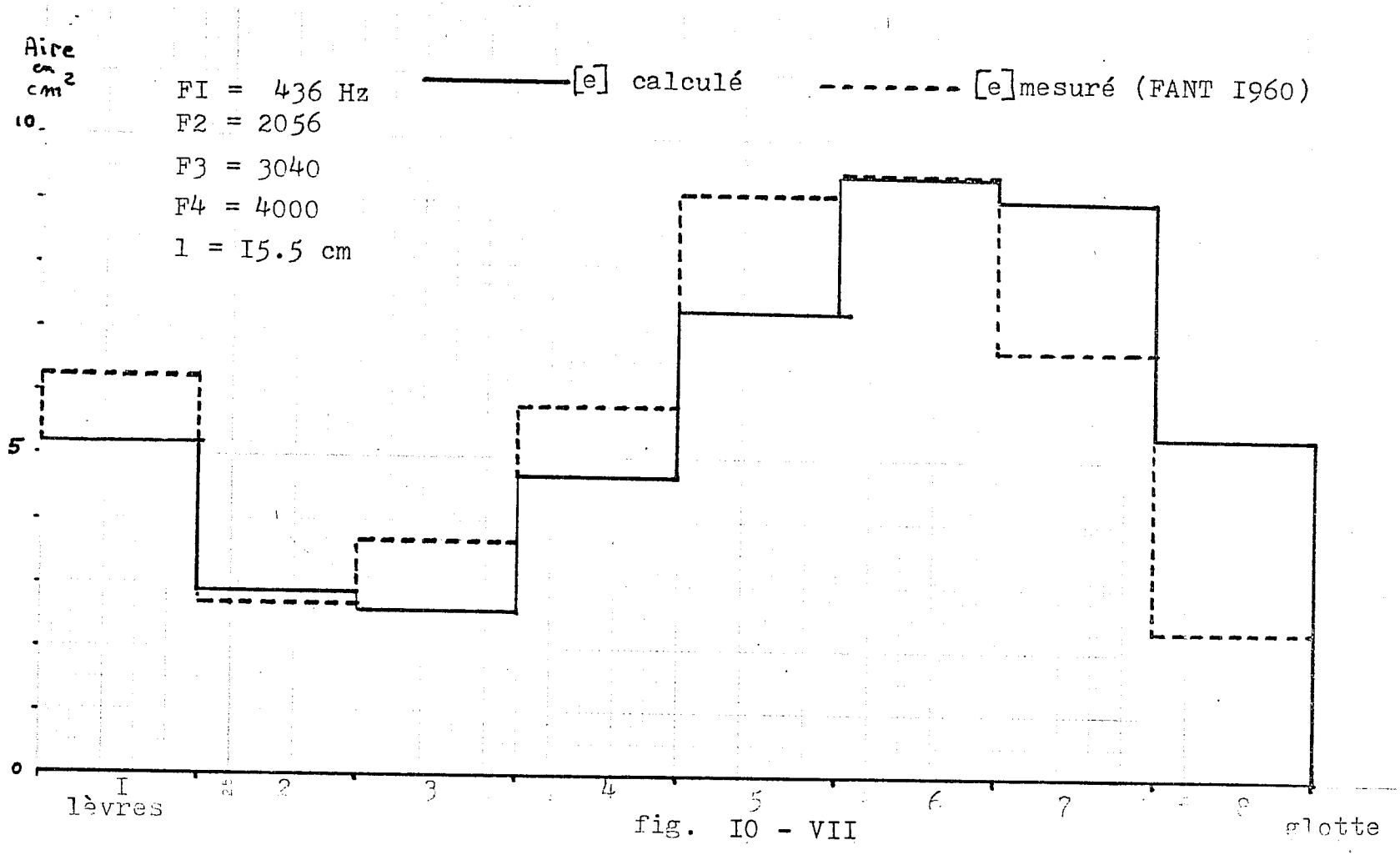

Aire
cm²
10
5
0
FI = 436 Hz
F2 = 2056
F3 = 3040
F4 = 4000
l = I5.5 cm
[e] calculé
[e]mesuré (FANT I960)
I
lèvres
2
3
4
5
6
7
8
glotte

fig. IO - VII

A titre de vérification, nous avons estimé à nouveau les fréquences de résonance du tube calculé. On obtient les résultats suivants :

	F1	F2	F3	F4
valeurs initiales	436	2056	3040	4000
valeurs recalculées	426	2057	3048	4017
% erreur	-2,35	0,05	0,26	0,43

L'accord est excellent, en particulier pour le deuxième formant. Ceci confirme que le facteur d'interpolation utilisé dans le calcul des formants par une équation de Webster discrétisée sous-estime légèrement F1 et surestime très légèrement F3 et F4.

L'erreur normalisée entre les deux formes de conduit atteint 17,4 %. Cette valeur est supérieure à celle qu'obtient WAKITA (1974). L'erreur sur la 8ème section est relativement importante. Il faudrait vérifier que le calcul des fréquences de résonance, à partir de la fonction d'aire décrite par 33 sections, ne provoque pas une augmentation importante de l'erreur. Le rôle des largeurs de bande en synthèse doit également être précisé.

On constate néanmoins que la méthode de WAKITA permet de retrouver les caractéristiques articulatoires essentielles de la voyelle [e] : aire aux lèvres assez importante, constriction maximale antérieure, importance de la cavité postérieure par rapport à la cavité antérieure.

VII - CONCLUSIONS

Ce chapitre a présenté le deuxième "outil" que nous utiliserons pour étudier le rôle des contraintes articulatoires. Il permet, avec une précision acceptable, d'extraire directement de l'onde de pression des renseignements sur la forme du conduit vocal, et par là sur les mécanismes articulatoires mis en jeu. Nous présentons en annexe le listing Fortran d'un programme d'ordinateur permettant d'effectuer le calcul de la fonction d'aire.

ANNEXE A

Nombres complexes

Les nombres complexes ont été introduits pour donner, dans tous les cas, une solution à des polynômes du type

$$x^2 + 1 = 0. \qquad x^2 = -1$$

On pose comme solution dans ce cas $x = \pm j$ où j est un nombre "complexe".

Il n'existe pas en effet de nombre réel qui, multiplié par lui-même, donne un nombre négatif. On est amené à poser de ce fait $j.j = -1$ et $-j.-j = -1$

Par définition, on considère qu'un nombre complexe z est la somme d'une partie réelle (qui peut être nulle) et d'une partie complexe $j.b$ (b réel).

$$z = a + j.b$$

Une représentation géométrique commode est la représentation vectorielle suivante (fig. 11-VII).

Par définition, l'addition de deux nombres complexes

$z_1 = a_1 + j.b_1$ et $z_2 = a_2 + j.b_2$ est donnée par

$$z_1 + z_2 = (a_1 + a_2) + j.(b_1 + b_2)$$

Pour définir la multiplication, nous noterons d'abord que le vecteur z de la figure 11-VII peut être décrit également par les deux paramètres suivants :

(fig. 12-VII) - l'angle θ que fait z avec l'axe des réels
- le module (longueur) R du vecteur

On notera $z = R\angle\theta$ $\qquad z_1 = R_1\angle\theta_1 = a_1 + j.b_1$

$z_2 = R_2\angle\theta_2 = a_2 + j.b_2$

Par définition, le résultat de la multiplication de z_1 par z_2 est égale à $R_1 R_2 \angle \theta_1 + \theta_2$ (fig. 13-VII).

Calculons formellement $(a_1 + j.b_1)(a_2 + j.b_2)$: on obtient

$$z_1 . z_2 = a_1 a_2 + a_1 jb_2 + a_2 jb_1 + jb_1 jb_2$$

$$= a_1 a_2 - b_1 b_2 + j(a_1 b_2 + a_2 b_1)$$ en utilisant la propriété $j.j = -1$

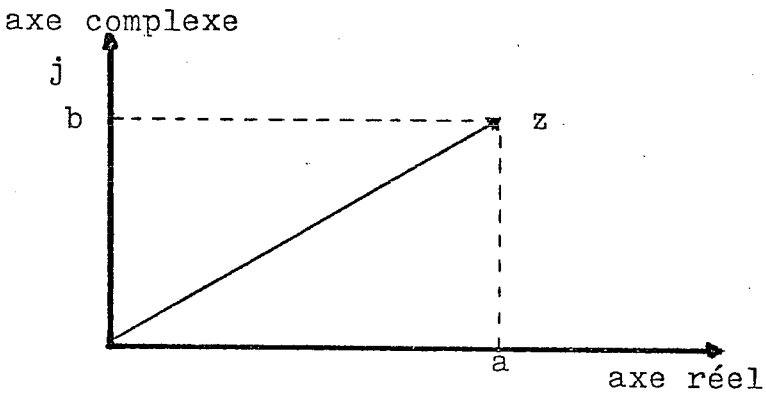


Fig. II - VII

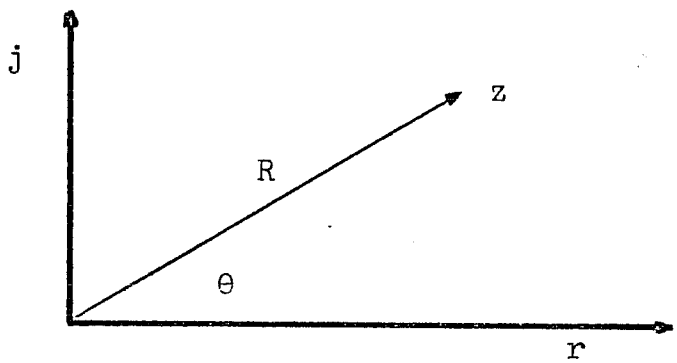


Fig. I2 - VII

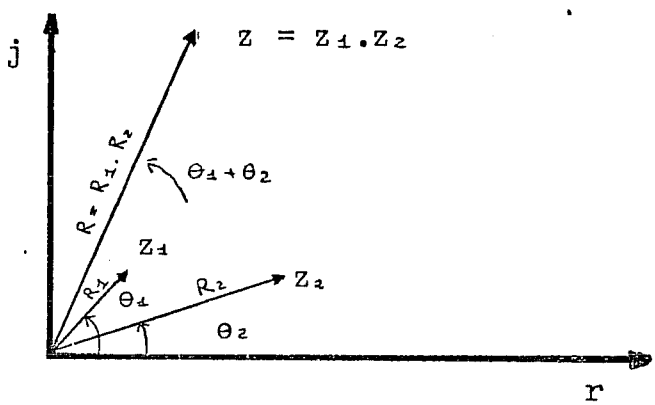


Fig. I3 - VII

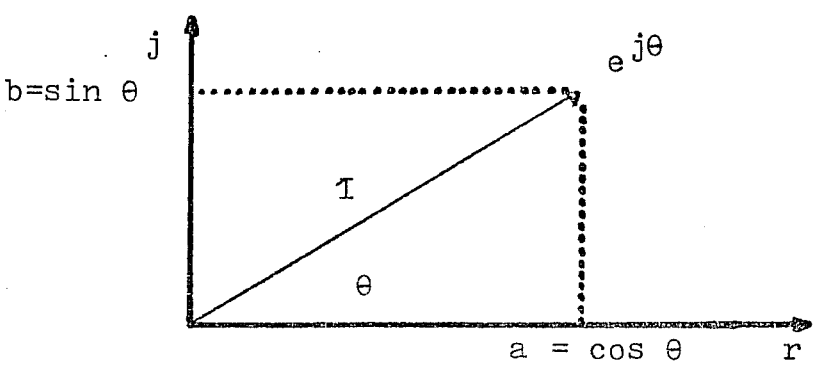


Fig. I4 - VII

- La partie réelle de z est $R_1 R_2 \cos(\theta_1 + \theta_2) = R_1 R_2 (\cos\theta_1 . \cos\theta_2 - \sin\theta_1 . \sin\theta_2) = R_1 \cos\theta_1 \; R_2 \cos\theta_2 - R_1 \sin\theta_1 \; R_2 \sin\theta_2)$.

$R_1 \cos\theta_1$ n'est autre que la partie réelle de $z_1 = a_1$
$R_2 \cos\theta_2$ est la partie réelle de $z_2 = a_2$
$R_1 \sin\theta_1$ est la partie imaginaire de $z_1 = b_1$
$R_2 \sin\theta_2$ est la partie imaginaire de $z_2 = b_2$

La partie réelle de $z_1 . z_2$ est donc égale à $R_1 R_2 \cos(\theta_1 + \theta_2)$
$= (a_1 a_2 - b_1 b_2)$.

- La partie imaginaire de $z_1 . z_2$ est $R_1 R_2 \sin(\theta_1 + \theta_2)$
$= R_1 R_2 (\sin\theta_1 \cos\theta_2 + \sin\theta_2 \cos\theta_1)$
$= R_1 \sin\theta_1 R_2 \cos\theta_2 + R_1 \cos\theta_1 R_2 \sin\theta_2$
$= b_1 a_2 + b_2 a_1$

$z_1 . z_2$ est donc égal à $(a_1 a_2 - b_1 b_2) + j(a_1 b_2 + a_2 b_1)$
On voit que les 2 définitions concordent si on interprète $j.j$ comme égale à -1.

ANNEXE B

Exponentielle complexe

Par définition, on note $e^{j\theta} = \cos\theta + j \sin\theta$.

$e^{j\theta}$ correspond au vecteur z de la figure 14-VII. Le choix de la notation e peut être considéré comme purement mnémotechnique. Elle rappelle que $e^{j\theta}$ et e^{x} obéissent à la même règle d'addition des exposants :

$$e^{x} . e^{y} = e^{x+y} \; ; \; e^{j\theta} . e^{j\omega} = e^{j(\theta+\omega)}$$

d'après les définitions $e^{j\theta} . e^{j\omega} = (\cos\theta + j \sin\theta)(\cos\omega + j \sin\omega)$
$= \cos\theta \cos\omega - \sin\theta \sin\omega + j(\sin\theta \cos\omega + \sin\omega \cos\theta)$
$= \cos(\theta + \omega) + j \sin(\theta + \omega)$

cette dernière expression est bien égale à $e^{j(\theta+\omega)}$, d'après la définition.

ANNEXE C

Notion de "Phaseur"

Nous traduisons par "phaseur" le terme anglais "phasor" popularisé par STEIGLITZ (1974) dans "An Introduction to Discrete Systems" , Wiley, New York. D'une façon surprenante pour le néophyte, il est plus aisé mathématiquement d'utiliser une fonction complexe plutôt qu'une fonction sinus ou cosinus (réelle) pour calculer l'effet d'un filtre sur l'amplitude ou la phase d'une sinusoïde.
Une sinusoïde discrétisée, c'est-à-dire définie numériquement à certains points k du temps régulièrement espacés, a pour équation :

$$F(k) = A \cos(\omega k + \phi) \quad \text{(fig. 15-VII)}$$

A : amplitude maximale du signal ; ω fréquence en radians par seconde $F = 2\pi\omega$, ϕ est la phase (qui définit l'amplitude à l'instant k = 0) ; k : variable entière.

Le phaseur Φ (k) correspondant est

$$A\, e^{j(\omega k + \phi)} = A \cos(\omega k + \phi) + j A \sin(\omega k + \phi)$$

On voit que F (k) est la partie entière de Φ (k), c'est-à-dire la projection de Φ (k) sur l'axe des réels. Le phaseur (k) est présenté à la fig. 16-VII pour k = 0, 1, 2....

Le phaseur peut être considéré comme un vecteur tournant d'un angle ω dans le sens inverse des aiguilles d'une montre chaque fois que k augmente d'une unité. L'extrémité du vecteur se déplace le long de la circonférence d'un cercle de rayon A. Si ω est petit, c'est-à-dire si la fréquence est basse, le vecteur tourne lentement. Considérons la représentation d'une fréquence égale à π + a (fig. 17-VII).
On suppose qu'à l'instant 0, le vecteur tournant est sur l'axe des réels (phase nulle). A l'instant k = 1, le vecteur a tourné de $\pi + a$. Mais on peut également considérer qu'il a tourné d'un angle $\pi - a$ dans le sens négatif (sens des aiguilles d'une montre). A l'instant k = 2,le vecteur est en 2, c'est-à-dire un demi-tour + a plus loin, ou un demi-tour - a dans le

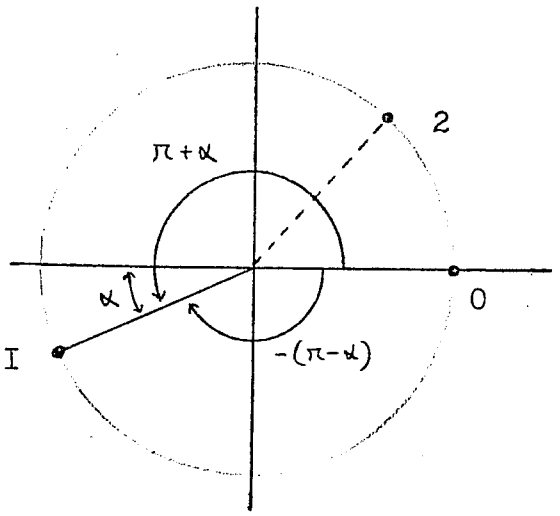


Fig. I7 - VII

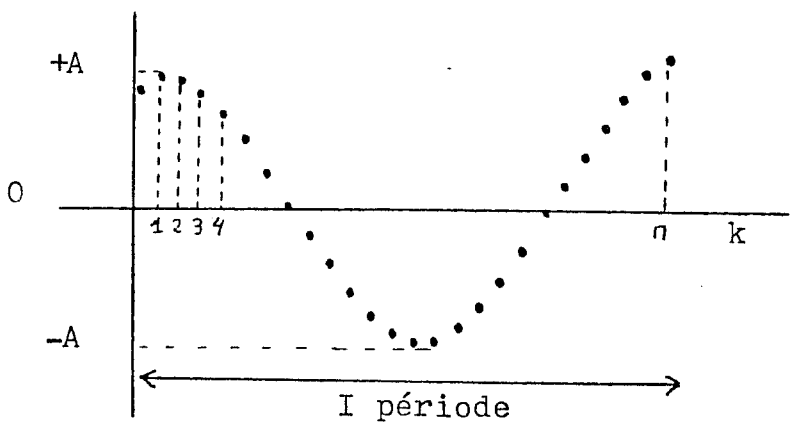


Fig. I5 - VII

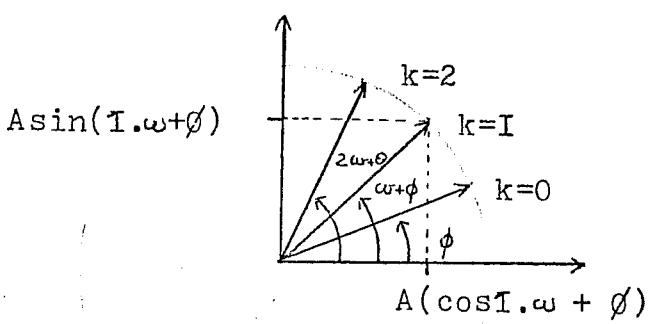


Fig. I6 - VII

sens négatif. Or le signal réel (projection du vecteur sur l'axe réel) est le même, que le vecteur tourne dans le sens positif ou dans le sens négatif. La fréquence la plus élevée que l'on peut représenter par un phaseur d'une façon non ambiguë correspond à un vecteur tournant de π radians entre chaque échantillon. Cette fréquence correspond à la moitié de la fréquence d'échantillonnage. Un signal de fréquence supérieure à π ($\pi + a$ par exemple) apparaît comme un phaseur de fréquence $\pi - a$. La fréquence égale à π radians est appelée "fréquence de Nyqvist". Son dépassement provoque le "repliement" du spectre (fig. 18-VII). Ceci impose le filtrage des fréquences supérieures à la fréquence de Nyqvist.

ANNEXE D

TRANSFORMEE en Z : Définition et théorème fondamental

Le logarithme est un exemple de "transformée" qui transforme en particulier les produits en sommes :
$\log(x.y) = \log(x) + \log(y)$

Définition : Soit un signal échantillonné $x(k)$ défini aux instants $k = 0, 1, \ldots\ldots\ldots, \infty$

Par définition, la transformée en z de ce signal (notée $X(z)$) est : $X(z) = x(0)\, z^{0} + x(1)\, z^{-1} + x(2) z^{-2} + x(3) z^{-3} + \ldots\ldots$

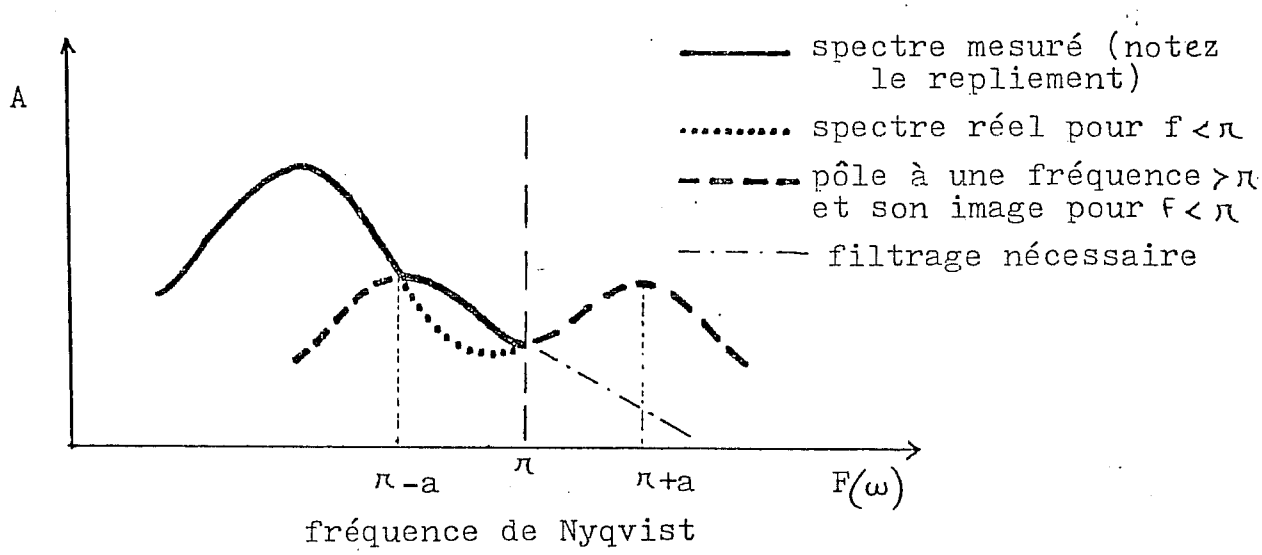
$$= \sum_{k=0}^{\infty} x(k)\, z^{-k}$$

z est un nombre complexe quelconque : l'annexe E montre qu'il est utile de lui donner la valeur $e^{j\theta}$ pour le calcul de la fonction de transfert d'un filtre.

Considérons un signal digitalisé $x(k)$ qui est nul pour $k < 0$,
et un autre signal défini par $y(k) = x(k-1)$. On a donc

$y(0) = x(-1)$
$y(1) = x(0)$
$y(z) = x(1)$ etc....

Fig. I8 - VII



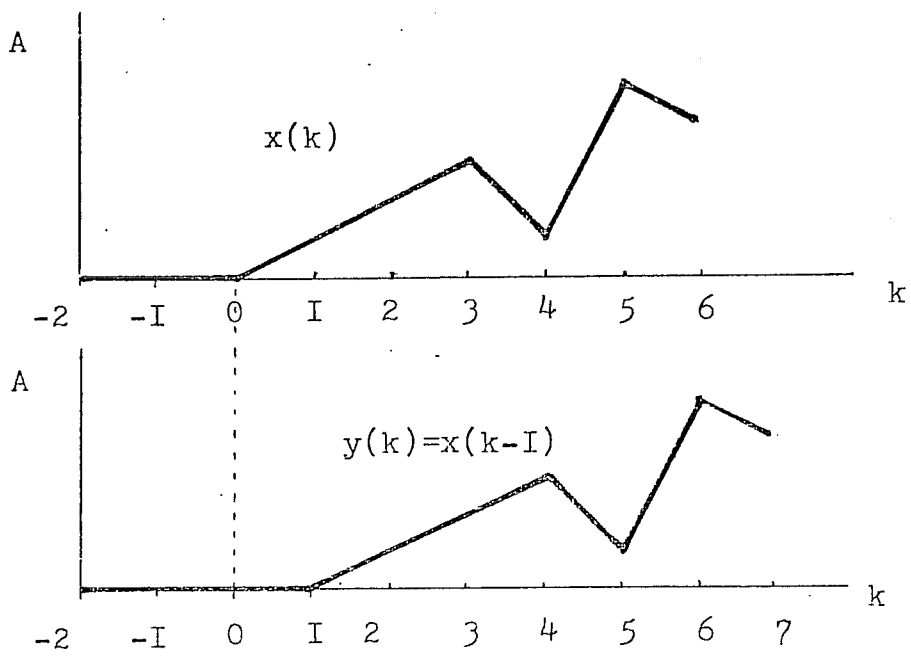


Fig. I9 - VII

le signal y(k) est identique à x(k) à l'exception du fait qu'il est décalé d'un échantillon sur l'axe des temps (fig.19-VII). En utilisant la définition de la transformée en z, on peut écrire :

$$y(z) = y(0)z^{0} + y(1)z^{-1} + y(z)z^{-2} + y(3)z^{-3} + \ldots\ldots$$
$$= x(-1)z^{0} + x(0)z^{-1} + x(1)z^{-2} + x(z)z^{-3} + \ldots\ldots$$

Or $x(k) = 0$ pour $k < 0$

$$y(z) = x(0)z^{-1} + x(1)z^{-2} + x(z)z^{-3} + \ldots\ldots$$

Ceci peut s'écrire, en remplaçant $z^{-a} . z^{-b}$ par $z^{-(a+b)}$

$$= z^{-1}(x(0)z^{0} + x(1)z^{-1} + X(z)z^{-2} + \ldots\ldots)$$

On a donc démontré que le signal $y(k) = x(k - 1)$ a pour transformée en z

$$z^{-1}(X(z))$$

Notons que la démonstration dépend du fait que $x(k) = 0$ pour $k < 0$.

On généralise aisément : la transformée en z de $x(k - m)$ est égale à

$$z^{-m} X(z)$$

ANNEXE E

Exemple du calcul d'une fonction de transfert (spectre) à l'aide de la transformée en z.

FANT (1960, p. 44-45) a montré que l'effet de la radiation aux lèvres est de rehausser le spectre d'environ 6 dB/octave. Cette correction est analogue au redressement spectral ("high-shaping") effectué par le spectrographe. On peut considérer ce redressement comme un filtrage passe-haut avec une pente de 6 dB par octave. Nous allons montrer qu'au plan temporel, cet effet est produit par une simple différenciation. Pour des intervalles de temps très courts, une différenciation est égale à la différence entre la valeur d'un

échantillon et la valeur précédente. Cette procédure est adoptée en synthèse et en analyse par prédiction linéaire. On écrit formellement

$$X(i) = x(i) - x(i-1)$$

La transformée en z est

$$X(i) = X(i) - X(i-1) = X(i) - z^{-1} X(i)$$
$$= X(i)\ (1 - z^{-1})$$

par application du théorème fondamental .

Pour $z = e^{j\omega}$, on peut calculer la fonction de transfert. Pour l'évaluation numérique de $1 - e^{-j\omega} = 1-(\cos\omega - j \sin\omega)$ $= 1 - \cos\omega + j \sin\omega$, supposons une fréquence d'échantillonnage de 10 000 Hz. La fréquence limite supérieure est donc de 5000 Hz ($\omega = 2\pi F$)

Le module $|H(\omega)|$ est égal à la racine carrée de la somme du carré de la partie réelle $(1 - \cos\omega)^2$ et du carré de la partie imaginaire $(\sin\omega)^2$: $|H(\omega)|$

$$= \sqrt{(1 - \cos\omega)^2 + \sin\omega^2} = \sqrt{1 - 2\cos\omega + \cos^2\omega + \sin^2\omega}$$
$$= \sqrt{2 - 2\cos\omega}$$
$$= \sqrt{2\ (1 - \cos\omega)}$$

La table 1-VII présente la valeur du module pour quelques valeurs de ω , exprimée également en Hz pour la fréquence d'échantillonnage retenue. La dernière colonne indique la valeur du module en dB par rapport à un niveau arbitraire, calculée par $H(\omega)_{dB} = 20 \text{ Log}_{10} (H(\omega)/1)$.

La fig. 20-VII présente le module de la fonction de transfert (c'est-à-dire le gain du filtre équivalent). On vérifie qu'il s'agit bien d'un filtre passe-haut. La fig. 21-VII présente la fonction de transfert en dB en fonction de la fréquence (sur une échelle logarithmique). On note que la pente est approximativement de + 6 dB par octave, sauf pour les fréquences les plus élevées.

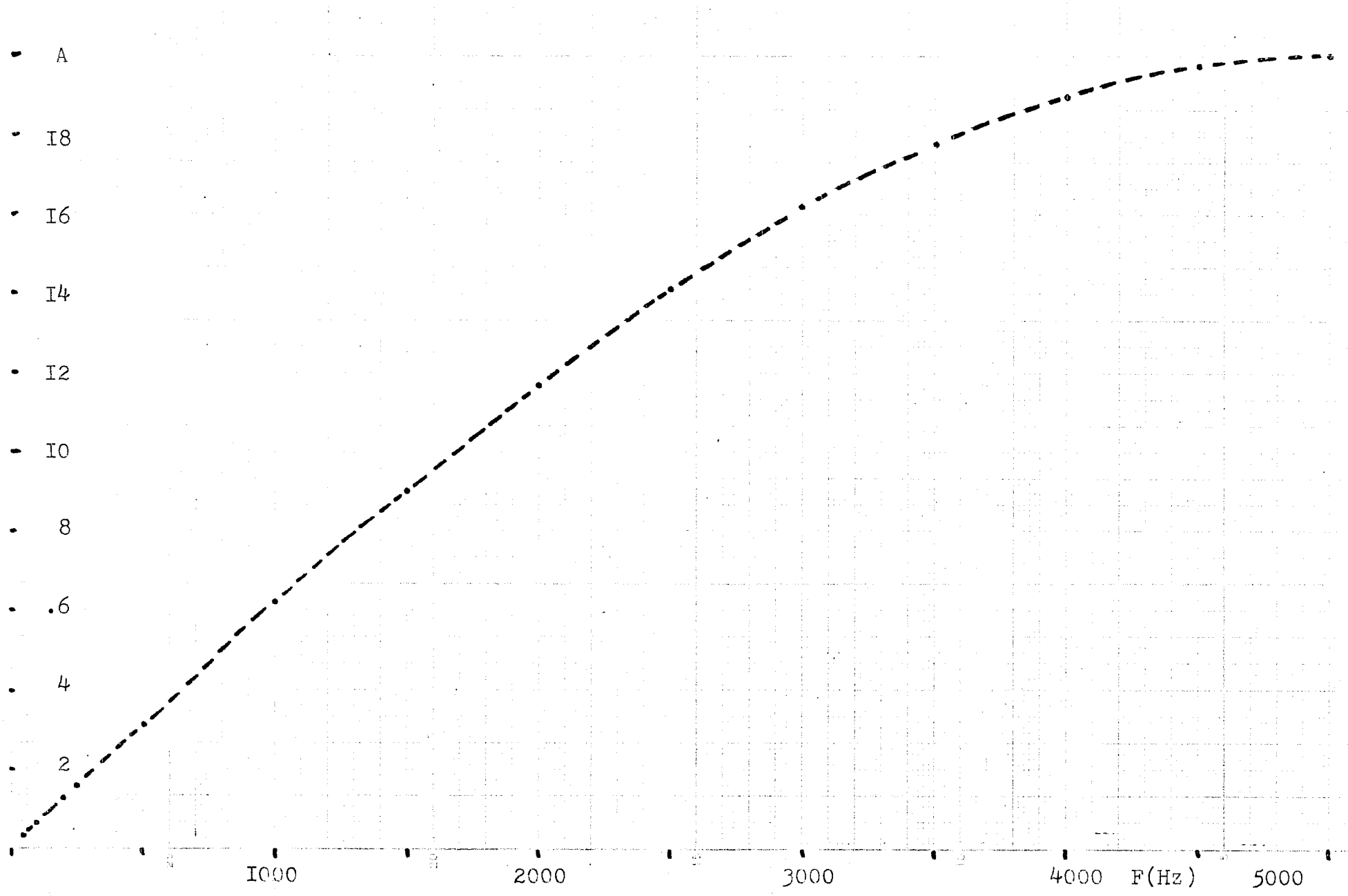


Fig. 20 - VII

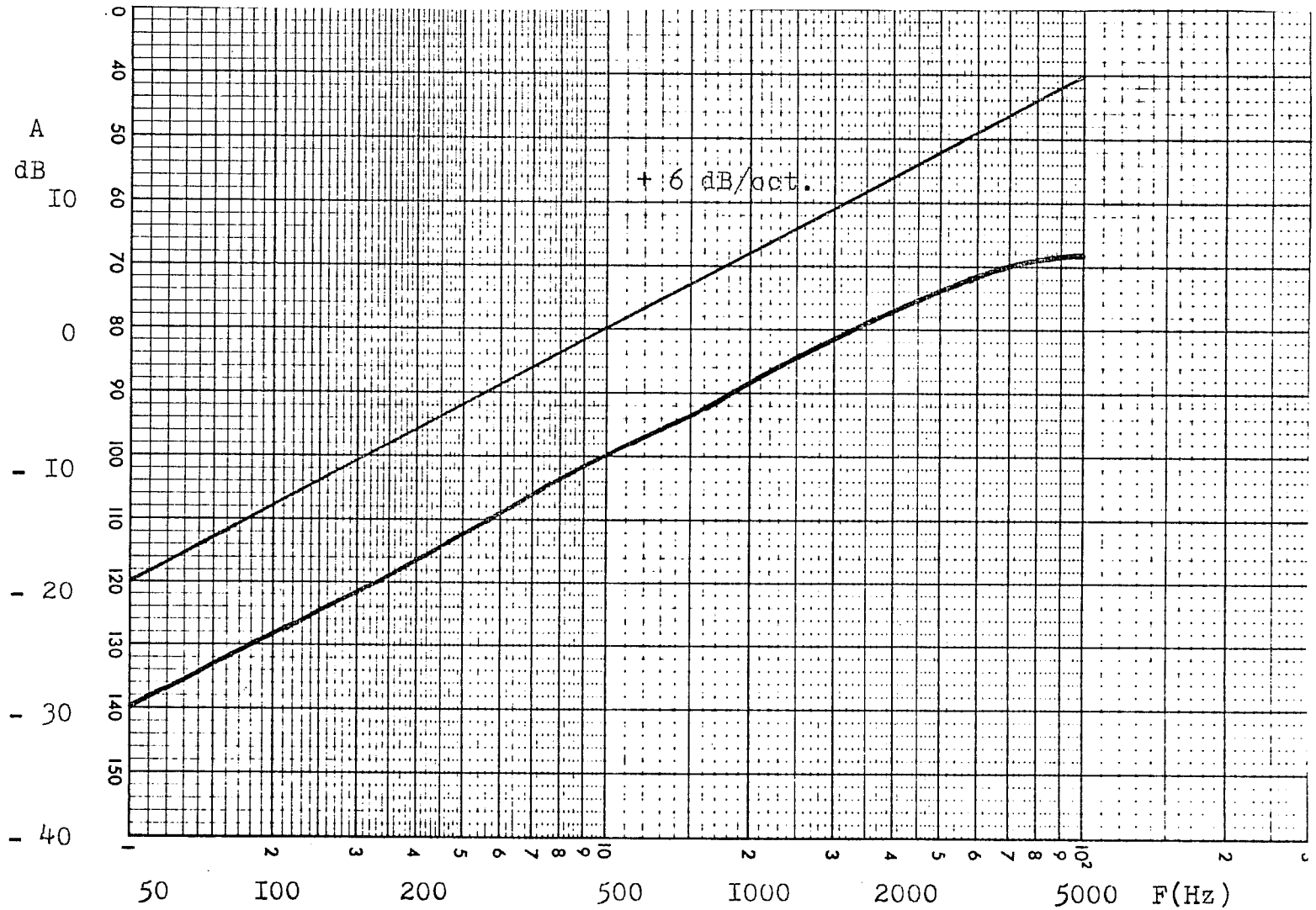


Fig. 2I - VII

F (Hz)	$\omega \times \pi$	$\lvert H(\omega) \rvert$	$\lvert H(\omega) \rvert_{dB}$
50	0.01	0.031	-30.1
100	0.02	0.063	-24.0
200	0.04	0.126	-18.0
250	0.05	0.157	-16.1
500	0.1	0.313	-10.1
1000	0.2	0.618	- 4.2
1500	0.3	0.908	- 0.84
2000	0.4	1.176	1.40
2500	0.5	1.414	3.01
3000	0.6	1.618	4.18
3500	0.7	1.782	5.02
4000	0.8	1.902	5.58
4500	0.9	1.975	5.91
5000	1.0	2.000	6.02

Table I-VII

A N N E X E F

Programme de calcul de la forme du conduit vocal par prédiction linéaire

Ce programme a pour objet le calcul d'un tube décrit par M sections à partir des échantillons temporels de l'onde de pression. Il est fondé sur les travaux de WAKITA (I973). La fonction de transfert du tube calculé est égale à celle du filtre qui modélise la production de la voyelle. Le nombre de sections (M) doit être choisi de façon adéquate par l'utilisateur. Il doit être égal à deux fois le nombre de pôles conjugués complexes (formants) présents dans le signal. La longueur du conduit est déterminée par la relation

$$l = cMT/2 = cM/2Fs \ .$$

l = longueur du conduit; M: nombre de sections; c: vitesse du son
Fs = I/T = fréquence d'échantillonnage du signal.

ENTREE

I - N: nombre d'échantillons ($\geqslant$ I28)
M : nombre de sections décrivant le tube acoustique

2 - X(I): lecture de l'onde de pression échantillonnée

PROGRAMME

3 - Dérivation du signal après création de N x 2 points. Seuls les 256 premiers points seront utilisés. La dérivation assure une égalisation approximative du spectre.

4 - Fenêtre de HAMMING.

5 - Algorithme de DURBIN.

SORTIE

6 - Ecriture des M coefficients Ai

7 - Ecriture des M coefficients de corrélation partielle (PARCOR).

8 - Ecriture des M aires: la première est l'aire aux lèvres.

```
      DIMENSION X(400),R(20),A(20),E(20),AA(20),C(20)
      REAL K(20)
      E2PI=2*3.14159
      READ(105,10)N,M                    I
   10 FORMAT(I3,1X,I2)
      READ(105,20)(X(I),I=1,N)           2
   20 FORMAT(9(F6.4,1X))
      DO90T=1,2
      DO90J=1,N
      X(100*T+J)=X(J)
   90 CONTINUE
      N=256
      N=N-1
      NN=N-1
      DO1I=1,N
      X(I)=X(I+1)-X(I)                   3
    1 CONTINUE
      DO2I=1,N
      X(I)=X(I)*(0.54-(0.46*COS(E2PI*(I-1)/NN)))   4
    2 CONTINUE
      DO3KK=1,M+1
      R(KK)=0.
      DO4I=1,N-(KK-1)
      R(KK)=R(KK)+X(I)*X(I+(KK-1))
    4 CONTINUE
    3 CONTINUE
      RT=R(1)
      DO7KK=1,M+1
      R(KK)=R(KK)/RT
    7 CONTINUE
      RZ=R(1)
      DO5KK=2,M+1
      R(KK-1)=R(KK)
    5 CONTINUE
      EZ=RZ
      K(1)=-R(1)/EZ                      5
      A(1)=K(1)
      E(1)=(1-K(1)**2)*EZ
      DO100I=2,M
      S=0
      DO200J=1,I-1
      S=S+A(J)*R(I-J)
  200 CONTINUE
      K(I)=-(R(I)+S)/E(I-1)
      A(I)=K(I)
      DO300J=1,I-1
      C(J)=A(J)+K(I)*A(I-J)
  300 CONTINUE
      DO400J=1,I-1
      A(J)=C(J)
  400 CONTINUE
      IF(I.EQ.M)GOTO500
      E(I)=(1-K(I)**2)*E(I-1)
  100 CONTINUE
  500 WRITE(108,60)(A(I),I=1,M)          6
   60 FORMAT(5(2X,F9.4))
      WRITE(108,70)(K(I),I=1,M)          7
   70 FORMAT(5(2X,F9.4))
      AA(MP+1)=1.
      DO406I=MP,1,-1
      AA(I)=(AA(I+1)+AA(I+1)*K(I))/(1-K(I))
  406 CONTINUE
      WRITE(108,408)(AA(I),I=1,MP)       8
  408 FORMAT(/,5X,' AA(I) ',/,10(1X,F8.4,2X))
      STOP
```

C H A P I T R E VIII

LES CONTRAINTES ARTICULATOIRES

forme du conduit la plus appropriée. Ils ont choisi, à priori, le critère suivant : le sujet utilise la forme de conduit la plus proche du tube acoustique de section uniforme. Cette contrainte peut être considérée comme un cas particulier du principe du moindre effort. Si cette contrainte est exacte, la longueur du conduit qui vérifie le critère de moindre écart sera identique à la longueur mesurée sur un cliché radiologique.

La discussion de la formalisation de ce critère étant très succincte dans la publication originale, nous avons jugé utile de présenter une dérivation complète.

B/ Formalisation du critère de moindre déviation par rappor au tube uniforme

Nous suivons PAIGE et ZUE (1970a) en considérant que la contrainte sur la longueur (1) consiste à minimiser un critère ε^2 défini par les équations (1) et (2).

$$(1) \quad |\varepsilon^2| = \frac{1}{L} \int_0^L \left[\text{Log } A(x)^2 \right] dx$$

$$(2) \quad \int_0^L \text{Log } A(x)\, dx = 0$$

Log A(x) : Log de l'aire au point x
l : longueur du conduit.

La forme donnée par PAIGE et ZUE est inutilement complexe. De plus, ces équations ne peuvent être appliquées directement aux données. L'équation (1) exprime formellement le fait que le conduit calculé doit être le plus proche possible du conduit uniforme. Si l'on utilise la méthode de WAKITA pour la prédiction de la forme, le conduit est considéré comme une succession de cylindres : (1) et (2) peuvent s'écrire :

$$(1) \quad |\varepsilon^2| = \frac{\Delta x}{l} \sum_{i=1}^{m} \left[\text{Log } Ai \right]^2$$

$$(2) \quad \sum_{i=1}^{m} \text{Log } Ai . \Delta x = 0$$

x : longueur du cylindre ; m : nombre de cylindres.
Ai : aire de la ième section.

On peut dériver (1) de la manière suivante : Soit C l'aire d'un tube uniforme. Le carré de la différence α entre le logarithme des aires (normalisées par unité de longueur) et le logarithme de l'aire du tube uniforme peut s'écrire :

$$\alpha^2 = \frac{1}{L} \sum_{i=1}^{m} \left[\text{Log } Ai - \text{Log } C \right]^2 . \Delta x$$

$$\alpha^2 = \frac{1}{L} \sum_{i=1}^{m} \left[\text{Log } Ai^{\,2} \Delta x - 2 \text{ Log } C \sum_{i=1}^{m} \text{Log } Ai \, \Delta x + (\text{Log } C)^2 \Delta \right.$$

comme $\sum_{i=1}^{m} \text{Log } Ai \, \Delta x = 0$, d'après (2) $\alpha^2 = \frac{1}{L} \sum_{i=1}^{m} (\text{Log } Ai)^2 \Delta x + (\text{Log } C)^2$

$(\text{Log } C)^2$ est une constante, d'où

$$\varepsilon^2 = \alpha^2 - (\text{Log } C)^2 = \frac{1}{L} \sum_{i=1}^{m} (\text{Log } Ai)^2 \Delta x = \frac{\Delta x}{1} \sum_{i=1}^{m} (\text{Log } Ai)^2$$

ce qui est bien l'expression (1). Le choix du logarithme des aires des sections conduit à accorder une plus grande importance aux constrictions qu'aux cavités.

Le nombre des sections servant à décrire une même fonction d'aire ne fait pas varier ε^2. Montrons-le sur un exemple.

Soit une fonction d'aire décrite, par 4 sections (fig. 1-VIII)

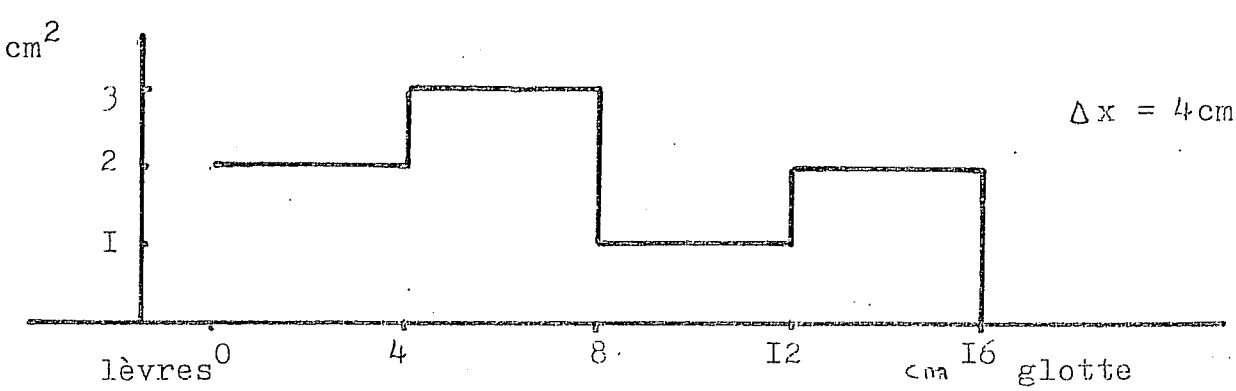


fig. I - VIII

On a

(3) $$\varepsilon^2 = \frac{1}{16} \cdot 4 \cdot \left[\mathrm{Log}(2)^2 + \mathrm{Log}\ (3)^2 + \mathrm{Log}(1)^2 + \mathrm{Log}(2)^2 \right] = 0.54196$$

Si cette même fonction d'aire est décrite par 8 sections (Δx = 2 cm), la première section sera décrite par A_{I} et A_{II} = 2 cm^2. Dans (3).

$(\mathrm{Log}\ A_1)^2 = (\mathrm{Log}\ 2)^2$ sera remplacé par $(\mathrm{Log}\ A_I)^2 + (\mathrm{Log}\ A_{II})^2 =$
$(\mathrm{Log}\ 2)^2 + (\mathrm{Log}\ 2)^2 = 2\ (\mathrm{Log}\ 2)^2 = 2\ (\mathrm{Log}\ A_1)^2$

La partie entre parenthèses dans (3) vaut maintenant le double de sa valeur antérieure. Mais comme Δx est divisé par 2, ε^2 possède bien la même valeur.

On sait d'autre part, que les fréquences formantiques ne dépendent que du <u>rapport</u> des aires (A_i/A_{i+1}). Pour cette raison, les conduits vocaux calculés ne sont connus qu'à une constante multiplicative près.

Calculons ε^2 à l'aide de (1) seulement pour deux tubes uniformes de 2 et 4 cm^2 :

$$\varepsilon^2(2\ \mathrm{cm}^2) = 0.4805$$

$$\varepsilon^2(4\ \mathrm{cm}^2) = 1.9218$$

Ces 2 tubes étant uniformes, il serait commode d'obtenir la même valeur pour ε^2. le choix de $\varepsilon^2 = 0$ est logique, car les tubes sont uniformes, et ε^2 représente l'écart par rapport au tube uniforme.

$$\text{posons } \frac{1}{1} \sum_{i=1}^{m} (\mathrm{Log}\ Ai)^2\ \Delta x = 0.$$

Les sections étant identiques, on a $(\mathrm{Log}\ A_i)^2 = 0$, d'où $\mathrm{Log}\ A_i = 0$ et $A_i = 1$. Ceci explique la raison d'être de l'équation (2).

Montrons que (2) implique une normalisation du type $A'(x) = A(x)/k$

Supposons $c = \int_0^L \mathrm{Log}\ A'(x)\ dx = 0$

et calculons $\int_0^L \mathrm{Log}\ k\ A'(x)\ dx$, c'est -à-dire la valeur de l'intégrale pour une fonction d'aire dérivée d'une première, qui satisfait (2), par multiplication des aires par une constante k.

Comme $\text{Log } k.A'(x) = \text{Log } k + \text{Log } A'(x)$, on a :

$$\int_0^L \text{Log } k\ A'(x)\ dx = \int_0^L \text{Log } k\ dx + \int_0^L \text{Log } A'(x)\ dx$$

Nous avons supposé que la 2ème intégrale est nulle, d'où :

$$\int_0^L \text{Log } k\ A'(x)\,dx = \int_0^L \text{Log } k\ dx = l \text{ Log } k = c$$

car k est une constante indépendante de dx.

$$\text{Log } k = \frac{c}{l} \qquad \text{d'où } \underline{k = e^{c/l}}$$

Vérifions que la division de $Ax = k\ A'(x)$ par k donne bien

$$\int_0^L \frac{\text{Log } k\ A'(x)}{e^{c/l}}\ dx = 0$$

$$\int_0^L \text{Log } A'(x)\,dx + \int_0^L \text{Log } k\ dx - \int_0^L \text{Log } e^{-\frac{c}{l}}\ dx =$$

$$0 + l \text{ Log } k - \int_0^L c/l\ dx = l \text{ Log } k - c$$

Comme c est égal à l Log k, nous avons montré que (2) correspond en fait à une normalisation des Ai par un coefficient $k = e^{c/l}$. Pour que (2) soit vérifié, il suffit de diviser toutes les aires par k.

Pour normaliser une fonction d'aire, il suffit de calculer la constante c :

$c = \int_0^L \text{Log } A\ x)\,dx = \sum_{i=1}^{m} \text{Log } Ai\ \Delta x = \Delta x \sum_{i=1}^{m} \text{Log } Ai$, puis de diviser chaque Ai par $e^{c/l}$.

Vérifions sur l'aire décrite à la fig. 1-VIII.

$c = \Delta x \sum_{i=1}^{6} \text{Log } Ai = 9.9\ 396 \qquad e^{c/l} = 1.8616 \quad (l = 16)$

$A'i = Ai/1.8612$

$A'1 = 1.07$

$A'2 = 1.61$

$A'3 = 0.54$

$A'4 = 1.07$

On vérifie que :

$$\sum_{i=1}^{6} \text{Log } Ai = 1.5 \times 10^{-5} \cong 0$$

PAIGE et ZUE (1970a) ont testé cette hypothèse sur les 6 fonctions d'aire des voyelles russes données par FANT (1960). Le calcul des aires a été fait par une méthode différente de celle présentée au chapitre VII (PAIGE et ZUE 1970b). Leurs résultats sont présentés dans le tableau ci-dessous :

Voyelles	Lg. du conduit FANT 1960	Lg. calculée	% erreur
i	16.5	16.4	- 0.6
e	16.5	15.8	- 4.2
a	17.0	17.3	1.8
o	18.5	17.6	- 4.9
u	19.5	19.0	- 2.6
ɨ	19.0	18.4	- 3.2

Le pourcentage d'erreurs n'est pas supérieur à la précision des mesures, que FANT estime à ± 5 % (c'est-à-dire ± 1 cm). La méthode de calcul des aires utilisée n'étant pas la plus précise, nous avons décidé de faire une étude préliminaire sur les mêmes données par la méthode de WAKITA.

C/ Test de la contrainte sur les données de FANT (1960)

Afin de procéder à une première évaluation de cette contrainte, nous avons utilisé les conduits mesurés par FANT (1960). Les fonctions d'aire ont été calculées par la méthode de WAKITA (1973). Notre expérience a été réalisée avant que WAKITA (1977) ne publie des résultats portant sur le même ensemble de données. La comparaison des deux études révèle une différence significative que nous détaillerons plus loin.

Nous possédons donc des estimations de la longueur du conduit pour 6 voyelles russes [i, e, a, o, u, ɨ]. Les fonctions d'aire permettent le calcul des 4 premières fréquences formantiques. Si l'on utilise la méthode décrite au chapitre VI sans aucune correction, on obtient des fréquences pour un tube rigide uniforme, ce qui est conforme aux conditions d'utilisation de la méthode de calcul des aires par la prédiction linéaire.

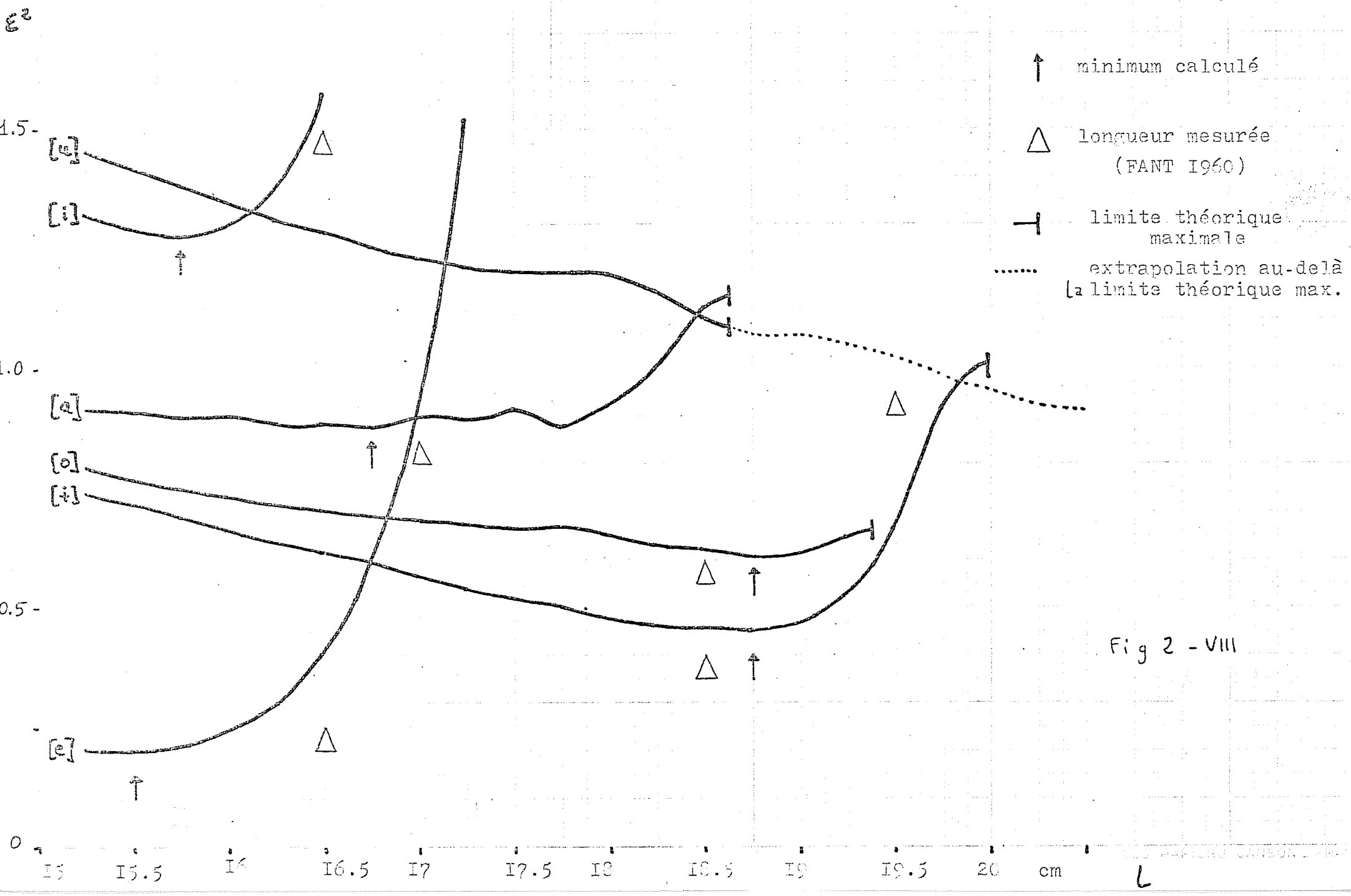

ε^2
1.5
1.0
0.5
0
[u]
[i]
[a]
[o]
[ɨ]
[e]
minimum calculé
longueur mesurée (FANT 1960)
limite théorique maximale
extrapolation au-delà la limite théorique max.
15
15.5
16
16.5
17
17.5
18
18.5
19
19.5
20
cm
L

Fig 2 - VIII

A partir de ces fréquences formantiques, qui sont présentées à la table ci-dessous, nous avons synthétisé numériquement une période de chaque voyelle.

	F1	F2	F3	F4	FANT (1974)			
i	229	2331	3330	4100	228	2230	2817	3553
e	436	2056	3040	4000				
a	697	1188	2590	3800				
o	544	948	2450	3650				
u	257	621	2410	3800				
ɨ	300	1822	2480	3550	285	1431	2302	3213

Les valeurs de FANT (1974) ont été calculées par une équation de WEBSTER, mais elles incluent l'influence de la radiation. Elles ne sont donc pas directement comparables avec les nôtres. On notera pourtant une bonne correspondance.

A partir des échantillons temporels des voyelles synthétisées, nous avons calculé, pour des longueurs variant de 15 à 20 cm, les fonctions d'aire correspondantes, ainsi que le critère de moindre écart présenté au paragraphe précédent. Rappelons que si notre contrainte est vérifiée, le critère doit être minimum pour la longueur du conduit réel.

La fig. 2-VIII présente les variations du critère en fonction de la longueur. Pour 5 voyelles sur 6, le critère présente un minimum que nous avons indiqué par une flèche verticale. La longueur mesurée est indiquée par un triangle. La table I-VIII compare les longueurs sélectionnées par le critère de moindre déviation par rapport au conduit uniforme et les longueurs réelles. Le problème posé par la voyelle [u] sera examiné plus loin. Cette table présente également les résultats de PAIGE et ZUE (1970a).

TABLE I-VIII

Voyelle	Lg. mesurée FANT (1960)	Lg. calculée	1	% erreur	PAIGE et ZUE (1970)
i	16.5	15.75	0.75	- 4.7	16.4
e	16.5	15.50	1.00	- 6.5	15.8
a	17	16.75	0.25	- 1.5	17.3
o	18.5	18.75	0.25	1.35	17.6
ɨ	18.5	18.75	0.25	1.35	18.4
u	19.5	(18.6)	(0.90)	(- 4.9)	19

Les résultats sont excellents pour 3 voyelles et bons pour deux autres. On se souvient que FANT n'estime pas la précision des mesures à plus de ± 5 %. WAKITA (1977, fig. 6) présente des résultats analogues, mais à une échelle telle qu'il est impossible d'apprécier l'erreur commise. Il indique cependant que l'erreur la plus forte est de 4.6 %. Il a employé une méthode très différente de la nôtre, qui n'exige pas d'hypothèses sur les largeurs de bande. C'est, comme nous l'avons indiqué au chapitre précédent, un défaut de notre méthode, qui passe explicitement par une synthèse numérique. Dans l'article cité, WAKITA montre également que la méthode de prédiction linéaire impose une longueur théorique maximale ; celle-ci est définie par la formule :

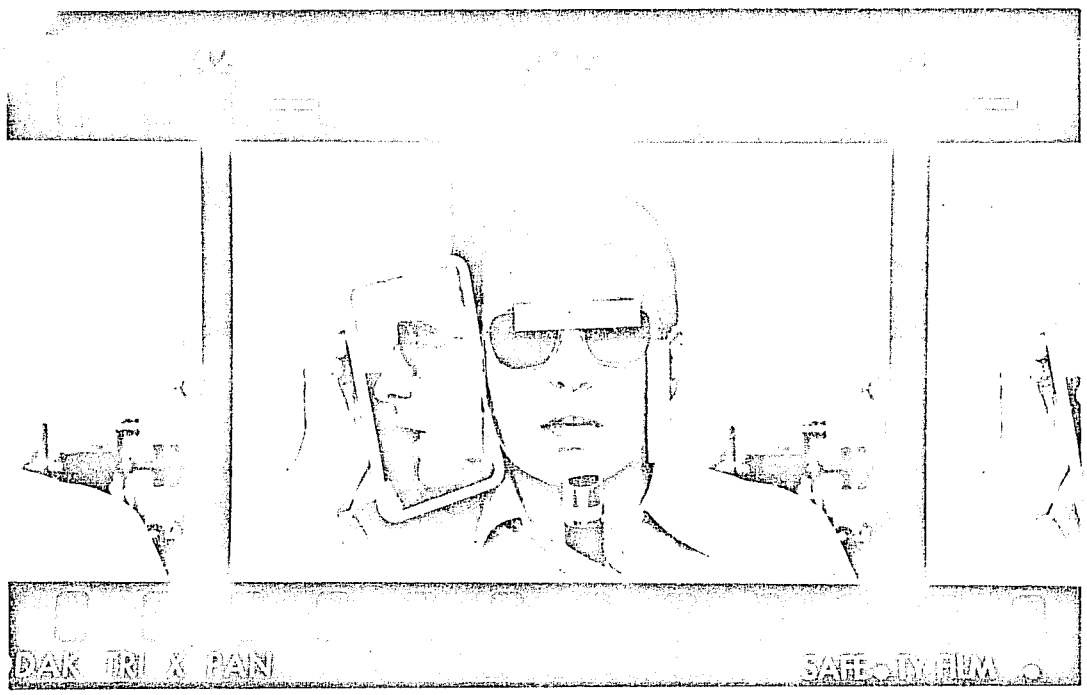
$$L\ max = Mc / 4\ Fh$$

Fh : fréquence du formant le plus élevé ; M : nombre de sections ; c : vitesse du son.

Pour les voyelles étudiées, la longueur théorique maximale atteint les valeurs suivantes :

i	17.22 cm
e	17.65
a	18.58
o	19.34
u	18.58
ɨ	19.89

Pour 4 voyelles, ces longueurs maximales sont portées à la figure 2-VIII. Nous avions constaté un comportement "aberrant" du critère au voisinage de cette longueur : celui-ci



voyelle [a] contexte ∂b – b

Fig. 3 - VIII

présente un maximum pour la longueur théorique maximale, puis décroît rapidement. Pour la voyelle [u], on constate que la longueur mesurée est supérieure à la longueur maximale. Si l'on adopte cette dernière comme meilleure estimation de la longueur calculée, on obtient la valeur portée entre parenthèses dans la table I-VIII. L'erreur n'est que de 4.9 %. La longueur mesurée impliquerait une fréquence de F4 égale à 3620 Hz ; ceci ne représente qu'un écart de 180 Hz avec la valeur calculée (3800 Hz).

La contrainte articulatoire semble donc vérifiée pour les données de FANT (1960). Nous avons ensuite procédé à une étude portant sur des voyelles françaises.

D/ Test de la contrainte sur 30 voyelles françaises*

La méthode utilisée dans ce test portant sur un corpus de 30 voyelles françaises ([i, ɛ, a, ɔ, u] prononcées par deux locuteurs dans trois contextes consonantiques [əb-b, əd-d, əg-g]), est la même que celle que nous avons décrite au paragraphe précédent. Nous ne possédions cependant pas les fonctions d'aire, mais seulement les fréquences formantiques des voyelles enregistrées lors de la séance de cinéradiographie. Nous envisagerons successivement les problèmes posés par l'estimation des longueurs sur les clichés radiographiques et par la correction des valeurs formantiques. Puis nous discuterons des variations du critère ε^2 en fonction de la longueur.

a) Mesure de la longueur du conduit vocal

L'extrémité "interne" du conduit est constituée par les cordes vocales. Celles-ci, ou plus précisément les aryténoïdes, sont visibles sur les clichés radiologiques. L'extrémité "externe" du conduit est souvent décrite comme étant le plan tangent aux lèvres. Mais comme le remarque FANT (1960),

---

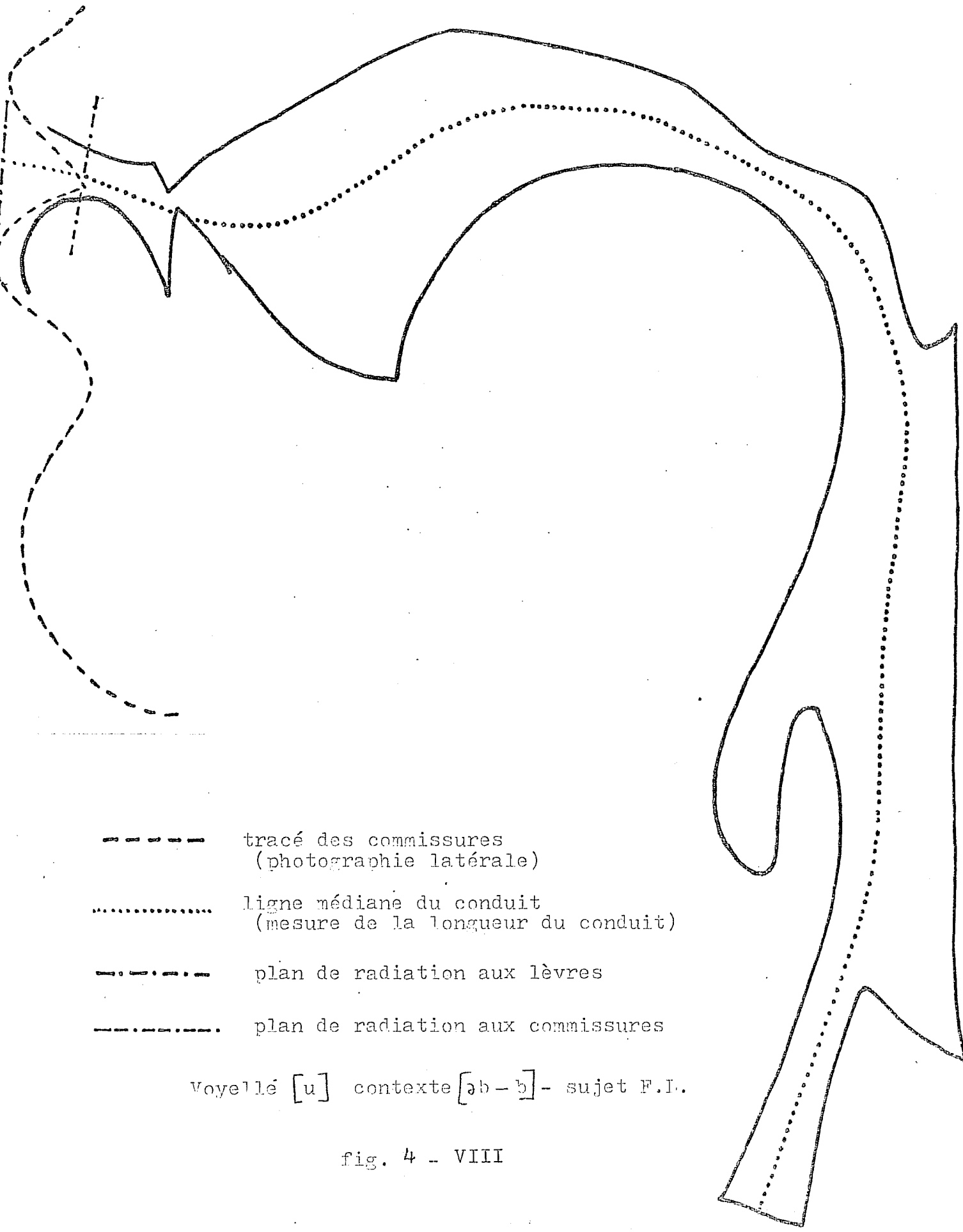
* Cette étude a été réalisée en collaboration avec J.P. ZERLING. Il a bien voulu nous communiquer les mesures de longueur du conduit qu'il a effectuées pour sa thèse. Les recherches photographiques ont été menées en commun. Qu'il soit remercié pour le sacrifice du coin droit de sa moustache lors de la mesure de la position des commissures des lèvres.

ce critère n'est vraisemblablement pas le meilleur : "Le point exact où se place le plan de radiation (c'est-à-dire ce que nous avons appelé l'extrémité "externe") n'est pas facilement définissable, car le front d'onde du son aux lèvres ne se limite pas à une surface plane mais peut s'étendre plus ou moins latéralement. La fermeture labiale a pour effet de masquer les ouvertures latérales en direction des joues, et provoquera un certain allongement, même si les lèvres ne sont pas projetées en avant. Pour les sons non-labialisés, le plan de radiation a été estimé ne jamais se trouver à plus de 0.5 cm en avant des dents (p. 98)". LINDBLOM et SUNDBERG (1969, 1971b) proposent d'utiliser la commissure des lèvres comme position du plan de radiation. Ce critère devrait, selon eux, être plus représentatif de la labialisation qui se marque plus, pour les voyelles non-fermées, par un avancement des commissures que par la projection labiale. Pour obtenir quelques données préliminaires, une étude photographique a été entreprise sur les deux sujets ayant fourni les tracés radiologiques. La fig. 3-VIII présente à titre d'exemple, le cliché photographique de la voyelle [ɑ] dans le contexte [ə b-b] . Ce cliché permet de calculer l'aire (frontale) aux lèvres et surtout la position de la commissure par rapport aux incisives supérieures et inférieures. Les 3 bandes de carton de 10 cm permettent une estimation des dimensions réelles. Afin de préserver une élocution naturelle tout en assurant la prise du cliché dans la phase centrale de la voyelle, la procédure suivante a été suivie : le sujet répétait 3 fois la voyelle dans son contexte et allongeait la voyelle lors d'une 4ème répétition pour permettre la prise du cliché. Le bruit de déclenchement de l'appareil est visible sur les spectrogrammes de contrôle, et permet de vérifier que le cliché a été pris pendant la phase centrale.

Après agrandissement, le tracé latéral des lèvres a été reproduit sur le tracé radiologique tiré d'une vue radiographique* sélectionnée pendant la partie stable de la voyelle.

---

* Vitesse 50 images/seconde. Les détails techniques et le détail de la procédure seront présentés dans ZERLING (1979 : à paraître). Un bon centrage de l'écran de brillance a permis d'obtenir une vue complète de la cavité orale, jusqu'au larynx.



Voyelle [u] contexte [əb – b] - sujet F.I.

fig. 4 - VIII

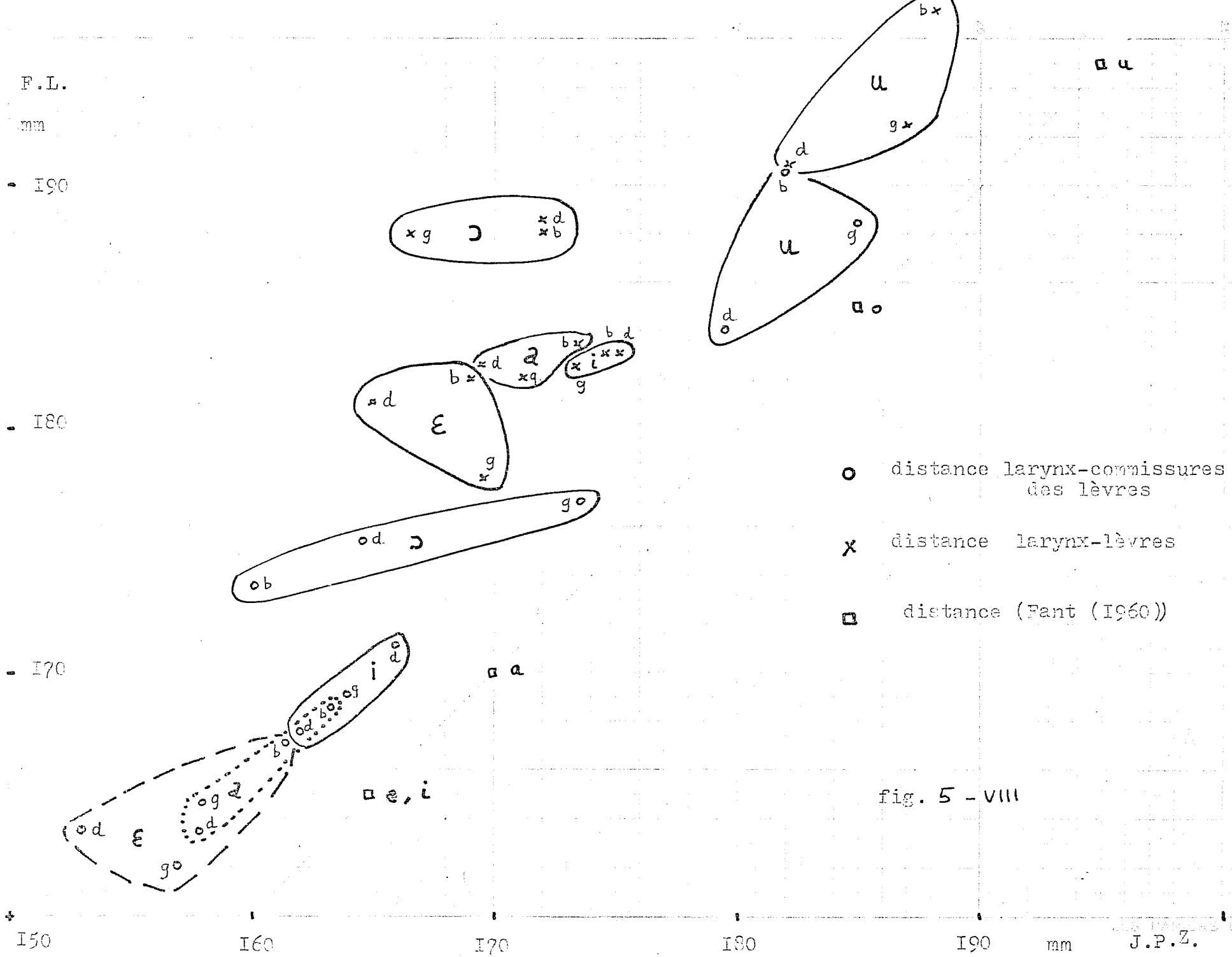

F.L.
mm
I90
I80
I70
I50
I60
I70
I80
I90
mm
J.P.Z.
distance larynx-commissures des lèvres
distance larynx-lèvres
distance (Fant (I960))

fig. 5 - VIII

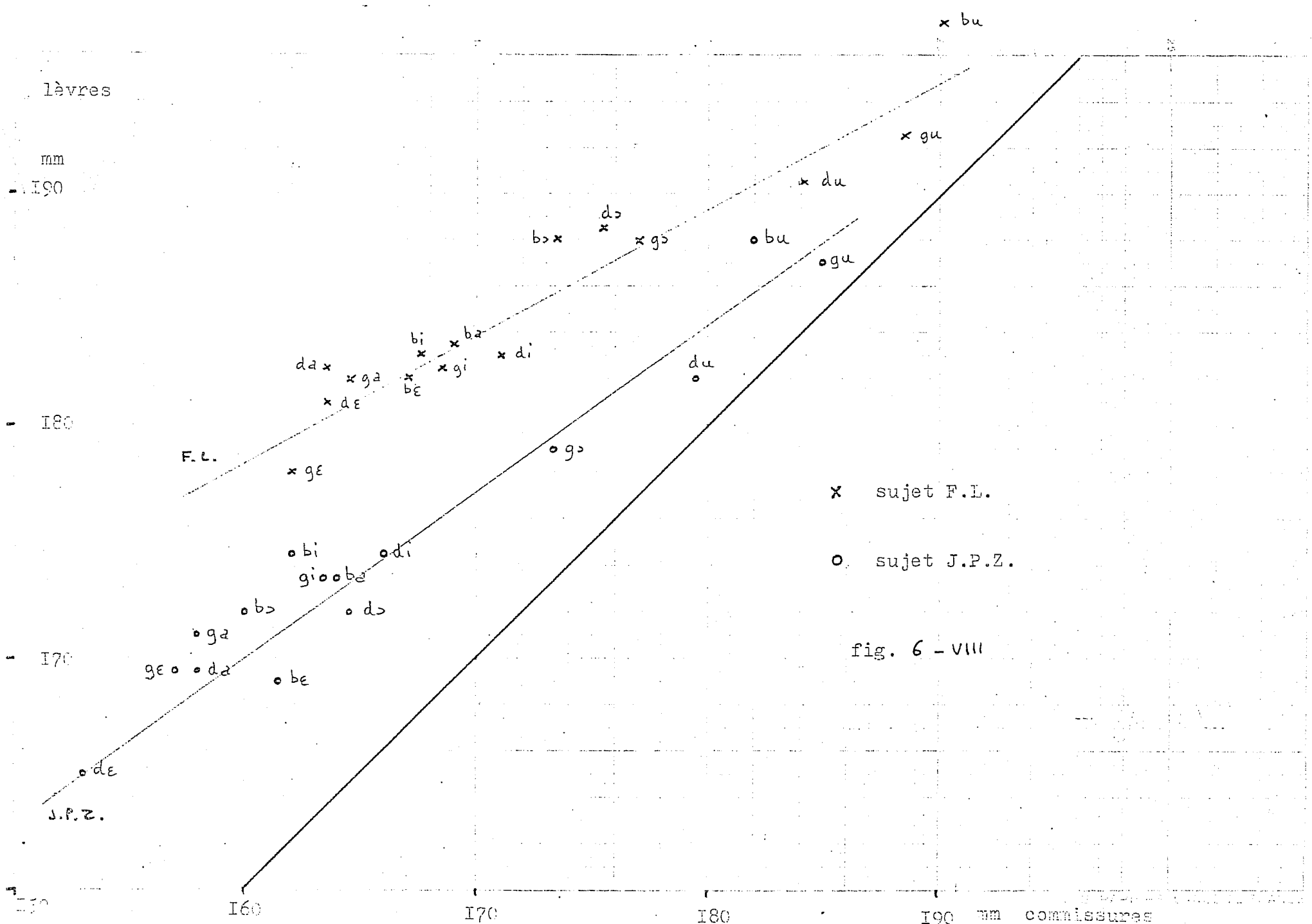

lèvres
mm
I90
I80
I70
I60
I70
I80
I90
mm commissures
F.L.
J.P.Z.
sujet F.L.
sujet J.P.Z.

fig. 6 - VIII

La concordance a été assurée en utilisant la position des incisives supérieures et inférieures ainsi que la forme du menton. L'angle maxillaire et le tracé des lèvres, sans coïncider parfaitement, se sont révélés assez semblables pour permettre la superposition. La fig. 4-VIII illustre les 2 mesures de la longueur du conduit vocal qui ont été effectuées à l'aide d'un curvimètre le long de la ligne médiane : distance larynx - commissure et distance larynx - plan tangent aux lèvres. L'estimation de la ligne médiane n'est pas aisée dans le cas d'une variation brusque de l'aire. C'est le cas dans la zone buccale antérieure sur la fig. 4-VIII.

La fig. 5-VIII présente les résultats des mesures sur les 30 voyelles prononcées par les deux sujets. La faible dispersion des points est relativement encourageante. La longueur du conduit est systématiquement plus longue pour le sujet F.L. La fig. 6-VIII compare les résultats des deux séries de mesures. On notera que l'écart est plus important pour les voyelles non-labialisées. Les valeurs tabulées seront présentées lors de la discussion de l'expérience. Ces 30 estimations de la longueur du conduit constituent un corpus dont nous ne connaissons pas l'équivalent pour une autre langue.

b) Analyse spectrographique

Les mesures des fréquences formantiques ont été effectuées sur les voyelles enregistrées lors de la séance de cinéradiographie. Elles ont été faites indépendamment par 2 phonéticiens sur des sections et des tracés en bande large réalisés sur deux spectrographes (KAY ELECTRIC 6061 B et VOICEPRINT modèle 700). L'erreur sur la calibration, vérifiée par un fréquencemètre est nettement inférieure à 1 %. Tout écart entre les deux séries de mesure a été contrôlé sur un 3ème tracé spectrographique. Trois sections différentes ont été réalisées pour chaque voyelle.

Dans les cas les plus favorables, l'écart entre les deux séries de mesure est de l'ordre de 30 Hz ($F_o/4$). Le tableau II-VIII (a et b) présente la moyenne des deux séries d'estimation

Sujet JPZ

	FI mesuré	FI corrigé	F2 mesuré	F2 corrigé	F3 mesuré	F3 corrigé	F4 mesuré	F4 corrigé	
bi	345	30I	2I00	2II0	2900	3050	3350	3550	Hz
di	340	296	2I00	2II0	2850	3000	3500	3700	
gi	3I5	267	2I30	2I40	2780	2930	3350	3550	
bɛ	580	558	I750	I835	2580	2780	3300	3625	
dɛ	6I0	588	I800	I885	2500	2700	3350	3675	
gɛ	580	558	I780	I865	2520	2720	3300	3625	
ba	720	727	I225	I350	2400	2480	3200?	3425	
da	735	742	I285	I4I0	2400	2480	3225	3450	
ga	730	737	I350	I475	2250?	2330	3300	3525	
bɔ	6I0	588	I000	II50	2470	2590	3200	3280	
dɔ	6I0	588	II00	I250	2250	2370	3I50	3230	
gɔ	605	583	II00	I250	2300	2420	3I50	3230	
bu	350	3I4	850	850	2050?	2050	3200?	3230	
du	335	297	I000	I000	2300	2300	3200	3230	
gu	390	359	880	880	2I50	2I50	3200?	3230	

Table II - VIII a

Sujet FL

Voyelle	FI mesuré	FI corrigé	F2 mesuré	F2 corrigé	F3 mesuré	F3 corrigé	F4 mesuré	F4 corrigé
bi	295	244	2020	2030	2850	3000	3720	3920
di	255	I93	2075	2085	2775	2925	3560	3760
gi	245	I8I	2I00	2II0	2700	2850	3420?	3620
bɛ	555	533	I620	I705	24I0	26I0	3620	3945
dɛ	550	528	I630	I7I5	2430	2630	3475	3800
gɛ	480	457	I735	I820	2300	2500	3450	3775
ba	660	668	I250	I375	2500	2580	33I0?	3535
da	645	653	I335	I460	2400	2480	3450	3650
ga	670	678	I390	I5I5	2270	2350	3450	3675
bɔ	555	533	I070	I220	2500	2620	3290	3370
dɔ	525	502	I250	I400	2400	2520	3450	3500
gɔ	525	502	I090	I240	2300	2420	3300	3380
bu	3I0	268	8I0	8I0	2320	2320	3420?	3460
du	335	297	I040	I040	2275	2275	3350	3380
gu	340	303	950	950	2250	2250	3375	3405

Table II - VIII b

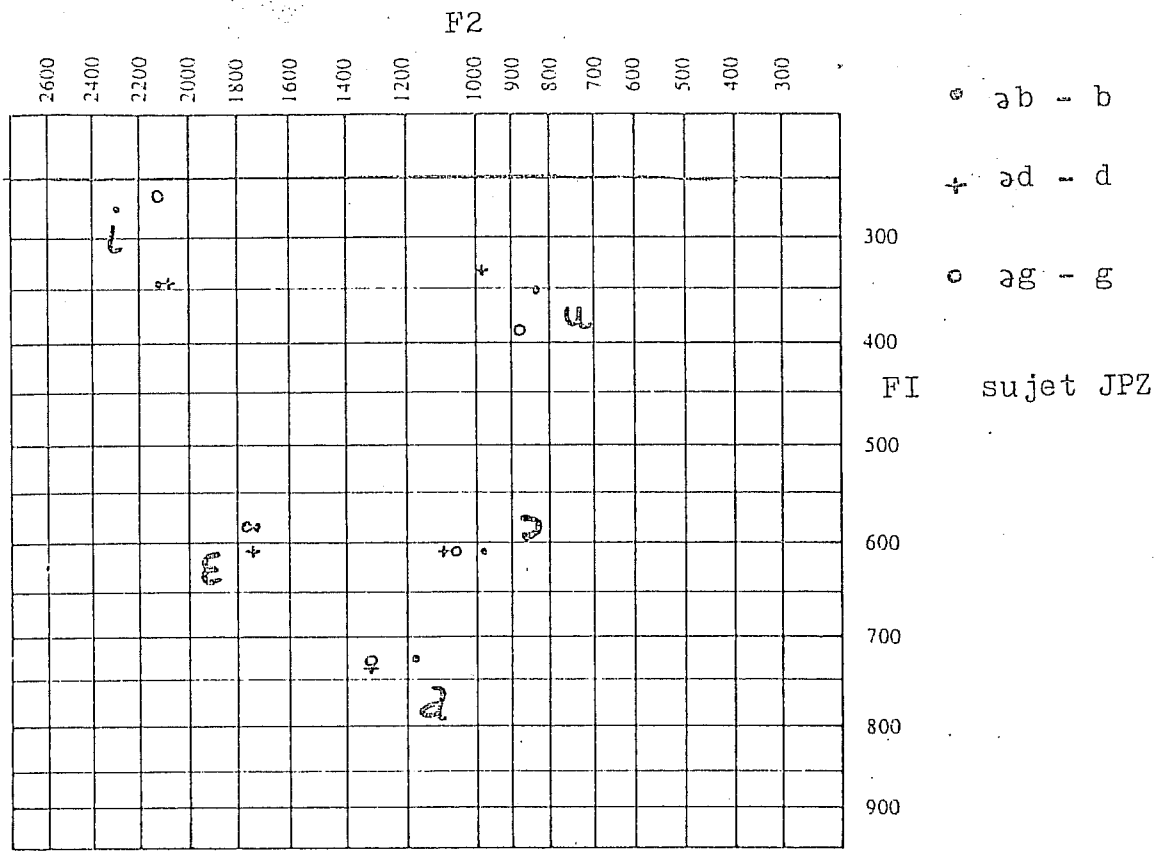


Fig. 7 - VIII

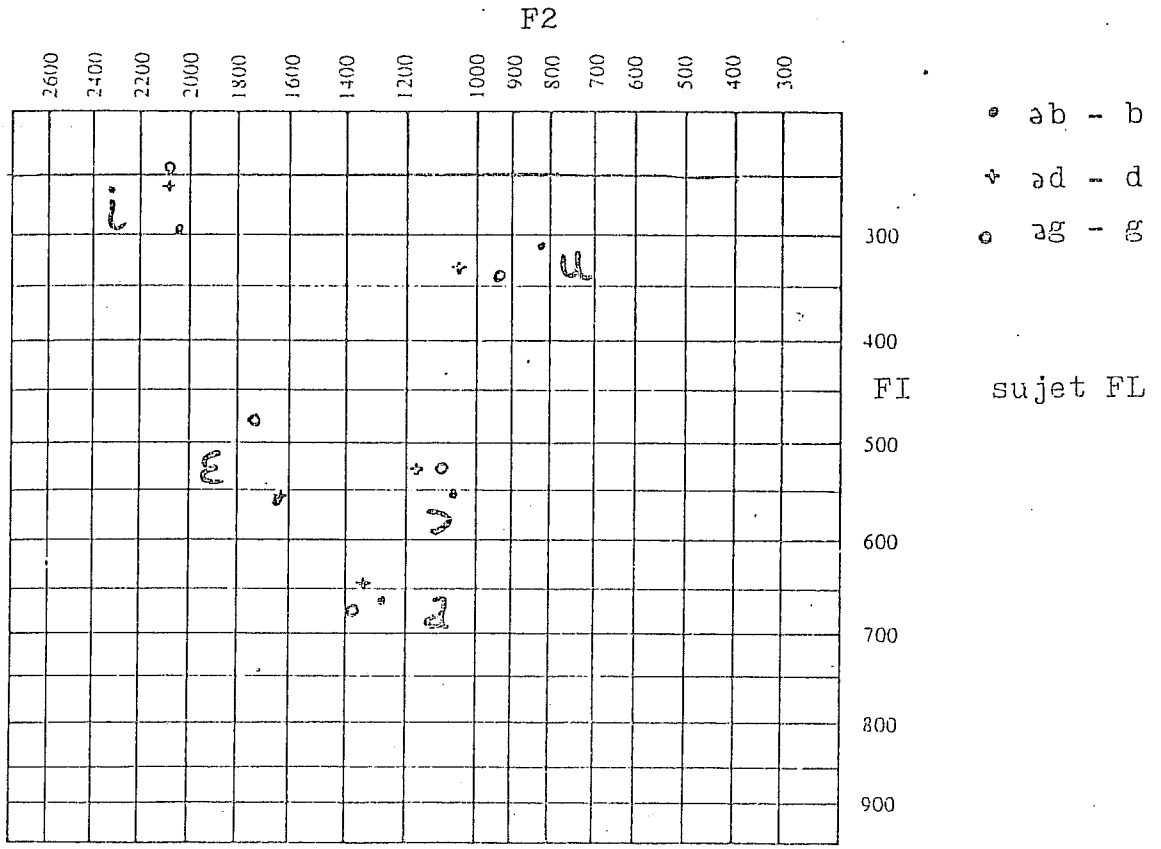


fig. 8 - VIII

Un point d'interrogation indique une valeur difficile à mesurer. Le fondamental était nettement plus élevé chez le sujet JPZ. Les fig. 7-VIII et 8-VIII donnent la position des voyelles dans le plan F1-F2. Les voyelles du sujet F.L. semblent plus sensibles à la coarticulation. Mais les voyelles [u] et [ɔ] pour les deux sujets, ont un F2 nettement plus élevé dans le contexte

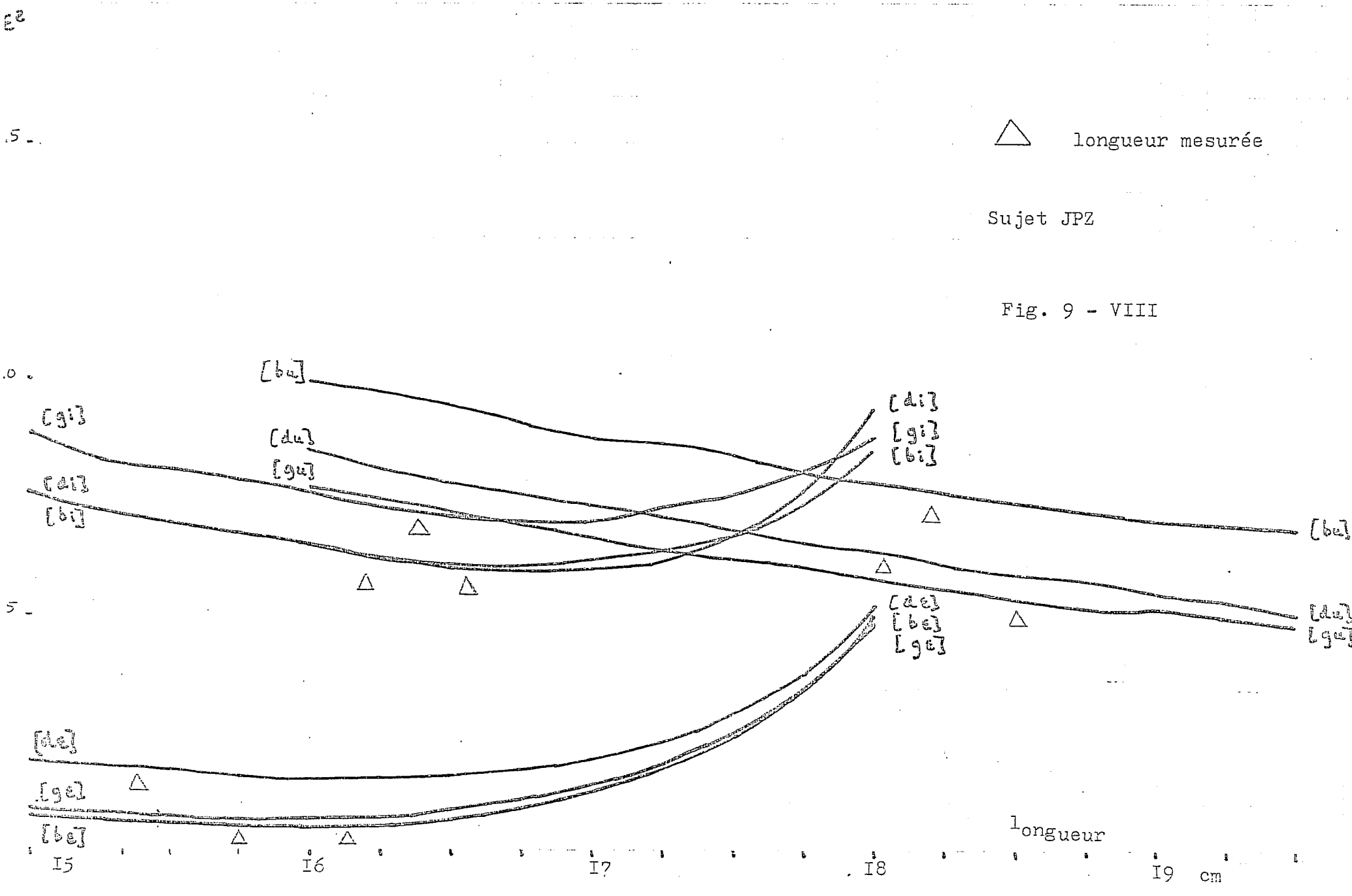
Ces mesures ne concernent pas des tubes rigides sans pertes. Il est nécessaire de les corriger afin de supprimer l'influence des pertes acoustiques, et de se placer ainsi dans les conditions qu'exige la méthode de calcul des fonctions d'aire par prédiction linéaire. La correction due à l'effet des parois a été effectuée en sens "inverse" à l'aide de la formule paramétrée présentée au chapitre VI. Pour la correction de l'effet de radiation nous avons utilisé la table suivante pour les deux sujets. Cette table dérive des valeurs numériques présentées au chapitre VI. Les valeurs pour F4 sont approximatives.

	F1	F2	F3	F4	
i	+ 6	10	150	200	Hz
ɛ	20	85	200	325	
a	50	125	80	225	
ɔ	20	150	120	80	
u	13	0	0	30	

Le tableau II-VIII présente en parallèle les valeurs originales et les valeurs corrigées.

c) Evolution du critère ε^2 en fonction de la longueur

Le critère ε^2 a été calculé à partir des fonctions d'aire obtenues par la méthode décrite pour le test préliminaire sur les données de FANT. L'évolution en fonction de la longueur est présentée aux fig. 9-VIII, 10-VIII, 11-VIII, 12-VIII. Le triangle indique la longueur mesurée pour chaque voyelle. Il est apparu immédiatement que la correspondance entre la longueur calculée et la longueur larynx - commissure était meilleure que



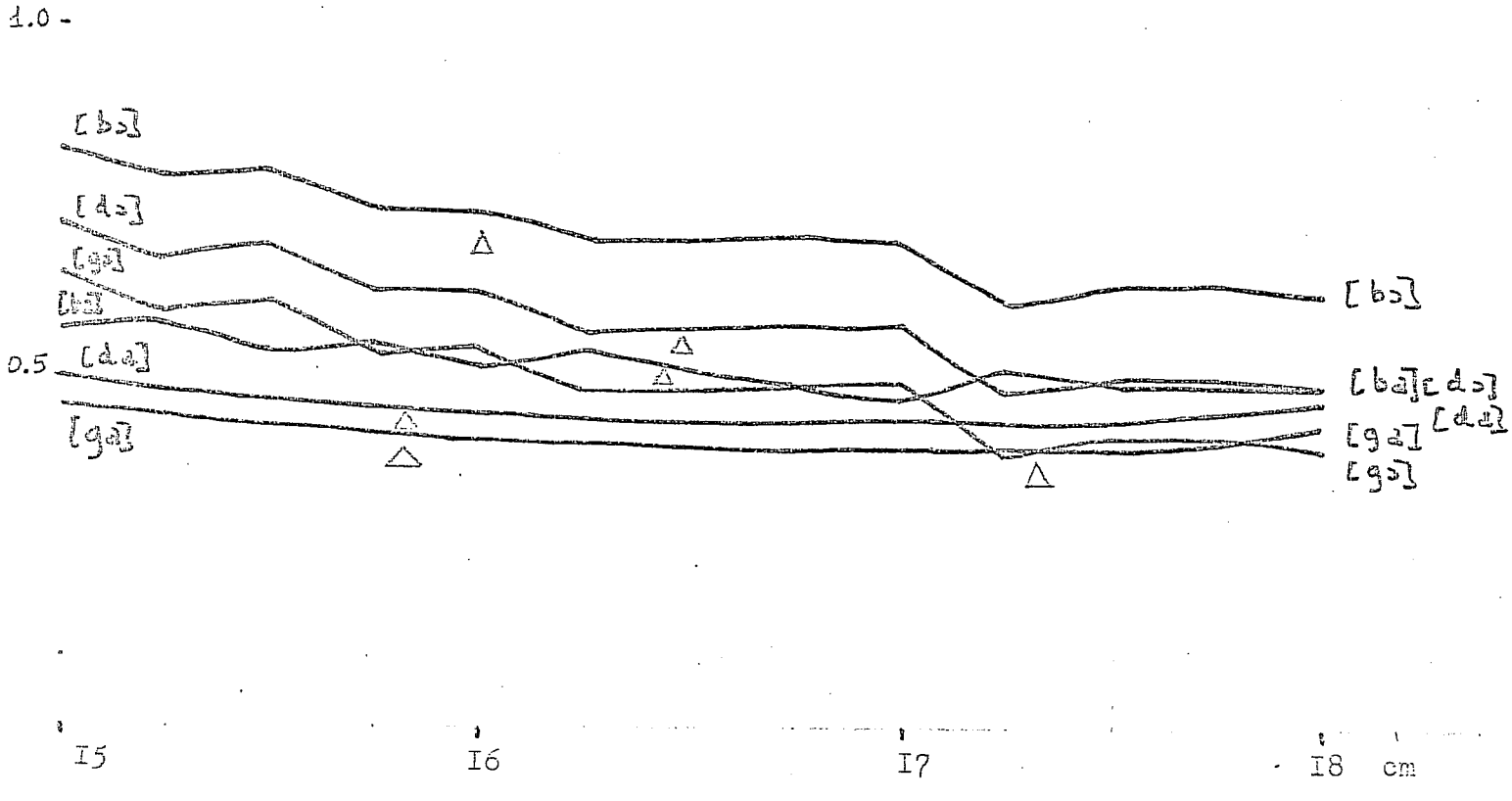
△ longueur mesurée

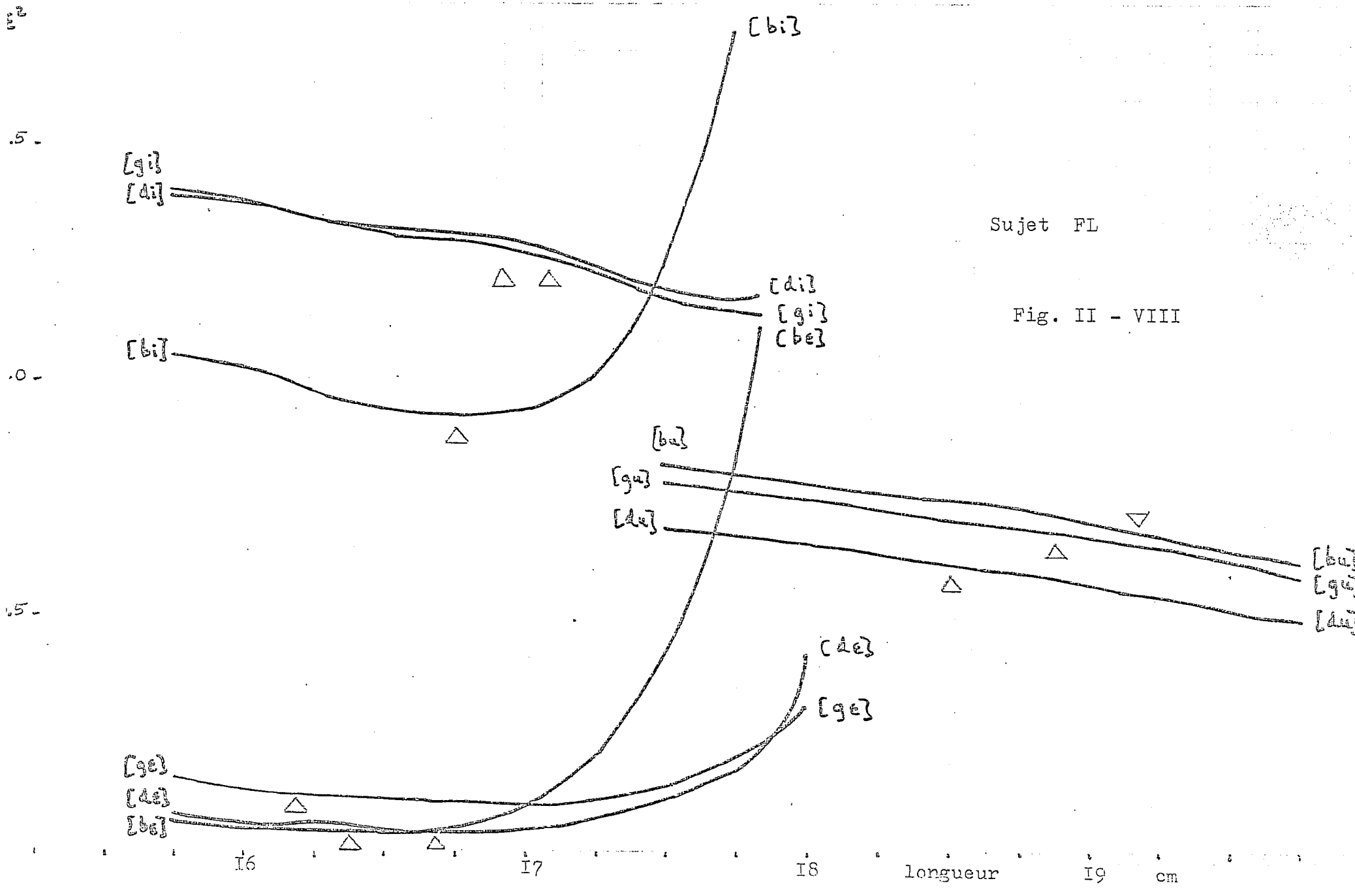
Sujet JPZ

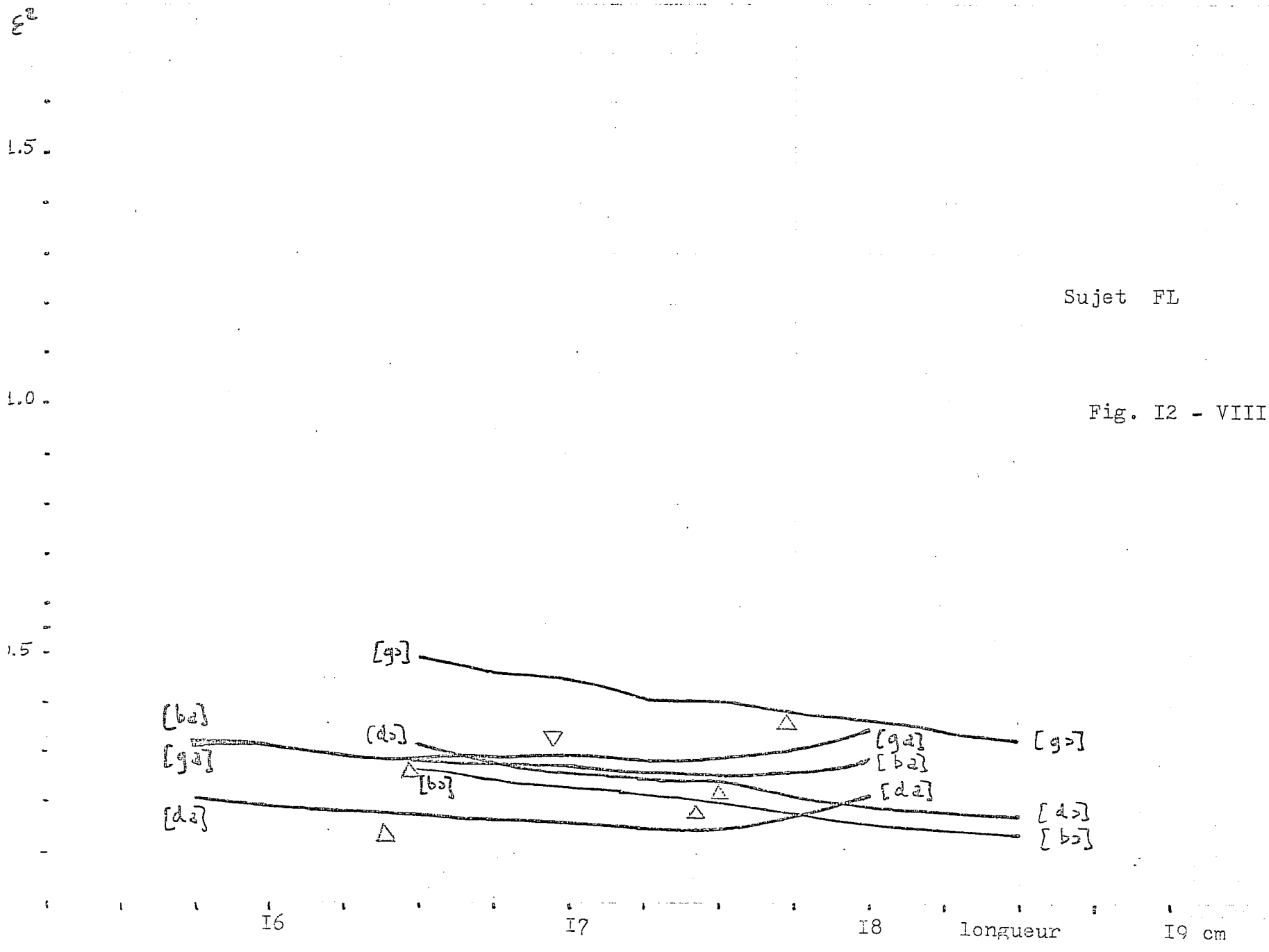
Fig. 9 - VIII

Sujet JPZ

Fig. IO - VIII




Sujet FL
Fig. II - VIII
[bi]
[gi]
[di]
[di]
[gi]
[bi]
[be]
[bu]
[gu]
[du]
[bu]
[gu]
[du]
[de]
[ge]
[ge]
[de]
[be]
.5
.0
.5
I6
I7
I8
longueur
I9
cm



Sujet FL

Fig. I2 - VIII

Sujet FL

Voyelles	longueur (mm) larynx-commissures	longueur larynx-lèvres	longueur calculée	% erreur
bi	I67.5	I83	I67.5	0.0
b	I67	I82	I65	-I.2
ba	I69	I83.5	I72.5	2.I
b	I73.5	I88	-	-
bu	I90.5	I97.5	-	-
di	I7I	I83	I77.5	3.8
d	I63.5	I8I	I67.5	2.4
da	I63	I82.5	I72.5	5.8
d	I75.5	I88.5	-	-
du	I84	I90.5	-	-
gi	I68.5	I82.5	-	-
g	I62	I78	I70	4.9
ga	I64.5	I82	I72.5	4.8
g	I77	I88	-	-
gu	I88.5	I92.5	-	-

Table III - VIII b

Sujet JPZ

Voyelles	longueur (mm) larynx-commissures	longueur larynx-lèvres	longueur calculée	% erreur
bi	I62	I74.5	I65	I.9
bɛ	I6I.5	I69	I60	-0.9
ba	I64	I73.5	I70	3.7
bɔ	I60	I72	-	-
bu	I82	I88	-	-
di	I66	I74.5	I67.5	0.9
dɛ	I53	I65	I60	4.6
da	I58	I69.5	I65	4.4
dɔ	I64.5	I72	-	-
du	I79.5	I82	-	-
gi	I63.5	I73.5	I67.5	2.8
gɛ	I57	I69.5	I60	I.9
ga	I58	I7I	I67.5	6.0
gɔ	I73.5	I79	-	-
gu	I85	I87	-	-

Table III - VIII a

celle obtenue par la distance larynx - lèvres. La discussion ne portera donc que sur cette première méthode d'estimation des longueurs. Les résultats étant voisins pour les deux sujets, nous ne les présenterons que voyelle par voyelle. On considérera que l'accord est satisfaisant si l'erreur ne dépasse pas 5 % (= ± 1 cm). Les résultats sont présentés à la table III et fig. I3

- Voyelle [i] : On observe un accord satisfaisant pour 5 voyelles sur 6. Les écarts en pourcentage sont les suivants :

sujet JPZ	bi	+ 1.85 %	sujet F.L.	bi	+ 2.76 %
	di	0.90		di	0.00
	gi	2.76		gi	——

Un pourcentage positif indique que la valeur calculée est supérieure à la longueur mesurée.

Pour le sujet FL, l'erreur n'est importante que pour le contexte [əg-g]. Aucun minimum n'est atteint pour une longueur de 18 cm. L'examen des fréquences formantiques révèle que la fréquence de F4 n'a pu être déterminée avec précision. La valeur retenue est sans doute trop faible, ce qui pourrait expliquer le comportement du critère.

- Voyelle [ɛ]

Le pourcentage d'erreurs est inférieur à 5 % pour toutes les voyelles. Nous n'avons pas d'explication pour l'erreur plus forte constatée pour les contextes [ədɑ-dɑ] et [əg-g] chez les sujets JPZ et FL.

sujet JPZ		sujet FL	
bɛ	- 0.94 %	bɛ	- 1.21 %
dɛ	4.58	dɛ	2.45
gɛ	1.91	gɛ	4.94

- Voyelle [ɑ]

Les variations du critère sont particulièrement faibles pour les deux sujets. Les longueurs calculées sont systématiquement plus fortes, et les valeurs ne sont que marginalement acceptables. Ceci pourrait provenir d'une estimation incorrecte de la ligne médiane.

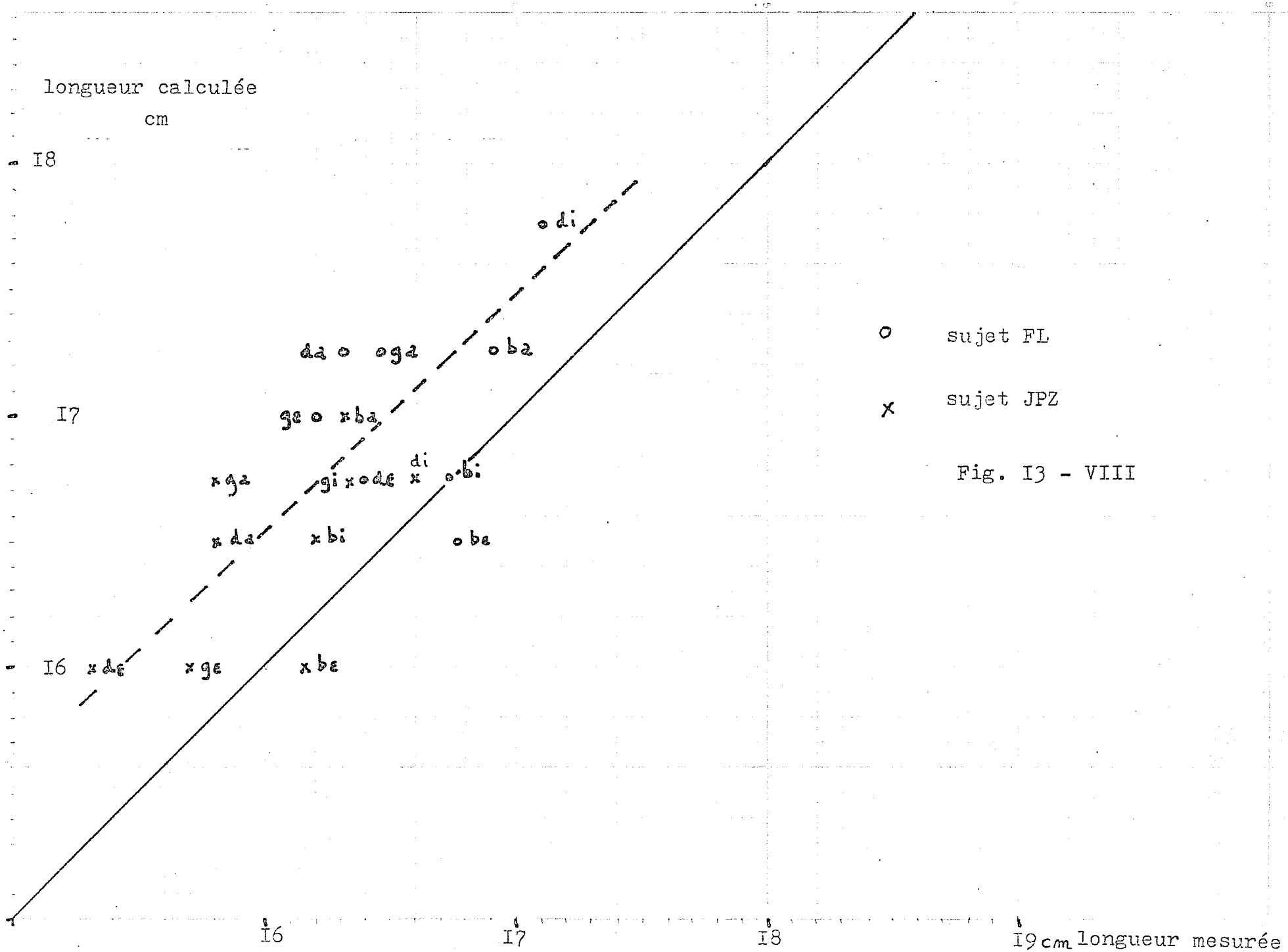

longueur calculée
cm
I8
I7
I6
di
da
ga
ba
ge
ba
di
ga
gi
de
di
bi
da
bi
ba
de
ge
be
I6
I7
I8
I9 cm longueur mesurée
sujet FL
sujet JPZ

Fig. I3 - VIII

sujet JPZ		sujet FL	
ba	3.66 %	ba	2.07
da	4.43	da	5.83
ga	6.01	ga	4.86

Une variation importante de la longueur du conduit ne provoque pas une variation sensible de l'écart moyen par rapport au conduit uniforme. On pourrait conclure qu'il est possible de produire un même résultat acoustique dans la région de [a] avec des positions articulatoires différentes. Il suffirait de faire varier la longueur du conduit vocal. Or il est évident que ceci n'est pas possible en pratique car le principal moyen qui permet cette variation de la longueur n'est pas disponible : il est en effet impossible de labialiser la voyelle [ɑ] . C'est pourtant ce qu'indiquent les fonctions d'aire correspondant aux longueurs de 16.75 et 17.75 cm.

sections	1	2	3	4	5	6	7	8
L= 16.75 cm	3.95	10.00	6.40	5.09	2.66	2.62	2.40	1.95
L= 17.75	6.39	8.05	9.00	5.65	3.42	2.71	2.83	2.07

Pour le conduit "court", l'aire de la première section (qui englobe l'aire aux lèvres et qui dépend fortement de l'angle maxillaire) est largement inférieure à celle du conduit "long".

Le conduit "long" exige donc à la fois une plus grande longueur (ce qui implique une certaine labialisation) et une plus forte aperture labiale ; Ces deux exigences sont incompatibles.

On doit supposer qu'il n'y a pas d'ajustement possible de la longueur du conduit pour la voyelle [a]. Notre résultat montre que, quelque soit la longueur effective, il existe une possibilité de réaliser la voyelle [a] d'une façon économique, c'est-à-dire sans écart trop important avec le tube uniforme. On peut donc prévoir que des réalisations acoustiques proches seront réalisées par des formes articulatoires relativement différentes si le conduit n'a pas la même taille pour les sujets. Nous n'avons pas trouvé de travaux publiés confirmant ou infirmant cette hypothèse. LADEFOGED et al. (1972) indiquent que les

sujets utilisent des stratégies différentes pour l'articulation des voyelles antérieures et en particulier pour [æ] . Mais ils ne présentent pas de mesures fréquentielles permettant de savoir si ces différences s'accompagnent de larges différences acoustiques. Une faible variation est vraisemblable pour les voyelles ouvertes, car STEVENS (1972) a montré que la voyelle [a] était particulièrement insensible aux variations articulatoires. Ceci est un des aspects essentiels de sa "théorie quantale de l'articulation". Ce résultat est vérifié pour les deux voyelles suivantes du sujet JPZ, où une variation de l'angle maxillaire de 35 % (distance incisive supérieure - incisive inférieure : 6.5 - 10 mm) ne se traduit pratiquement pas sur le plan acoustique.

	F1	F2	F3	F4	lg.
[da]	735	1285	2400	3225	15.8 cm
[ga]	730	1350	2250?	3300	15.8

Les remarques précédentes permettent de faire une hypothèse sur l'origine des stratégies articulatoires différentes selon les sujets. Pour atteindre un résultat acoustique similaire, ceux-ci doivent utiliser des positions articulatoires différentes en raison de la différence des longueurs de conduits vocaux. De plus, toutes ces positions ont approximativement le même "coût", car elles présentent des écarts voisins par rapport au conduit acoustique uniforme. Il serait intéressant de vérifier si les sujets utilisent bien cette possibilité de compensation.

- Voyelle [ɔ]

Pour le sujet FL, on ne constate pas de minimum pour la voyelle dans la gamme des longueurs analysées. Chez le sujet JPZ, la courbe présente une irrégularité inexplicable. Celle-ci ne se place pas à la longueur mesurée, sauf pour [gɔ]

- Voyelle [u]

Pour les deux sujets, le critère ne présente pas de minimum. Contrairement à ce que nous avions constaté pour les données de FANT, la longueur maximum théorique ne peut être invoquée, puisqu'elle est supérieure à 20.5 cm dans le cas le plus favorable. Il semble donc que nos résultats ne confirment pas les conclusions de PAIGE et ZUE (1970a) et de WAKITA (1977) :

Le critère d'écart par rapport au tube uniforme ne permet pas de sélectionner le tube acoustique optimum pour les voyelles non-antérieures.

Pour les voyelles postérieures, nos résultats indiquent simplement que PLUS LE CONDUIT VOCAL EST LONG, PLUS LA FORME ARTICULATOIRE NECESSAIRE SERA PROCHE DU TUBE ACOUSTIQUE UNIFORME.

Ceci implique que les sujets auront intérêt à labialiser la voyelle [u] ou la voyelle [o] le plus possible. L'articulation la plus économique pour les voyelles postérieures est une articulation labialisée, avec, éventuellement, un abaissement du larynx pour augmenter la longueur du conduit (cf. RIORDAN 1977).

La première contrainte articulatoire donne une explication de l'affinité constatée par les phonologues entre le trait "postérieur" et le trait "labial". Le principe du moindre effort musculaire lingual implique la labialisation des voyelles postérieures. Une discussion détaillée de l'importance de cette constatation pour une théorie des traits distinctifs sera présentée au chapitre XI. Le paragraphe suivant présente la deuxième contrainte articulatoire que nous avons également étudiée au moyen de fonctions d'aire calculées.

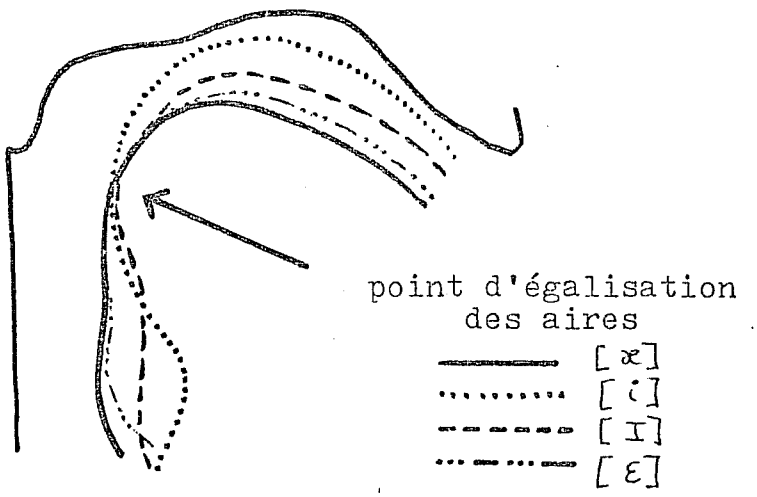
II - UNE CONTRAINTE ARTICULATOIRE SUR LES "AXES" DE L'ESPACE ACOUSTIQUE

La caractérisation de cette deuxième contrainte articulatoire est plus difficile. Aussi n'en donnerons-nous une définition précise qu'à la fin du chapitre. Précisons simplement que, selon cette contrainte, les voyelles se répartissent de façon privilégiée sur deux "axes", [i - ɑ] et [y - œ], et peut être également sur l'axe [ɑ - ɯ]. La raison serait que ce ne sont que sur ces axes que l'action d'un muscle unique produit un effet acoustique proportionnel à la contraction.

L'idée d'une telle contrainte est née de la juxtaposition de deux études indépendantes (GUNNILSTAM 1974 ; WOOD 1977)

WOODS montre que la musculature linguale est conçue de manière à réaliser des constrictions dans quatre zones distinctes : zone palatale pour [i, y, e, ø, ɛ] , zone vélaire pour [ɨ, u, ʊ, ɯ] , partie supérieure du pharynx pour [o, ɔ, ɤ] et partie inférieure du pharynx pour [a, ɑ, ʌ, æ] . Nous ne sommes pas d'accord avec le regroupement des voyelles [ɨ] et [u] dans une même classe. WOOD indique en effet que le champ articulatoire de [u] est limité par l'action de muscles quasi antagonistes et que les variations sont en conséquence limitées (styloglossus et palatoglossus). Les tracés de FANT (1960) indiquent une position linguale très différente pour [ɨ] et [u] (cf. fig. 17-VIII). On trouvera une figure décrivant l'action des muscles dans l'articulation de [u] au chapitre suivant.

GUNNILSTAM propose une théorie de la structure vocalique qui prévoit en substance que seront privilégiées les zones articulatoires où il existe une relation proportionnelle entre un mouvement et un effet acoustique : "Seul un très petit nombre de changements potentiels des organes de parole produisent une variation acoustique traduisant, d'une manière proportionnelle et adéquate (?), les mouvements de l'organe de parole en question (op. cit., p. 97). Le détail de la démonstration de GUNNILSTAM est de toute évidence incorrect. L'hypothèse de deux points d'articulation (à 4 et 11 cm de la glotte) le conduit à traiter [ɯ] avec les voyelles labialisées. Mais l'idée générale est intéressante ; elle explique en particulier la facilité avec laquelle nous apprenons à siffler. Il existe en effet une relation linéaire entre le son produit et la distance de la constriction linguale aux lèvres. Cette relation pourrait également faciliter l'apprentissage chez l'enfant car il s'agit d'une relation simple entre le produit de l'action d'un muscle et un résultat auditif. Les implications diachroniques ne sont pas négligeables. Les évolutions se feraient de façon privilégiée le long de ces axes, et exceptionnellement perpendiculairement à eux. On pense tout de suite au "déficit" de voyelles centrales que nous avons constaté dans l'étude du principe de différenciation maxima.



(d'après PERKELL 1969 p. 55)

fig. 14 - VIII

Une première vérification de l'existence d'une telle contrainte peut être faite en comparant les gestes articulatoires permettant de passer de la voyelle [i] à la voyelle [ɛ] et ceux nécessaires pour aller de [u] à [i]. Nous faisons l'hypothèse que le passage de [i] à [ɛ] se fait le long d'un axe "naturel" et qu'il ne nécessite que la modification de la contraction d'un muscle. L'axe [u - i], au contraire n'est pas un axe "naturel" dans la mesure où ce passage nécessite une réorganisation complète de l'activité musculaire. L'évaluation de l'axe [ɑ - u] se heurte à une difficulté. WOOD a montré que [u] et [o] possèdent deux lieux de constrictions distincts, qui sont atteints par des activités musculaires différentes. Au chapitre suivant nous montrerons pourtant que le passage de [u] à [o], en particulier, se fait bien le long d'un axe "naturel". Mais il ne dépend pas principalement d'une action musculaire linguale.

Si la démonstration d'une liaison étroite entre l'action du génioglossus et la position des voyelles sur l'axe [i - ɛ] est relativement facile, l'absence de données articulatoires précises sur les voyelles entre [y] et [u] pose un problème. C'est pourquoi nous avons eu recours au calcul de la fonction d'aire par prédiction linéaire. Ces résultats ont été comparés aux quelques données disponibles.

L'utilisation des fonctions d'aire pose néanmoins un problème. Celles-ci ne sont connues qu'à une constante multiplicative près, ce qui rend difficile les comparaisons. Ce problème a fait l'objet d'une étude récente (BROOKS et FALLSIDE 1976), qui portait sur la transformation des fonctions d'aire en "vues" latérales dans un système de rééducation d'enfants mal-entendants. Mais la solution que nous proposons nous semble supérieure.

Si l'on compare (fig. 14-VIII) les tracés latéraux pour les voyelles antérieures, on s'aperçoit qu'il existe un point relativement fixe pour lequel les fonctions d'aire seront identiques. Nous avons appelé ce point, le "point d'égalisation des aires". Nous avons utilisé les tracés de PERKELL (1969).

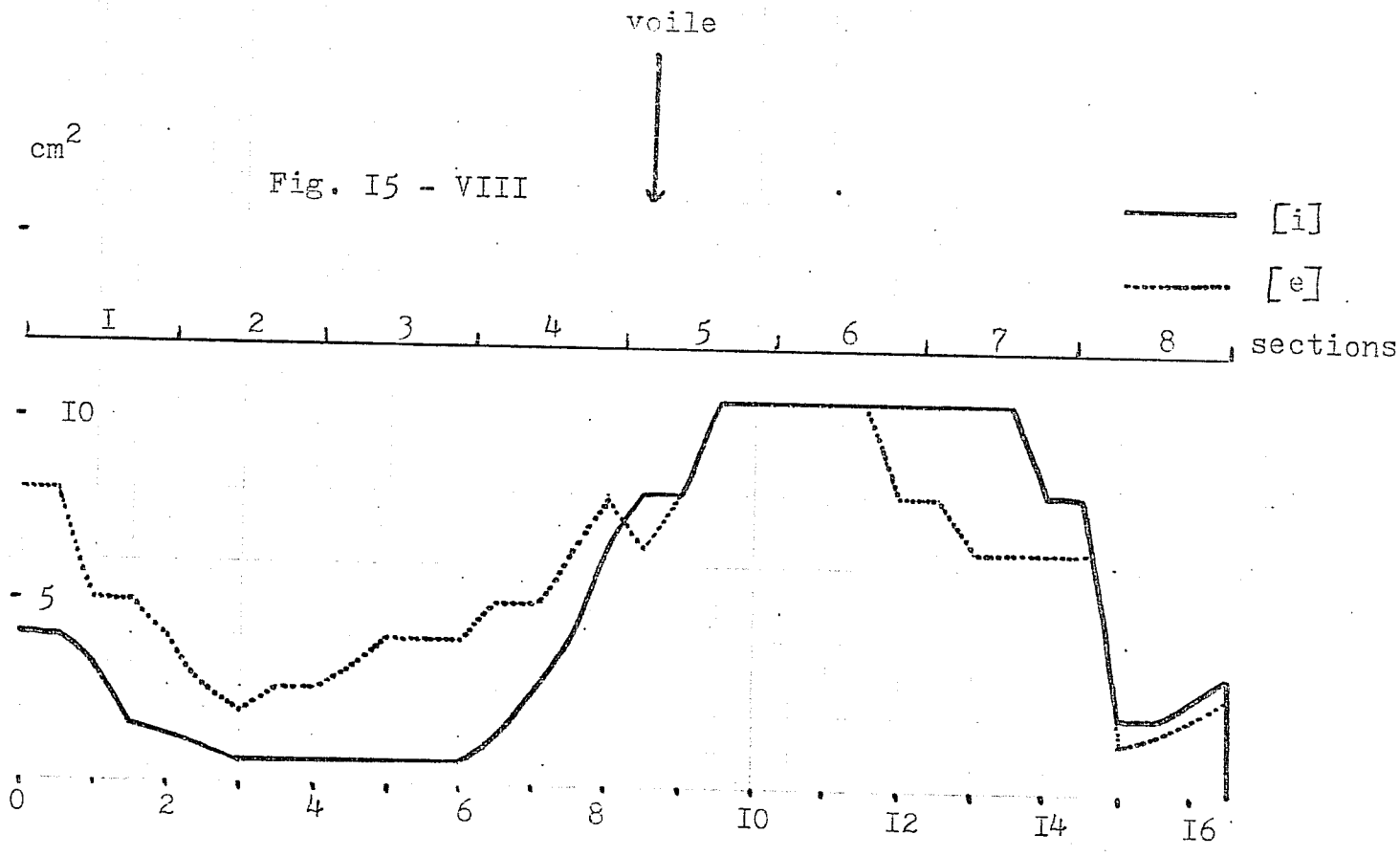


Fig. I5 - VIII

Certains tracés de BRICHLER (1970) semblent indiquer une rotation de la tête dans un plan vertical. Ceci ne permet pas de superposer les conduits. D'autre part il peut s'être produit une aberration optique sur les bords de l'écran de brillance.

Ce point d'égalisation des aires correspond approximativement à notre cinquième section, comme le prouve la superposition des fonctions d'aires de FANT (1960) pour [i] et [e] La position du voile est indiquée à la fig. 15-VIII . Les fonctions d'aire obtenues par prédiction linéaire ont été normalisées de manière que l'aire de la 5ème section soit approximativement fixe. Ce problème sera repris plus bas pour les voyelles non-antérieures.

A/ Evolution articulatoire entre [i] et [ɛ]

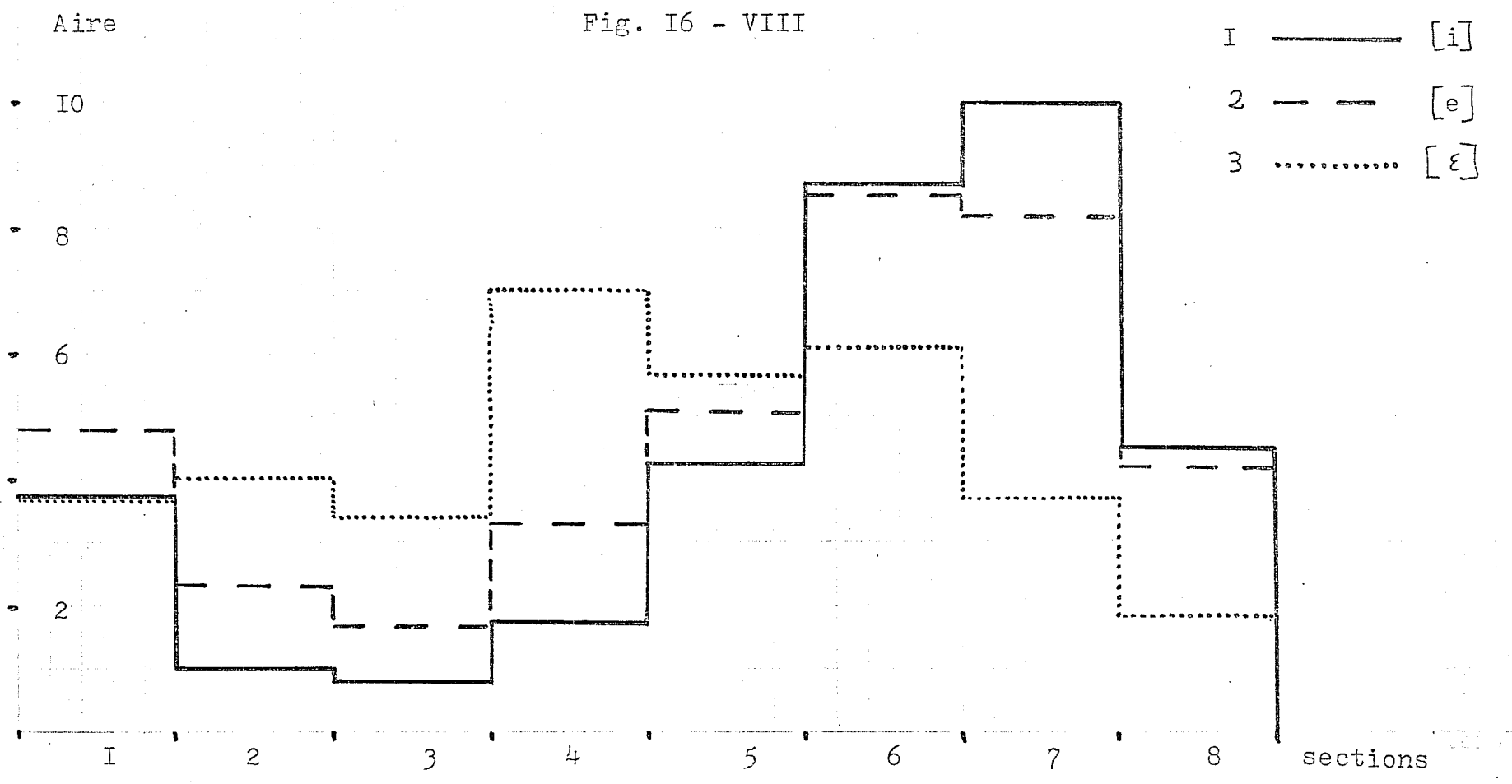
Nous avons calculé les fonctions d'aires pour 3 voyelles correspondant approximativement à [i], [e] et . (fig. 16 - VIII). Les valeurs numériques des formants et les aires sont données à la table IV-VIII. Le "facteur de cadrage" est le coefficient multiplicateur adopté pour égaliser approximativement les aires de la 5ème section. La comparaison entre les fonctions d'aire de FANT (fig. 15-VIII) et les aires calculées fig. 16-VIII est bonne.

- La section n° 3, qui correspond au point de constriction présente les voyelles dans l'ordre [i, e, ɛ] attendu ; cet ordre est conforme au critère traditionnel d'aperture. Cette relation vaut également pour les sections 2 et 4. En revanche, l'aire aux lèvres (section n° 1) ne reflète pas une aperture croissante.

- La cavité pharyngale est nettement plus petite pour [ɛ]. Mais elle décroît légèrement de [i] à [e].

Notre discussion de l'activité musculaire sera incomplète, car elle ne tient pas compte de l'angle maxillaire. Mais en première approximation, on peut dire, que la variation de la fonction d'aire s'explique aisément à partir d'une contraction différente du genioglossus. On trouvera dans WOOD (1977) un excellent résumé des données électromyographiques. MIYAWAKI et

Fig. I6 - VIII



VOYELLE N° I [i] 250 - 2100 - 2900 - 3600 Hz l=17.5 cm.
50 80 100 120 Hz

lèvres	3.73
	0.97
	0.70
	1.73
	4.26
	8.78
	10.00
glotte	4.46

l'aire la plus grande a été arbitrairement fixée à 10.00

VOYELLE N° 2 [e] 350 - 1900 - 2750 - 3500 Hz l = 17.5 cm

facteur de cadrage 0.85

lèvres	5.62	4.78
	2.76	2.35
	1.99	1.69
	3.87	3.29
	6.00	5.10
	10.00	8.50
	9.67	8.22
glotte	4.92	4.18

VOYELLE N° 3 [ɛ] 500 - 1750 - 2650 - 3400 Hz l = 17.25 cm

facteur de cadrage 0.70

lèvres	5.26	3.68
	5.76	4.02
	4.85	3.40
	10.00	7.00
	8.06	5.64
	8.66	6.06
	5.23	3.66
glotte	2.64	1.85

Table IV - VIII

al. (1976), que WOOD ne cite pas, ont établi qu'en japonais, l'activité des fibres antérieures du génioglossus est inférieure de moitié pour [e] par rapport à [i]. Les fibres postérieures, inactives pour [e], sont actives pour [i].

La traction opérée par le génioglossus (cf. fig 3-XI au chapitre suivant pour la voyelle [y]) a pour effet d'agrandir la cavité pharyngale et d'élever la masse de la langue par un effet de "gonflement" mécanique, résultant de la compression opérée. Le génioglossus est le seul muscle actif lors de l'articulation des voyelles antérieures. Plus la contraction est importante, plus la cavité pharyngale est grande et plus la langue est élevée. Il y a donc lieu de penser que la position alignée des voyelles antérieures dans l'espace acoustique est bien fonction du degré de contraction d'un muscle unique. Ceci s'applique également aux voyelles [y, ø, œ] qui sont des antérieures labialisées.

B/ Evolution articulatoire entre [u] et [i]

a) Description

La fig. 17-VIII montre que le point d'égalisation des aires, déterminé à partir des voyelles [i , ɨ] et [u] du russe (FANT 1960), est légèrement plus antérieur que celui des voyelles antérieures. Il se situe sur la 4ème section de nos approximations en 8 sections. Nous avons calculé les fonctions d'aire de 5 voyelles approximativement équidistantes entre [u] et [y]. Les fréquences formantiques et les aires normalisées sont données à la table V-VIII

- La voyelle n° 1 (280 - 850 - 2300 Hz) [u] (fig. 18-VIII) se caractérise par une forte labialisation, une cavité antérieure de grande taille et par un faible diamètre de la cavité pharyngale. La comparaison entre la voyelle [u] naturelle (FANT 1960, fig. 17-VIII) montre une nette différence au niveau du pharynx. D'autre part, le rapport entre l'aire buccale et l'aire de la constriction est incorrect dans notre simulation. Une variation de la longueur n'amène aucune amélioration. Pour une

Fig. I7 - VIII

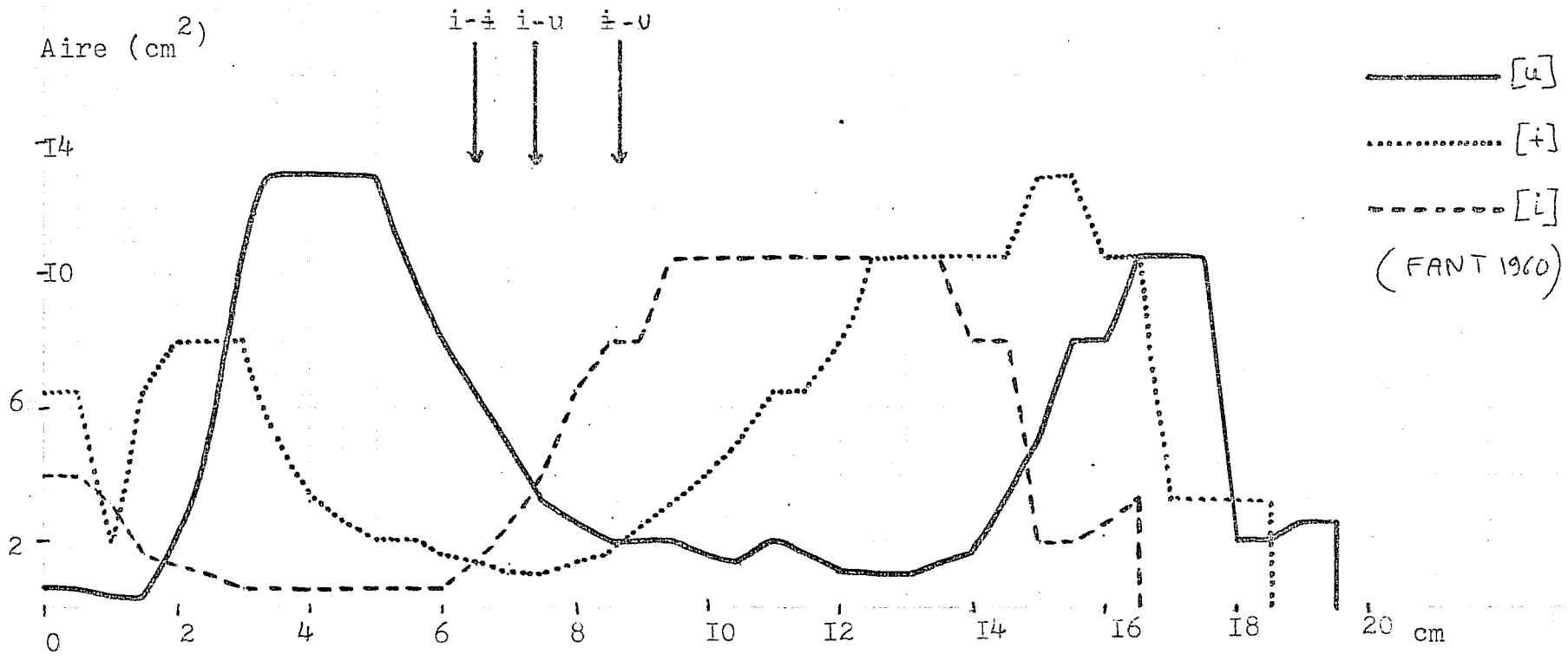


Fig. I8 - VIII a

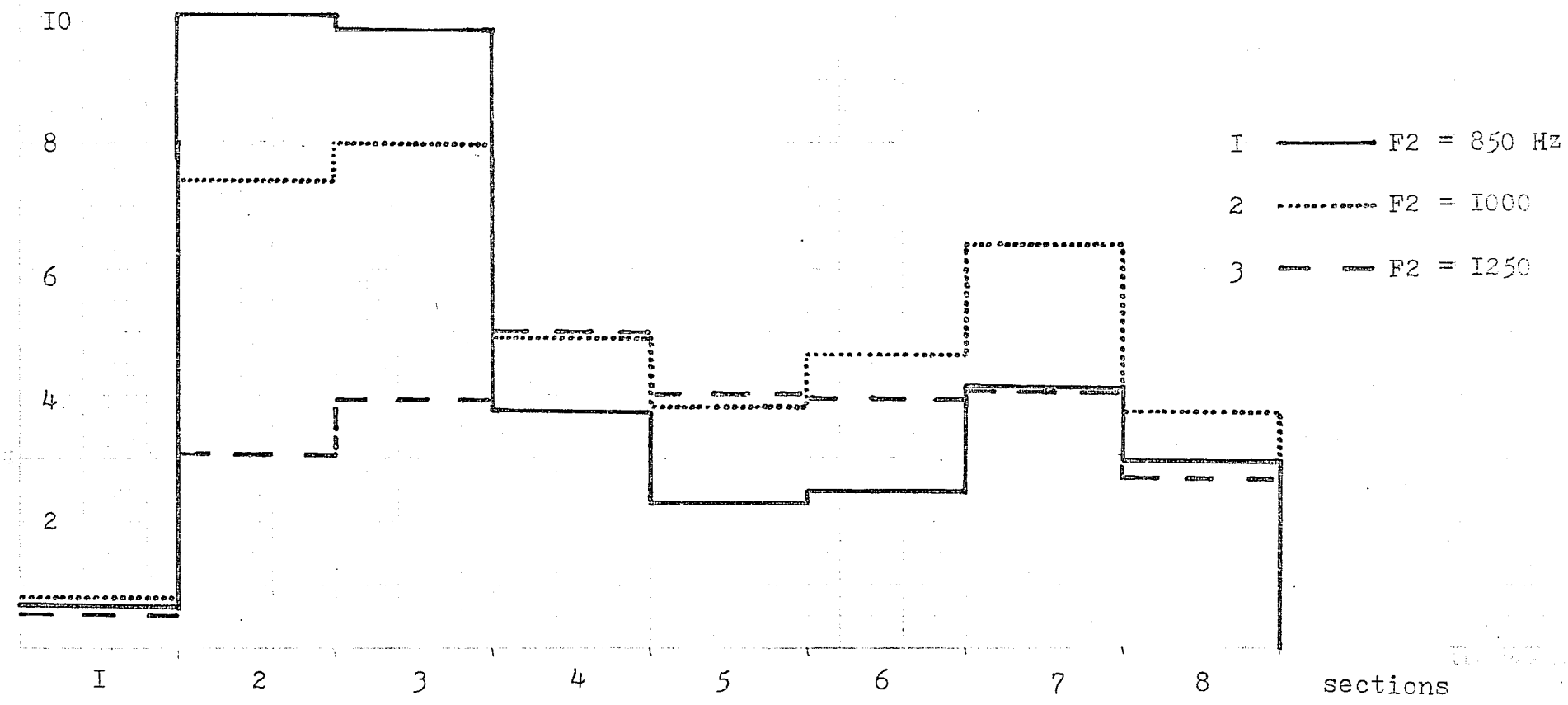
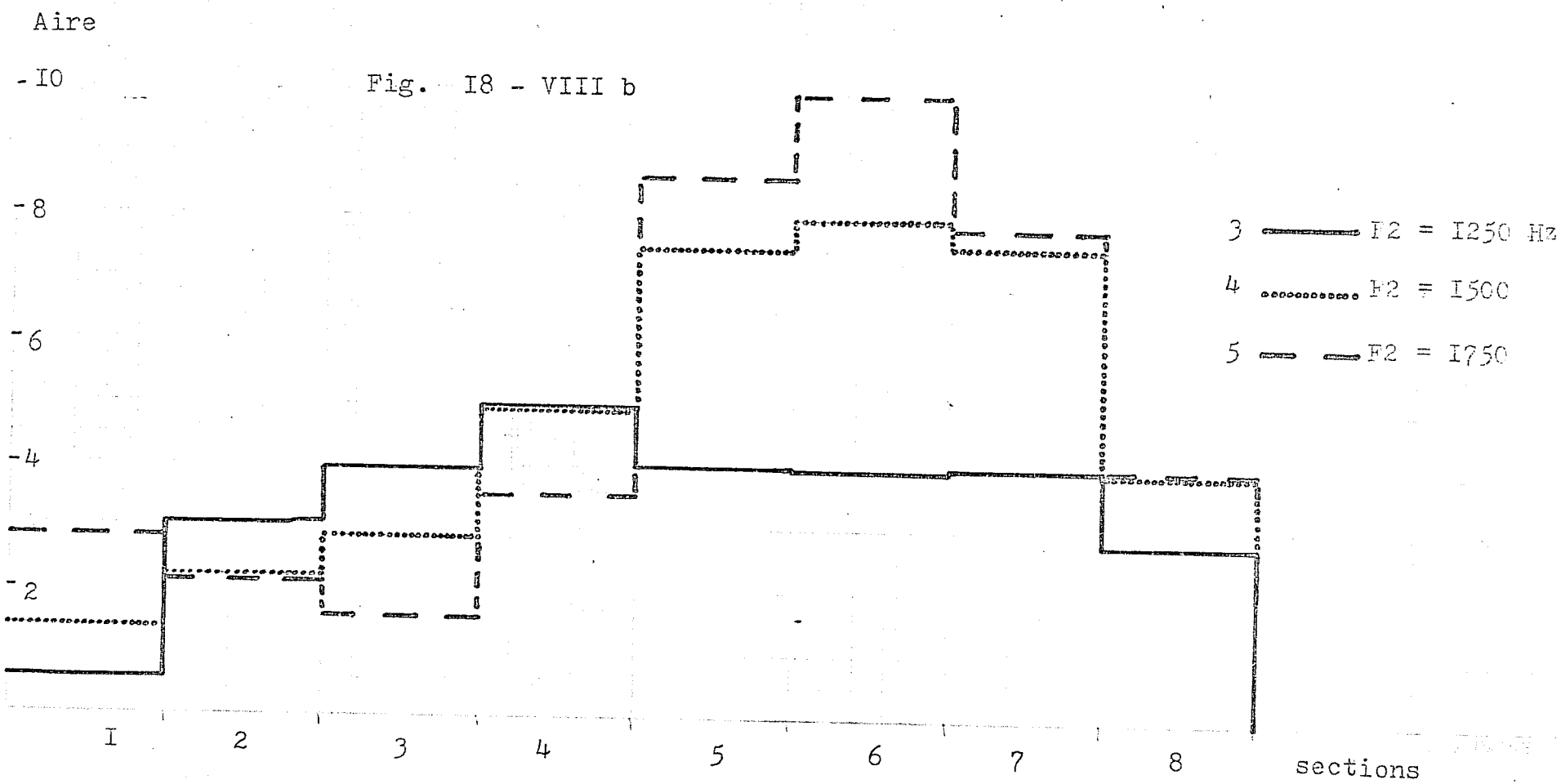


Fig. I8 - VIII b



VOYELLE N° I 280 850 2300 3200 Hz l = I9.75 cm

Lèvres	0.68
	I0.00
	9.8I
	3.76
	2.29
	2.48
	4.I4
glotte	2.96

l'aire maximale a été arbitrairement fixée à I0.00

VOYELLE N° 2 280 I000 2300 3200 Hz l = I9.5 cm

facteur de cadrage 0.8

lèvres	0.99	0.79
	9.25	7.40
	I0.00	8.00
	6.I4	4.9I
	4.79	3.83
	5.80	4.64
	7.98	6.34
glotte	4.63	3.74

VOYELLE N° 3 280 I250 2300 3200Hz l = I9.0 cm

facteur de cadrage 0.5

lèvres	I.I3	0.57
	6.07	3.04
	7.9I	3.96
	I0.00	5.00
	8.03	4.0I
	7.87	3.94
	8.II	4.06
glotte	5.34	2.67

VOYELLE N° 4 280 I500 2300 3200 Hz l = I9.5 cm

facteur de cadrage 0.8

lèvres	I.73	I.38
	2.75	2.20
	3.57	2.85
	6.I8	4.94
	9.38	7.50
	I0.00	8.00
	9.44	7.55
glotte	4.87	3.90

VOYELLE N° 5 280 I750 2300 3200 Hz l = I9.5 cm

facteur de cadrage I.00

lèvres	2.87
	2.24
	I.6I
	3.54
	8.65
	I0.00
	7.86
glotte	3.94

Table V - VIII

longueur de 18 cm, la cavité buccale devient énorme, et le "tube" pharyngal se rétrécit encore.

- La voyelle n° 2 (280 - 1000 - 2300 Hz) fait au contraire apparaître une cavité pharyngale importante, conforme aux mesures de FANT. Mais le rapport aire maximale / aire de la constriction est encore plus incorrect. Nous n'avons pas d'explications pour ces divergences. L'articulation de la voyelle n° 2 exige une action du génioglossus afin d'obtenir une large cavité pharyngale et une traction du styloglossus et du palatoglossus (cf. fig. 2-IX, chapitre suivant) pour postérioriser la masse de la langue.

- La voyelle n° 3 (280 - 1250 - 2300) possède une forme singulière. les formants F1 et F2 sont approximativement ceux de la voyelle japonaise [ɯ] (KANAMORI 1974, fig. 9-IX). La comparaison avec la fonction d'aire de cette voyelle (fig. 19-VIII CHIBA et KAJIYAMA 1941) ne montre aucun point commun. La voyelle naturelle possède une structure de voyelle postérieure avec une cavité antérieure plus petite révélant une légère antériorisation de la langue. La forme calculée ressemble à un tube "gonflé" en son milieu. On doit se demander si l'articulation prévue est physiquement possible : il faudrait que la langue épouse la forme de la voûte palatale et du pharynx, avec un léger "creux" au niveau du voile, ce qui paraît peu plausible. La forme calculée est également très différente de celle de la voyelle [ɨ] (fig. 17-VIII).

- La voyelle n° 4 (280 - 1500 - 2300 - 3200) possède une structure de voyelle antérieure labialisée : il n'y a plus de cavité antérieure, et la cavité postérieure est importante. Notons l'absence de point de constriction. Cette voyelle ne correspond sans doute à aucune voyelle existante. Les formants du [ɨ] russe calculés pour un tube rigide (cf. supra) sont plus élevés. Quant au deuxième formant de [ʉ], il atteint 1640 Hz (FANT 1959).

- La voyelle n° 5 (280 - 1750 - 2300 - 3200) possède une constriction palatale bien visible. On peut vérifier, par comparaison avec la fig. 16-VIII que cette voyelle est proche d'un [i] labialisé.

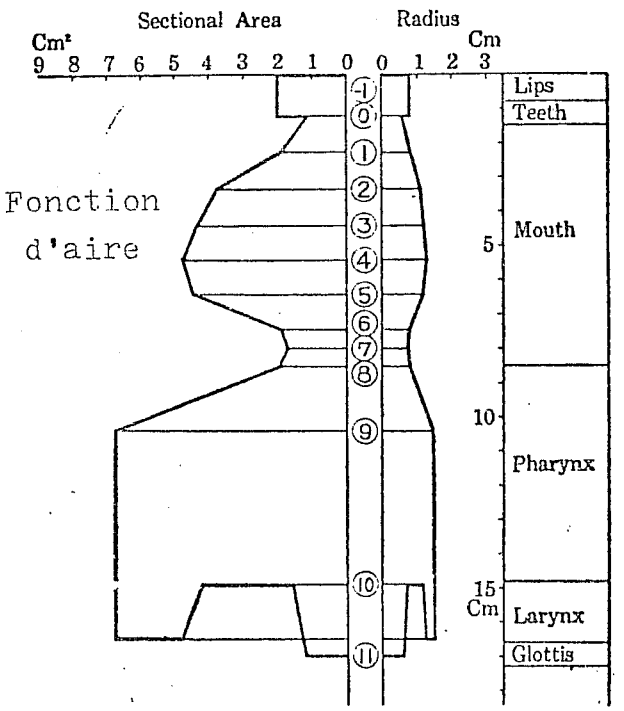


Fig. 77. Diagrammatical View of Vocal Resonator of う (u)

Fig. I9 - VIII

(CHIBA et KAJIYAMA I94I)

b) conséquences

Cette présentation des aires calculées montre que l'évolution [u > y] sera complexe.

Si la voyelle n° 2 (280 - 1000 - 2300 Hz) est représentative de l'articulation de la voyelle [u], on note qu'il est physiquement impossible de faire passer F2 à 1250 Hz. Que la voyelle [ɯ] japonaise ait un formant dans cette zone signifie qu'elle utilise une articulation postérieure mais fortement délabialisée, ce qui est indiqué par l'élévation du premier formant, comme nous le montrerons au chapitre suivant (F1 = 400 Hz ; KANAMORI 1974). La zone entre 1000 et 1500 Hz peut être considérée comme une "zone interdite", si l'on exige de plus un premier formant inférieur ou égal à 300 Hz.

Pour F2 compris entre 1500 et 1750 Hz, il est possible de réaliser deux voyelles: [ɨ] et [ʉ]. Cette dernière n'est, comme nous le verrons au chapitre suivant, qu'un [y] fortement labialisé. L'articulation est centrale pour [ɨ] comme le montre la fig. 17-VIII ; La voyelle [ʉ] possède une articulation antérieure.

D'un point de vue musculaire, il est facile de montrer que le passage de [u] à [y] nécessite une réorganisation importante de l'action musculaire. Si l'on suit l'élévation de F2 à partir de [u], on constate que les ajustements suivants sont nécessaires :

- la voyelle [u] exige une forte activité du styloglossus, du palatoglossus et du génioglossus ainsi qu'une labialisation intense.
- Le passage à la voyelle [ɯ], pour contourner la "zone interdite" vers 1250 Hz pour un premier formant bas, nécessite une délabialisation, un relâchement du styloglossus et un accroissement de l'effet du palatoglossus, ce qui tendra à antérioriser la masse de la langue.
- L'articulation de [ɨ/ʉ] puis de [y] nécessitera la cessation de l'activité du styloglossus, la diminution de l'action du palatoglossus de [ɨ] à [y] et la reprise de la labialisation.

Il est clair qu'une langue n'effectuera ces changements articulatoires lors de la palatalisation spontanée de [u] à [y] que si elle y est "contrainte". Le chapitre IX est consacré à l'étude des facteurs qui ont pu déterminer cette évolution.

c) Conclusions

Ce paragraphe a tenté de montrer l'existence d'une seconde "contrainte" articulatoire, qui permet de définir des axes de l'espace acoustico-perceptif sur lesquels les évolutions seront plus ou moins "naturelles". Une évolution "naturelle" implique une relation linéaire entre l'action d'un muscle, ou d'un groupe de muscles, et les variations dans l'espace acoustico-perceptif. Il semble que les structures phonologiques étudiées au chapitre V possèdent un plus grand nombre de voyelles sur les axes naturels qu'en dehors de ceux-ci. On a montré que l'axe joignant [i] à [ɛ] résulte de l'importance de l'activité du génioglossus. Il s'agit donc d'un axe "naturel". En revanche, l'axe [u - i] ne peut être parcouru par la seule variation du degré d'activité d'un muscle. Le passage de [i] à [u] exige une réorganisation de l'effort musculaire lingual et un contrôle du degré de labialisation. Il nous semble également qu'il est impossible de placer une voyelle dans une zone acoustique limitée par une fréquence du premier formant inférieur à 300 Hz et un deuxième formant compris entre 1500 et 1000 Hz.

Ces résultats ont été obtenus par l'utilisation des fonctions d'aire calculées par prédiction linéaire.

III - CONCLUSIONS

L'étude des contraintes articulatoires au niveau de la production résulte de l'absence de paradigmes expérimentaux permettant de les mettre en évidence au niveau perceptif. L'utilisation d'informations articulatoires dans le processus de perception, par le biais d'un centre intégrateur sensori-moteur, est sans doute nécessaire pour la normalisation des voyelles en contexte.

La première contrainte étudiée révèle que la labialisation des voyelles postérieures n'est qu'une conséquence d'un principe du moindre effort. Nous avons vérifié sur des voyelles françaises que, pour une articulation s'écartant le moins du tube acoustique uniforme, il était nécessaire de délabialiser les voyelles antérieures et de labialiser les voyelles postérieures. Pour la voyelle [ɑ] , nous faisons l'hypothèse que la forme du conduit sera variable en fonction de la longueur. Mais toutes les articulations possèdent un coefficient voisin d'écart par rapport au tube uniforme.

Cette contrainte de moindre écart par rapport au tube uniforme permet d'expliquer la liaison entre les traits distinctifs de labialisation et de postériorisation. Dans l'espace acoustico-perceptif, les voyelles semblent se répartir de façon privilégiée sur des axes où le résultat acoustique est proportionnel à l'activité d'un muscle. Ceci constitue la deuxième contrainte articulatoire étudiée. Nous avons vérifié, par l'étude de fonctions d'aire calculées, que tel était bien le cas pour les axes [i - ɛ] et [y - œ] . L'articulation de ces voyelles ne dépend que de l'activité du génioglossus.

En revanche, nous avons montré que le passage de [ʊ] à [y] implique une réorganisation du geste articulatoire. Ce passage suppose une délabialisation préalable et une avancée articulatoire résultant d'une diminution de l'activité du styloglossus, d'une variation de l'activité du palatoglossus, et d'une "relabialisation" pour atteindre [y] .

Cet axe n'est pas "naturel", au sens où la variation acoustique n'est pas le résultat de la variation du potentiel d'activité d'un muscle unique.

Le chapitre suivant présentera un examen détaillé de la "catastrophe" linguistique provoquant la mutation [u > y] le long d'un axe non-naturel.

C H A P I T R E IX

VERS UNE EXPLICATION ARTICULO - PERCEPT IVE

DE LA MUTATION U > Y

I - INTRODUCTION

Le chapitre V a mis en évidence que la notion de différenciation maxima dans l'espace du timbre ne pouvait à elle-seule rendre compte de la structuration linguistique de l'espace vocalique. Le chapitre précédent a présenté deux aspects des contraintes articulatoires qui pèsent sur la production et qui sont peut-être utilisées au niveau de la perception : le degré de labialisation est fonction de la règle du moindre écart par rapport à la forme neutre du conduit ; certaines "dimensions" articulatoires peuvent être considérées comme plus "naturelles" que d'autres.

Nous avons, dans ce chapitre, cherché à comprendre le mécanisme sous-jacent à la mutation u > y. Plus généralement, nous nous sommes intéressés à l'ensemble des voyelles "centrales" arrondies ou non arrondies. Nous pensons en effet que le mécanisme de création de ces voyelles est commun, bien que nous ne soyions pas en mesure de le prouver. De même, la création des voyelles "centrales" à partir de [i] ou [e] par labialisation ou rétraction articulatoire nous paraît une forme secondaire de formation, qui ne touche que quelques mots d'une langue, sans jamais être un mécanisme général. Là-aussi, une preuve supposerait une étude bien plus longue que celle à laquelle nous nous sommes livrés.

Il s'agit, en bref, de comprendre pourquoi une langue transforme radicalement l'une des voyelles extrêmes, qui devrait être des plus stables. [i], à l'inverse de [u], est une voyelle extrêmement stable : l'évolution de cette voyelle de l'Indo-Européen aux différents dialectes contemporains en est une preuve irréfutable. De plus cette mutation, comme l'indique le chapitre précédent, ne résulte pas d'un mécanisme articulatoire naturel ; elle exige une réorganisation importante du geste articulatoire. Elle ne peut donc se comprendre que par un bouleversement, une "catastrophe" linguistique.

II - LA PALATALISATION [u] > [y] EN ANCIEN FRANÇAIS

La recherche d'une explication causale à la "palatalisation" de [u] en [y] * en ancien français a conduit les linguistes à émettre de nombreuses hypothèses. On observera que, si nous conservons le terme consacré par l'usage de "palatalisation", nous n'ignorons pas qu'il faudrait plutôt parler d'une "avancée articulatoire". Même si les controverses sont aujourd'hui moins vives, les spécialistes observent toujours une très grande prudence. Un manuel récent (CARTON 1974, p. 186) écrit par exemple : "... Les avis sont partagés sur l'origine de cette transformation.. Nous préférons nous rallier à l'idée d'une évolution,..., dont la cause semble être un affaiblissement articulatoire." Avec toutes les réserves d'usage, nous proposons une explication qui, pour être, du moins le croyons-nous, originale n'en est pas moins compatible avec la théorie de l'affaiblissement articulatoire, mais également avec l'explication structurale de HAUDRICOURT & JUILLAND (1949). Il ne s'agit bien sûr que d'une esquisse. L'organisation du chapitre sera la suivante : Après avoir caractérisé rapidement cette palatalisation, dans son cadre linguistique mais également géographique et diachronique, nous examinerons et critiquerons les deux principales explications proposées à ce jour. Puis nous présenterons la nôtre, qui insiste sur le rôle acoustique de la délabialisation de [u] sous l'effet d'un affaiblissement articulatoire. Nous chercherons des indices favorables à notre thèse dans le domaine gallo-roman, et plus particulièrement en wallon. Nous examinerons enfin le cas d'autres langues comme l'alsacien, le grec archaïque, le suédois ou l'anglais australien qui ont connu ou connaissent une situation comparable.

L'apparition d'un son [y] palatal s'opposant à un [u] vélaire est une caractéristique essentielle des mutations vocaliques de l'ancien français. A la différence de nombreux

---

* Ce chapitre mêle des sources diverses : phonétique et phonologie diachroniques ; phonétique acoustique. L'utilisation systématique des crochets en transcription signifie que le son est vu sous son aspect phonétique.

autres changements, la palatalisation de [u] n'est pas conditionnée par l'entourage consonantique ou vocalique, ou par l'accent, ou même par le type de syllabe. "Tout ū latin, libre ou entravé, portant ou non l'accent, est passé en français du son grave u (celui de notre diphtongue graphique ou) à un son aigu ü (qui a continué à être noté u)" (BOURCIEZ 1967, p. 94). Pour les quelques exceptions comme moule < muscūlum, joute < jūxta, lourd < lūridum, coupe < cūpa, touffe < tūfa, on doit poser, dès le latin vulgaire, une forme en u bref. Mais leur nombre extrêmement réduit ne doit pas masquer le fait que l'immense majorité des formes, y compris les mots d'emprunt récent comme "nature", "juste" ou "lustre", a connu ce développement. Cette évolution s'est produite sur tout le territoire de l'ancienne Gaule, sauf à l'extrême-nord du domaine, en pays liégois (Wallonie), dans le Valais et en franco-provençal (pour une mise au point récente, cf. TUAILLON 1972). En ce qui concerne la datation, on peut noter les points suivants (pour l'ensemble des données, cf. FOUCHE 1958, p. 203-5).

- Si le changement de [u] à [y] avait été complet dès l'origine, le [u] de cūpa (>cuve) aurait sans doute palatalisé la consonne [p].
- Si l'on admet que le catalan s'est séparé du provençal au VIIe siècle, il apparaît que la palatalisation ne pouvait être complète dans le midi à cette date, car le catalan a conservé [u]
- GAMILLSCHEG a noté dès 1917 que l'évolution d'un mot comme pulice > puce par les étapes pults > puuts (vocalisation de l) > puts > pyts montre que l > u antérieurement à la palatalisation, ce qui implique que celle-ci ne peut être antérieure au VIIe siècle.

Comme le fait remarquer STRAKA (1964) citant VON WARTBURG"... dans ce laps de temps, (trois ou quatre siècles avant le VIIe) [u] a très bien pu se déplacer progressivement en direction de [y], avec de longs arrêts sur les stades intermédiaires... comparables à.. l'articulation centrale du russe ou du polonais". Mais dans sa discussion de TUAILLON (1972), il déclare " (j'ai écrit que) le premier stade intermédiaire entre [u] et [y]

pouvait peut-être remonter jusqu'au IIIe siècle, tandis que le résultat final ne peut être antérieur au VIIe ou VIe siècle. Mais aujourd'hui,..., je ne tiens plus tellement à ce stade intermédiaire". TUAILLON (1972) dans son étude sur le franco-provencal, propose le VII ou VIIIe siècle. MEYER-LÜBKE faisait remarquer que la palatalisation avait pu se produire à différentes époques, selon les différentes régions.

L'explication de la palatalisation par un substrat celtique est aujourd'hui abandonnée, tout comme l'hypothèse d'une évolution analogique à partir de [u] palatalisé au contact de [i] (SCHÜRR 1954). De même est abandonnée, sans parler du substrat altaïque (FOUCHE 1952, p. 43-44), une autre suggestion du même auteur concernant l'effet d'une très forte labialisation de [u]. On voit mal, en effet, le rapport entre une forte labialisation et une avancée articulatoire. Ces références mineures sont discutées dans BEYER (1964).

Il reste en présence une explication physiologique et une explication structurale.

III - LA THEORIE PHYSIOLOGIQUE DE G. STRAKA

L'explication physiologique est due à G. STRAKA (cf. parmi ses nombreux travaux, STRAKA 1964). La théorie articulatoire du professeur STRAKA repose sur le concept d'énergie articulatoire. Cette énergie articulatoire est étroitement liée à l'intensité des contractions musculaires nécessaires à l'articulation des différents sons. Pour STRAKA, les voyelles sont articulées essentiellement par les muscles "abaisseurs" de la langue. En effet, une "loi physiologique absolue" régit l'articulation des voyelles et des consonnes : prononcée avec énergie, une voyelle s'ouvre et une consonne se ferme ; prononcée faiblement, une voyelle se ferme et une consonne s'ouvre. Si les voyelles sont articulées principalement à l'aide des muscles abaisseurs, et si l'augmentation de l'énergie articulatoire est liée à une plus grande contraction des muscles, il découle que les voyelles articulées de façon

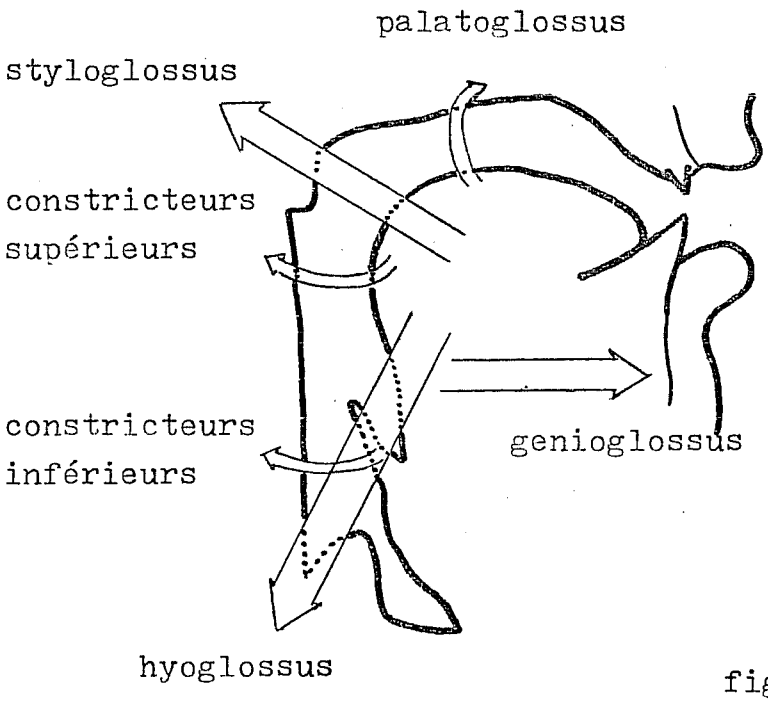


fig. I - IX

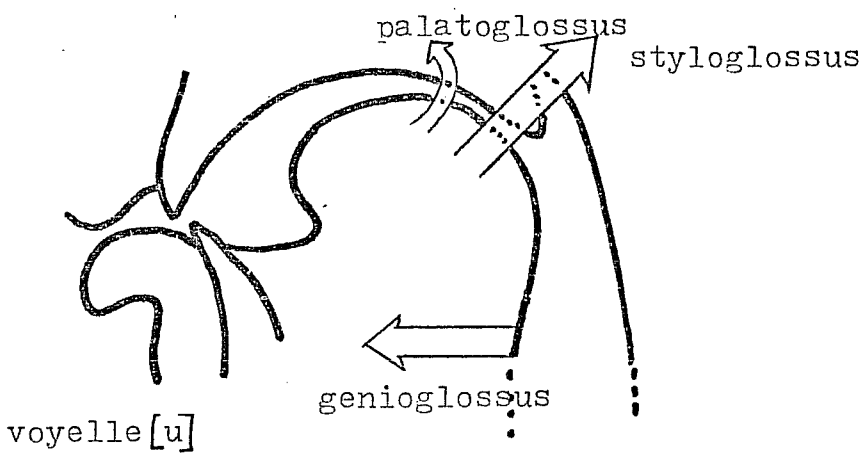
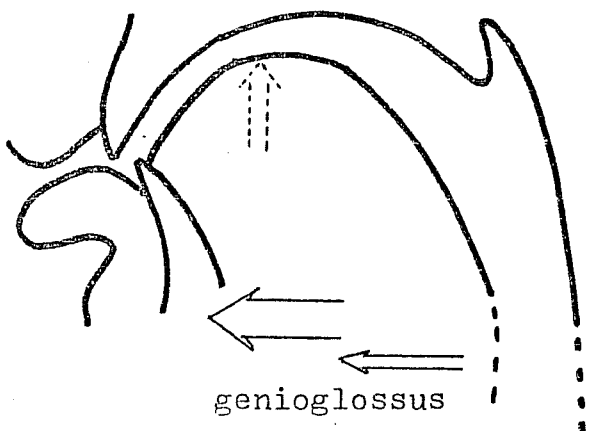


Fig. 2 - IX

effet de "gonflement"



genioglossus
antérieur postérieur

Voyelle [y]

(d'après BRICHLER 1970)

fig. 3 - IX

énergique doivent s'ouvrir. Une intervention secondaire des muscles élévateurs de la langue (génioglossus, palatoglossus et pharyngoglossus pour l'élévation et la postériorisation des voyelles d'arrière) est nécessaire pour expliquer le comportement des voyelles de faible aperture. Les voyelles centrales (par ex. [ɨ]) sont articulées sans intervention des muscles élévateurs ; la voyelle [u] qui est articulatoirement plus centrale que [o] nécessite une moins grande participation des muscles élévateurs postérieurs. Le mouvement articulatoire requis pour [u] demande, par rapport à celui de [o], moins d'énergie musculaire, en particulier moins de contraction des muscles abaisseurs, mais également moins de contraction des muscles élévateurs d'arrière (styloglossus et pharyngoglossus). "Pour u , la langue recule moins et se place davantage dans la région centrale de la cavité" (STRAKA 1964, p. 38). L'explication de la palatalisation de [u] en une voyelle centrale (dans un premier temps) découle immédiatement : cette palatalisation est le reflet d'une diminution de l'énergie articulatoire. "La palatalisation de u suppose sans aucun doute un affaiblissement de l'effort musculaire lingual" (op. cit., p. 39). La moindre contraction des muscles élévateurs d'arrière assure à [u] une position centrale, qui ne nécessite pas, selon STRAKA, d'efforts articulatoires des muscles élévateurs;"..... La labialité diminue de u à y , et cet aspect reflète à son tour un amoindrissement des contractions des muscles labiaux allant de pair avec celui des muscles de la langue" (op. cit., p. 39). Le passage de la voyelle centrale à [y] se fera à une époque ultérieure par une augmentation de l'énergie articulatoire des muscles élévateurs d'avant.

L'évaluation de l'hypothèse de STRAKA est difficile, car les travaux anatomiques récents ne confirment pas la dichotomie des muscles abaisseurs opposés aux muscles élévateurs. Les muscles actifs pour l'articulation des voyelles sont présentés à la fig. 1-XI. Ces renseignements proviennent de travaux utilisant l'electromyographie (EMG). On trouvera un excellent résumé dans WOOD (1977). Il faut y ajouter les références

suivantes : HIROSE et al. 1977a, 1977b ; MIYAWAKI et al. 1976.

Pour la voyelle [u], la position linguale est obtenue à l'aide des muscles suivants (fig. 2-IX d'après BRICHLER 1970, p.102 phase centrale de pu).

- Le genioglossus antérieur ne joue aucun rôle;

- Le genioglossus postérieur provoque l'élargissement de la cavité pharyngale et l'élévation par "gonflement" de la masse de la langue ("bunching");

- Le styloglossus provoque la rétraction et l'élévation de la partie postérieure de la langue;

- Le palatoglossus assure le placement proprement vélaire de l'arrière de la langue. Son activité est variable selon les sujets.

- On notera enfin l'importance de la projection labiale.

Pour la voyelle [y] (fig. 3-IX d'après BRICHLER 1970, p. 153) l'EMG permet de faire les hypothèses suivantes :

- Le genioglossus antérieur provoque le mouvement d'élévation par "gonflement" de la masse de la langue;

- Le genioglossus postérieur joue un rôle d'appoint;

- Il est probable que la musculature intrinsèque de la langue joue un rôle dans le positionnement exact de la partie antérieure de la langue.

Pour la voyelle [ʉ] (cf suédois), les données limitées dont on dispose montrent que l'activité des deux parties du genioglossus n'est que très légèrement inférieure à celle observée pour [y]. Ceci n'est pas anormal dans la mesure où [ʉ] n'est que légèrement plus ouvert que [u] ou [y]. On peut supposer que l'activité du palatoglossus est supérieure, alors que celle du styloglossus doit être inférieure, à ce que l'on observe pour [u].

Si l'affaiblissement articulatoire de STRAKA se traduit par une moindre tension des muscles, on peut faire

l'hypothèse que, pour la voyelle [u], cet affaiblissement se traduira par une diminution de la cavité pharyngale et un affaiblissement de la masse de la langue qui prendra une position légèrement plus antérieure que [o]. Cette dernière voyelle exige en effet un effort des muscles constricteurs du pharynx supérieur. Le mouvement d'ouverture sera plus important que le mouvement d'avancée articulatoire. Il n'existe pas de trace diachronique d'une telle ouverture (confusion [u-o], ou [y-ø] par exemple. Le mouvement de palatalisation exige un accroissement de l'effort musculaire du genioglossus antérieur, le maintien de l'activité du palatoglossus, et la diminution de l'activité du styloglossus.

L'absence de renseignements sur le rôle des muscles intrinsèques de la langue exclut cependant toute affirmation tranchée. Pourtant, il ne semble pas que les conséquences d'un affaiblissement articulatoire sur les muscles mis en jeu pour l'articulation linguale soient de nature à expliquer la palatalisation de la voyelle postérieure. Nous verrons qu'il n'en est pas de même pour l'affaiblissement articulatoire au niveau labial. Les preuves accumulées par STRAKA sont si nombreuses et si cohérentes qu'on ne peut nier l'existence de ce phénomène général. Précisons que nous ne nions pas la réalité de l'affaiblissement articulatoire, mais simplement l'explication causale simple postulée par STRAKA.

IV - L'EXPLICATION STRUCTURALE DE HAUDRICOURT ET JUILLANT

L'explication structurale est due à HAUDRICOURT et JUILLAND (1949). Après avoir critiqué la théorie du substrat celtique, ils établissent que le changement [u > y] ne s'est produit que dans les systèmes possédant 4 degrés d'ouverture. Les exemples examinés incluent le suédois, le grec ancien et un dialecte portugais des Açores. Pour ce dernier, les données fragmentaires n'ont pas été complétées (à notre connaissance), ce qui ne permet aucune discussion satisfaisante. Dans tous ces cas, [u > y] et [o] se ferme en [u]. HAUDRICOURT et JUILLAND concluent que le nombre de voyelles est une condition

nécessaire. Mais l'explication elle-même s'appuie principalement sur des considérations physiologiques. L'espace articulatoire dévolu aux voyelles antérieures serait presque deux fois plus important que celui réservé aux voyelles postérieures (op. cit., p. 110). Ceci résulterait d'une "configuration asymétrique des organes de parole". HAUDRICOURT et JUILLAND se fondent sur le schéma vocalique des voyelles cardinales de D. JONES, qu'ils citent explicitement. La palatalisation s'explique par le fait que la "surface articulatoire" nécessaire à la voyelle [u] deviendrait trop étroite dans le cas de 4 voyelles postérieures. "La seule issue possible est d'élargir cet espace, afin d'assurer entre les différents phonèmes la "distance" nécessaire à la netteté de l'articulation (p. 110)

Or LADEFOGED (1971, p. 68-69) a montré, à partir de ses propres recherches radiographiques, que le point le plus élevé de la langue, c'est-à-dire le critère retenu par JONES, ne se place pas sur un trapèze dont le côté "postérieur" serait deux fois plus petit que le côté "antérieur", mais plutôt sur une ellipse. Sur cette ellipse, on constate que la distance entre [i] et [a] n'est pas supérieure à la distance entre [a] et [u].

Il n'est pas certain d'ailleurs que le critère articulatoire du point le plus élevé de la langue soit un critère utile. Quoi qu'il en soit, la critique de l'explication structurale ne peut s'appuyer uniquement sur des considérations articulatoires. Notre critique la plus fondamentale porte sur l'existence de 4 voyelles postérieures comme condition de la palatalisation spontanée. Nous montrerons sur des exemples que cette condition ne tient pas pour un nombre important de systèmes.

Pour finir, on peut tenter de traduire l'explication d'HAUDRICOURT et JUILLAND en termes perceptifs. La palatalisation naîtrait d'une trop grande proximité des voyelles dans l'espace des timbres. Nous verrons que cette hypothèse est sans doute vérifiée, mais que la présence de 4 voyelles n'est pas une condition nécessaire.

V - POUR UNE EXPLICATION ARTICULO-PERCEPTIVE

Il convient maintenant de présenter notre hypothèse. Celle-ci ne s'applique pas au cas de la palatalisation par "umlaut", due à la proximité d'une voyelle ou d'une semi-voyelle palatale. Dans ce cas, la palatalisation est conditionnée. Pour la palatalisation "spontanée", nous faisons l'hypothèse que la phase d'avancée articulatoire a été précédée d'une phase de délabialisation, due à un affaiblissement articulatoire. Cette délabialisation correspond articulatoirement à une diminution de la projection labiale, c'est-à-dire à un raccourcissement de la longueur du conduit vocal. Elle correspond également à une augmentation de l'aire aux lèvres, due à un accroissement de la séparation verticale et à un écartement des commissures. Le conduit vocal restant identique par ailleurs, l'effet acoustique de cette délabialisation est d'amener la voyelle [u] à une position proche de [o]. Pour éviter cette confusion et pour sauvegarder la " marge de sécurité", le point d'articulation est déplacé en avant. Ce déplacement conduit à une centralisation de la voyelle qui prend un timbre [ɯ] ou [ɨ] selon l'importance de cette avancée. La voyelle [y] n'apparaîtra que si cette langue connaît une phase de relabialisation (renforcement articulatoire). En résumé, l'avancée articulatoire, qui caractérise la palatalisation spontanée de [u], est un effort pour conserver la distinction [u ~ o] mise en péril par la délabialisation. Cette délabialisation touche bien sûr tout le système vocalique. Mais il faut noter que la voyelle [o] est moins labialisée que u . C'est en particulier le cas en français moderne. "... les lèvres sont plus avancées pour [u] (que pour [o])" écrit BRICHLER (1970, p. 95). De plus, on peut compenser l'effet de la délabialisation de [o] par une postériorisation de l'articulation linguale, et un resserrement de l'angle maxillaire. C'est le cas du japonais que nous examinerons pous tard. On comprend que la délabialisation ne touche pratiquement que [u].

On voit que cette théorie doit à STRAKA en ce qui concerne l'affaiblissement articulatoire. Mais nous ne faisons

aucune hypothèse sur l'effet de cet affaiblissement sur l'articulation linguale. On voit également ce que cette théorie doit à HAUDRICOURT et JUILLAND. La palatalisation est vue comme un mécanisme ayant pour but d'éviter la confusion entre [o] et [u], inacceptable pour des raisons de rendement phonologique. La présence d'une voyelle [o] est nécessaire, mais on peut envisager une palatalisation spontanée en l'absence de quatre voyelles d'arrière, c'est-à-dire avec la seule présence d'un "o" moyen, en cas de délabialisation très prononcée. Il est certain, néanmoins, que la présence de quatre voyelles d'arrière, impliquant la présence d'un [o] fermé, rend indispensable une forte avancée articulatoire même en cas de faible délabialisation.

Il nous semble que le point de départ et le point d'arrivée également labialisés en français ont contribué à masquer la possibilité d'une phase de délabialisation. Mais l'exemple du russe, que nous étudierons plus loin est intéressant. La voyelle [ɨ] vient également d'un [u] vélaire.

Notre hypothèse permet de traiter ensemble ces deux types d'avancée articulatoire.

Notre hypothèse de délabialisation sera également renforcée si nous montrons que les consonnes labiales ont subi des transformations à l'époque de la palatalisation spontanée.

Avant de passer à l'examen détaillé des évolutions diachroniques et de la structure synchronique de quelques langues, il nous faut démontrer l'effet de la délabialisation, de façon quantitative, sur la structure formantique de la voyelle fermée d'arrière.

L'analyse de l'anglais contemporain permet une première approche. On sait que l'anglais possède deux voyelles de type u : un [u] dit tendu et son homologue relâché [ʊ], sans que nous prenions parti sur la réalité physiologique qu'impliquent ces termes. PERKELL (1969) a montré (p. 43) que "... la protrusion pour la voyelle [ʊ] est plus faible... on observera la protrusion maximale pour la voyelle [u] ". En ce

qui concerne l'aperture labiale on peut déduire des données présentées page 46 de la monographie citée que l'aperture labiale est inférieure de 3 mm pour [u] par rapport à [ʊ] . Ceci est confirmé par FROMKIN (1964) qui indique pour [u] une aire labiale de 2 à II mm^2.
alors que cette même mesure atteint de 7 à I8 mm^2 pour [ʊ] .
Si l'on néglige la différence entre les configurations linguales, les fréquences formantiques nous donnent une première indication sur l'effet d'une moindre labialisation.
(PETERSON et BARNEY 1952, STEVENS et HOUSE 1963).

exemple	F1	F2	Hz F3	PETERSON & BARNEY : contexte h-d
who'd	300	870	2240	
hood	440	1020	2240	
ball	570	840	2410	
who'd	320	880	2220	STEVENS et HOUSE : contexte h-d
hood	490	1180	2220	
-	320	820	2260	STEVENS et HOUSE : voyelles isolées
-	460	1040	2200	-

Il semble bien que la moindre labialisation de [ʊ] conduit à une notable élévation de F1 et F2. Il se rapproche alors de [ɔ] . Les valeurs pour l'anglais britannique (WELLS 1962) présentent le même phénomène, bien que la différence soit moins marquée, ce qui laisse présager une plus grande similitude entre le degré de labialisation des deux voyelles :

	F1	F2	
ɑ:	300	940	WELLS (1962) voyelles isolées
ʊ	380	950	
o:	450	740	cette voyelle est beaucoup plus fermée que la voyelle américaine correspondante.

Le travail de WOOD (1975) sur les voyelles tendues et relâchées confirme cette analyse. Une simulation à l'aide du L.E.A. (Line Electrical Analogue ; simulateur électrique analogue du conduit vocal) de G. FANT (1960) fait apparaître (p. 125) qu'"un léger relâchement de la constriction labiale

(pour [u]) augmente F1 et F2 de 80 à 90 Hz", ceci pour une longueur et une forme articulatoire constantes.

Ces premiers indices encourageants pour notre hypothèse nous ont conduit à réaliser une simulation sur ordinateur, afin de contrôler les divers paramètres : séparation labiale, projection labiale et mouvement lingual. Nous avons utilisé un programme dérivé de celui de ZERLING (1974), présenté au chapitre VI de ce travail. Ce programme permet le calcul des quatre premiers formants à partir de la forme du conduit vocal décrite par la fonction d'aire.

Il n'existe pas d'étude portant sur les fonctions d'aire des voyelles du français. Aussi avons-nous utilisé le travail de MRAYATI et GUERIN (1976) qui proposent, à partir des données recueillies par FANT (1960) et à la suite de patientes modifications par essais et erreurs, des formes de conduit vocal qui possèdent des fréquences de résonance conformes aux évaluations spectrographiques pour des voyelles françaises naturelles. Les fréquences formantiques calculées par MRAYATI et GUERIN pour la voyelle [u] (fig. 4-IX) s'établissent à 296, 783 et 2286 Hz pour les trois premiers formants. Pour des voyelles naturelles, DEBROCK et FOREZ (1976) obtiennent les moyennes et écart types suivants : 289 (σ =28) ; 856 (σ = 91) ; 2270 (σ = 157). La correspondance est excellente pour F1 et F3. La fréquence de F2 pour des voyelles naturelles est un peu plus élevée que celle obtenue par MRAYATI et GUERIN, (75 Hz), mais l'écart type de 91 Hz indique que cette différence n'est pas significative.

Dans une première étape, nous avons calculé, à l'aide du programme présenté au chapitre VI, les fréquences des 3 premiers formants pour la forme du conduit de la voyelle [u]. Le programme utilisé ne comportait pas de correction pour l'effet de radiation. Ces corrections sont très faibles dans le cas de [u].

	F1	F2	F3	Hz
MRAYATI et GUERIN	296	783	2286	
nos résultats	312	773	2237	
(DEBROCK et FOREZ)	(289	856	2270)	Voyelle naturelle.

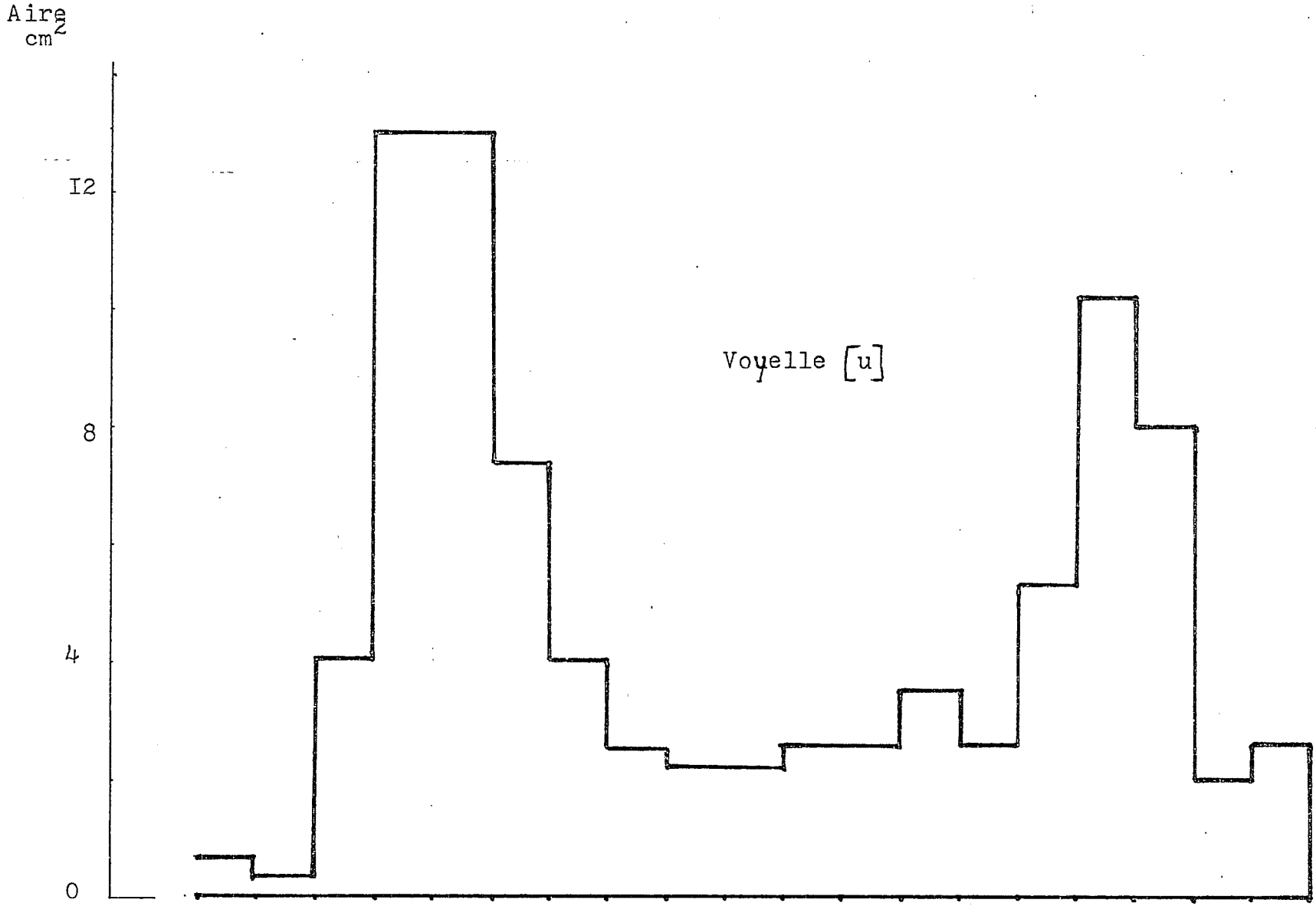


fig. 4 - IX

Nos résultats serviront de valeurs de référence. Chaque modification sera affectée d'un indice numérique. Le conduit non modifié recevra l'indice 0.

Dans une deuxième étape, nous avons étudié l'effet d'une augmentation de l'aperture labiale, qui se traduit par une augmentation de l'aire aux lèvres. ΔFi est la différence entre les fréquences formantiques des conduits modifiés et du conduit non modifié. Les modifications de l'aire aux lèvres ne dépassent pas les 2/3 de la différence entre [u] et [o].

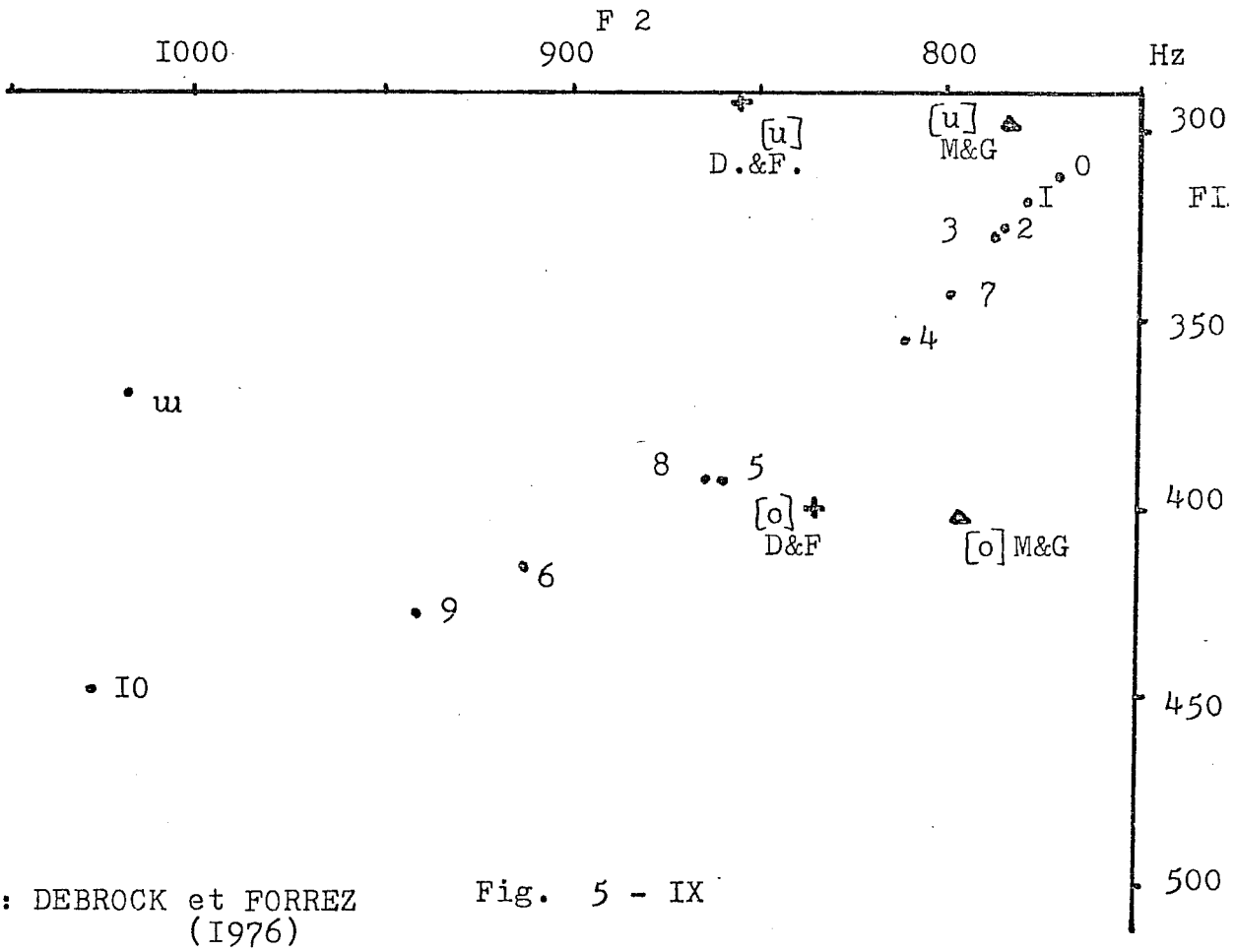
Indice	Aire aux lèvres cm^2	F1	ΔF1	F2	ΔF2	F3	ΔF3
0	0,65	312		773		2237	
1	1,00	321	9	780	7	2238	1
2	1,50	326	14	786	13	2239	2
3	2,00	328	16	788	15	2239	2

On constate que les écarts sont faibles. Ce résultat s'explique par la valeur très faible de l'avant-dernière section (région dentale) : 0,32 cm^2. Si l'on fait varier l'aire des deux dernières sections, on obtient au contraire des variations importantes pour F1 et F2. L'ensemble des résultats de la simulation est présenté à la fig. 5-IX.

Indice	Aire aux lèvres	Avant dernière section	F1	ΔF1	F2	ΔF2	F3	ΔF3
0	0.65	0.32	312		773		2237	
4	1.00	0.60	355	43	811	38	2239	2
5	1.50	1.00	393	81	862	89	2244	7
6	2.00	1.50	418	106	913	140	2248	11

Nous pensons que cette augmentation des deux dernières sections est licite, car il est difficile de savoir si les dents forment ou non une constriction. La valeur extrême de F2 chez MRAYATI et GUERIN est due à l'utilisation d'une aire très faible au niveau dental. Pour les autres voyelles fermées, ils utilisent des valeurs beaucoup plus importantes :

[u] 0.32 [i] 1.40 [y] 1.70 [o] 3.20 [ø] 4.55 cm^2



D&F: DEBROCK et FORREZ (I976)

M&G: MRAYATI et GUERIN (I976)

Fig. 5 - IX

Indice	lg. (cm)	section 19	section 18	section 17	FI	F2	F3 Hz
0	I9	0.65	0.32	4.00	3I2	773	2237
I	I9	I.00	0.32	4.00	32I	780	2238
2	I9	I.50	0.32	4.00	326	786	2239
3	I9	2.00	0.32	4.00	328	788	2239
4	I9	I.00	0.60	4.00	355	8II	2239
5	I9	I.50	I.00	4.00	393	862	2244
6	I9	2.00	I.50	4.00	4I8	9I3	2248
7	I8	-	0.32	4.00	342	799	2239
8	I8	-	0.60	4.00	393	862	2245
9	I8	-	I.00	4.00	428	94I	2253
I0	I8	-	I.50	4.00	450	I020	226I

L'augmentation relative de F1 atteint 34 % pour une aire aux lèvres qui reste bien inférieure à l'aire de la voyelle [o]. La comparaison entre les formants de la voyelle [u] delabialisée et de la voyelle [o] est instructive :

	F1	F2	F3	
[o] DEBROCK et FORREZ	400	833	2232	Hz
[o] MRAYATI et GUERIN	402	797	2417	
5	393	862	2244	
6	418	913	2248	
[o] nos résultats	375	871	2475	

Nos voyelles [u] délabialisées ont un premier formant égal à celui de [o]. Quant à F2, il est légèrement supérieur. La correction de radiation pour F2 dans cette zone, qui est de l'ordre de -80 Hz, rapproche encore la structure formantique des [u] délabialisés de celle de [o]. La conclusion de cette première simulation est claire. La délabialisation de la voyelle a pour effet de placer [u] à proximité immédiate de la voyelle [o] dans le plan F1-F2 (fig. 5-IX).

Jusqu'à présent nous avons considéré la délabialisation comme une simple augmentation de l'aire aux lèvres. Mais il est probable qu'elle implique également une moindre protrusion labiale. Nous avons testé l'effet acoustique dû au raccourcissement du conduit vocal en supprimant une section (1 cm) et en faisant varier l'aire aux lèvres.
Cette modification conduit à des fréquences formantiques proches de celles de la voyelle [ɔ]. Rappelons que l'articulation linguale est toujours celle de la voyelle [u].

	aire aux lèvres (cm²)	F1	ΔF1	F2	ΔF2	F3	ΔF3
7	0.32	342	30	799	26	2239	2
8	0.60	393	81	863	90	2245	8
9	1.00	428	116	942	169	2254	17
10	1.50	450	138	1020	247	2261	24
DEBROCK & FORREZ		514		1162		2545	
MRYATI & GUERIN		540		1007		2480	

L'effet conjugué d'un raccourcissement du conduit vocal et d'une augmentation de l'aire aux lèvres conduit à une augmentation rapide de la fréquence de F1 et à une augmentation plus modérée de F2.

Dans une dernière étape, nous avons voulu vérifier l'effet d'une avancée du point d'articulation lingual. Nous avons modifié arbitrairement la fonction d'aire du conduit comme l'indique la fig. 6-IX. Cette modification correspond à une augmentation légère du volume pharyngal et à une diminution du volume buccal. Pour une forte labialisation, on observe les modifications suivantes des fréquences formantiques :

	F1	Δ F1	F2	Δ F2	F3	Δ F3
[u]	312		773		2237	
[u] modifié	369	57	1020	247	2218	-18

On constate qu'une avancée articulatoire provoque bien une augmentation de F2 (+ 250 Hz) et un faible accroissement de F1 (+ 57 Hz). On a donc montré que l'antériorisation de l'articulation permet d'éviter la confusion de [u] et de [o]. L'ampleur de l'effet acoustique dû à la délabialisation suggère que l'avancée articulatoire sera nécessaire même dans le cas où le système vocalique ne comprend qu'une voyelle postérieure intermédiaire entre [o] et [ɔ]. Ceci est contraire à l'hypothèse d'HAUDRICOURT et JUILLAND.

Notre explication de la palatalisation spontanée de [u] est la suivante :

- Ce phénomène se produit lorsqu'une langue connaît une phase d'affaiblissement articulatoire qui se traduit, en particulier, par une délabialisation.
- La palatalisation, ou plus exactement l'avancée articulatoire, se produit afin d'éviter la confusion auditive entre la voyelle [u] délabialisée et la voyelle postérieure d'aperture supérieure.

Afin d'étayer cette hypothèse, il convient de rechercher les indices de délabialisation dans les langues ou un phénomène de palatisation spontanée a été mis en évidence.

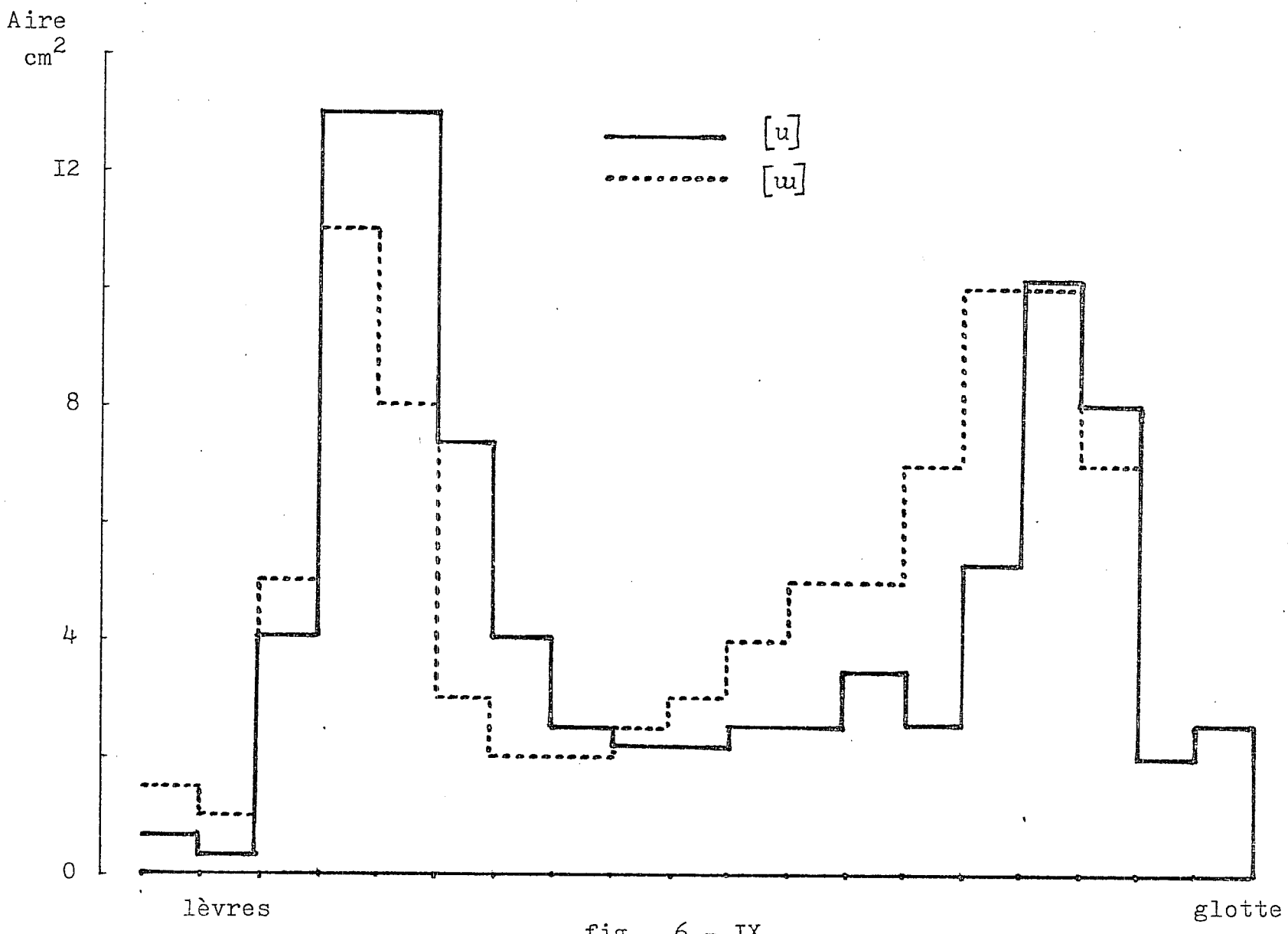


fig. 6 - IX

VI - ANCIEN FRANÇAIS

Nous ne reviendrons pas sur les faits eux-mêmes, que nous avons exposés dans l'introduction de ce chapitre. Mais il nous a paru intéressant, à la lumière de notre hypothèse, d'étudier le développement diachronique des voyelles postérieures. Il ne s'agit que d'un survol rapide. Les datations sont tirées de BOURCIEZ (1967 ; B), STRAKA (1964 ; S) et FOUCHE (1958 ; F). Elles ne sont pas toujours, comme on le sait, concordantes. Nous ne les utiliserons que pour une chronologie relative grossière. Nous négligerons l'évolution de ɑ(:) et les développements conditionnés au contact de [l], [j] et des consonnes nasales, pour ne retenir que les positions toniques libres et entravées.

Situation aux Ier - IIe siècles :

libre	entravée
palatal / non palatal	palatal / non palatal
u:	u
o:	o
ɔ:	ɔ

C'est l'étape de l'allongement des voyelles accentuées libres : S : Ier siècle ; F : déjà commencée au VIe.

Situation à la fin du IIIe - IVe siècles :

u:	ú
o:	o
uo	ɔ

On observe la diphtongaison de [ɔ:] . STRAKA y voit l'effet d'une prononciation énergique. Cette évolution est en effet contraire à ce que notre hypothèse de délabialisation laisserait prévoir. S : fin IIIe - début IVe. F : 2ème moitié du Ve - Ière moitié du VIe siècle.
Au Ve, selon STRAKA, [au] > [o:]. Ce fait, par contre s'explique aisément si l'on postule une faible labialisation pour [u] : [au] > [ao] > [o:].

Situation à la fin du VIe - VIIe siècle

ɯ :	ɯ
ou	o
uo	ɔ

Nous plaçons (arbitrairement) à cette date le début de l'avancée articulatoire de [u:] et [u], sous l'effet d'une délabialisation. Le symbole [ɯ] représente une voyelle centralisée faiblement labialisée, très différente de [y].
S : fin VIe - début VIIe ; F : 2ème moitié VIIIe - IXe siècles. C'est à la même époque que B, F et S placent la diphtongaison de [o:] en [ou]. Cette diphtongaison ne nous paraît pas incompatible avec l'antériorisation de [u]. La labialisation plus marquée de [o] "protège" en quelque sorte le [u] dans la diphtongue et permet sa conservation temporaire aussi bien dans [ou] que dans [uo]. Les graphies de la fin du IXe sont principalement ou ; Cantilène d'Eulalie : bellezour < *bellatiōne ; Jonas : correcious < *corruptiōsu. Mais il est bien connu que les graphies postérieures (XI et XIIe) sont en majorité de type o (et u en Normandie). Doit-on supposer dans ce cas une réduction de [u], deuxième élément de diphtongue à [o] sous l'effet de la délabialisation croissante ? L'évolution a pu être différente selon les régions en raison de degrés divers de délabialisation.

Situation aux IXe - et XIe siècles

ɯ :	ɯ
ɣ :	o
ɯɣ	ɔ

C'est au IXe pour S, ou au XIe pour B et F, que [uo] passe à une articulation palatalisée, que nous avons notée symboliquement par [ɯɣ]. La délabialisation importante n'a pas permis à [o] de conserver une articulation postérieure. On peut supposer une première phase où [uo] passe à [ɯo], puis une antériorisation en [ɯɣ]. Le deuxième élément de la diphtongue n'était sans doute pas très différent de [ə]. A la même époque, l'ancien [ou] a également été palatalisé. Ceci

nous amène à envisager une <u>deuxième palatalisation spontanée</u> de [o:] à [ɣ:], parallèle au passage de [u:] à [ɯ:], après réduction de [ou] à [o:]. Que le passage de [o:] à [ɣ:] soit lié à la délabialisation semble confirmé par la conservation d'une voyelle postérieure devant une consonne labiale. On observe en effet les évolutions suivantes :

cŭbat > couve ; lŭpa > louve ; Lŭpare > Louvre ; rōbur > rouvre ; scrōba > écrou ; et d'après F, *doga > *douja > *douwa > douve

D'autre part, comme nous le verrons plus loin, le wallon (mais pas le lorrain) a connu cette deuxième palatalisation, alors qu'il n'a pas connu la première. Une analyse complète exigerait une analyse structurale de l'ensemble des systèmes vocaliques des différents dialectes, ce qui dépasse largement le cadre de ce travail. Mais nous émettons néanmoins l'hypothèse que cette deuxième palatalisation a été facilitée par la présence d'une voyelle centrale de type [ɸ] ou [ə] qui serait issue de ē (>wa) (HAUDRICOURT 1947). Il pourrait s'agir de la "capture" d'une réalisation assez antérieure de [o:] par une voyelle centrale préexistante. Quoi qu'il en soit, il y a matière, nous semble-t-il, à une réanalyse structurale de ces mutations.

<u>Situation aux XII-XIIIe siècles</u>

y:	y	u
ɸ:		
yɸ ~ ɸ:		ɔ

Cette époque est marquée, selon S, par un renforcement articulatoire, qui pourrait expliquer une labialisation plus intense, et par conséquent, la naissance de véritables palatales arrondies : [y], [ɸ]. C'est également à cette époque que [o] passe à [u].

Le retard entre l'avancée articulatoire de [u] et la fermeture de [o] doit être fixé selon nous à près de 400 ans. Il s'explique d'ailleurs assez bien dans le cadre de notre hypothèse. Tout d'abord, la palatalisation de [u] n'est complète qu'au XII-XIIIème. La fermeture n'a pu se produire

que lorsque la voyelle centrale faiblement labialisée est passée à une position antérieure arrondie. Mais le passage de [o] à [u] implique une labialisation marquée, incompatible avec la phase de délabialisation des siècles antérieurs.

Nous sommes conscients que cette brève exposition ne rend pas compte de la complexité des faits. Notre but principal était de placer notre hypothèse dans une perspective plus large. Nous avons voulu indiquer un certain nombre de questions posées par cette nouvelle explication.

Pour l'examen du domaine consonantique nous renvoyons le lecteur au paragraphe suivant, traitant du dialecte wallon oriental. L'examen des mutations consonantiques spécifiques de ce dialecte met en évidence, à contrario, la faible résistance des consonnes ayant une composante articulatoire labiale en ancien français.

VII - WALLON

On sait que la Wallonie orientale (région de Liège) n'a pas connu la palatalisation spontanée (archaïsme latéral). Les faits wallons sont bien connus par les travaux de REMACLE (1948, 1953, 1972). Nous ne donnerons que trois exemples en conservant les notations de REMACLE (1948) (ou = [u] , u = [y] v̂ = v :)

latin	perdutu >	pièrdou	français	perdu
	nūdu >	nou		nu
	cūpa >	couve		cuve

Il nous a paru intéressant de vérifier si l'évolution des voyelles postérieures du wallon pouvait s'expliquer, à contrario, par le maintien d'une forte labialisation. Mais les arguments les plus importants nous seront donnés par l'étude du système consonantique. L'évolution diachronique des voyelles postérieures peut se résumer ainsi, en dehors des conditionnements contextuels :

Syllabe ouverte	Syllabe fermée		Syllabe ouverte	Syllabe fermée
u: 1	u 1	>	u 1	u 1
o: 2	o 3		ø 2	o 3
ɔ: 4	ɔ 5		u 4	wɛ 5

Le chiffre suivant chaque voyelle renvoie au paragraphe correspondant.

1) Nous faisons l'hypothèse que la forte labialisation a permis le maintien de [u]. Il n'y a palatalisation qu'au contact de [r] et [s], qui sont des consonnes articulées dans la partie antérieure de la cavité buccale : dūru > deûr fr. dur

On retrouve une voyelle postérieure, mais ouverte devant [n] et [m] ūnu > on [ɔ̃] fr. un

La présence d'une voyelle atone provenant d'un a final latin a exercé une action ouvrante. Une semi-consonne labio-vélaire [w] s'est intercalée. On peut considérer que son rôle ancien était de réduire le hiatus. Mais [w] peut également être la trace d'une diphtongaison postulée par quelques auteurs (par ex. FERGUSON 1976). La palatalisation devant [r] et les phénomènes d'ouverture laissent à penser que la palatalisation spontanée n'a été évitée que de justesse.

nūda >	nowe	fr.	nue
	towe		tue
	rumowe		remue

Les seuls [y] originaux de ce dialecte correspondent au français [ɥi] .

2) La palatalisation de la voyelle [o:] est un "problème énigmatique" selon REMACLE (1948). Nous pensons qu'il faut faire l'hypothèse d'une deuxième phase de palatalisation spontanée, comme nous l'avons indiqué dans le paragraphe consacré à l'ancien français.

	flōres >	fleûr	fr. fleur
suffixe	-ōsu >	-eû	fr. eux

Une réponse précise exigerait une analyse structurale de l'ensemble du système vocalique. On notera que $\bar{e}$ libre était certainement déjà passé à [φ(ʃ)] au XIIIème (graphie oi), ce qui rend moins improbable l'idée d'une capture. L'examen des faits lorrains, où cette palatalisation ne s'est pas produite, serait d'un intérêt capital.

3) Pour la voyelle [o] en syllabe fermée, le résultat est le même qu'en français. Pourtant on ne trouve pas la fermeture vers [u]. Ceci semble confirmer que la fermeture de [o] en [u] est liée à la palatalisation en français. La persistance du [u] en wallon est la cause probable de cette absence de fermeture.

mŭsca > mohe	fr.	mouche
* totta > tote		toute

4) La voyelle [ɔ:] en syllabe ouverte reste postérieure.

cŏr > coûr	fr.	coeur
bŏve > boû		boeuf
Mŏsa > Moûse		Meuse

La diphtongue [uo] n'a donc pas été palatalisée. L'accent tombait, selon REMACLE, sur la première syllabe. On peut y voir la cause de cette conservation. Le traitement de cette voyelle est donc identique à celui de [u].

5) Pour [ɔ] en syllabe fermée, il n'y a palatalisation que devant [r] et [s]. On notera la ressemblance avec l'évolution de [u].

porta > pwète	fr.	porte
costa > cwisse, cwasse	fr.	cuisse

La cohérence interne des évolutions exclut l'hypothèse d'un retour de l'ancien [u] à une position postérieure après une phase de palatalisation, comme l'envisage BEYER (1964). Mais les évolutions vocaliques ne fournissent pas d'indices déterminants pour notre thèse.

Le domaine consonantique, au contraire, fournit de nombreux exemples de résistance des consonnes ayant une composante articulatoire labiale.

1) La forte labialisation pourrait expliquer la persistance du [w] germanique.

germ.	wardon > wârder	fr.	garder
	* want > want		gant

Notons que plusieurs de ces indices consonantiques sont communs dans une zone géographique beaucoup plus large. Ceci permet de penser que beaucoup de dialectes n'ont connu la palatalisation spontanée que tardivement, comme nous essaierons de le montrer sur un dialecte lorrain.

2) On observe la conservation de l'élément labial dans les groupes [kw] et [gw]

quando > cwand [kwã]	fr.	quand
lingua > linwe [lẽ:w]		langue

3) Il y a persistance d'une occlusive labiale dans le groupe latin [pj]. Cette persistance pourrait être la preuve de la fermeté de l'occlusion labiale en wallon.

sapiat > sèpe	fr.	sache
* hapja > hèpe		hache
appropiare > aprépi		approcher

- il y a conservation d'une labio-dentale dans les groupes latins [bj] et [vj]

gobione > govion	fr.	goujon
* rubeola > rovioûle		rougeole
cavea > tchève		cage

4) Le traitement du groupe latin ab(u)la > âv, ôl serait une trace précieuse de la forte labialité du [u]. Celui-ci aurait entraîné l'évolution b (ou β^{w}) > w > v.

tabula > *tawla > tôle	fr.	table
tâve [ta:f]	(liégeois)	

Cette évolution n'est cependant pas certaine au vu de
ab(i)le > âve [a:f], ôve fr -able
ébrave [ebrɑ:F] = fr. ouvrable

Si on ne recourt pas à l'analogie, il faut supposer une évolution directe [b > bw > w]. Le dialecte lorrain que nous étudierons plus loin connaît encore des formes en [bw].
-ib(u)l a connu un traitement identique : [b(u) > bwu > wu > u]

tribulare > troûler, trûler

5) L'absence de consonne intercalaire [b] dans l'exemple
sim(u)lare > sonler fr. sembler

pourrait indiquer une forte labialisation de [m] au contact de [u], avec le [i] précédent passant de ce fait à [u]; [m > n] au contact de [l], puis ouverture [u] > [o] au contact de la nasale.

6) Pour les formes de l'imparfait (et ceci est vrai dans tout le Nord-Est) les flexions -abam ou -ebam donnent régulièrement -ève ou -ive, ce qui implique une forte résistance de la labiale qui s'amuit totalement en français.

7) La flexion -ve des passés définis en -ui montre sans doute la trace d'un ancien [w], conservé sous la forme d'un [v].

volui > vôve fr. voulus
* venui > vinve vins

En résumé, un grand nombre d'évolutions observées en wallon oriental s'expliquent par une résistance importante (par rapport au français) des occlusives labiales et de la semi-consonne labio-vélaire [w].

Les évolutions consonantiques du wallon oriental révèlent donc des indices importants, favorables à notre hypothèse d'une liaison directe entre le degré de labialisation et la palatalisation spontanée.

VII - LORRAIN

En dehors des faits wallons, les patois lorrains orientaux, moins connus, sont également des sources intéressantes pour l'étude de la palatalisation. En réponse à une question sur l'étendue de l'aire de non-palatalisation, TUAILLON (1972) précise que l'A.L.F. ne donne qu'une attestation de "lona" (=lune) pour toute la Lorraine. Mais l'examen de la monographie de G. AUB-BÜSCHER (1962) sur le parler de Ranrupt (Bas-Rhin), à quelques kilomètres de la frontière linguistique au niveau de Sélestat, ainsi qu'une enquête personnelle à Charbes (commune de Lalaye Bas-Rhin), à 4,5 km au Sud-Est de Ranrupt, chez un témoin née en 1912, nous permet d'affirmer qu'il existe des traces beaucoup plus conséquentes de non-palatalisation.

On notera tout d'abord, dans le lexique d'AUB-BÜSCHER et dans le parler de notre témoin, les formes suivantes que nous éliminerons :

	Ranrupt	Charbes	
1-	[(l')aluat]	[(La)Lyɛt]	la luette ; confusion avec alouette ?
2-	[bosu]	[bɔsy]	bossu ; erreur typographique ?
3-	[ckurã]	[XKyRã]	écureuil- analogie avec vb. courrir ? (c = IPA [χ])

Les formes suivantes pourraient indiquer une pénétration des formes françaises dans le parler de notre témoin, qui a 24 ans de moins que le plus vieux témoin d'AUB-BÜSCHER.

4-	[muzẽ]	[myze]	museau
5-	[muzlẽr]	[myzLjɛ:R]	muselière
6-	[muzat]		musaraigne ; Forme inconnue

chez notre témoin qui utilise [Rpɛt(χpɛt?)mus]. La forme est douteuse en raison de l'allemand "maus". Mais la forme alsacienne devrait être palatalisée.

Nous ne connaissons pas la forme réellement utilisée à la frontière linguistique (3 km de Charbes ; 0,5 km de Lalaye).

Par contre les formes suivantes sont bien attestées.

7- [punèl] [pumɛl] : prunelle
8- [pun ã] [pumi] : prunier
9- [kruk] [KRuK] : cruche
10- [ctrãbuehi] [tRãbuʃi] : trébucher

On notera que les formes 4, 5, 6, 7, 8 et 10 comportent une consonne labiale devant le [u] maintenu. Pour la forme 9, la voyelle est encadrée de deux vélaires.

Les deux formes suivantes pourraient être des hypercorrectismes :

11- [pysĩ] [pysĩ] : poussin
12- [vãtyzé ?] [vãtuze] : "ventouser" = poser des ventouses.

Quant à [robic] "robuste", il pourrait s'agir d'un emprunt à l'époque où le patois ne connaissait (ou ne pouvait) prononcer un [y] devant une consonne labiale.

On trouve également de nombreux exemples où, à [y] français, correspond un [o^u] ou un [o^w] chez AUB-BÜSCHER. A Charbes, on a plutôt un [o] très fermé avec labialisation très importante. Ces voyelles correspondent souvent à un [u] en hiatus avec une ancienne voyelle atone < a latin. On retrouve ici une caractéristique du wallon. Cette prononciation est en concurrence avec la forme [y] à Charbes ; elle pourrait être sentie comme un archaïsme.

[bro^u] [bRo] : verrue
[earo^u] [ʃaRo] : charrue
[ro^u] [Ro^(w)] : rue (dans une ville comme Strasbourg - [Ry] est le terme commun pour "ruisseau")
[kro^u] [KRo] : (viande) crue
[to^u] [to^(w)] : (je) tue
[pjo^u] [pjo] : pluie (mais parapluie = [paRaply])
[ɛmo^uwè] [ɛRmowɛ] : remuer
[molo^w] : moulue
[mudo^w] : mordue
[pedo^w] [pedo^(w)] : perdue
[vno^u] : venue
[bivno^w] : bienvenue
[ckuro^u] : secouée

Une des caractéristiques principales du parler de Ranrupt et de Charbes est la forte labialisation des consonnes labiales. Cette forte labialisation explique sans doute l'apparition d'un "glide" labiovélaire après [ω p , b] et [m], devant les voyelles [ɑ] et [o]

[pwɑə]	: poche	[bwɑrp]	barbes(épi)	[mwɑrəã]	marchand
[pwɑrã]	: parent	[bwõ]	bon	[mwɑrèn]	marraine
[pwɑrmi]	: parmi	[bwõnur]	bonheur	[mwõt]	morde,morte
[pwõ]	: porc				

Les formes [grɑwus] pour "écrevisse" ou [ʒɑybi] pour "gibier" s'expliquent peut-être de la même manière.

ō accentué donne régulièrement [u] en particulier dans le suffixe -atōre > -eur ou -ōsu > eux en français

[saju] = scieur, [hãtu] = honteux, [ūr] = heure, [nu] = noeud, [Ko] = queue

l'influence du français fait rapidement disparaître ces anciens suffixes. Les faits lorrains s'opposent ici nettement à la palatalisation du wallon. Une étude diachronique structurale, à partir des matériaux recueillis pour l'atlas de la Lorraine romane, serait la bienvenue pour éclairer les causes des ressemblances et des divergences entre les faits lorrains et wallons. Cette étude devrait en particulier étudier les conséquences de l'hypothèse de la centralisation de ẹ (qui passerait à [ə]) (HAUDRICOURT 1947) et de la "capture" de ọ.

L'étude de ces quelques faits lorrains confirme le rapport existant entre la palatalisation de [u] et les problèmes de labialisation. L'entourage labial a freiné la palatalisation, ce qui confirme l'hypothèse d'une délabialisation comme cause première de la palatalisation.

IX - FRANCO-PROVENÇAL

Le problème de la palatalisation spontanée en franco-provençal a fait l'objet récemment d'une remarquable étude (TUAILLON 1972). La conservation des vélaires [u] apparaît bien

comme un conservatisme et non comme une régression. L'examen détaillé de la région du Bugey confirme que la palatalisation est conditionnée, alors que la conservation du timbre vélaire est le traitement général. TUAILLON montre également de façon convaincante que l'hypothèse du substrat celtique est intenable. Les zones de conservatisme absolu sont entourées d'une zone beaucoup plus large où le timbre vélaire ne se conserve qu'au contact d'une consonne nasale. Mais le travail le plus intéressant de notre point de vue est une étude déjà ancienne de A. DURAFFOUR (1927) sur l'introduction de la voyelle [y] dans le village de Vaux (Ain) à la lisière de la zone conservatrice du Bugey, au début du XXe siècle. DURAFFOUR note tout d'abord que "... l'inertie... des lèvres est un des faits qui frappent le plus l'observateur dans l'articulation des parlers de la région lyonnaise" (p. 77). L'innovation s'est produite d'abord dans la population féminine. Elle s'observe d'abord dans quelques mots d'emprunt :"abus", "salut". Par contre les suffixes français -ure et -ule sont prononcés avec un [i]. D'après DURAFFOUR, l'introduction de la palatalisation, qui est un phénomène non conditionné, a été facilitée par "... la création d'une semi-voyelle [ɥ] à côté d'un [w]" au contact d'un [i] ou d'un [e] (p. 78).

[tɥe]	: tué	[sãsɥi]	: sangue
[adɥire]	: apporter	[Krɥi]	: croix
[sãɥi]	: charrue	[mɥi]	: noix

Mais, pour notre hypothèse, le fait capital est que "... toutes les fois que [w] était précédé d'une consonne labiale, l'inertie naturelle de l'organe (les lèvres) a été un obstacle à l'évolution en [ɥ]".

[fwīre]	: fuir
[zarbwi]	: (clématite du lat, albucium)
[pwi]	: puits

On note également un autre phénomène conditionné : une ancienne diphtongue [jau] (ou [jaw]) a donné [jy].

[mjy] : mieux ; [vjy] : vieux ; [pjy] : pou.

Une dernière remarque (p. 79) est d'un intérêt capital. La palatalisation [u] > [y] s'accompagne d'une palatalisation de [o] en [ø]. C'est également une innovation féminine.
[ʃø] : chaud ; [Fø] : (il)faut ; [Vø] : "Vaux" ;
[øtø] : auto(mobile).
Cette évolution donne quelque vraisemblance à l'hypothèse d'une deuxième palatalisation spontanée en ancien français. Les faits wallons et lorrains que nous étudierons plus loin sont également intéressants de ce point de vue. En conclusion, on voit que les faits franco-provençaux confirment l'hypothèse d'une liaison entre le degré de labialisation et la palatalisation.

X - ALSACIEN

La palatalisation spontanée de l'alsacien (et du badois) est bien connue grâce aux travaux de BEYER (1964) et de PHILIPP (1965, 1968). Celle-ci n'est pas mentionnée par HAUDRICOURT et JUILLAND qui n'y voyaient sans doute qu'un sous-produit de la palatalisation romane. La palatalisation alsacienne est à distinguer soigneusement de l'apparition antérieure de voyelles palatalisées arrondies issues d'anciennes vélaires à la suite d'un "umlaut" par i. Ces voyelles palatalisées (ü, ö, oe, öu, üe : notation des germanistes), phonologisées par la réduction des voyelles atones à /e/, n'avaient qu'un faible rendement phonologique, mais jouaient un rôle capital dans la morphologie.

Selon BEYER (1964), cette mutation a touché :

a) la voyelle vélaire fermée longue du moyen-haut allemand (MHA), notée û par les germanistes.

MHA	All. mod. (graphie)	alsacien	traduction
bûr	bauer	[by : r]	paysan
hûs	haus	[hy : s]	maison
mûl	maul	[my : l]	gueule
dûme	daumen	[dy : mə]	pouce
brûchen	brauchen	[bry : xə]	avoir besoin
snûven	schnauben	[ʃny : fə]	souffler

"La palatalisation de û apparait, en règle générale, comme n'ayant été liée ni à la position de cette voyelle, ni à son entourage phonétique" (BEYER 1964, p. 26). La voyelle de l'alsacien moderne peut être brève ou longue selon la nature "faible" ou "forte" de la consonne suivante, et ceci de façon variable selon les dialectes. Les radiographies et photographies publiées (BEYER, fascicule annexe p. 42-47) montrent clairement que le [y] alsacien est faiblement labialisé ("paresse labiale"). Il serait également moins antérieur et plus ouvert que le son correspondant du français et de l'allemand. Il existe une zone en pays de Bade où le représentant du MHA û est articulé comme une voyelle centrale. Un processus de régression apparaît nettement dans tout le pays de Bade. Les villages isolés se montrent conservateurs. A l'ouest du domaine alsacien (vallée de la Thur), 3 îlots conservateurs résistent encore à la palatalisation, et conservent des timbres vélaires.

b) certains u brefs (notés u par les germanistes ; API[ʊ]) lorsque cette voyelle s'est fermée devant une consonne "faible".

Jude	Jude	[jyt]	juif

c) L'élément initial de l'ancienne diphtongue uo. Les résultats actuels sont nombreux : [yːə, øːə, yː, øː]
PHILLIP (1964) suppose une monophtongaison préalable en [ʊː]

vuoʒ	fuss	[fyː (ə)s]	pied

d) La semi-consonne w, deuxième élément de diphtongue, ou provenant d'un g postvocalique relâché.

boum	baum	bøʏm	arbre
wagen	wagen	vøːʏə	voiture

Il faut noter l'absence de toute palatalisation pour ô et â. Pour BEYER, la palatalisation n'a eu lieu que pour les voyelles et les semi-consonnes les plus fermées.

Les emprunts anciens au français (BEYER 1964, p. 189-196) montrent très souvent une palatalisation de la voyelle vélaire. Ceci est un témoignage important de la force du phénomène. A l'époque de l'emprunt, une voyelle d'arrière fermée était impossible à prononcer.

boutique > bydiK tour > tyːr couvert(ure) > gyvert
cartouche > Kårdyʃ soupape > sybåp soutane > sydåm
bouteille > budal bouquet > buge boule > bul
å = ɔ gourmand > gurmã

Les quatre derniers mots sont des emprunts plus récents qui ne connaissent pas la palatalisation spontanée.

Les emprunts anciens de mots français comportant la voyelle [y] montrent souvent un passage à [i]. Plusieurs hypothèses sont possibles : Le [y] français plus antérieur et plus fermé n'aurait pas été confondu avec la voyelle plus centrale et délabialisée de l'alsacien ; Les emprunts auraient été faits par l'intermédiaire de dialectes alsaciens ayant palatalisé tardivement [u] ; Les emprunts auraient été faits à des dialectes romans ayant connu une palatalisation tardive.

buffet > biFɛ luzerne > lisɛrm perruque > bårek

les emprunts récents conservent le [y] :

gibus > ʃibys jus > ʃy salut > såly

quelques [o] et [ɔ] français sont rendus en alsacien par un [v]. Ceci souligne le caractère particulièrement ouvert du [v] alsacien. La plupart des dialectes alsaciens ne connaissent pas le [u] fermé.

fauteuil > Fvdɛl hotte > hvt coffre > Kʰv Fr̥
côtelette > Kʰvdlɛt

L'explication de la palatalisation par une "marge de sécurité" trop étroite est envisagée avec prudence par M. PHILIPP. Mais nous pensons que notre explication fournit une vue plus cohérente des évolutions diachroniques.

Il se pose de délicats problèmes de chronologie. Nous pensons que beaucoup d'évolutions, ayant une origine commune, ont pu se produire à la même époque, à des rythmes certes différents. Les informations suivantes ont été puisées dans BEYER (1964), PHILIPP (1965, 1968).

- On observe tout d'abord une délabialisation des anciennes palatales arrondies (provenant d'un "umlaut" par i). Cette délabialisation, qui précède de peu la palatalisation spontanée, est un des indices les plus importants que nous connaissions en faveur de notre hypothèse. Cette délabialisation entraîne la confusion des voyelles [y] et [i], ainsi que [ø] et [e] .

- On peut voir dans l'ouverture des brèves fermées [i] > [I] , [u] > [ʊ] la conséquence d'un affaiblissement articulatoire. L'ouverture de [u] s'explique facilement par des considérations acoustiques.

- On observe également une monophtongaison de la diphtongue uo. Pour l'alsacien de Strasbourg (PHILIPP 1968), l'évolution semble être en [ʊ:] . Dans la plupart des autres dialectes, on doit supposer une transformation en [uə] . La différence est importante, car elle suggère un système à 4 voyelles postérieures longues en strasbourgeois, mais un système à 3 voyelles postérieures seulement dans d'autres dialectes. L'évolution du strasbourgeois est compatible avec l'hypothèse d'une faible labialisation provoquant une ouverture de la première partie de la diphtongue. Pour les autres dialectes, il faut sans doute y voir une délabialisation plus radicale du 2ème élément.

- L'hypothèse d'une chaîne de pression repose exclusivement sur la mutation suivante : â > ô . BEYER (1964) conteste cette possibilité "...Ce mouvement, en admettant même qu'il se soit produit avant la palatalisation, ce qui est peu probable, n'a pu exercer aucune pression sur la voyelle extrême, car u a toujours conservé intégralement son champ d'articulation (la fermeture de [o] n'est jamais allée au-delà de [u]). L'idée d'une chaîne de pression ne saurait

donc être retenue (p. 237)". Notons que dans les dialectes non-strasbourgeois, la présence de 3 voyelles d'arrière et la coalescence de â et ô n'expliquent en aucune manière la poursuite de la fermeture vers [u]. Celle-ci ne se comprend qu'à la suite de la palatalisation, afin d'occuper l'espace laissé libre. La faible labialisation est à nouveau évidente dans le fait que la voyelle extrême ne dépasse pas le stade [ʊ]. Pour les dialectes alsaciens qui comportent 4 voyelles postérieures, les deux voyelles [ɑ] et [ʊ], ce qui représente une différence essentielle avec l'hypothèse de la chaîne de pression d'HAUDRICOURT et JUILLAND, passent à [y] et [Y].

- Dans le domaine consonantique, sans entrer dans le labyrinthe des évolutions complexes, on peut noter un relâchement de [b] intervocalique en une spirante bilabiale, ainsi que la délabialisation de [w] dans un grand nombre de contextes. Celui-ci passera finalement à [v].

On constate que l'évolution de l'alsacien fournit des indices importants qui confortent notre hypothèse.

XI - GREC ANCIEN

En ce qui concerne le grec ancien, rappelons d'abord l'analyse d'HAUDRICOURT et JUILLAND (1949, p. 107-108), empruntée pour les faits à LEJEUNE (1946). Les voyelles brèves et longues se présentent de la manière suivante pour le grec archaïque :

brèves		longues	
ι	υ	ῑ	ῡ
ε	ο	η	ω
α		ᾱ	

Il n'y a que trois degrés d'aperture pour les voyelles vélaires. Mais dès l'apparition des premiers textes, un signe différent de η est employé en corinthien et en attique pour noter un nouveau phonème, contraction de e + e, ou résultat d'allongements compensatoires. C'est un [e] très fermé, intermédiaire entre ῑ et η . A la même époque, un signe

différent de ω sert à noter un nouveau phonème, résultat de la contraction de o + o ou d'allongements combinatoires. Ce nouveau phonème est un [o] très fermé, intermédiaire entre ῡ et ω . Aux VI-VIIe siècles, le système des longues est donc le suivant :

ῑ	ῡ
ει	ου
η	ω
ᾱ	

A l'époque classique, nous obtenons :

ῑ	ῦ	ῡ		i:	y:	u:
ει			c'est-à-dire	e:		
η		ω		ε:		o:
	ᾱ				a:	

L'articulation de ῡ s'est déplacée vers l'avant pour aboutir à ῦ . De même ου s'est fermé en ῡ . COLLINGE (1961) s'est élevé contre le concept de symétrie phonétique ou phonologique à propos du grec ancien. Mais sa conception de la symétrie est trop formelle. Il considère en effet le schéma vocalique du grec classique comme asymétrique (même si ǫ passe ǫ) parce qu'il possède quatre voyelles d'avant et trois voyelles d'arrière. Mais c'est justement l'argumentation d'HAUDRICOURT et JUILLAND qu'une structure à 4 voyelles d'avant et trois voyelles d'arrière est plus stable, même si elle est en apparence moins symétrique, qu'une structure ayant quatre voyelles d'avant et d'arrière. Il insiste par contre, à juste titre, sur un point important. La fermeture de ŏ en ŭ n'a sans doute jamais eu lieu, ce qui donne pour les voyelles brèves un schéma sans [u] qui n'est pas très fréquent dans les systèmes vocaliques. Ce dernier point, ainsi que l'ensemble des faits exposés plus bas, nous amènent à penser que le grec n'a jamais connu de voyelle [y] (comme le français ou l'allemand moderne) mais plutôt une voyelle d'arrière délabialisée de type [ɯ] pour la brève et une voyelle centrale délabialisée [ɨ] pour la longue. Ceci expliquerait tout d'abord que ŏ ne passe pas à ŭ.

Notons dès à présent que les passages rapides de [y:] ou [ɨ:] à [i:] et de [ɯ] à [i] présupposent une faible labialisation. L'important travail de S.T. TEODORSSON (1974) apporte des précisions phonétiques et chronologiques importantes. Ce travail est fondé sur l'analyse des erreurs graphiques que l'on peut déceler dans l'ensemble du corpus des inscriptions grecques des origines à 300 av. J.C. A l'aide d'une méthode rigoureuse d'analyse de la validité des exemples, l'auteur confirme le passage [ou] > [o:] > [u:] dès 600. Ce passage peut paraître curieux alors que [u] > [ɨ:] . On peut émettre l'hypothèse que [ou] passe à [ow] afin de contrecarrer l'effet de la délabialisation et pour maintenir un écart suffisant entre les deux voyelles. L'effet moins marqué de la délabialisation sur [o] protège d'ailleurs la diphtongue. [u:] > [y:] très tôt, certainement avant le milieu du VIe siècle. Les traces graphiques sont peu nombreuses. Ce n'est que grâce à la présence de Ϙ (koppa) dans les plus anciens documents que le caractère palatal de Υ graphique est confirmé. On note également la persistance d'une graphie béotienne ΟΥ correspondant à l'attique-ionique Υ après l'introduction de l'alphabet ionique vers 400-410 av. J.C. On peut poser comme "terminus post quem" de cette évolution (autour de 700) le moment où ΕΥ et ΑΥ sont utilisés pour noter [eu] et [au].

Alors qu'ALLEN (1968, p. 65) place le passage de [y] à [i] vers le 2ème ou 3ème siècle ap. J.C. au vu de papyrus égyptiens, en n'y voyant d'ailleurs qu'une particularité régionale, l'analyse de TEODORSSON montre de nombreuses utilisations de Ι pour Υ dès le début du VIe av. J.C. Au voisinage des consonnes sonantes, le changement semble terminé dès avant 450 et pour les autres contextes vers 350. En outre, le passage graphique de ΥΙ à Υ avant le VIe siècle témoigne de la grande similitude entre Υ et Ι. Ceci confirmerait le caractère délabialisé de Υ qui passe rapidement à [i]. On peut encore observer (sans que les indications soient très claires) des graphies Υ pour ΟΙ, sporadiquement, dès

le VIe jusqu'à la première moitié du IVe. Ceci pourrait s'expliquer par un [o] relativement avancé et délabialisé. (cf. API [ɣ] , comme en vietnamien) avec ou sans aide de la palatalisation par [j] au cours de la monophtongaison. La fréquence de ces cas augmente après 350. Plus intéressant est le fait que l'on trouve de nombreuses graphies EB correspondant à [eu] , dès le VIe et en nombre plus important au IVe. [au] est touché, mais à un moindre degré. Ceci suggère une prononciation [eω] qui tend à se délabialiser très tôt en [eΦ] devant une sourde, et [eβ] devant une sonore. Ces graphies sont particulièrement nombreuses devant [s] et les occlusives sourdes. Elles indiquent sans équivoque le caractère rapidement déliabialisé de [u] second élément de diphtongue. Le développement ultérieur en [f] ou [v] confirme cette interprétation. Pourtant, contrairement à de nombreux auteurs (LEJEUNE, LUPAŞ, etc...), qui datent le passage de [b] à [v] vers le IIIe siècle et le passage de [ph] > [f] vers le IVe, TEODORSSON soutient que [b] est encore intact vers 350. Ce fait semble contraire aux phénomènes d'affaiblissement articulatoire de [b] en [β] et de délabialisation que nous venons de voir. TEODORSSON (p. 291) parle d'une "attraction latente de la position centrale libre sur l'ancien u", mais reprend à son compte également l'explication par une "chaîne de pression". [a:] aurait donné [æ:] alors que le [ɑ:] original serait passé à [ɑ:] vélaire. Cette poussée aurait été transmise par [ɔ:] , passée ensuite à [o:] . Il se pose néanmoins un problème de chronologie, de l'aveu même de l'auteur. Ce nouvel [ɑ:] , repoussant le [ɑ:] ancien en [ɑ:] n'est attesté qu'entre le VIe et la fin du IVe. Or on a vu que [u: > ɨ:] certainement avant le milieu du VIe siècle. Pour nous, ce passage de [ɑ:] à [ɑ:] a plutôt été rendu possible par l'avancée de [u:] . Il ne peut en être la cause. En ce qui concerne les voyelles brèves, [u] est également passé à [y], alors qu'il n'y avait pas quatre voyelles d'arrière. A moins de recourir à une explication fondée sur l'analogie de [u:] et de [u], l'explication purement structurale ne tient pas.

Les données du grec ancien nous semblent confirmer l'hypothèse d'une délabialisation due à un affaiblissement articulatoire comme cause première de l'avancée articulatoire de [u], et à nouveau nous trouvons un certain parallélisme entre les évolutions de la voyelle [u] et des consonnes labiales.

XII - ANGLAIS AUSTRALIEN

Les travaux sur l'anglais australien confirment également les résultats de notre simulation. On dispose d'études cinéradiographiques, (BERNARD 1970a), acoustiques (BERNARD 1970b) et auditives (PLATT 1970, PILCH 1971). BERNARD (1970a) écrit : ".. Les spécifications acoustiques pour /u/ montrent que celui-ci est antériorisé d'une façon tout à fait inhabituelle (par rapport à l'anglais britannique standard)". Les valeurs formantiques au point cible sont les suivantes pour les trois sociolectes australiens :

a- "broad" (populaire), b- "general", c- "cultivated" (cultivé)

	F1	F2	F3
a-	365	1660	2405
b-	350	1610	2365
c-	335	1515	2270

cf. fig. 7-IX reproduite de BERNARD (1970b) p. 116, fig. 1.

"Même ces chiffres ne reflètent pas complètement la réalité articulatoire : le tracé de la langue pour /u/ est souvent aussi antérieur que pour la voyelle /ɛ/" (cf. BERNARD 1970a, p. 148, fig. 25, reproduite à la fig. 8-IX). Si la position linguale est identique, l'aperture labiale est plus faible pour /u/. Mais BERNARD (1970a) confirme qu'il n'y a pas de protrusion labiale. Ce fait a été noté par de nombreux observateurs, et on parle couramment de "lip laziness" (paresse labiale) à propos de l'australien.

L'analyse des diphtongues de l'australien et des voyelles de l'anglais néo-zélandais confirme cette caractéristique.

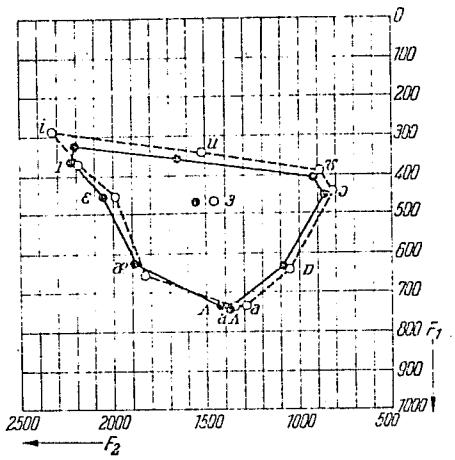


Fig. 7 - IX

(BERNARD 1970 b)

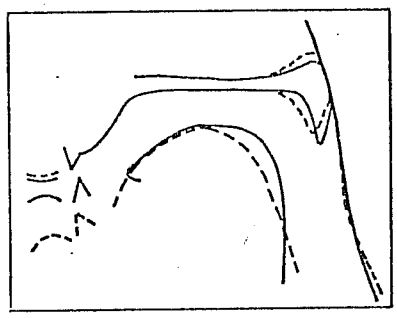


Fig. 5. Subject 6. ——— [u] at target; ------ [ε] at target.

Fig. 8 - IX

(BERNARD 1970 a)

La voyelle /ʊ/ occupe la position la plus postérieure et la plus fermée. Mais on constate une confusion entre la voyelle /ʊ/ et la voyelle /ɔ/ (cf. fig. 7-IX). Cette voyelle /ɔ/ se réalise d'ailleurs à mi-chemin entre un [o] et un [ɔ] français. La seule différence entre /ʊ/ et /ɔ/ est que cette dernière est une voyelle longue alors que /ʊ/ est bref. Le système des voyelles fermées de l'anglais australien est donc le suivant en API (nous interprètons le /u/ antériorisé comme un [ɨ]) :

i ɨ

o o:

La table suivante présente les valeurs formantiques des voyelles /ʊ/ et /ɔ/ pour les trois sociolectes :

/ʊ/	F1	F2	F3		/ɔ/	F1	F2	F3	
a-	405	930	2385	Hz	a-	450	865	2140	Hz
b-	400	910	2390		b-	440	825	2400	
c-	395	890	2290		c-	440	820	2340	

Les descriptions de /u/ dans les études de phonétique auditive sont plus "timorées". PLATT (1970), qui parle néanmoins de palatalisation de /t/ devant des voyelles d'arrière (c'est-à-dire /u/), décrit la voyelle /u/ comme un "glide" d'une position centrale à une position plus postérieure. Parmi les allophones possibles, il indique [ᵊɯ] . PILCH (1971) note que l'opposition entre /i/ et /u/ est neutralisée en /u/ en syllabe prétonique : ex. giraffe - july : [dʒu-] . Il nous semble plus probable que la neutralisation soit en [ɨ].

Il paraît difficile d'expliquer cette palatalisation spontanée de /u/ par une chaîne de pression. /u/ étant une voyelle longue, seules les voyelles longues (et éventuellement les diphtongues) peuvent jouer un rôle. L'anglais britannique standard ne montre aucune trace de palatalisation, du fait de la présence des voyelles [ɑ:, ɔ: et o:] et des diphtongues [ɑʊ, oʊ, ʊə]. Notre hypothèse, par contre, rend parfaitement compte de la situation observée. L'anglais australien connaît, pour des raisons que nous ne pouvons expliquer, une phase de délabialisation qui touche son système vocalique. Les

voyelles postérieures fermées, en se délabialisant, devraient se confondre avec la voyelle /ɔ/ qui est relativement fermée. Cette confusion des timbres n'est admissible que pour la brève [ʊ] qui se distingue encore de /ɔ/ par la durée. Pour éviter la confusion des deux longues, /u/ est "contraint" de se palataliser : une modification de l'articulation linguale est nécessaire pour éviter la confusion. L'exemple de l'anglais australien est donc une illustration parfaite de notre hypothèse. On remarquera aussi que cette palatalisation se réalise presque à l'insu des intéressés, qui n'éprouvent pas de difficultés à converser avec des sujets britanniques ayant conservé un [u] postérieur. Ceci rend plausible l'hypothèse de stades de développement divers en ancien français.

XIII - SLAVE COMMUN

Le slave commun a connu une mutation ū > ï vers l'an 500 de notre ère. Le russe a conservé cette voyelle, dont les fréquences formantiques attestent du caractère central : (FANT 1960, p. 109).

	F1	F2	F3	
[ï]	300	1480	2230	(Hz)

Sans que nous prenions position sur l'origine de cette évolution dans un domaine qui ne nous est pas familier, on peut noter (cf. MILEWSKI 1967) que celle-ci est contemporaine de la monophtongaison des diphtongues au et eu :

au > ou > ū
eu > jau > jou > jū

On peut supposer que, par suite d'une tendance à la délabialisation, les anciens ū et les nouveaux aient eu tendance à se confondre, ce qui aurait amené les anciens ū à se palataliser. Le système ne comportant que deux voyelles postérieures (ā, ū), l'hypothèse d'une chaîne de pression n'est pas envisageable. Certains auteurs supposent une évolution ū > uï > ï par dissimilation. Cette hypothèse s'accorde mal avec la tendance à la monophtongaison.

XIV - JAPONAIS

Le système vocalique du japonais moderne est une preuve indirecte de la validité des résultats de notre simulation. On sait que le japonais possède une voyelle d'arrière fermée faiblement labialisée [ɯ], (BLOCH 1950). Les données formantiques (KANAMORI 1974) montrent que, pour les voyelles, isolées, [ɯ] possède un premier formant nettement plus élevé (355 Hz) que le premier formant de [i] (260 Hz). La valeur du deuxième formant indique une avancée articulatoire certaine. Il s'établit en effet à 1250 Hz.

On voit qu'en japonais, la délabialisation de /u/, phonétiquement [ɯ], coïncide avec une importante avancée articulatoire, et une aperture labiale importante. Selon CHIBA et KAJIYAMA (1941), l'aire aux lèvres pour [o] (1.4 cm^2) est plus petite que pour [ɯ] (2 cm^2). Même si les valeurs absolues ne sont pas comparables, on opposera ces valeurs à celles mesurées par FANT (1960) pour le [u] (0.65 cm^2) et le [o] russe (3.2 cm^2). L'avancée labiale est inexistante pour le japonais où [o] et [ɯ] ont la même longueur de conduit vocal (16.8 cm), alors que ces longueurs sont respectivement de 19.5 et 18.5 cm pour le [u] et le [o] russe. Le point de constriction maximale se trouve au 64/100e de la longueur totale pour le [u] russe (à partir des lèvres) mais au 48/100e pour le [ɯ] japonais. Ces chiffres confirment l'avancée articulatoire que les valeurs formantiques laissaient prévoir. La tendance est encore plus nette pour les voyelles en contexte. Le premier formant de [ɯ] atteint 400 Hz, ce qui est supérieur à la valeur du premier formant de [o] en français. Quant au deuxième, il atteint 1400 Hz. Ces valeurs formantiques caractérisent une voyelle un peu plus ouverte que le [ɨ] russe, dont les formants sont à 300 et 1480 Hz (FANT 1960, p. 109). Lorsqu'on considère les valeurs formantiques pour [o], il apparaît nettement que l'avancée articulatoire est essentielle pour préserver la différence entre les deux voyelles (fig. 9-IX). (moyennes F1 et F2 pour [o] en contexte : 450 - 1000 Hz). Dans le domaine consonantique, la présence des allophones [f],

連続音声中の母音のホルマント周波数の特徴

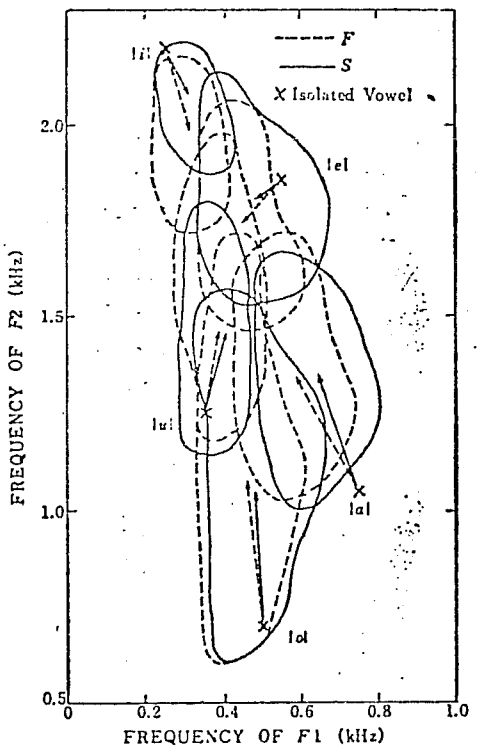


Fig. 9 - IX

(KANAMORI 1974)

[Φ] et [h] pour /f/ devant /u/ confirme les rapports entre la délabialisation vocalique et l'affaiblissement consonantique que nous avons postulés en ancien français. Les données synchroniques du japonais semblent donc confirmer le rôle de l'avancée articulatoire comme procédure ayant pour objet de rétablir la différence entre /o/ et /u/, différence mise en péril par la délabialisation.

XV - SUEDOIS

En ce qui concerne le suédois, ou plus précisément le scandinave central, le passage de [ʉ] à [ʉ̞], analysé en détail par HAUDRICOURT et JUILLAND (1949, p. 105-107) ne semble pas ressortir d'un phénomène de délabialisation. La voyelle [ʉ̞] est antérieure et fortement labialisée (FANT 1971). Cette palatalisation ne s'observe qu'en suédois ; Le norvégien, dialecte voisin au système vocalique presque identique, ne l'a pas connue (cf. SOMMERFELT 1960). Les traces de délabialisation du système vocalique sont inexistantes. Le système du suédois ancien possédait un [ʏ] provenant de [u] par "umlaut". Ce [ʏ] se conserve lors de la création de . On notera la différence avec le comportement de l'alsacien. Tout au plus peut-on noter que le [y] suédois moderne est moins labialisé que la voyelle correspondante en français et en allemand : les données de KITZING (1971) indiquent que le troisième formant de [y], qui indique le degré de labialisation, est beaucoup plus élevé qu'en allemand. En plus de la différence de labialisation, les voyelles [ʏ] et [ʉ̞] se différencient par deux "off-glides" consonantiques spécifiques : [ᵝ], qui représente un geste de fermeture labiale pour [ʉ̞ᵝ] et [uᵝ] ; [ʲ] qui indique une fermeture palatale, avec élévation de F2 et F3 pour [yʲ] (McALLISTER, LUBKER et CARLSON 1974 ; FANT 1971).

A l'époque de la palatalisation (vers 1400), on n'observe pas de trace de délabialisation pour les consonnes : HAUGEN (1976, p. 267) mentionne explicitement que [hw] reste [w] en suédois. Nous sommes sans doute en présence d'une

véritable chaîne de pression. Les faits sont bien documentés (HAUGEN 1976, p. 257). Vers 1200, $\bar{a} > \mathring{\bar{a}}$, qui se confond avec $\bar{ɔ}$. Plus tard (1400 ?) $\bar{o} > \bar{u}$ [uw] et $\bar{u} >$ [ʉʷ] Cette période (scandinave moyen) se caractérise par des renforcements articulatoires ([θ]>[t],[ð]>[d]par exemple) à l'inverse de ce que l'on observe en alsacien ou en ancien français. Ce même renforcement articulatoire pourrait également être à l'origine des "diphtonguaisons" ou plus précisément des adjonctions d'"off-glides" consonantiques. En conclusion, l'analyse d'HAUDRICOURT et JUILLAND nous semble parfaitement justifiée pour les mutations vocaliques du suédois.

XVI - CONCLUSIONS

Ce chapitre a présenté une nouvelle explication de la soit-disant "palatalisation spontanée" de [u] en [y]. Celle-ci serait fondée sur une avancée articulatoire ayant pour but d'éviter la confusion de [u] et [o] amenée par une délabialisation.

Ce chapitre invite à deux conclusions théoriques :

- la première est que l'espace perceptif des voyelles peut être mis en évidence, de façon indirecte, par une étude diachronique. On a donc montré que les études diachroniques pouvaient servir de "révélateur" aux études portant sur la perception des voyelles.

- la deuxième conclusion est que l'étude linguistique de l'utilisation des indices perceptifs met en pleine lumière l'interaction entre les indices articulatoires et perceptifs. Ce point capital sera repris au chapitre XI.

CHAPITRE X

LES VOYELLES NASALES

I - INTRODUCTION

Ce chapitre présente quelques résultats préliminaires concernant la production et la perception des voyelles nasales du français. Il n'est pas difficile de justifier la présence dans ce chapitre d'une étude sur la production, et plus précisément sur les caractéristiques acoustiques des voyelles nasales. En effet, à l'inverse des voyelles orales, il reste difficile d'identifier, dans une représentation spectrale, ces différentes caractéristiques. Dans un premier temps, nous avons effectué une analyse détaillée de 48 voyelles nasales prononcées par deux locuteurs. Cette analyse avait pour objet de vérifier que nous étions capables d'identifier les formants et antiformants (pôles et zéros) des voyelles nasales et d'accumuler quelques résultats quantitatifs en vue de l'analyse perceptive. Nous avons également réalisé une simulation sur ordinateur de la synthèse de voyelles nasales pour vérifier les résultats de notre analyse. Le temps nous a malheureusement manqué pour effectuer l'analyse perceptive que nous voulions mener par une technique de synthèse sur mini-ordinateur, en raison de la longueur de mise au point de la branche nasale de notre synthétiseur. Nous ne pourrons présenter que quelques résultats préliminaires portant sur des voyelles synthétisées sans zéros.

II - MODELE THEORIQUE DE PRODUCTION ET ANALYSE DESCRIPTIVE DES VOYELLES NASALES

Il n'est pas aisé de présenter le modèle de production des voyelles nasales, en raison de sa complexité mathématique. Il est même très difficile de présenter et d'utiliser les résultats des travaux descriptifs en raison des choix différents de paramètres descriptifs chez les différents auteurs.

D'une façon tout à fait singulière, le nombre d'études descriptives sur les voyelles nasales ne cesse de croître, alors que nos connaissances théoriques n'ont guère progressé depuis 1960. Mais souvent, ce cadre théorique, pourtant ancien,

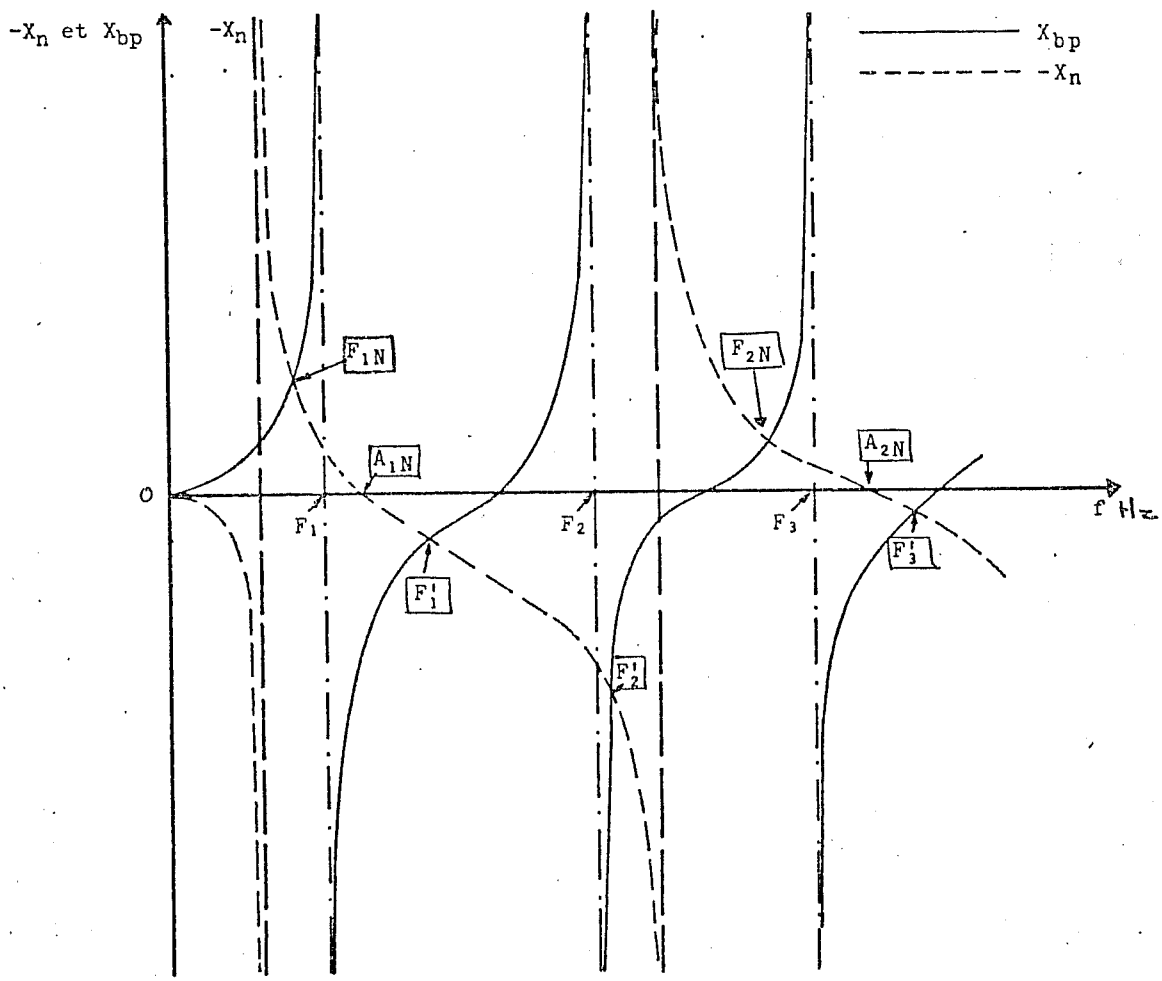


Fig. I - X

n'est pas utilisé, ce qui conduit au mieux à une description très pauvre et au pire à des erreurs d'identification des paramètres pertinents. Dès 1960, FUJIMURA (1960) a proposé une méthode graphique originale permettant de déterminer les pôles et les zéros d'un son nasal. Cette méthode graphique a été légèrement modifiée par MRAYATI (1975, 1976), auquel nous empruntons la fig. 1-X. Sur cette figure, les abscisses où Xbp (c'est-à-dire la réactance du conduit buccal associé au pharynx, au point de couplage du conduit nasal) est infinie correspondent aux fréquences des formants du conduit sans couplage nasal. Les fréquences pour lesquelles -Xn est infinie correspondent aux pôles (formants) propres du conduit nasal. Ces fréquences dépendent d'ailleurs du degré de couplage, c'est-à-dire de la surface du passage ouvert par l'abaissement du voile. Ceci amènera à considérer une famille de courbes -Xn pour différents degrés de couplage. Les points où la courbe -Xn coupe l'axe des fréquences donnent la position des zéros (ou antiformants). Les points d'intersection des courbes Xbp et -Xn définissent les fréquences de résonance (ou formants) de la voyelle nasale. D'un point de vue terminologique, nous appèlerons F'i les fréquences de résonance des formants associés à la cavité orale ; nous noterons par Fni, les formants nouveaux issus du couplage du conduit nasal (formants de nasalité) et par Ai (Antiformant) les zéros introduits par le couplage.

Les conclusions générales tirées par FUJIMURA (1960) ont été reprises pratiquement inchangées 11 ans plus tard dans une excellente vérification expérimentale à l'aide d'une méthode de mesure de la fonction de transfert par balayage acoustique (FUJIMURA et LINDQVIST 1971).

Ces conclusions peuvent se résumer ainsi :

a) - Entre 100 et 3000 Hz, les voyelles nasales possèdent deux formants supplémentaires de nasalité (F1n, F2n), associés à deux zéros (A1, A2). Quant aux formants correspondant à ceux de la voyelle orale (F'i), ils sont décalés vers les

fréquences hautes d'une quantité variable. L'absence de nasalité peut être considérée comme l'annulation des deux formants de nasalité par les deux zéros. La fig. 2-X présente l'effet spectral d'un formant associé à un zéro.

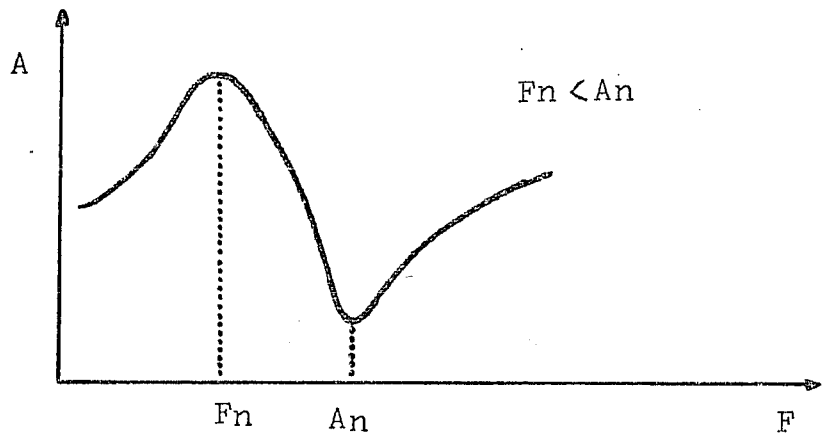


Fig. 2 - X

b) - Pour les voyelles nasales françaises, c'est-à-dire dans le cas où la voyelle nasale n'est pas du degré le plus fermé, le premier formant est F1n.

c) - Une variation dans le degré de nasalisation préserve l'ordre relatif des formants, tout en faisant varier les fréquences individuelles d'une façon négligeable. Par contre, les deux zéros peuvent, pour un certain degré de nasalité, détruire un formant oral décalé (F'i). (cf. fig.3 -X). Les fig. 4-X et 5-X illustrent la suppression de formants par la présence de zéros. La fig. 4-X présente la simulation sur ordinateur de la voyelle neutre [ə] dont les formants sont à 500, 1500, 2500, 3500 et 4500 Hz. La fig. 5-X présente cette même voyelle après adjonction de deux zéros à 1500 et 3500 Hz. On observe que les formants 2 et 4 ont été annihilés par les deux zéros.

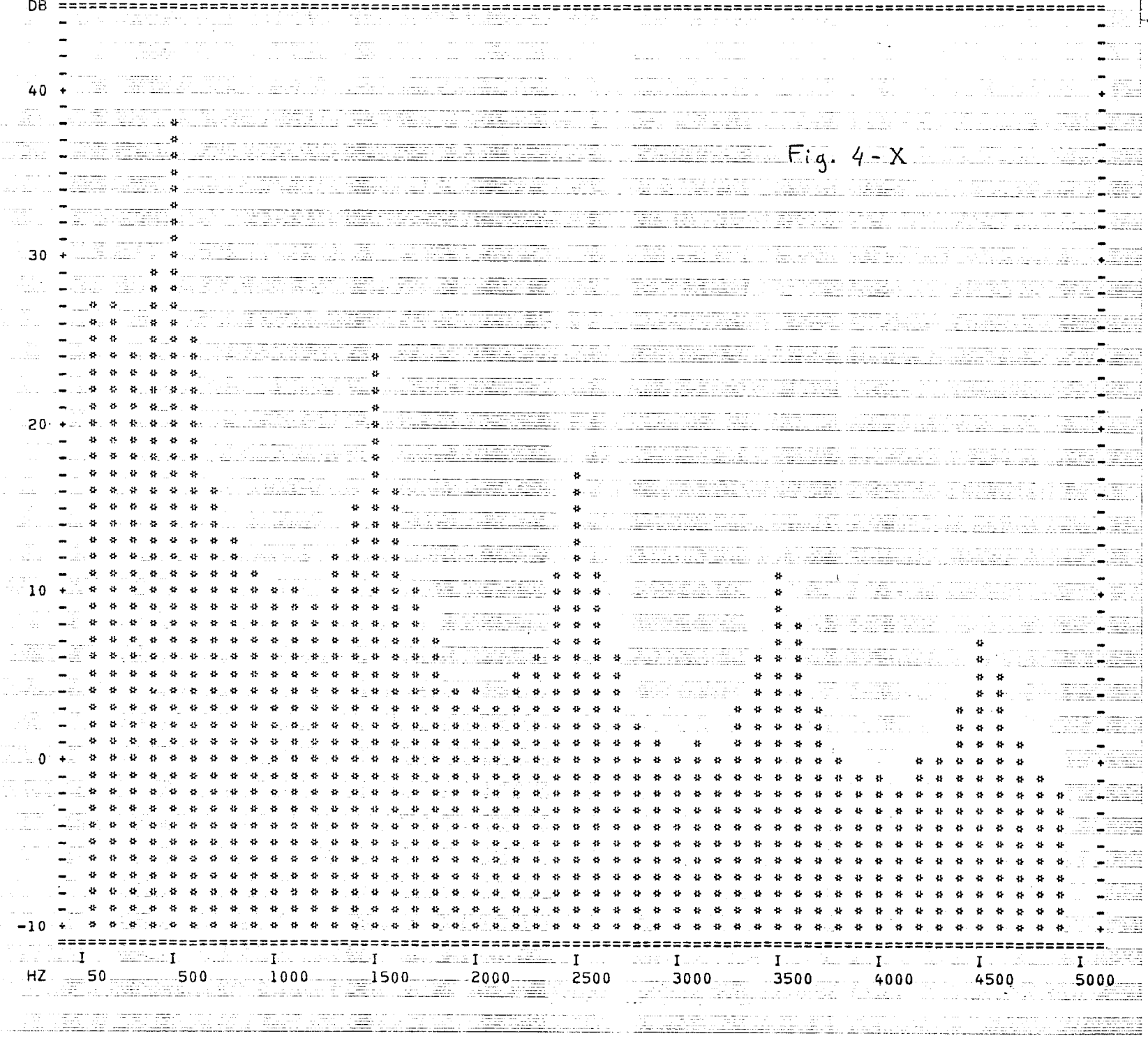

DB
40
30
20
10
0
-10
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Fig. 4-X

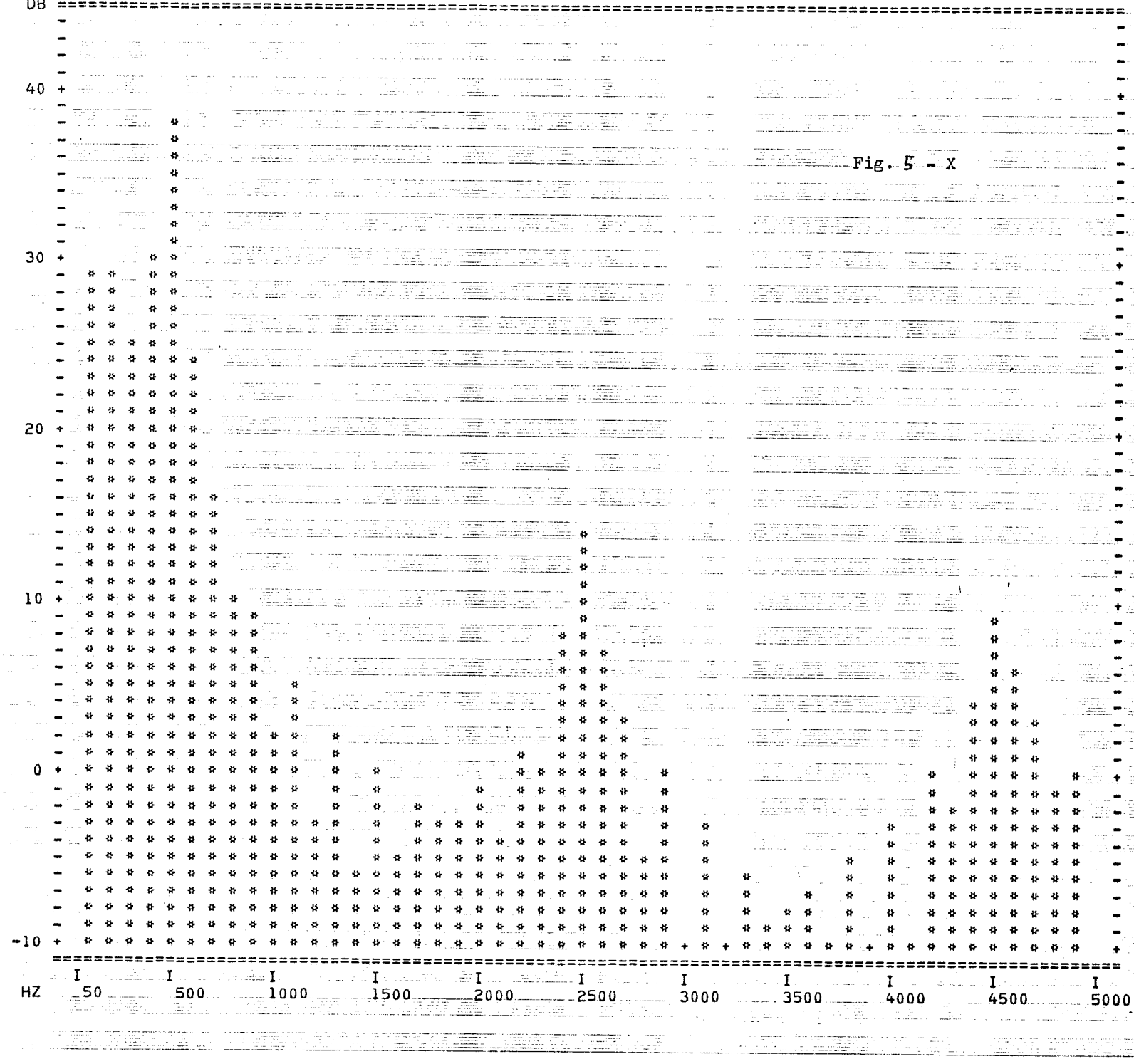

DB
40
30
20
10
0
-10
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Fig. 5 - X

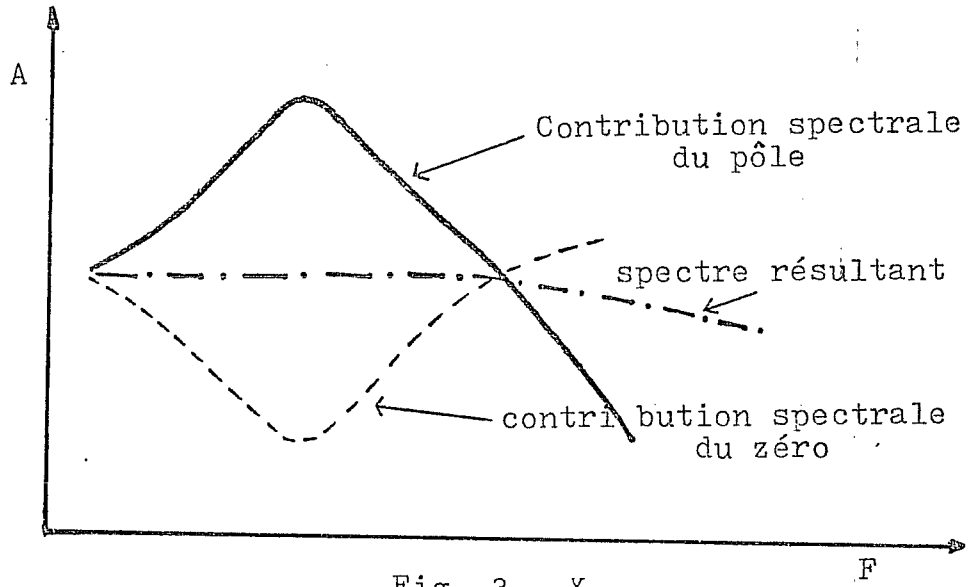


Fig. 3 - X

d) - L'ordre des pôles et des zéros est le suivant dans la plupart des cas :

F1n, AI, F'1, F'2, F2n, A2, F'3. La voyelle [ɔ̃] pourrait constituer une exception à cette règle générale.

Ces conclusions dépendent d'hypothèses concernant la forme du conduit nasal (BJUGGREN et FANT 1964). Certaines différences légères entre les résultats expérimentaux et le modèle théorique doivent être attribuées à la simplification de la forme du conduit dans le modèle théorique. Si l'on admet (LINDQVIST-GAUFFIN et SUNDBERG 1976) que le conduit nasal est constitué de deux conduits qui ne sont pas forcément symétriques, cette asymétrie peut produire un pôle et un zéro supplémentaires au-dessus de 1000 Hz. De plus, les cavités formées par les sinus frontaux et maxillaires peuvent introduire des pôles et des zéros supplémentaires entre 200 et 2000 Hz. Il n'est pas impossible que quelques exemples de ces paires pôle-zéro supplémentaires soient visibles dans nos spectrogrammes.

Dans cette brève revue des travaux portant sur la structure des voyelles nasales, il convient de discuter en détail les conclusions dégagées par DELATTRE (1954, 1965, 1968) en raison de leur diffusion importante. Pour DELATTRE, les caractéristiques essentielles des voyelles nasales se résument à deux aspects essentiels :

a) - L'extrême faiblesse d'intensité du premier formant (F'1) : Ce trait constituerait le paramètre perceptif essentiel. L'examen des spectrogrammes révèle que cette description n'est que partiellement exacte, en particulier pour [ɔ̃] où il n'est pas possible de discerner un pic que l'on puisse attribuer à F'1, bien que la simulation témoigne de son existence (cf. infra). Il est plus correct de considérer que la réduction d'intensité du premier formant est l'un des résultats de la présence d'un zéro dans la même zone fréquentielle. La réduction d'intensité sera variable en fonction de la position du zéro, qui dépend en dernière analyse du degré de nasalisation. En bref, DELATTRE a observé un résultat sans en dégager la cause. Les fig. 6-X et 7-X illustrent la réduction d'intensité de F'1 résultant de la proximité d'un zéro. La figure 6-X est le résultat de la simulation d'une voyelle nasale correspondant approximativement à [ɛ̃] sans inclure le premier zéro. La fig. 7-X présente la même voyelle en incluant l'effet du premier zéro. On note tout d'abord une nette réduction de l'intensité de F'1. La réduction observée est de -8 dB. Ce chiffre est voisin des -12 dB mentionnés par DELATTRE. L'ampleur de la réduction dépend à la fois de la position du zéro par rapport à F1 et de sa largeur de bande. On peut noter également un deuxième effet spectral dû à la présence d'un zéro en basse fréquence : Il s'agit de la réduction importante de la pente moyenne du spectre. On observera que la présence de ce zéro a pour effet de rehausser de plus de 10 dB le niveau de F'2, mais également de tous les formants supérieurs. Cette égalisation de la pente moyenne du spectre semble être un effet qui n'a pas été noté jusqu'à ce jour.

b) - L'apparition d'un formant nasal à 250 Hz.

Dans l'ensemble de nos spectrogrammes, il est certain que la première harmonique (à Fo x 2) est plus intense que le fondamental ou que la 2ème harmonique. FANT (1960, p. 160) suggère que ceci est le résultat de l'effet du contrôle AGC (gain automatique) du Sonagraphe. L'argument ne nous

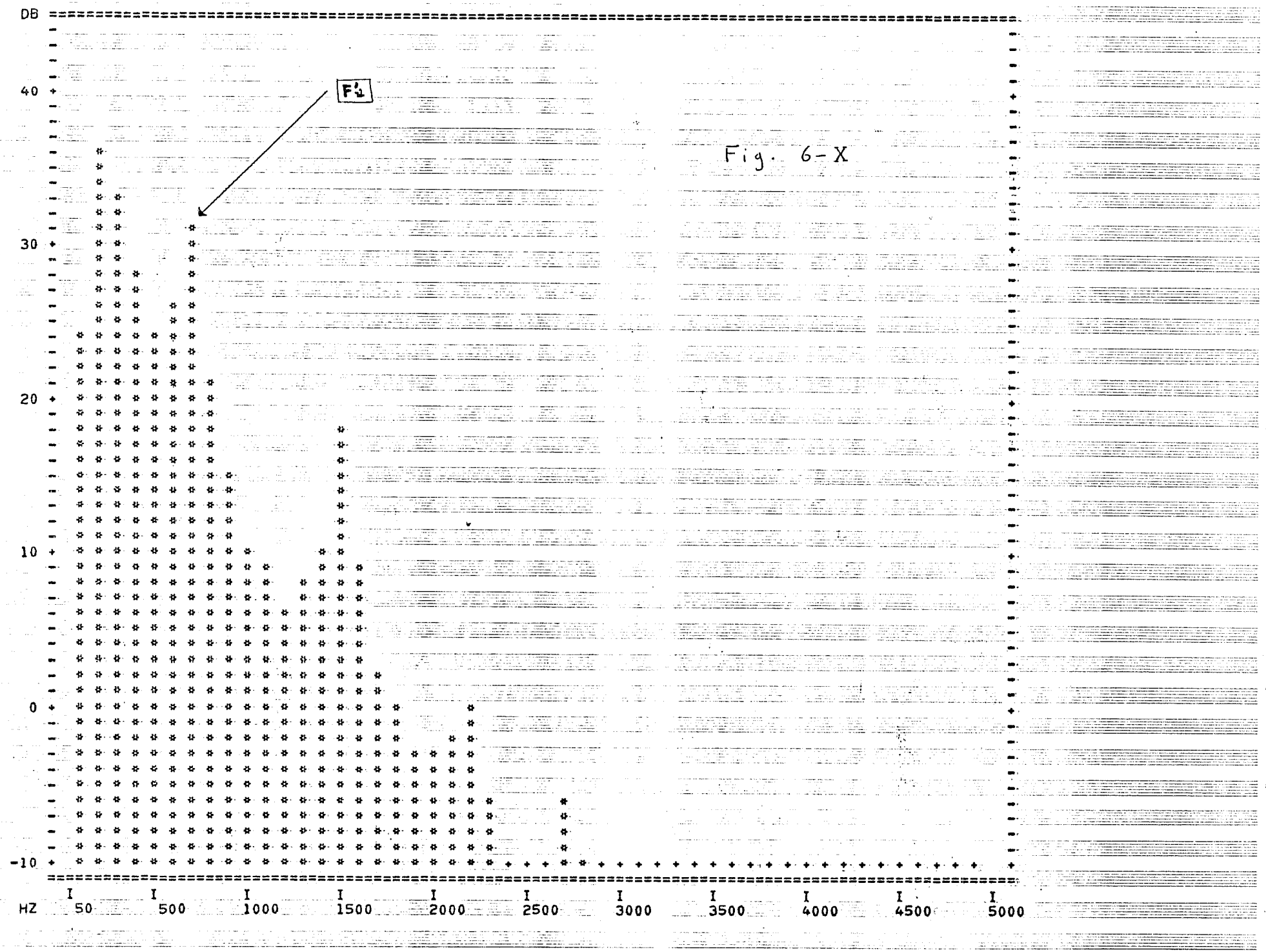

DB
40
30
20
10
0
-10
F2
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Fig. 6-X

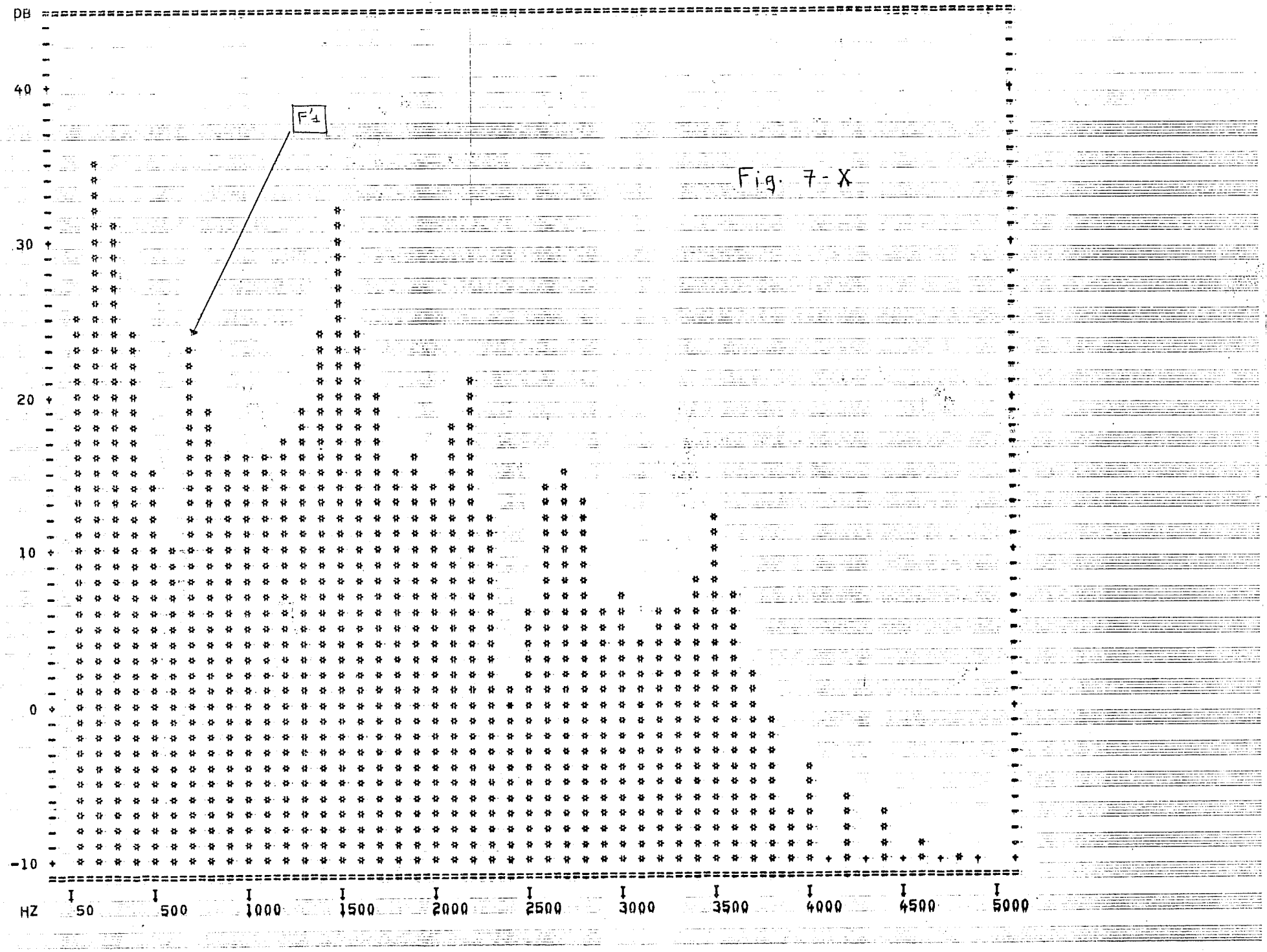

DB
40
30
20
10
0
-10
F'1
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Fig. 7-X

paraît pas convaincant car les exemples d'analyse numérique par transformée de Fourier donnent un résultat identique. Pourtant, le travail de MRAYATI (1975) suggère que le formant nasal F1n devrait se trouver à une fréquence supérieure (au-delà de 400 Hz). Les tracés publiés par FUJIMURA et LINDQVIST (1971) confirment ce fait. Nous avons observé, sur d'autres spectrogrammes que ceux soumis ici à l'analyse, des exemples où F1n semble placé vers 450-500 Hz (3ème harmonique) et ceci pour un des sujets dont nous avons analysé les spectrogrammes (sujet F.L.). Il ne peut donc s'agir d'une caractéristique personnelle. La fig. 8-X présente deux voyelles nasales ([ɛ̃] et [ã]) qui présentent cette caractéristique. L'hypothèse d'une variation liée au degré de nasalité pourrait être retenue. Un travail complémentaire serait nécessaire pour éclairer ce problème, car cette hypothèse n'est pas conforme aux conclusions de MRAYATI (1975).

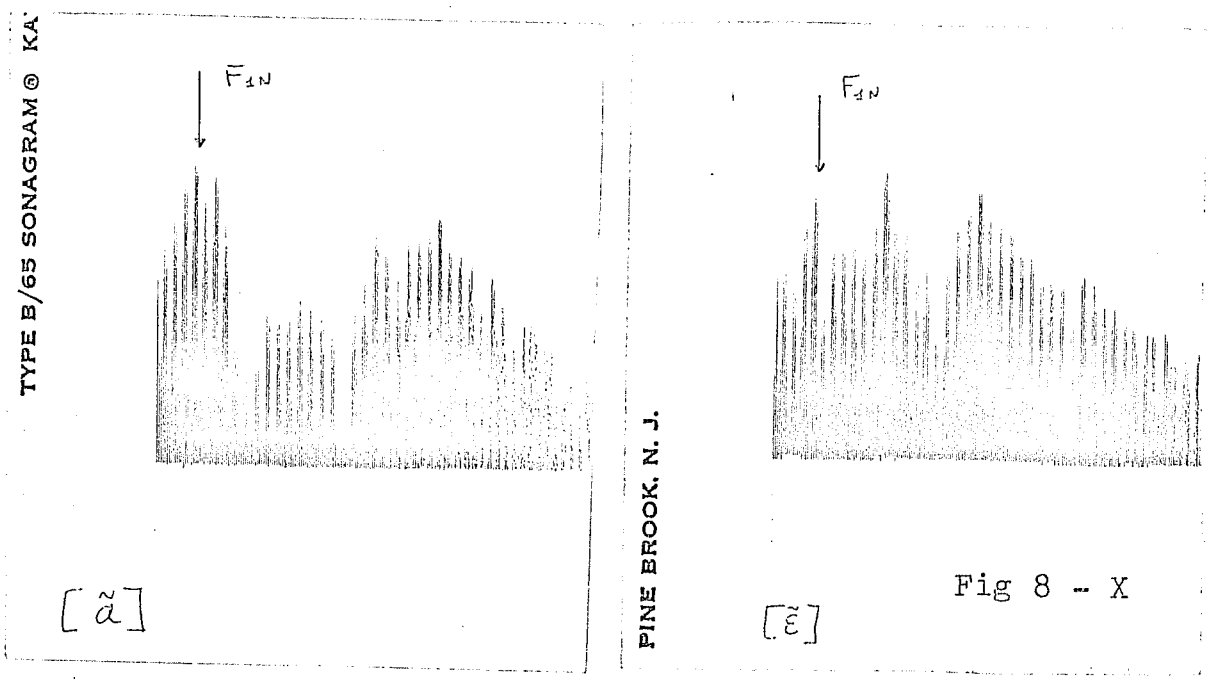


Fig 8 - X

c) - Indices mineurs

DELATTRE cite également 4 indices mineurs :

- apparition d'un formant nasal vers 2000 Hz : ceci est confirmé par le modèle théorique et les tracés spectrographiques.

- Invariance de F2 : ce point est également confirmé.

- Abaissement de F3 et élévation de F4 : ces deux derniers points ne sont pas confirmés par l'analyse théorique.

FANT (1960) consacre un chapitre très dense aux problèmes de la nasalisation. Pour les voyelles, les résultats obtenus par simulation sur un analogue du conduit vocal confirment largement les prédictions théoriques de FUJIMURA. Les résultats détaillés seront présentés lors de l'analyse des voyelles de notre corpus. Deux autres études plus anciennes contiennent également des données intéressantes : HOUSE et STEVENS (1956) ; HATTORI, YAMAMOTO et FUJIMURA (1958). La première insiste sur l'affaiblissement de F1, qu'elle attribue à une augmentation de la largeur de bande. La deuxième souligne la présence d'un formant nasal vers 250 Hz et la présence d'un zéro vers 500 Hz.

Il n'existe qu'une étude approfondie sur les caractéristiques acoustiques des voyelles françaises (MRAYATI 1975). Pour être complet, il faut encore mentionner les valeurs proposées par DELATTRE (1954, 1965, 1968), DEBROCK (1974) et GUTH (1975). Les contributions de LAFON (1961) et de HUSSON (1962) ne nous semblent pas reposer sur des bases scientifiques sûres. DEBROCK présente des résultats détaillés pour 3 sujets ayant prononcé une seule fois les 4 voyelles nasales. En plus d'une analyse spectrographique, les résultats ont été contrôlés par un analyseur spectral analogique. Les résultats de GUTH portent sur 4 sujets et une seule répétition des 4 voyelles, analysées au sonagraphe. Les résultats de MRAYATI sont des moyennes pour 5 locuteurs. Les mesures ont été effectuées sur la première voyelle de groupes CṼCṼ où C représente 3 contextes consonantiques différents : [p t k ; b d g ; m n ɲ] . Une méthode de prédiction linéaire a été utilisée. Nous discuterons l'emploi de cette méthode dans le cadre d'une étude sur les voyelles nasales à la fin de ce chapitre.

III - ANALYSE DESCRIPTIVE D'UN ECHANTILLON DE 48 VOYELLES NASALES

Devant la pauvreté des renseignements quantitatifs concernant les voyelles nasales, nous avons jugé nécessaire de réaliser une analyse préliminaire. Deux phonéticiens professionnels, possédant 4 voyelles nasales dans leur inventaire phonologique, ont enregistré en chambre sourde 6 échantillons des 4 voyelles nasales françaises dans un ordre aléatoire. Nous désirions disposer d'un nombre suffisant d'échantillons de chaque voyelle afin de séparer les caractéristiques permanentes d'éventuelles variations aléatoires. Les sujets devaient maintenir le fondamental constant à 120 Hz. Ils disposaient pour cela d'un son sinusoïdal de référence. De nombreuses mesures sur des harmoniques de rang élevé (> 25) nous ont permis de vérifier que F_0 s'établissait à 120 ± 2 Hz.

Devant les critiques adressées à la méthode spectrographique par de nombreux auteurs, (cf. MRAYATI 1975, p. 2 : "Malheureusement, l'imprécision de cet appareil est trop grande pour étudier un phénomène aussi complexe") nous avions prévu à l'origine de faire une analyse numérique par transformée de Fourier (LONCHAMP 1974). Pour cela, les voyelles nasales ont été digitalisées sur 12 bits à l'aide d'un programme écrit par M. LAMOTTE du Laboratoire d'Electricité et d'Automatique de l'Université de Nancy I. L'étude des tracés numériques nous a prouvé qu'il n'y avait aucune différence entre les spectres numériques et les sections sonagraphiques. Nous publions plus loin 4 exemples d'analyses numériques sur lesquels on peut constater cette grande similitude. Deux périodes consécutives ont été analysées, ce qui explique le fondamental apparent à $F_0/2$ (60 Hz). Aucune pré-emphase n'a été appliquée.

Toutes les sections spectrographiques sont publiées ci-dessous. L'échelle placée au bas des figures est approximative. Une calibration individuelle a été réalisée pour chaque section. Un numéro d'ordre arbitraire a été attribué à chaque

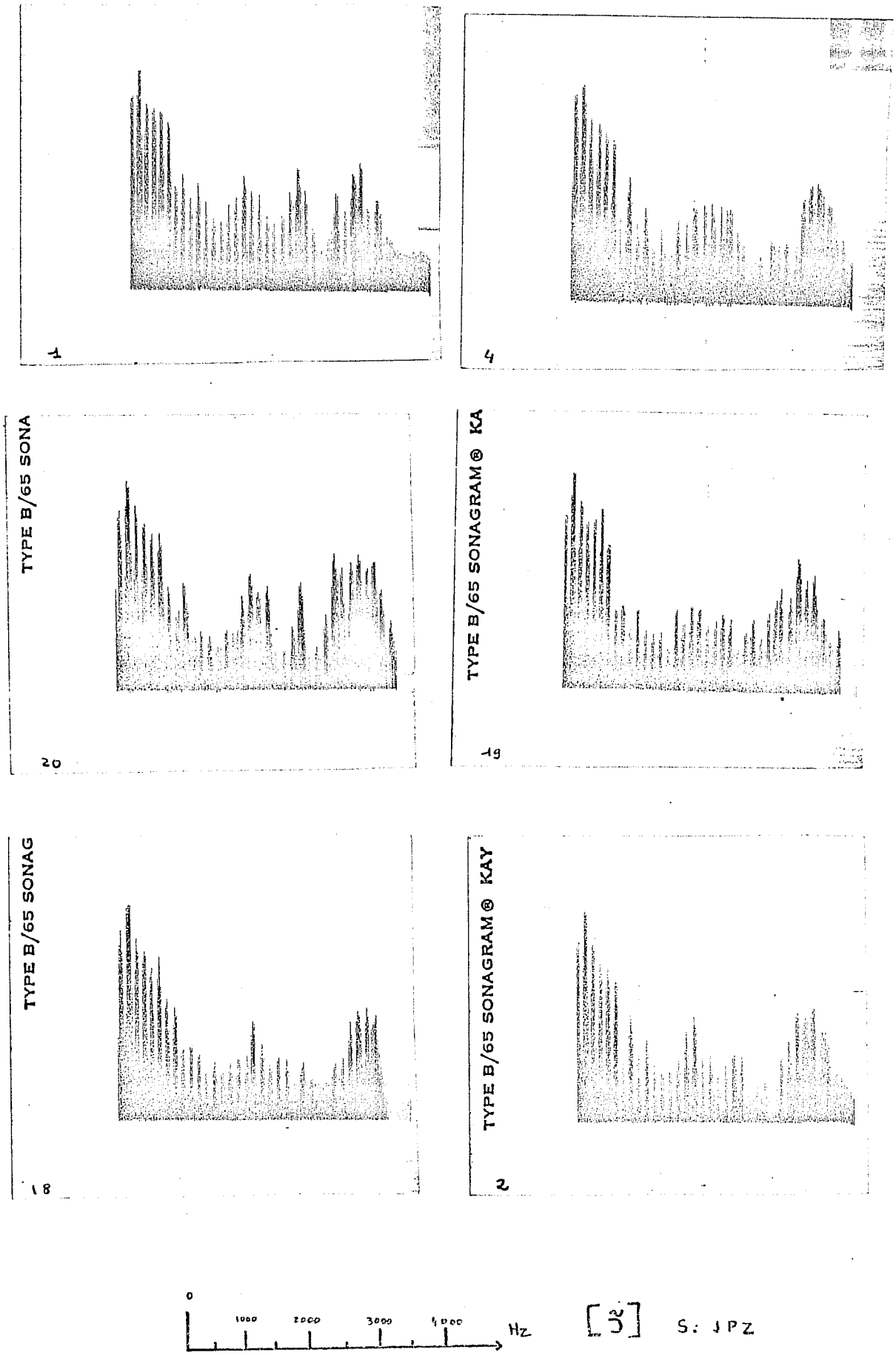

TYPE B/65 SONA
TYPE B/65 SONAGRAM® KA
TYPE B/65 SONAG
TYPE B/65 SONAGRAM® KAY
1
4
20
19
18
2
0
1000
2000
3000
4000
Hz
[3]
S: JPZ

Fig 9 - X

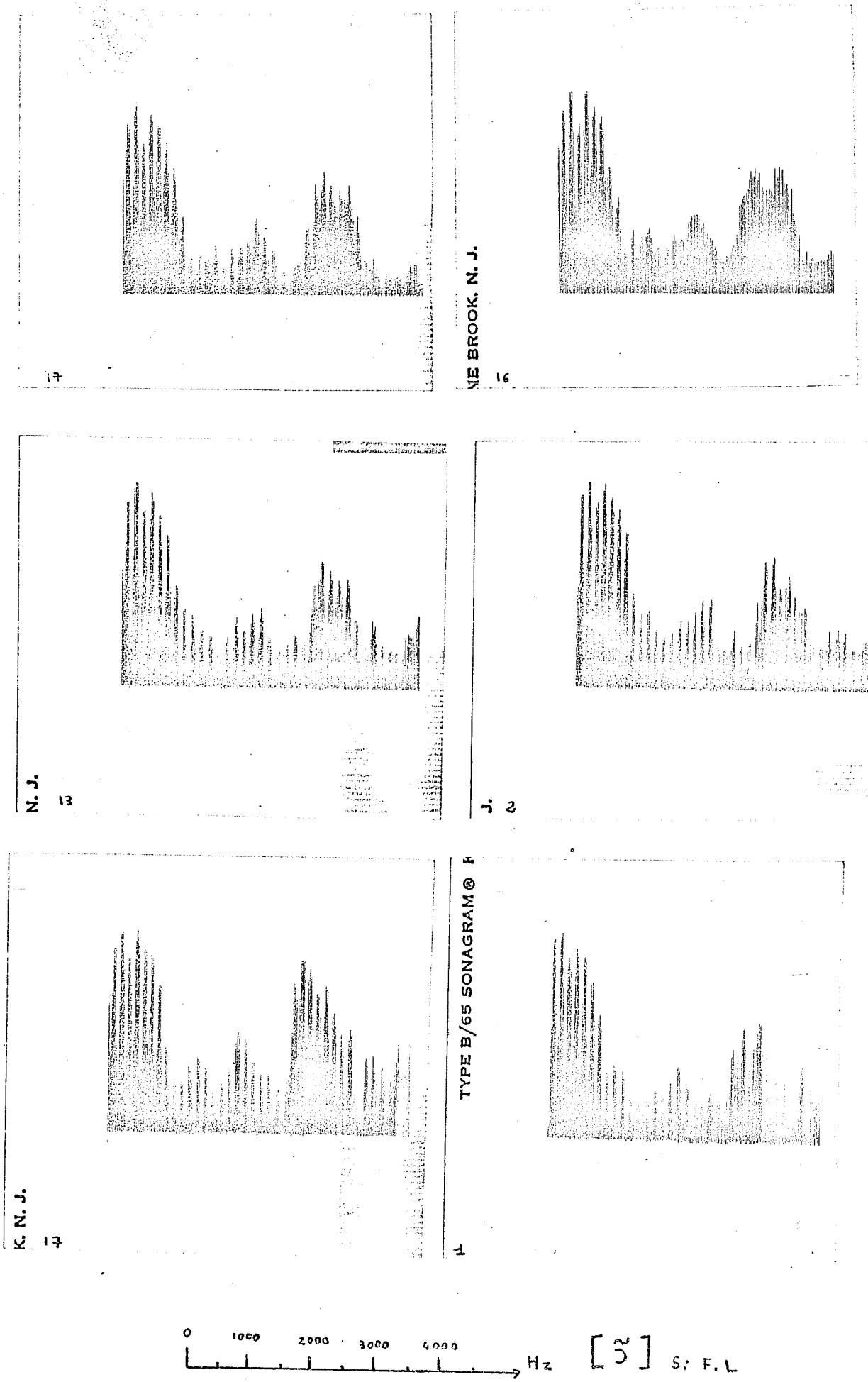

17
NE BROOK, N. J.
16
N. J.
13
J. 2
K. N. J.
17
TYPE B/65 SONAGRAM®
1
0
1000
2000
3000
4000
Hz
[5]
S: F. L

Fig. IO - X

section. Ce numéro apparaît sur chaque tableau et permet de repérer la section correspondant à chaque série de mesures. Une valeur entre parenthèses dans un tableau indique une légère irrégularité spectrale. Pour les fréquences élevées (> 2500 Hz), nous avons donné les fréquences de tous les maxima du spectre. Ces maxima sont symbolisés dans les tableaux par P1, P2... en fonction de la fréquence. Une erreur systématique semble affecter la mesure de P1, centré sur la première harmonique. F_o étant constant à 120 Hz, ce pic devrait se trouver vers 240 Hz. Cette erreur systématique est due à une forte non-linéarité de l'échelle spectrographique au-dessous de 500 Hz. Cette sur-estimation systématique de la fréquence de P1 n'a pas été corrigée sur les tableaux.

Pour chacune des voyelles, on présente d'abord une analyse des caractéristiques spectrales visibles sur les sections. Nous comparons ensuite nos mesures avec les valeurs publiées par différents auteurs. A partir des schémas théoriques publiés par MRAYATI (1975, p. 13-15), nous tentons d'identifier les pics. Puis nous présentons une simulation sur ordinateur du spectre moyen de la voyelle nasale. Cette simulation a été obtenue à l'aide du programme SIMNAS qui est présenté en annexe de ce travail. Tous les tracés ont été effectués avec un F_o théorique de 100 Hz. Une double différenciation temporelle a été utilisée pour rendre ces spectres directement comparables avec les sections sonagraphiques. Nous terminons chaque paragraphe par l'analyse numérique d'une voyelle nasale réelle.

IV - LA VOYELLE [ɔ̃]

A/ Description spectrale

1)- Sur les sections des fig. 9-X et 10-X, on remarque que la première harmonique (P1) est plus intense que le fondamental ou la 2ème harmonique. Cette constatation est valable pour les 12 sections.

2)- Les 4 harmoniques suivantes présentent un schéma régulier mais différent pour chacun des sujets. Pour le sujet

JPZ, l'intensité des harmoniques est décroissante dans 4 sections sur 6. Pour les deux autres, la dernière harmonique de ce groupe est plus intense. Nous avons choisi cette dernière harmonique pour la mesure de P2, qu'elle soit plus intense ou non. Chez le sujet FL, la deuxième harmonique du groupe est la plus intense, et a servi de base aux mesures. Ce choix différent explique l'écart non-négligeable des mesures entre les deux sujets.

3)- Les composantes spectrales de la zone s'étendant entre 1000 et 1800 Hz sont à niveau très faible. Dans quelques cas, on peut déceler la présence d'une ou de deux irrégularités spectrales (P3).

4)- Une zone d'harmoniques renforcées est visible vers 1850-2000 Hz chez le sujet JPZ et vers 2000-2200 Hz chez le sujet FL (P4). La présence de plusieurs valeurs dans le tableau indique la présence de plusieurs harmoniques intenses non contiguës (section JPZ n° 19 par ex.). Pour le sujet JPZ, cette zone est suivie d'une seconde zone d'harmoniques renforcées vers 2500 Hz (P4'). Cette zone n'apparaît pas chez le sujet FL, mais on peut noter que P4 pour ce sujet est à une fréquence intermédiaire entre P4 et P4'.

5)- On note enfin une zone complexe (P5) d'intensité relativement forte. Ce bloc est souvent bimodal chez le sujet FL.

Le tableau suivant présente les mesures effectuées sur les sections. Les moyennes ne sont indiquées que lorsque la dispersion des valeurs n'est pas trop forte.

Sujet JPZ	1	2	4	18	19	20	Moyenne
P1	275	305	280	250	280	285	280 H.
P2	700	780	750	725	700	770	735
P3	(915/1170)	(1000/1215)	(1000/1230)	(975/1230)	(1040/1250)	(1105)	
P4	1850	1955	1950	2160	1850	2090	1900
P4'	2660	2565	2450		2100 2520	2325 2850	
P5	3235 3550 3850	3565 3780	3175 3760	3760	3360 3660	3370 3670	

Sujet FL	14	16	13	2	17	1	Moyenne
P1	285	285	315	285	285	285	290 Hz
P2	540	520	565	520	520	520	530
P3	(1525)	(1230) (1460)	(1850)	(1210)	(1480)	(1125)	
P4	2150	2195	2220	2110	2070	2110	2150
P5	2950 3180 3580	3120 3495	3150 3915	3075 3370	3075 3455 3830	2970 3140 3350	

B/ Essai d'interprétation des caractéristiques spectrales

1)- Comparaison avec les travaux antérieurs

On observe un bon accord avec les valeurs de DEBROCK.

	S1	S2	S3	Moyenne JPZ	Moyenne FL
P1	200	-	200	280	290 Hz
P2	780	520	560 900	735	530
P4	2150	2090	1750	1900	2150
P5	3530	3210	3460	-	-

Par contre, les valeurs de GUTH se laissent plus difficilement ranger dans les différentes classes.

	S1	S2	S3	S4	Moyenne JPZ	Moyenne FL
P1	400	400	400	400	280	290 Hz
P2	1100	1000	900	1000	735	530
P4	1900	2000	2500	1600	1900	2150
95	3400	3200	3200	3200	-	-

Les valeurs de P2, en particulier, semblent trop élevées.

DELATTRE donne les valeurs suivantes : 250, 600 et 750 Hz. La faible séparation des deux dernières valeurs laisse présager que les deux formants seront souvent confondus.

On observe un bon accord avec les mesures de MRAYATI.

	contexte ptk	contexte bdg	contexte mnɲ	moyenne JPZ	moyenne FL
P1	431	442	419	280	290 Hz
P2	764	711	733	735	530
P3	3065	3049	3040	-	-

2- Essai d'interprétation

L'analyse théorique de MRAYATI (fig. 11-X) conduit aux valeurs de formants et de zéros suivantes pour 4 degrés de couplage. Les formants de la voyelle orale correspondante s'établissent à 500, 1000 et 2500 Hz.

Degré de couplage	F1n	A1	F'1	F'2	F2n	A2	F'3
0.27 cm^2	450	500	600	1000	2350	2350	2500
1.66 "	430	750	850	1000	2250	2350	2600
3.5 "	415	1000	1000	1000	2100	2350	2750
5.0 "	400.	1100	1000	1000	2000	2350	2875

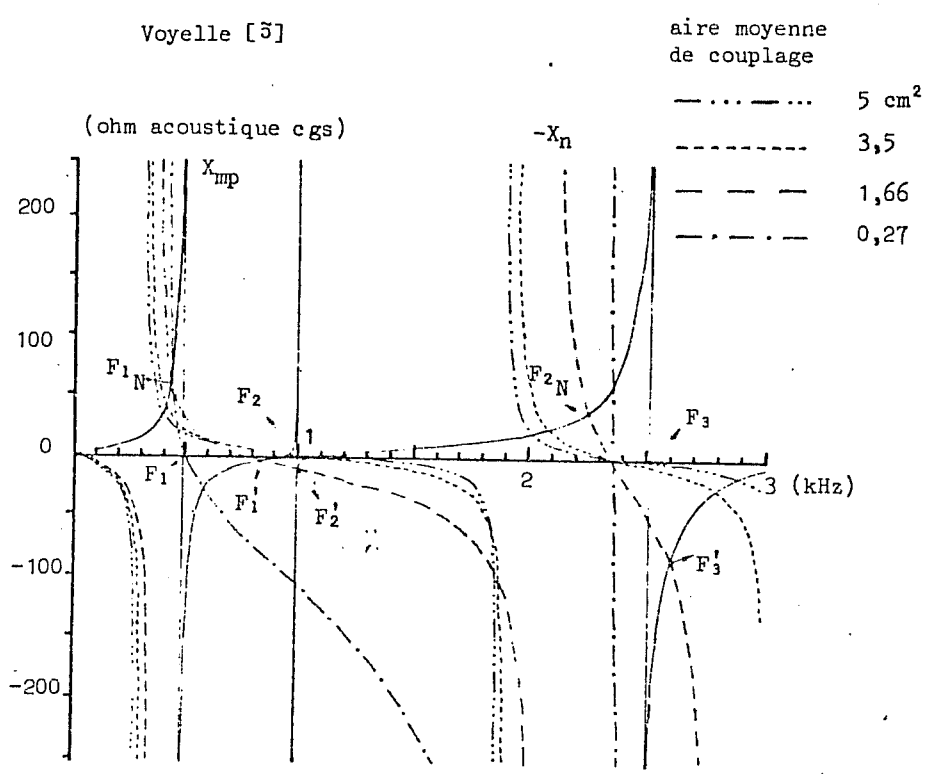


Fig -II - X

Pour les degrés de couplage importants, il est difficile d'estimer numériquement les valeurs de F'1, A1, et F'2. On peut néanmoins identifier les pics P du tableau de la manière suivante :

P1 correspond sans doute au premier formant nasal F1n. Quant à P2, il s'agit du pic formé par F'1 et F'2 qui sont très proches et partiellement masqués par le premier zéro A1. On comprend mieux les différences importantes constatées pour cette zone chez les deux sujets. Pour P3, les irrégularités spectrales ne correspondent pas à une caractéristique essentielle de nasalité. P4 correspond au deuxième formant nasal. P4' et P5 sont les formants F'3 et F'4.

La théorie prévoit la présence de deux zéros, mais ceux-ci n'apparaissent pas clairement sur les spectrogrammes. Le premier joue néanmoins un rôle essentiel dans la formation du pic P2. FUJIMURA et LINDQVIST (1971, fig. 15, p. 554) présentent le spectre d'un [s̃] obtenu par balayage fréquentiel. On note un premier pic (F1n) à 375 Hz et un zéro nettement visible à 475 Hz (A1). Le premier formant (F'1) n'apparaît que comme un léger ressaut à 600 Hz, dominé par un pic plus important à 850 Hz (F'2). Le deuxième formant nasal est bien visible à 2200 Hz, immédiatement suivi du deuxième zéro (A2) vers 2350 Hz et de F'3 (2500 Hz). Ces valeurs sont très proches de celles que nous avons mesurées.

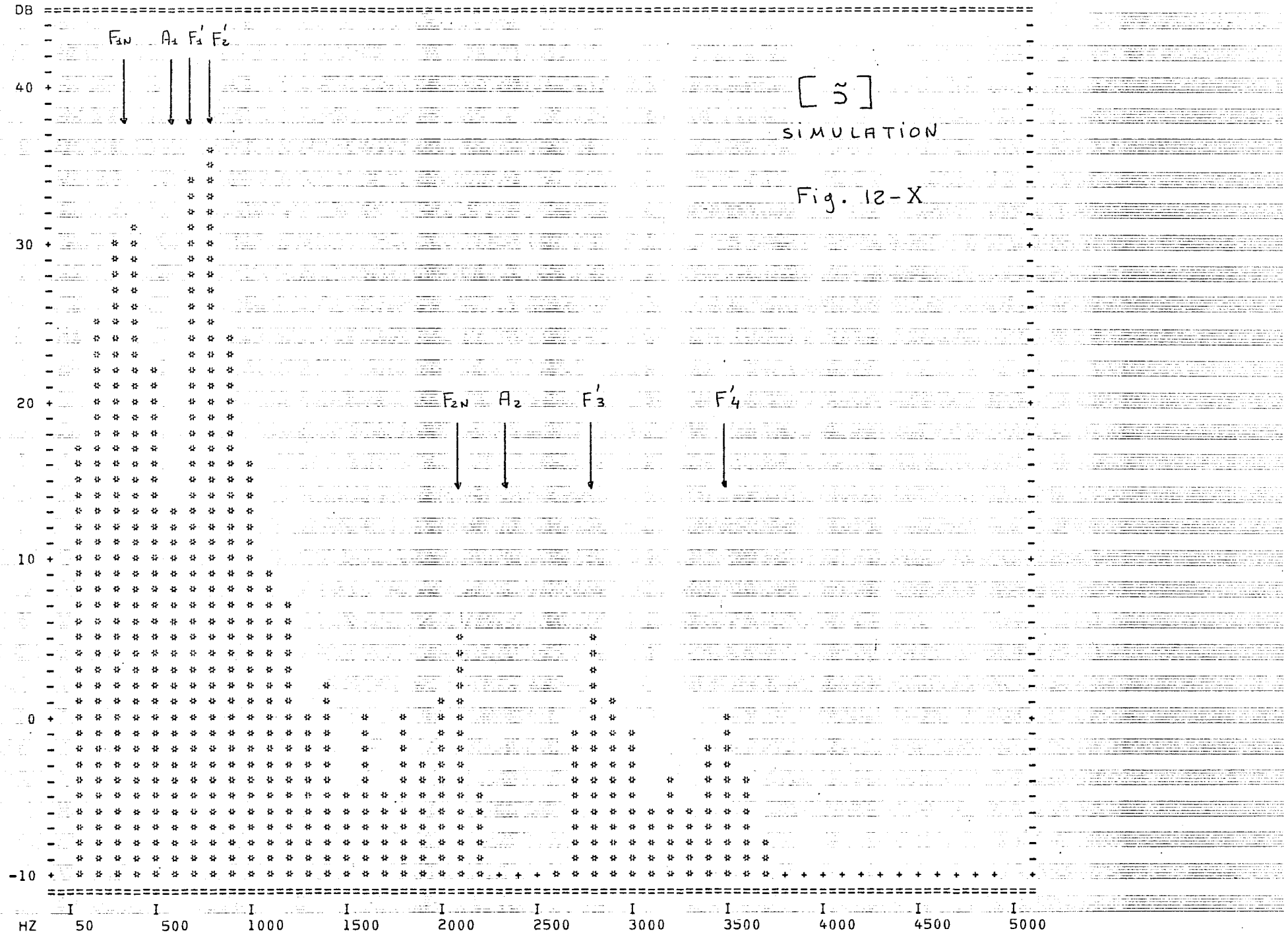
3)- Essai de simulation

Afin de vérifier les conclusions précédentes, nous avons réalisé plusieurs simulations en raison des divergences notées entre les deux sujets.

Il n'a pas été possible de produire une bonne approximation du spectre des voyelles du sujet FL. La meilleure simulation est présentée à la fig. 12-X. Les valeurs numériques sont les suivantes :

	F1n	F'1	F'2	F2n	F'3	F'4	A1	A2
Hz	350	700	800	2100	2800	3500	600	2350
Hz	65	80	80	100	120	180	100	120

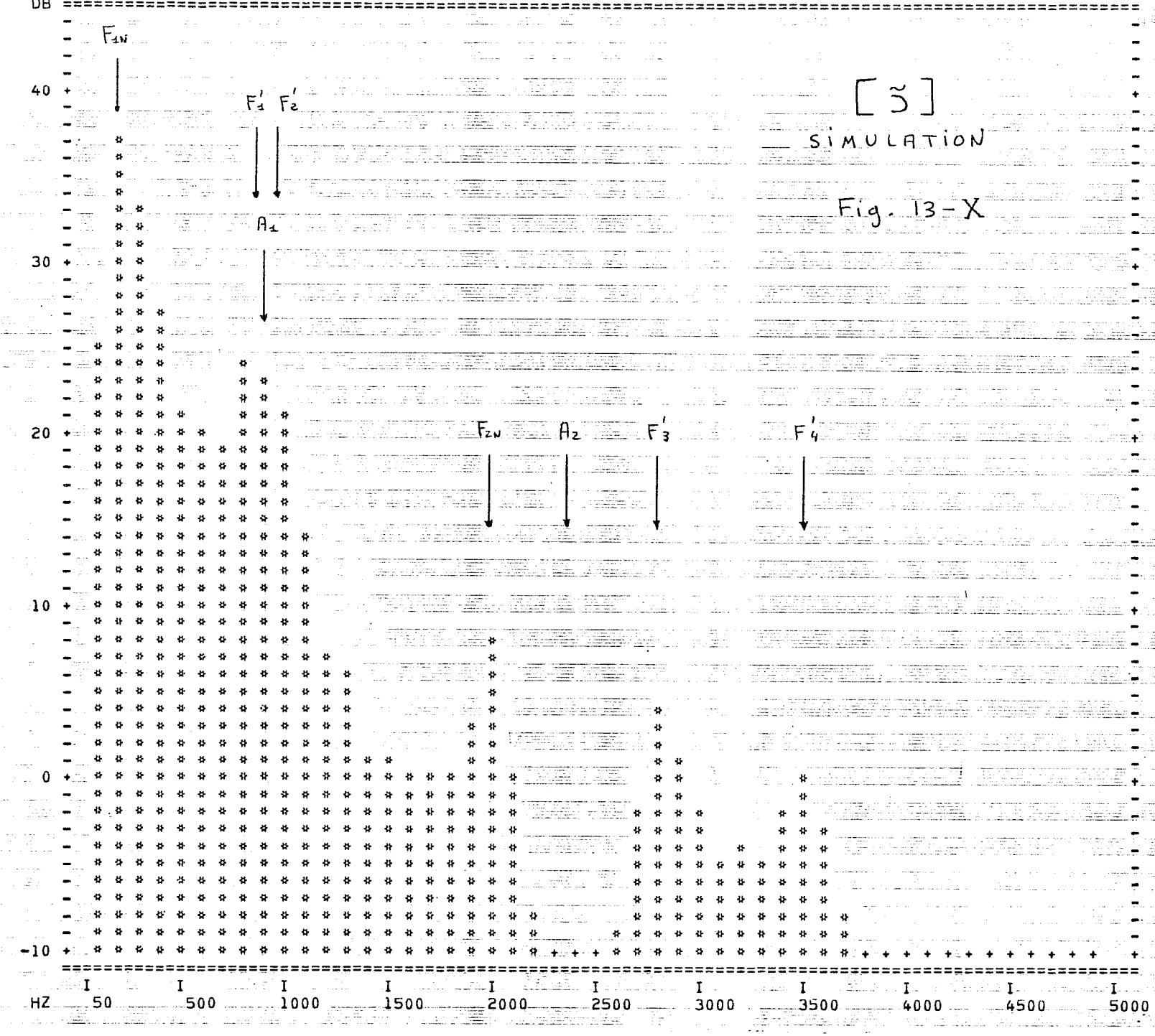
La deuxième ligne indique les largeurs de bande utilisées. Celles-ci sont identiques pour toutes les simulations.



[s̃]

SIMULATION

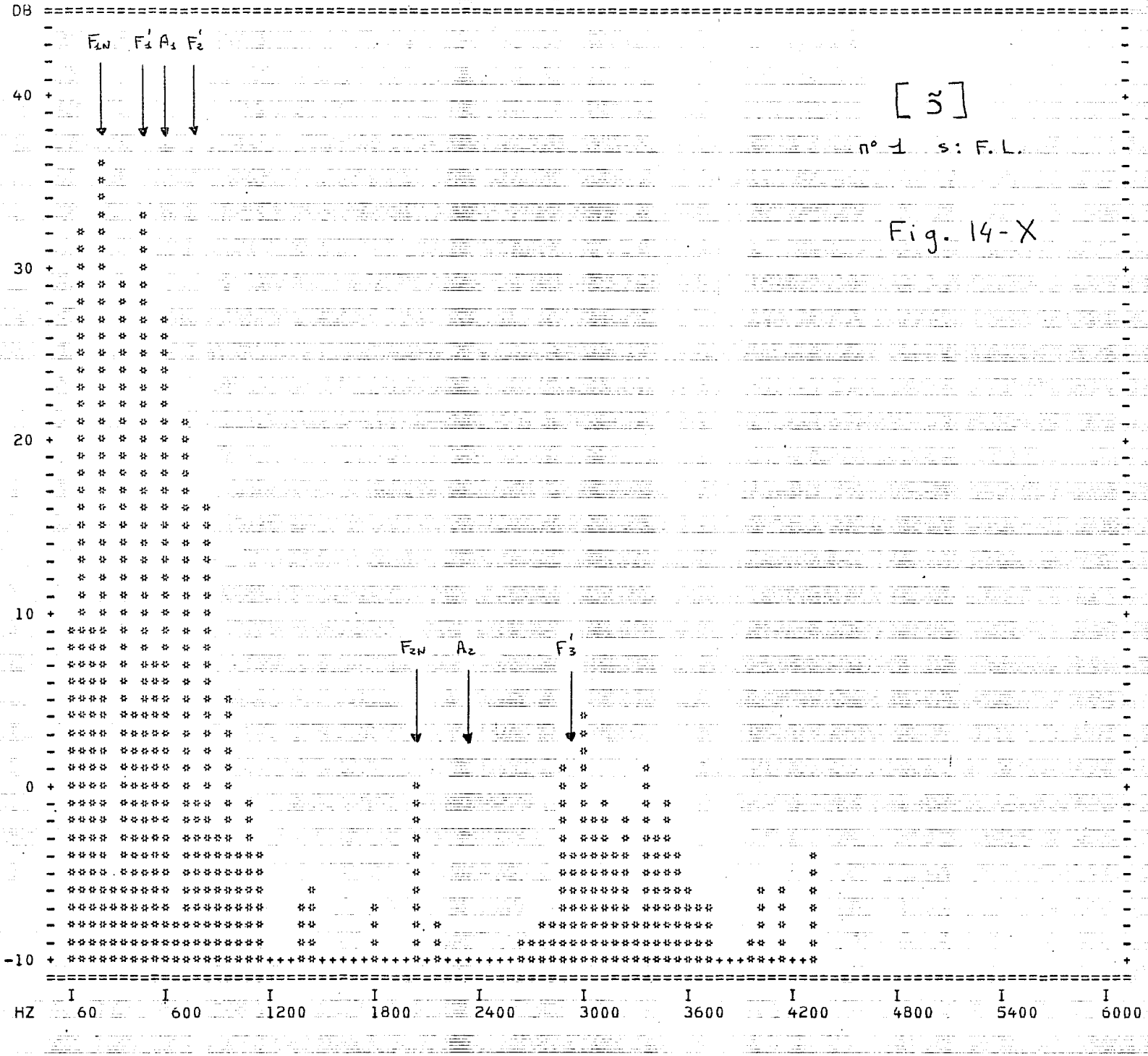
Fig. 12-X



[5]

SIMULATION

Fig. 13-X



[5]

n° 1 s: F. L.

Fig. 14-X

Si l'on compare les spectrogrammes de la fig. 10-X et les résultats de la simulation, l'erreur la plus visible concerne la zone fréquentielle contenant F'1 et F'2. Il faut admettre que, pour le sujet FL, F'1 est à une fréquence inférieure à 700 Hz, alors que le zéro et F'2 sont à la même fréquence, vers 800 Hz. Dans ce cas, l'analyse théorique de MRAYATI n'est plus conforme à la réalité, car A1 est plus élevé que F'1.
Pour le sujet JPZ, le spectre de la fig. 13-X semble relativement proche des sections spectrographiques (en particulier n° 19, fig. 9-X). Les valeurs numériques sont les suivantes :

	F1n	F'1	F'2	F2n	F'3	F'4	A1	A2
Hz	250	850	950	2000	2800	3500	900	2350

Cette simulation confirme les conclusions de MRAYATI (1975) concernant cette voyelle. "Le premier formant F'1 peut disparaître en raison de la proximité du premier zéro nasal A1 et de son rapprochement du 2ème formant F'2 avec lequel il se confond". (p. 14). Une étude plus détaillée est nécessaire pour préciser les fréquences relatives de F'1, F'2 et A1 et savoir s'il est théoriquement possible pour A1 d'être supérieur à F'1.

C/ Conclusions

Nous avons identifié sur la voyelle n° 1 du sujet F.L., analysée par transformée de Fourier, les caractéristiques spectrales essentielles de la voyelle nasale (fig. 14-X) On peut vérifier que la section et le spectre numérique sont bien identiques. La voyelle [ɔ̃] , en plus des deux formants de nasalité communs à toutes les voyelles nasales, se caractérise par un pic unique entre 500 et 800 Hz, composé de deux formants F'1 et F'2 et d'un zéro (A1).

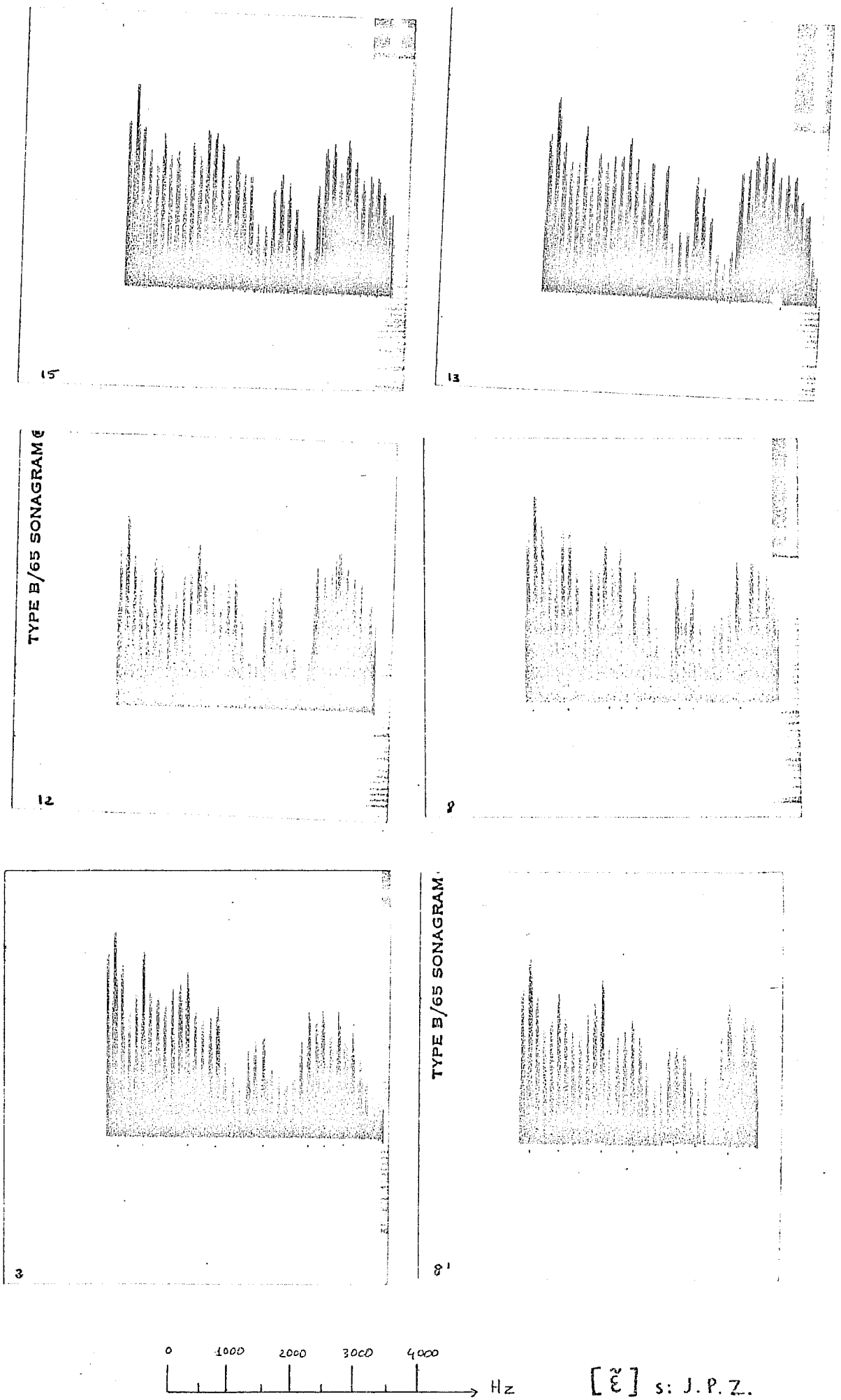
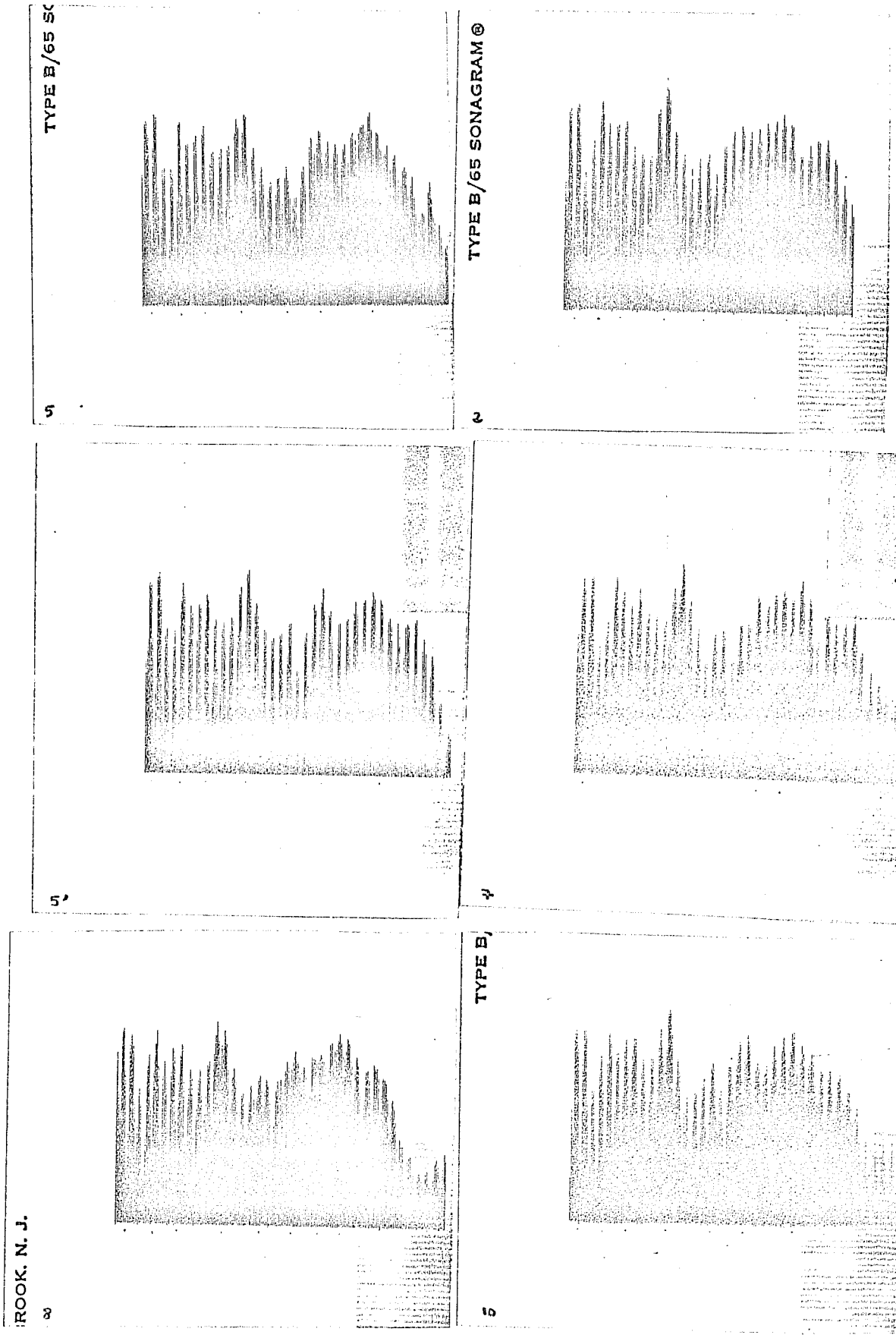
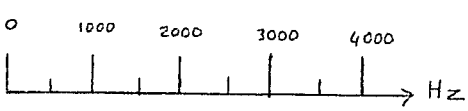

15
13
TYPE B/65 SONAGRAM
12
8
3
TYPE B/65 SONAGRAM
8'
0
1000
2000
3000
4000
Hz
[ɛ̃] s: J.P.Z.

Fig. I5 - X


TYPE B/65 SONAGRAM®
TYPE B/65 SC
ROOK, N. J.
TYPE B,
5
2
5'
4
8
6


0
1000
2000
3000
4000
Hz

[ξ] S: F. L.

Fig 16 - X

V - LA VOYELLE [ɛ̃]

A/ Description spectrale

Les 12 spectrogrammes sont reproduits aux fig. 15-X et 16-X.

1)- Chez le sujet JPZ, la première harmonique est la plus intense. Pour le sujet FL, le fondamental et la 1ère harmonique sont d'égale intensité.

2)- Si les sections du sujet FL présentent ensuite deux pics (P2 et P3) centrés sur la 4ème et 7ème harmoniques, celles de JPZ ne possèdent qu'un pic (P2) dans cette zone de fréquence. Il est centré sur la 5ème harmonique sauf pour les voyelles n° 13 et 15 où l'on retrouve une légère irrégularité spectrale au niveau de la 7ème harmonique.

3)- Le pic suivant (P4) est particulièrement intense et bien marqué, sur tous les tracés, vers 1400-1500 Hz.

4)- Chez le sujet JPZ on trouve ensuite une irrégularité spectrale dans 2 sections (P4'). Son statut formantique est douteux.

5)- Le pic P5, entre 1900 et 2200 Hz selon le sujet, est visible sur la totalité des sections.

6)- Quant au pic P6, il correspond au début d'une zone d'harmoniques renforcées relativement intenses, qui présente d'autres extrema à des fréquences plus élevées (P7).

On trouvera dans les deux tableaux suivants les mesures effectuées sur les spectrogrammes, ainsi que les moyennes de ces mesures.

Sujet FL	7	10	8	2	5	5'	Moyenne
P1	290	310	305	285	305	305	300 Hz
P2	625	665	640	640	645	645	645
P3	980	1020	1020	980	980	980	995
P4	1585	1520	1525	1565	1545	1585	1555
P5	2185	2145	2155	2175	2195	2195	2175
P6	2730	2645	2660	2680	2680	2655	2675
P7	3310 3685	3270 -	3285 3830	3265 3855	3390 4270	3390 4020	-

Sujet JPZ	13	12	8	8'	15	2	Moyenne
P1	275	235	255	265	265	250	260 Hz
P2	700	680	745	740	725	745	720
P3	960	-	-	-	960 (1170)	-	-
P4	1425	1380	1340	1450	1445	1420	1410
P4'	1785	-	1615 1850	-	-	-	-
P5	2000	1935	2065	1925	1890	1910	1965
P6	2530	2660	2530	2615	2585	2605	2585
P7	3595 3980	3255 3640	3470 3725	3390 3520	3365 3640	3325 3555	-

B/ Essai d'interprétation des caractéristiques spectrales

1)- Comparaison avec les travaux antérieurs

Il existe des différences non négligeables avec les valeurs de DEBROCK. Celles-ci concernent l'identification du pic (P2 ?) à 460 et 430 Hz, ainsi que l'identification des maxima entre 1800 et 1900 Hz (P4 ou P5 ?) .

	S1	S2	S3	Moyenne JPZ	Moyenne FL
P1	200	-	210	265	300 Hz
P2	700	460	430	720	645
P3	-	900	1110	-	995
P4	1900	1880	1800	1410	1555
P6	2600	2800	2680	2585	2675
P7	3450	3370	3850	-	-

De même l'interprétation des données de GUTH n'est pas chose aisée. Si les premières valeurs correspondent sans doute à P1, les valeurs suivantes sont très nettement supérieures à P2 ou P3 et pourraient correspondre à P4, sauf la

valeur du sujet S4. Pour la ligne suivante, on peut proposer P5. Quant aux valeurs supérieures, il ne nous a pas semblé possible de les identifier.

	S1	S2	S3	S4	Moyenne JPZ	Moyenne FL
P1	350	350	400	350	265	300 Hz
P4	1350	1300	1400	1000	1410	1555
P5	1900 2350 2900 3500	1700 2800 3500	2000 2800 3700	1800 2200 2700 3150	1965	2175

Les données de DELATTRE (250, 600 et 1750 Hz) correspondent sans doute à P1, P2 et P4 avec un large écart pour les valeurs de P4.
Par contre, les données de MRAYATI confirment dans une large mesure les chiffres que nous avons obtenus.

	Contexte ptk	Contexte bdg	Contexte mnɲ	Moyenne JPZ	Moyenne FL
P1	409	430	413	265	300 Hz
P2	735	718	754	720	645
P4	1663	1649	1697	1410	1555
P5	3155	3119	3004	-	-

En règle générale, on notera que P4 est systématiquement plus élevé dans les travaux antérieurs. La voyelle est un bon exemple des difficultés d'identification des paramètres pertinents observés lorsqu'on ne dispose pas d'un modèle théorique adéquat.

2)- Essai d'interprétation

L'analyse théorique de MRAYATI (fig. 17-X) porte sur une voyelle orale dont les formants s'établissent à 500, 1750 et 2450 Hz. On obtient les valeurs suivantes pour les différents degrés de couplage :

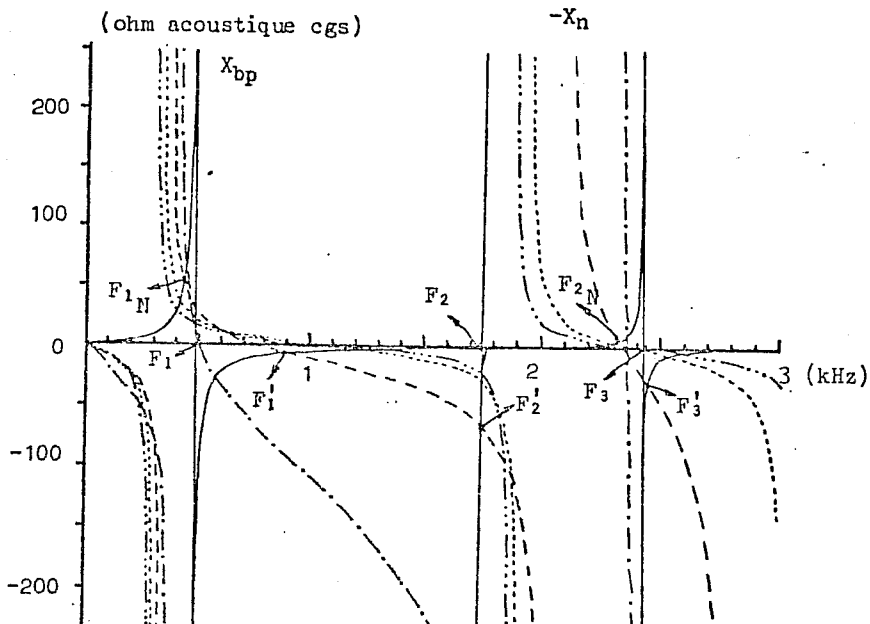
Degré de couplage	F1n	A1	F'1	F'2	F2n	A2	F'3
0.27 cm^2	450	500	575	1750	2250	2350	2450
1.66 "	435	750	900	1750	2250	2350	2450
3.5 "	415	1000	1250	1750	2250	2350	2525
5.0 "	410	1100	1300	1750	2250	2350	2600

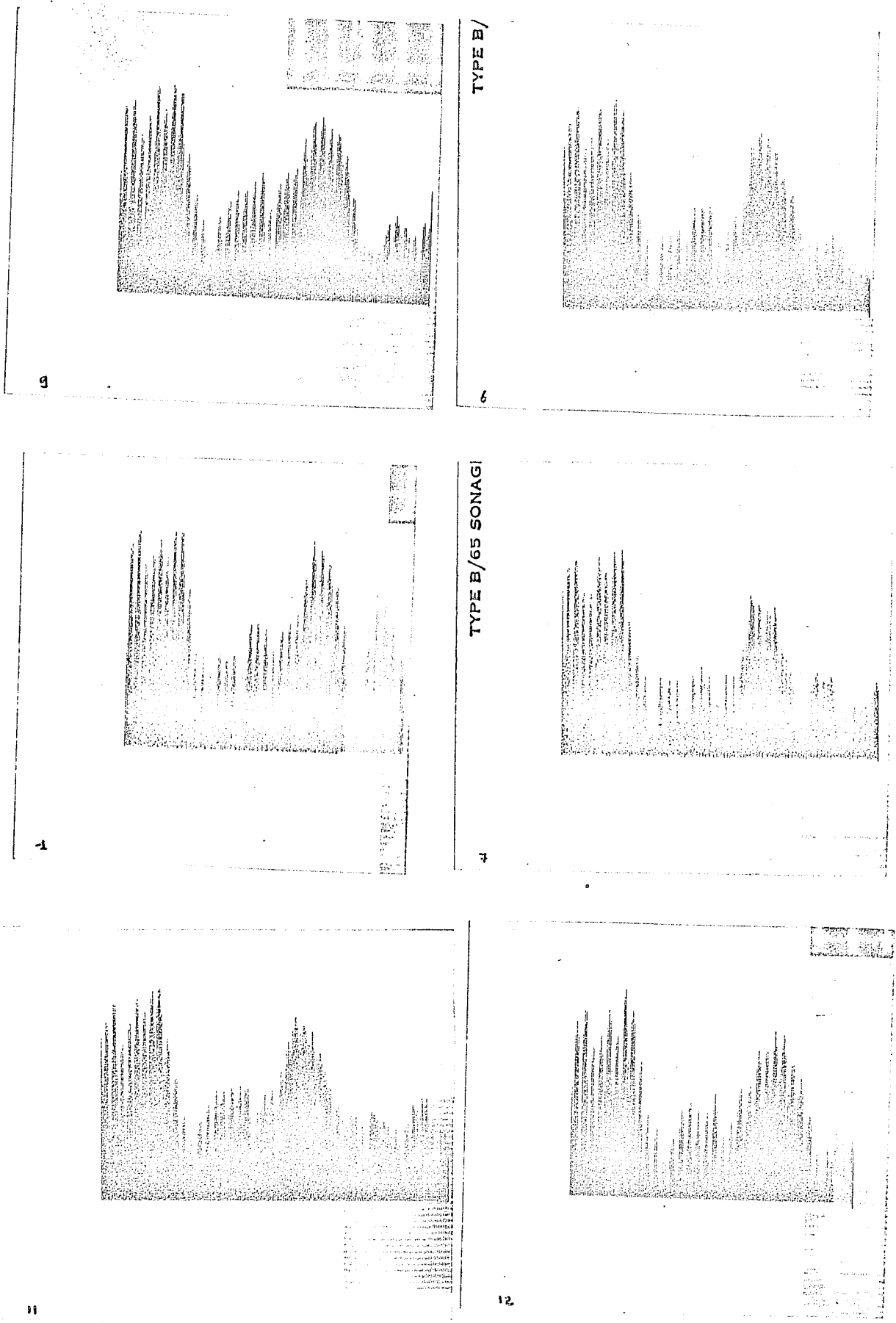
Comme pour la voyelle [ɔ̃] , il est difficile d'estimer avec précision la fréquence de A1 et F'1 pour les degrés de couplage élevés. Les chiffres ci-dessus conduisent à l'interprétation suivante des différents pics : P1 doit être identifié avec F1n. Quant à P2, on peut faire l'hypothèse qu'il correspond à F'1. Dans ce cas P3 ne constitue pas une caractéristique pertinente de la voyelle [ɛ̃] . P4 s'identifie alors avec F'2. Le deuxième formant nasal (F2n) est représenté par P5, alors que P6 n'est autre que F'3.

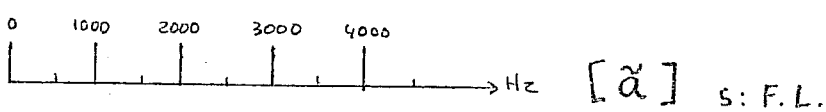
Pour une voyelle notée [ɛ̃] , les observations de FANT (1960, p. 154) ne correspondent pas à cette analyse théorique. La fréquence de F'1 passe de 420 à 490 Hz avec un élargissement de la largeur de bande. On n'observe pas de changement pour F'2 et les formants supérieurs. Il y a apparition d'un formant additionnel vers 1100 Hz, suivi d'un zéro vers 1350 Hz. Une aire de couplage plus faible diminue fortement l'importance de cette paire pôle/zéro supplémentaire. Aucune de ces caractéristiques ne se retrouve dans les sections analysées.

Voyelle [ɛ̃]

Fig. I7 - X







[ă] S: F.L.

Fig. 20 - X

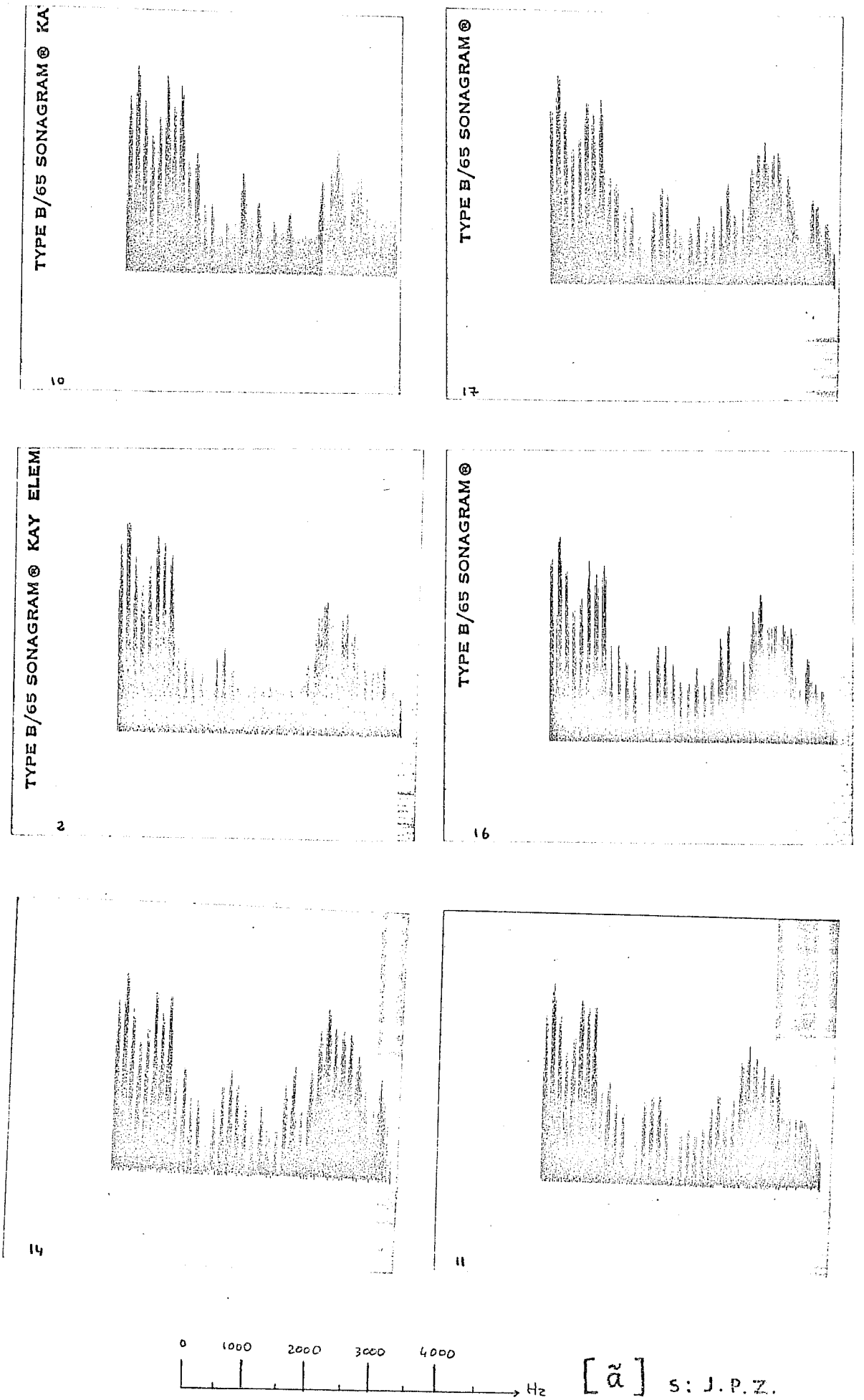

TYPE B/65 SONAGRAM® KA
10
TYPE B/65 SONAGRAM®
17
TYPE B/65 SONAGRAM® KAY ELEM
2
TYPE B/65 SONAGRAM®
16
14
11
0
1000
2000
3000
4000
Hz
[ã]
S: J.P.Z.

Fig. 2I - X

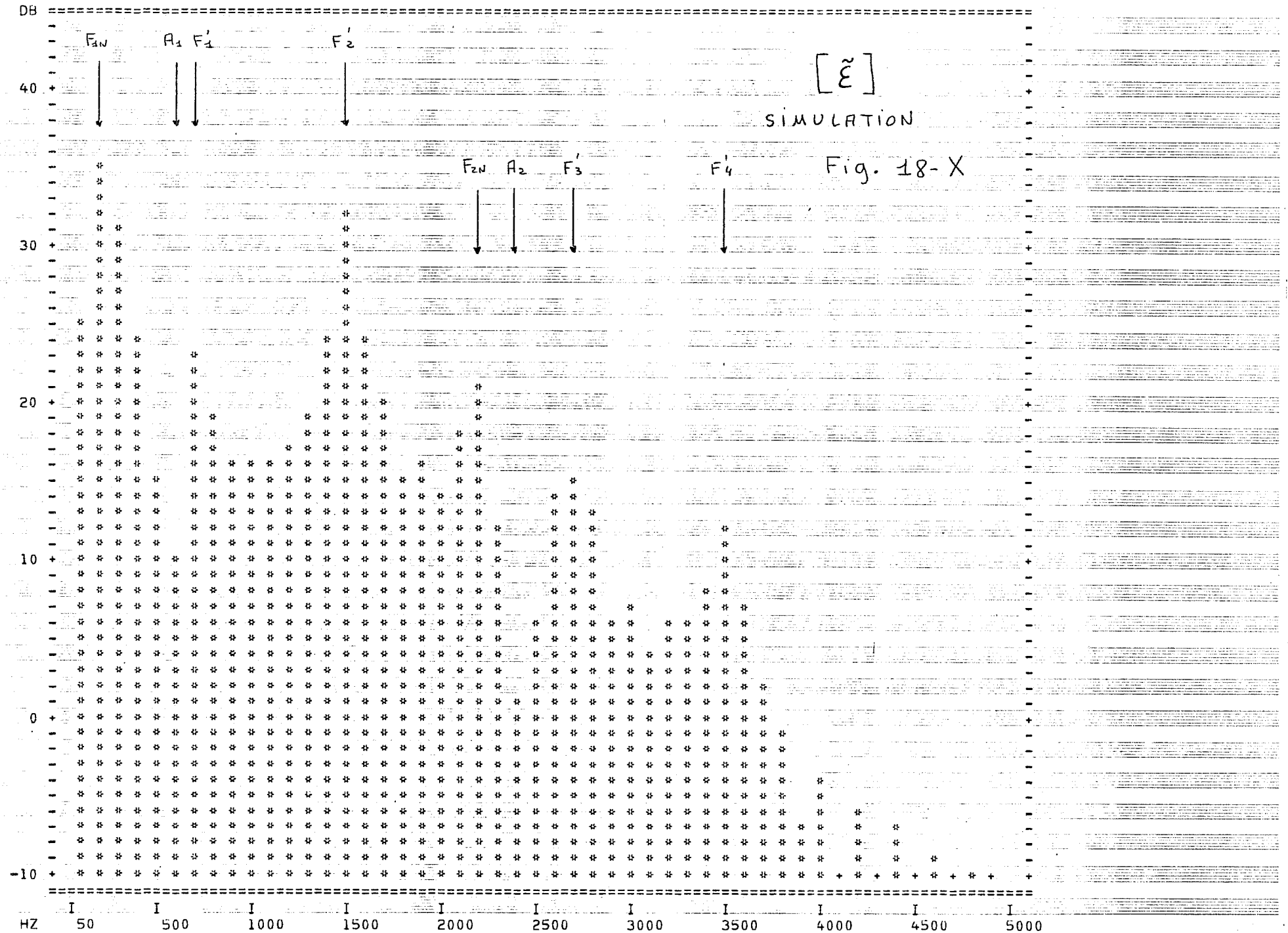
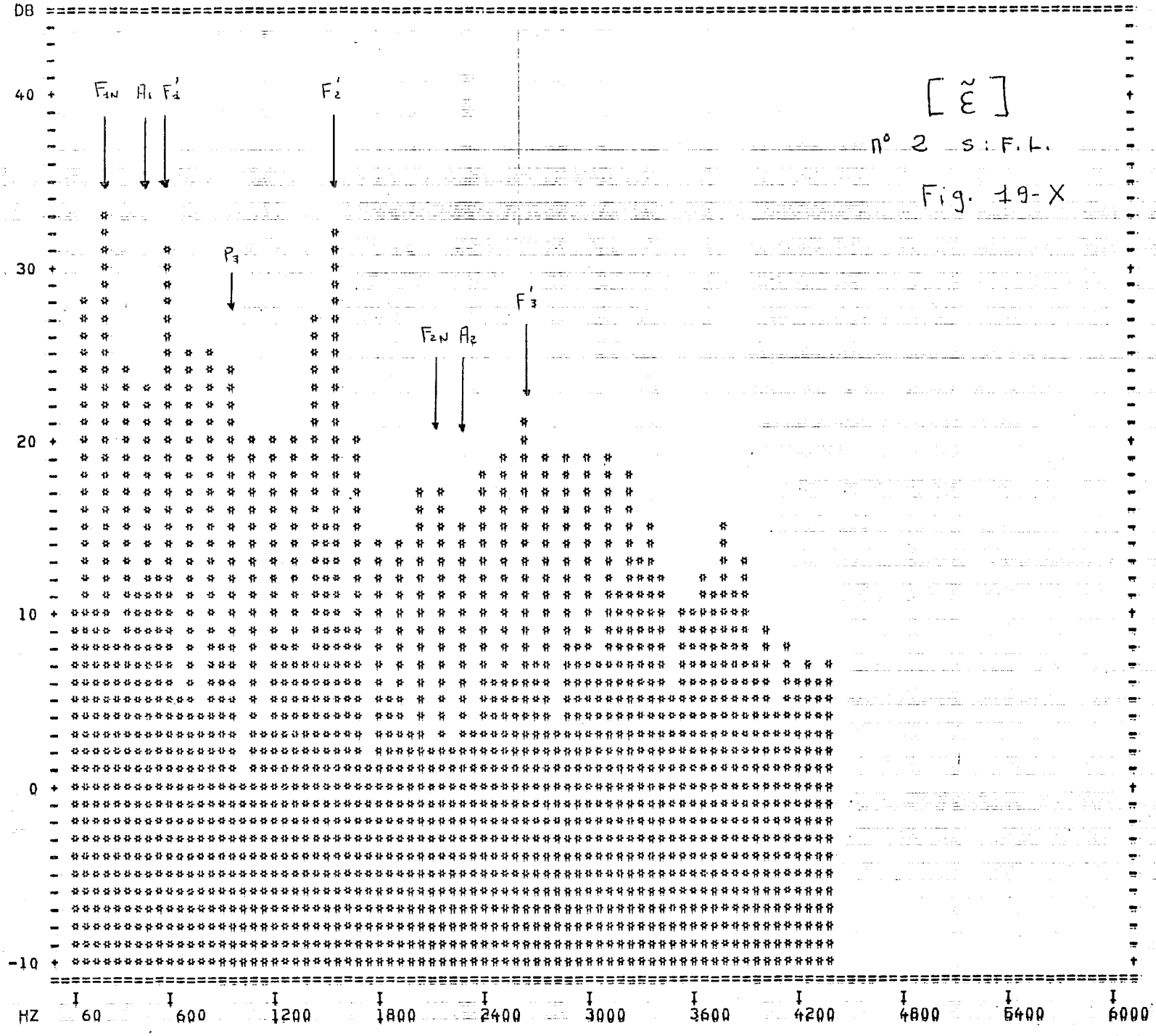

DB
40
30
20
10
0
-10
F1N
A1
F'1
F'2
F2N
A2
F'3
F'4
[ε̃]
SIMULATION
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Fig. 18-X



[ε̃]

n° 2 s : F.L.

Fig. 19-X

3)- Essai de simulation

Nous avons obtenu des résultats satisfaisants pour les valeurs numériques suivantes :

F1n	F'1	F'2	F2n	F'3	F'4	A1	A2
250	700	1500	2200	2700	3500	600	2400

La répartition de l'intensité des harmoniques est particulièrement proche de celle que l'on peut observer pour la voyelle n° 8' du sujet JPZ. Le pic P3 observé chez le sujet FL est sans doute dû à une paire pôle/zéro supplémentaire associée aux cavités sinusales. Sa faible intensité serait due à la proximité des fréquences du pôle et du zéro. Le résultat de la simulation est présenté à la figure 18-X.

C/ Conclusions

Nous avons identifié sur le spectre numérique de la voyelle n° 2 (sujet FL) les caractéristiques essentielles de la voyelle nasale (fig. 19-X). On note à nouveau une très bonne* spectrographique. La position des zéros n'est marquée que par des "vallées" peu profondes, mais leur présence se révèle indirectement par la faible intensité de F'1 et par la faible pente générale du spectre. On peut observer chez un sujet la trace d'une paire pôle/zéro supplémentaire (P3) associée aux cavités sinusales comme le prévoit le travail de LINDQVIST-GAUFFIN et SUNDBERG (1976). La voyelle [ɛ̃] se caractérise principalement par un premier formant (F'1) vers 600-700 Hz et par un deuxième formant relativement intense vers 1400-1550 Hz pour nos deux sujets. Il n'est pas impossible que cette deuxième fréquence soit un peu plus élevée (1600-1700 Hz) pour une population plus large.

VI - LA VOYELLE [ã]

A/ Description spectrale

Les 12 spectrogrammes relatifs à la voyelle sont présentés aux fig. 20-X et 21-X.

* correspondance avec le spectre obtenu par analyse numérique et la section

1)- Le premier pic (P1) est formé par la première harmonique dans tous les cas.

2)- On observe ensuite un groupe de 3 ou 4 harmoniques renforcées qui forment deux pics. Pour le sujet JPZ, P2 est centré sur la 5ème harmonique dans tous les cas. Le pic P3 est visible au niveau de la 7ème harmonique pour toutes les voyelles, à l'exception du n° 2. Pour le sujet FL, le pic P2 est centré sur la 4ème harmonique, alors que le pic P3 se place sur la 6ème ou entre la 6ème et la 7ème harmonique.

3)- Si l'on néglige une légère irrégularité spectrale surtout visible chez le sujet JPZ (voyelles n° 10, 14 et 11), les pics suivants (P4 et P5) sont caractérisés par une faible intensité. Ils sont peu visibles sur certaines sections, et leur fréquence est relativement variable.

4)- Cette zone instable est suivie par un groupe d'harmoniques très intenses, dans laquelle on distingue un nombre variable de maxima (P6).

Le tableau suivant présente les mesures fréquentielles des différents pics, ainsi que, le cas échéant, les moyennes calculées individuellement pour chaque sujet.

Sujet JPZ	17	16	10	11	14	2	Moyenne
P1	280	300	265	275	300	280	280 Hz
P2	700	760	725	725	750	750	735
P3	915	1000	960 (1190)	940 (1170)	1000 (1230)	960	965
P4	1890	1890	1900	1850	1930	1800	1875
P5	2435	2500	(2130)	2405	2435	-	2475
P6	2880 3450	2880 3390	3100 3310	2870 3340	2900 3410	3325	-

Sujet FL	7	12	11	9	6	1	Moyenne
P1	290	270	270	290	285	285	280 Hz
P2	645	645	645	625	625	580	630
P3	940	895	935	920	855	875	905
P4	(1690) 1900	1850	1810	1855	1820	1755	1835
P5	-	2230	2190	2230	2150	2005	2160
P6	2895 3230	2850 3100	3020	3060	2995	2890	-

B/ Essai d'interprétation des caractéristiques spectrales

1)- Comparaison avec les travaux antérieurs

Les valeurs publiées par DEBROCK s'inscrivent parfaitement dans les catégories P définies plus haut, à l'exception des valeurs 2150, 2200 et 1830 Hz dont il est difficile de préciser si elles correspondent à P4 ou P5.

	S1	S2	S3	Moyenne JPZ	Moyenne FL
P1	200	-	200	280	280 Hz
P2	700	630	600	735	630
P3	860	1050	920	965	905
P4/P5	2150	2200	1830	1875/2475	1835/2160
P6	3200	3300	3000	-	-

Au contraire, les valeurs de GUTH sont difficilement interprétables. Les valeurs de P2 semblent beaucoup trop basses. Aucun pic correspondant à P3 n'est donné pour le sujet n° 4 ; ce dernier atteindrait 1200 Hz pour S1.

	S1	S2	S3	S4	Moyenne JPZ	Moyenne FL
P2	600	450	500	400	735	630 Hz
P3	1200	1150	950	-	965	905
P4	1900	-	1900	1700		
	-	-	-	2100		
	-	-	2700	2700		
	-	3300	3100	3100		
	3500	-	3600	3500		

Les valeurs de DELATTRE coïncident parfaitement avec les valeurs que nous avons mesurées : 250, 600 et 950 Hz.

Pour P2, les valeurs de MRAYATI sont très proches de la valeur mesurée pour le sujet FL ; pour P3, l'écart entre la moyenne du sujet JPZ et la valeur de MRAYATI est inférieur à 100 Hz.

	contexte ptk	contexte bdg	contexte mn ɲ	moyenne JPZ	moyenne FL
P1	381	407	368	280	280 Hz
P2	619	597	587	735	630
P3	1057	1039	1055	965	905
P6	3274	3252	3180	-	-

A l'exception des valeurs de GUTH (1975), on notera le bon accord entre l'ensemble des valeurs publiées.

2)- Essai d'interprétation

L'analyse graphique de MRAYATI, reproduite à la fig. 22-X, conduit aux valeurs suivantes pour la nasalisation d'une voyelle orale dont les formants s'établissent à 600, 1250 et 2450 Hz.

Degré de couplage	F1n	A1	F'1	F'2	F2n	A2	F'3
0.27 cm^2	475	500	550	1250	2350	2350	2450
1.66 "	460	650	800	1250	2250	2350	2550
3.5 "	450	1000	1100	1250	2125	2350	2750
5.0 "	450	1100	1200	1250	2050	2350	2850

A nouveau, il est difficile d'estimer numériquement les fréquences de A1 et F'1. En outre, la valeur de F'2 semble trop forte.

L'interprétation des différents pics relevés sur les spectrogrammes ne pose guère de problèmes à l'exception de l'interprétation de P4 et de P5.

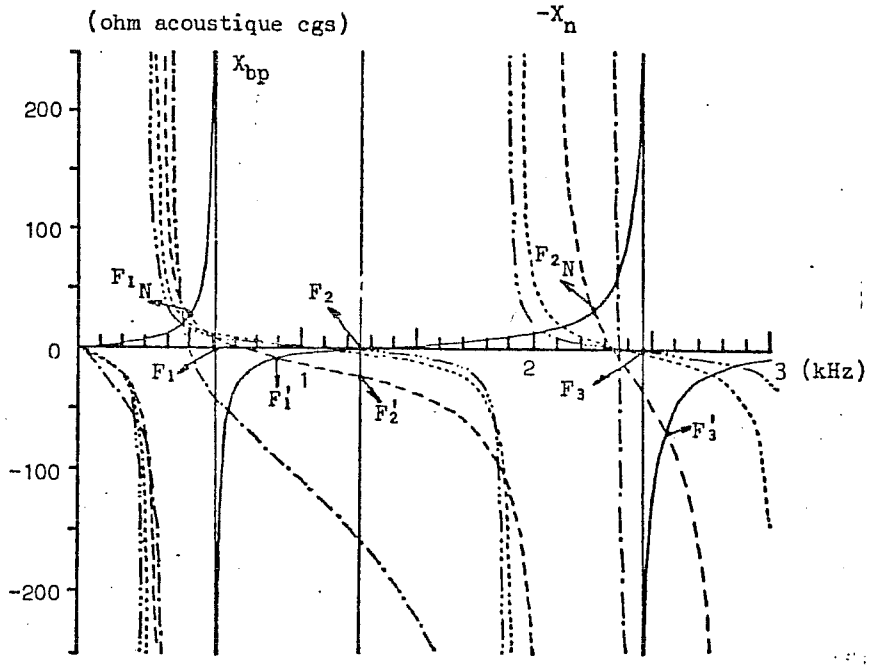


Fig. 22 - X

- Voyelle [ã]

Spectre calculé de la voyelle [ã]

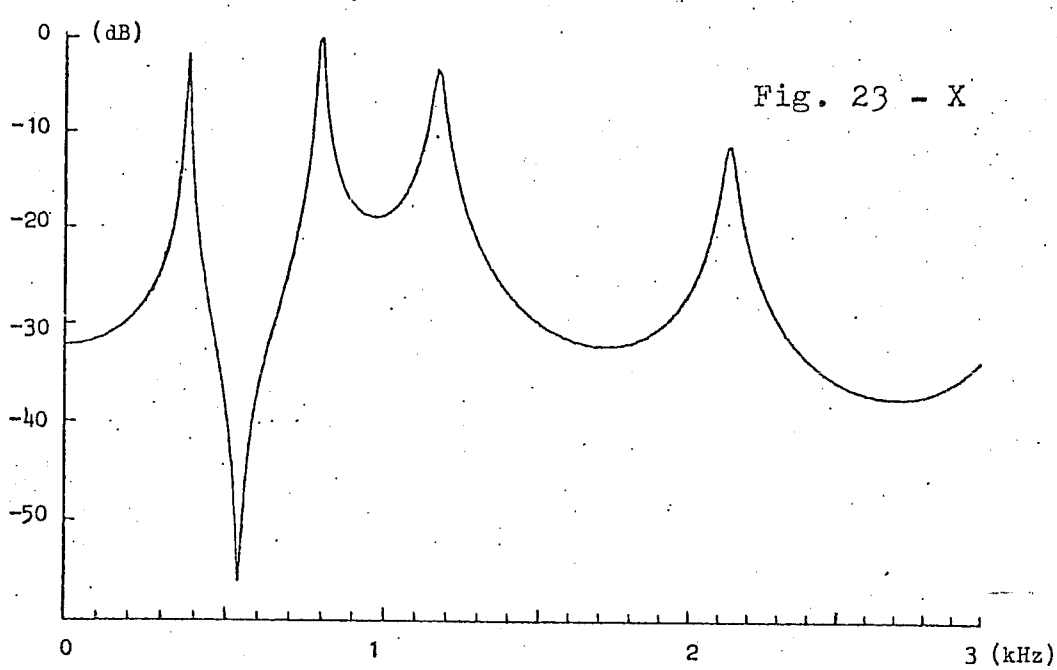


Fig. 23 - X

P1 correspond à F1n. Un zéro (A1) précède immédiatement F'1, qui apparaît comme P2 sur les sections. P3 correspond à F'2. Il est difficile de savoir s'il faut attribuer F2n à P4 ou P5. Nous pourrions être en présence, et ceci est vrai également pour les irrégularités spectrales du sujet JPZ vers 1200 Hz, d'une paire pôle/zéro supplémentaire due à l'asymétrie des conduits nasaux et aux cavités sinusales. Une autre solution consiste à considérer P4 comme la trace de F2n et d'attribuer P5 à F'3. Quant à P6, il s'agit, soit de F'3 et F'4 dont la proximité expliquerait l'intensité importante du groupe des harmoniques renforcées, soit de la combinaison de F'4 avec des formants de rang supérieur. Cette dernière solution est la plus plausible pour les autres voyelles nasales. Il faut sans doute expliquer ainsi l'intensité des harmoniques dans la zone fréquentielle supérieure à 3000 Hz. Certains de ces formants supérieurs pourraient être dus à la nasalité (F3n,...)

3)- Essai de simulation

MRAYATI (1975, p. 17) propose une simulation du spectre de la voyelle [ɑ̃] , que nous reproduisons à la fig. 23-X. Aucune précision n'étant donnée quant à la méthode ou aux valeurs numériques employées, il est difficile de savoir pourquoi la "vallée" produite par le premier zéro est tellement marquée alors que l'effet du deuxième zéro est invisible. L'intensité du premier formant est trop forte, ce qui laisse à penser que la fréquence et la largeur de bande du premier zéro n'ont pas été choisies de façon adéquate. Il nous semble que les deux exemples de simulations que nous présentons à la fig. 24-X et 25-X donnent une approximation plus correcte des tracés spectrographiques et de l'analyse numérique. Les valeurs numériques utilisées sont présentées ci-dessous :

	F1n	A1	F'1	F'2	F2n	A2	F'3	F'4
fig. 24	250	600	700	950	1850	2000	2300	3500
fig. 25	350	550	700	950	1850	2000	2300	3500

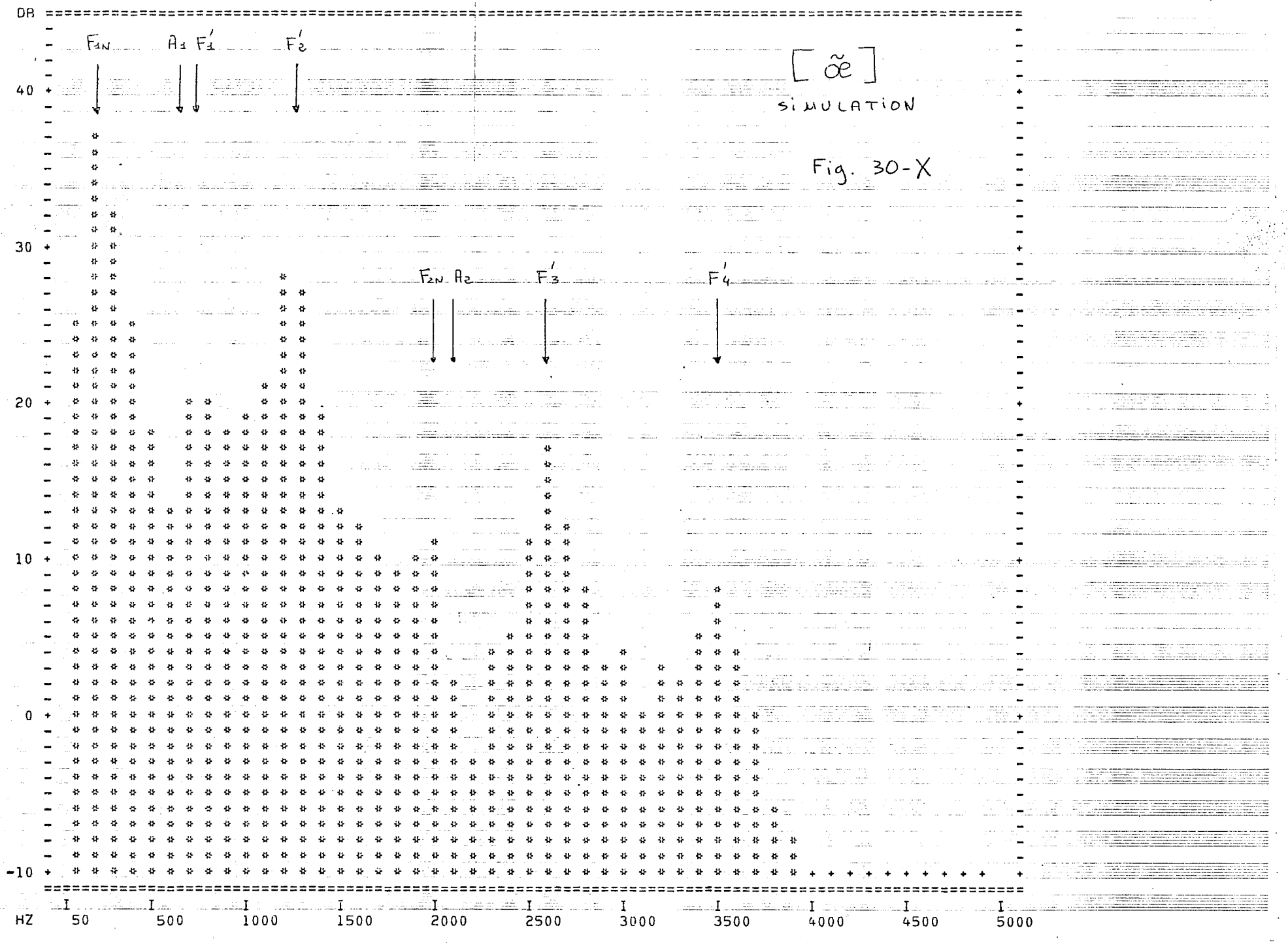

DB
40
30
20
10
0
-10
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
F1N
A1
F'1
F'2
F2N
A2
F'3
F'4
[œ̃]
SIMULATION

Fig. 30-X

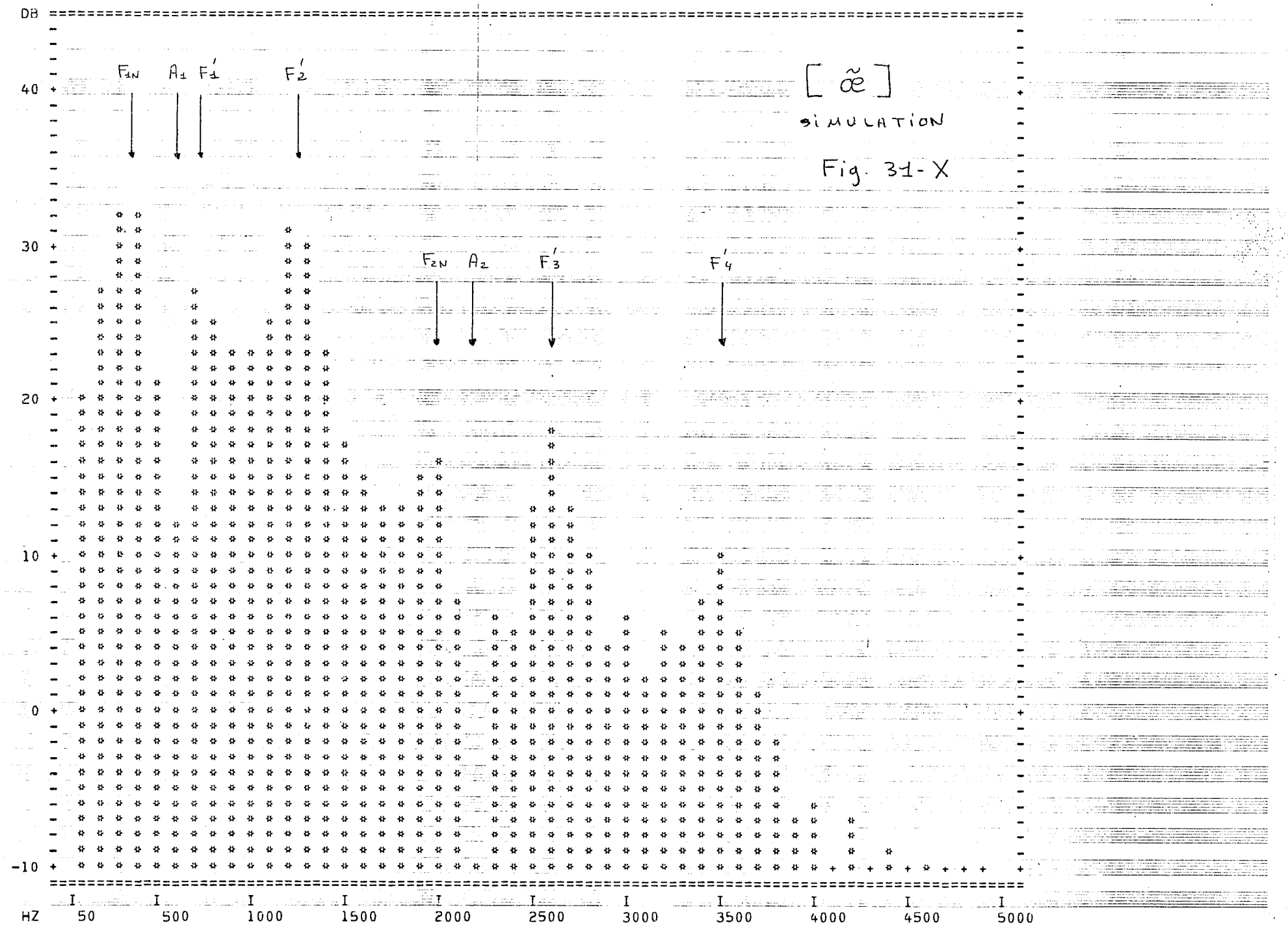
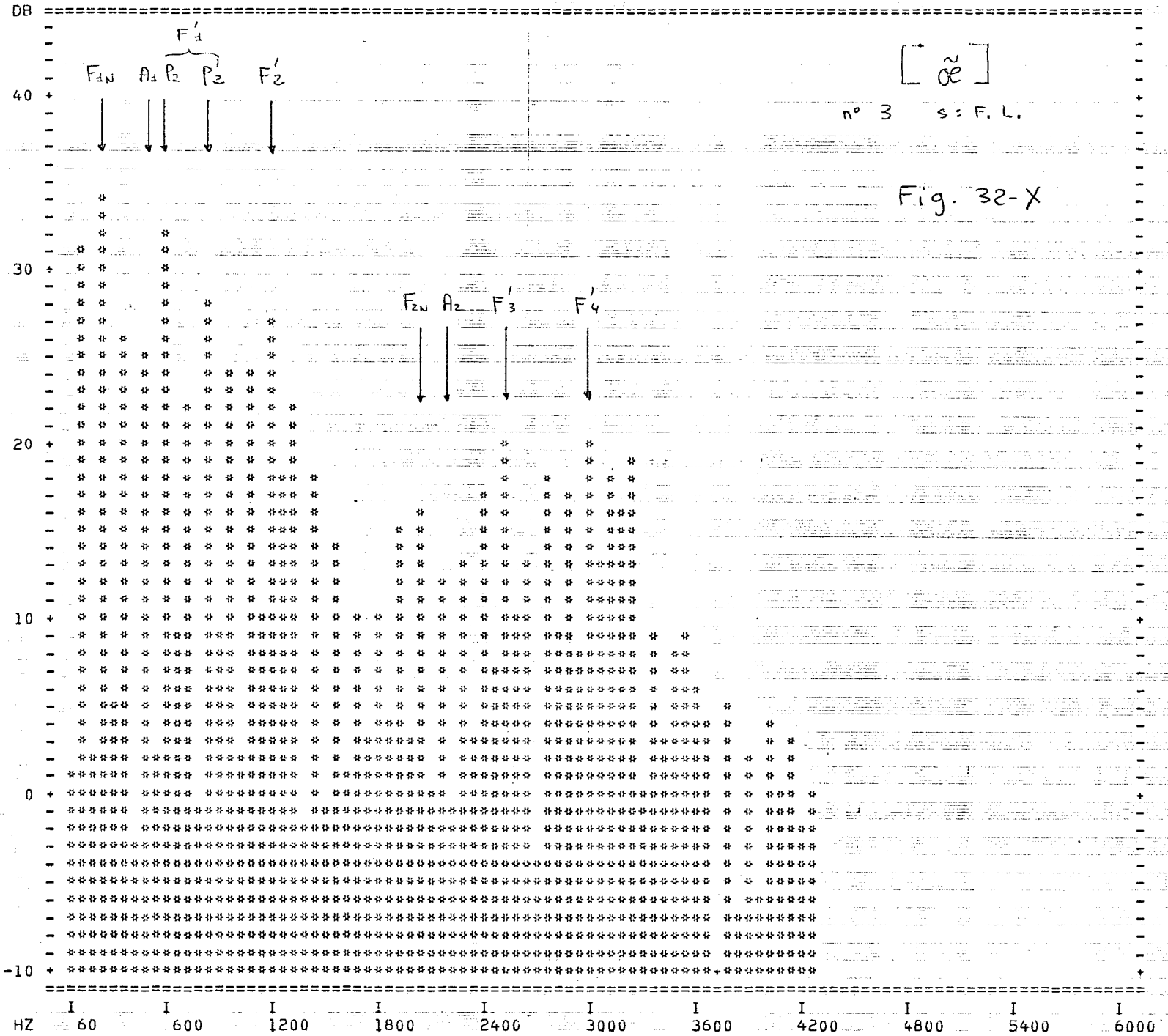

DB
40
30
20
10
0
-10
F1N
A1
F'1
F'2
F2N
A2
F'3
F'4
[œ]
SIMULATION
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

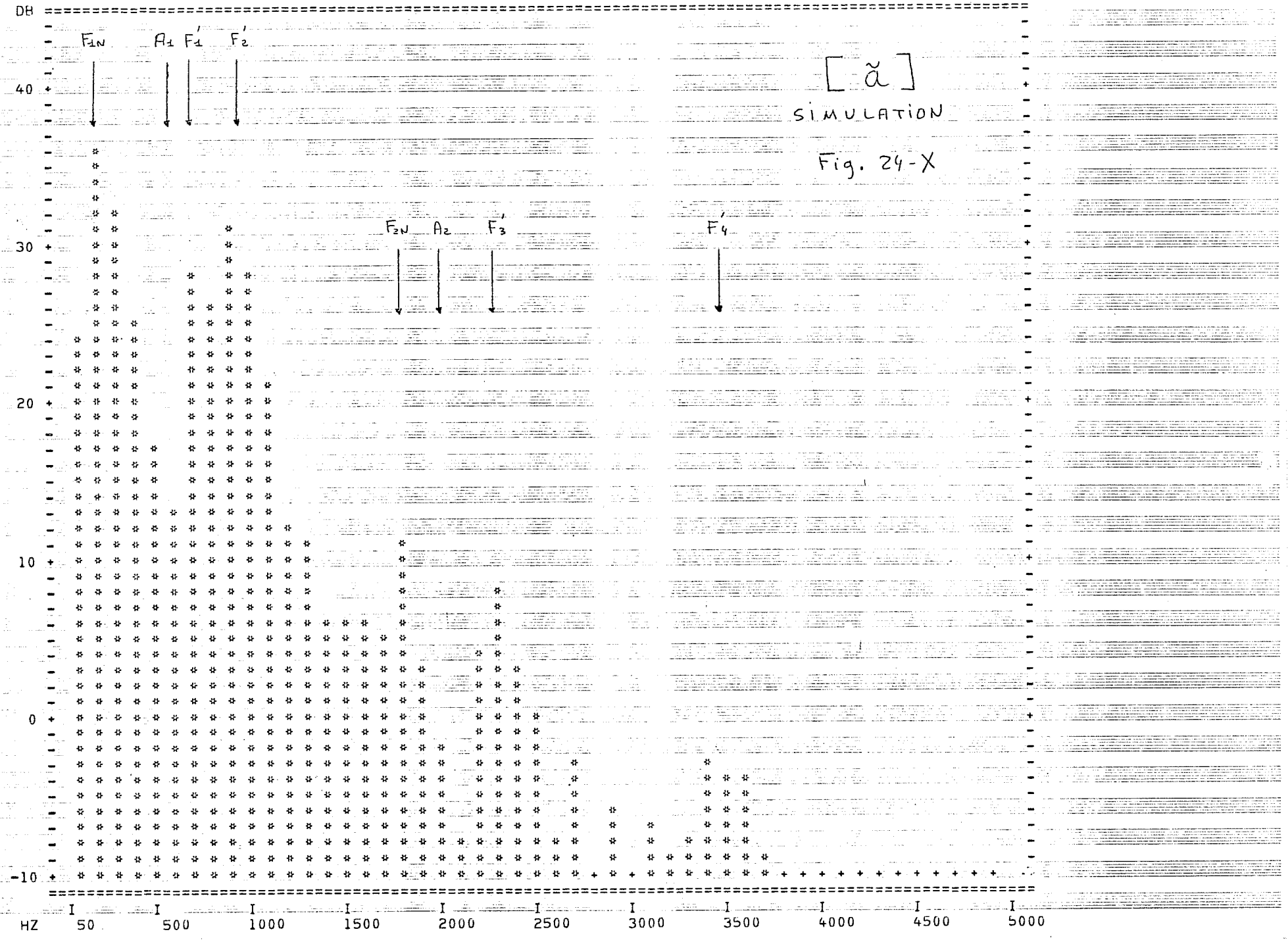
Fig. 31-X



[œ̃]

n° 3 s: F. L.

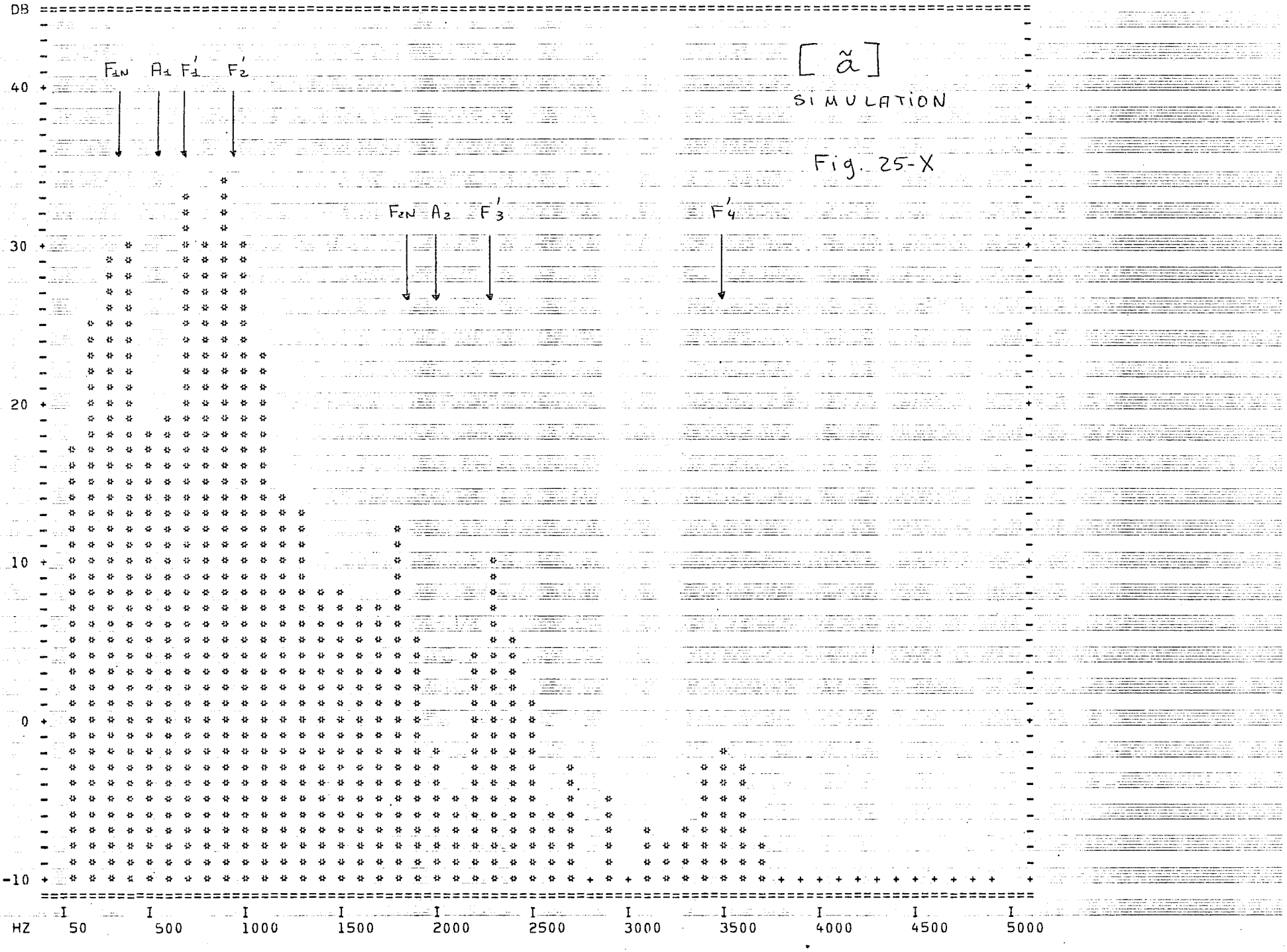
Fig. 32-X



[ă]

SIMULATION

Fig. 24-X


[ã]
SIMULATION
Fig. 25-X
F1N A1 F'1 F'2
F2N A2 F'3
F'4
DB
40
30
20
10
0
-10
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

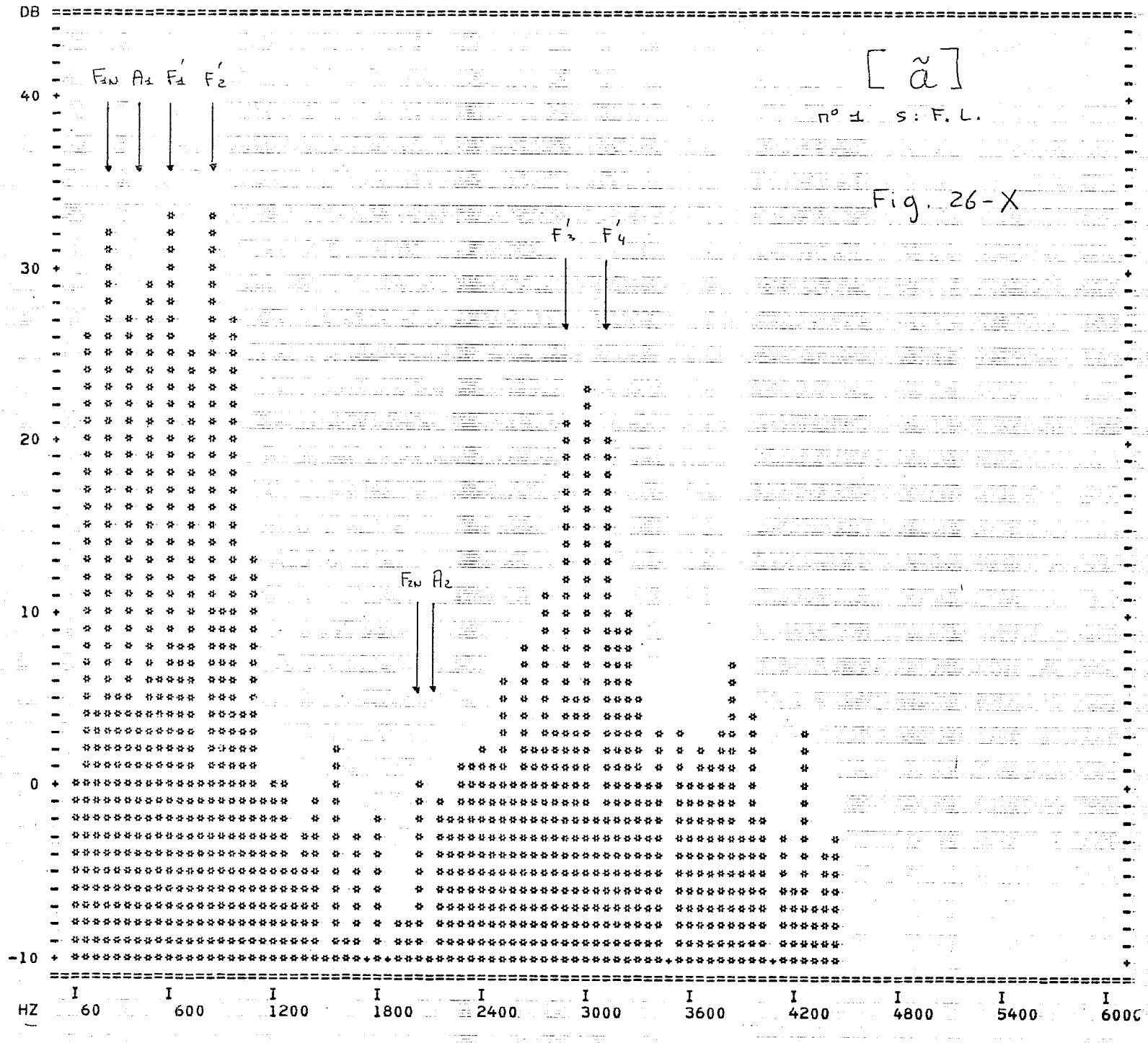

DB
40
30
20
10
0
-10
HZ
60
600
1200
1800
2400
3000
3600
4200
4800
5400
6000
FzN A1 F'1 F'2
F'3 F'4
FzN A2
[ã]
nº 1 S: F. L.

Fig. 26-X

. Dans les deux cas, l'amplitude relative des harmoniques de la zone P2-P3 (c'est-à-dire F'1-F'2) est bien reproduite. Par contre au-delà de 2000 Hz, l'amplitude des harmoniques est trop faible. Ceci indique, soit, la présence de pôles supplémentaires au niveau de F'4, soit que l'identification de P5 avec F'3 (que nous avons retenue pour cette simulation) est erronée.

FANT (1960, p. 153-155) a analysé en détail la voyelle [ã] . Un formant nasal (F2n) à 2000 Hz, ainsi qu'un zéro entre F2n et F'3, apparaissent nettement. Le premier formant nasal (F1n) apparaît faiblement vers 450 Hz, alors que F1 passe de 630 à 700 Hz sous l'effet de la nasalisation. F'2 reste stable. Ces valeurs numériques confirment les données de MRAYATI. F1n est cependant trop peu marqué dans la simulation de FANT. Le meilleur spectre publié pour la nasale [ã] apparaît dans FUJIMURA et LINDQVIST (1971). On distingue parfaitement F1n à 400 Hz, suivie de A1 vers 475 Hz ; F'1 et F'2 ont la même amplitude à 750 et 1100 Hz. F2n est visible vers 2100 Hz, tout comme A2 (2250 Hz) et F'3 (2500 Hz). La fréquence de F'2 apparaît un peu élevée. En outre, il ne faut pas perdre de vue que les sujets suédois utilisés ne possèdent pas de voyelles phonologiquement nasales et qu'ils ne possèdent donc pas une grande habitude de la nasalisation volontaire. On ne peut donc rien dire du degré de nasalisation. En ce qui concerne le lieu d'articulation de cette voyelle, on sait que le deuxième formant de ce sujet est placé à 1100 Hz. Ceci confirme l'invariance de F'2 par rapport à la nasalisation.

C/ Conclusions

Comme pour les voyelles précédentes, nous tentons d'interpréter à la fig. 26-X l'analyse numérique de la voyelle n° 1 du sujet FL. Si l'identification de F1n, F'1 et F'2 ne pose pas de problème, celle de F2n et A2 est plus problématique. On observera d'ailleurs, pour cette zone, une différence entre la section et l'analyse numérique, qui peut s'expliquer par le fait que les deux analyses n'ont pas été

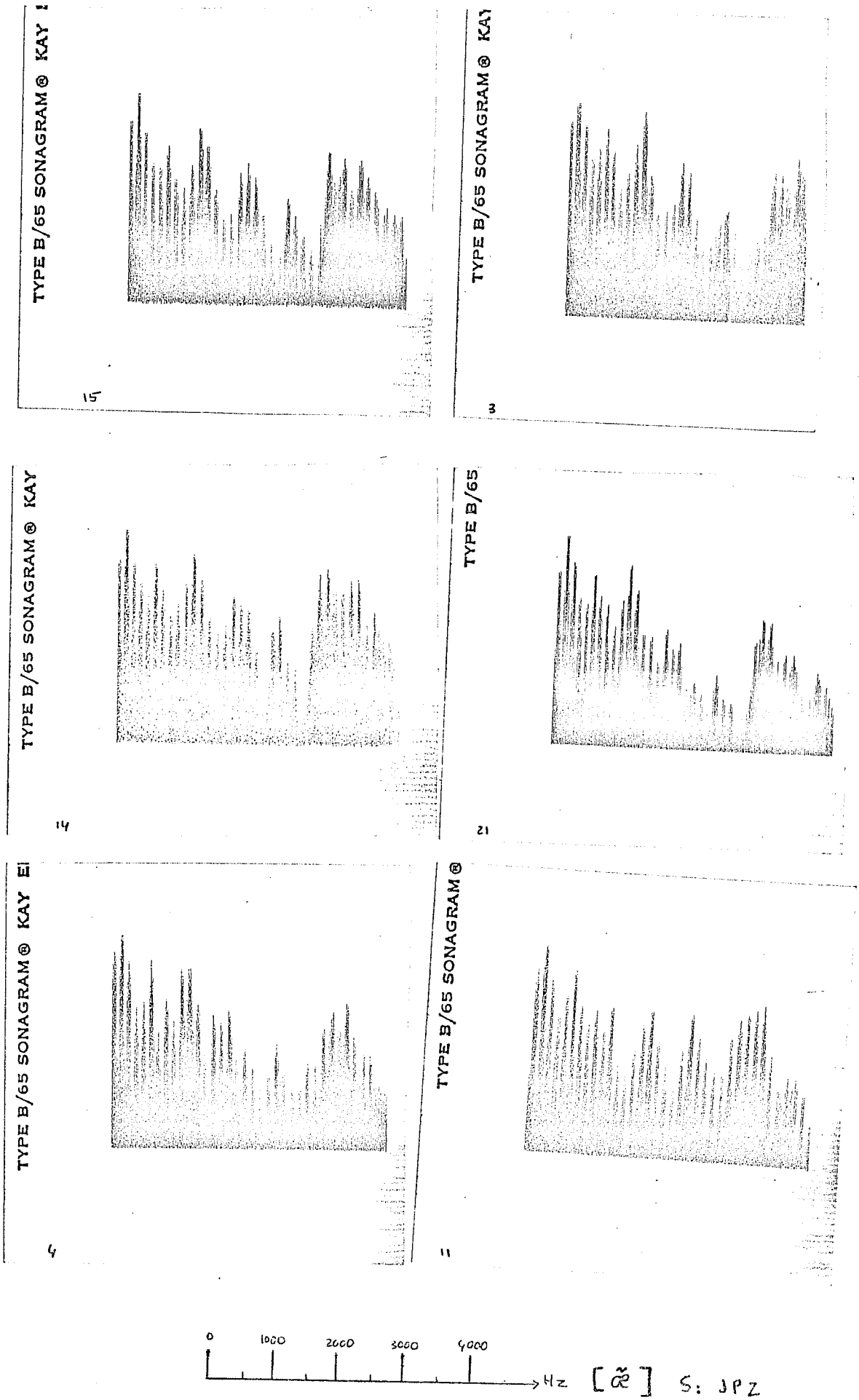

TYPE B/65 SONAGRAM® KAY
15
TYPE B/65 SONAGRAM® KAY
3
TYPE B/65 SONAGRAM® KAY
14
TYPE B/65
21
TYPE B/65 SONAGRAM® KAY
4
TYPE B/65 SONAGRAM®
11
0
1000
2000
3000
4000
Hz
[æ]
S: JPZ

Fig. 28 - X

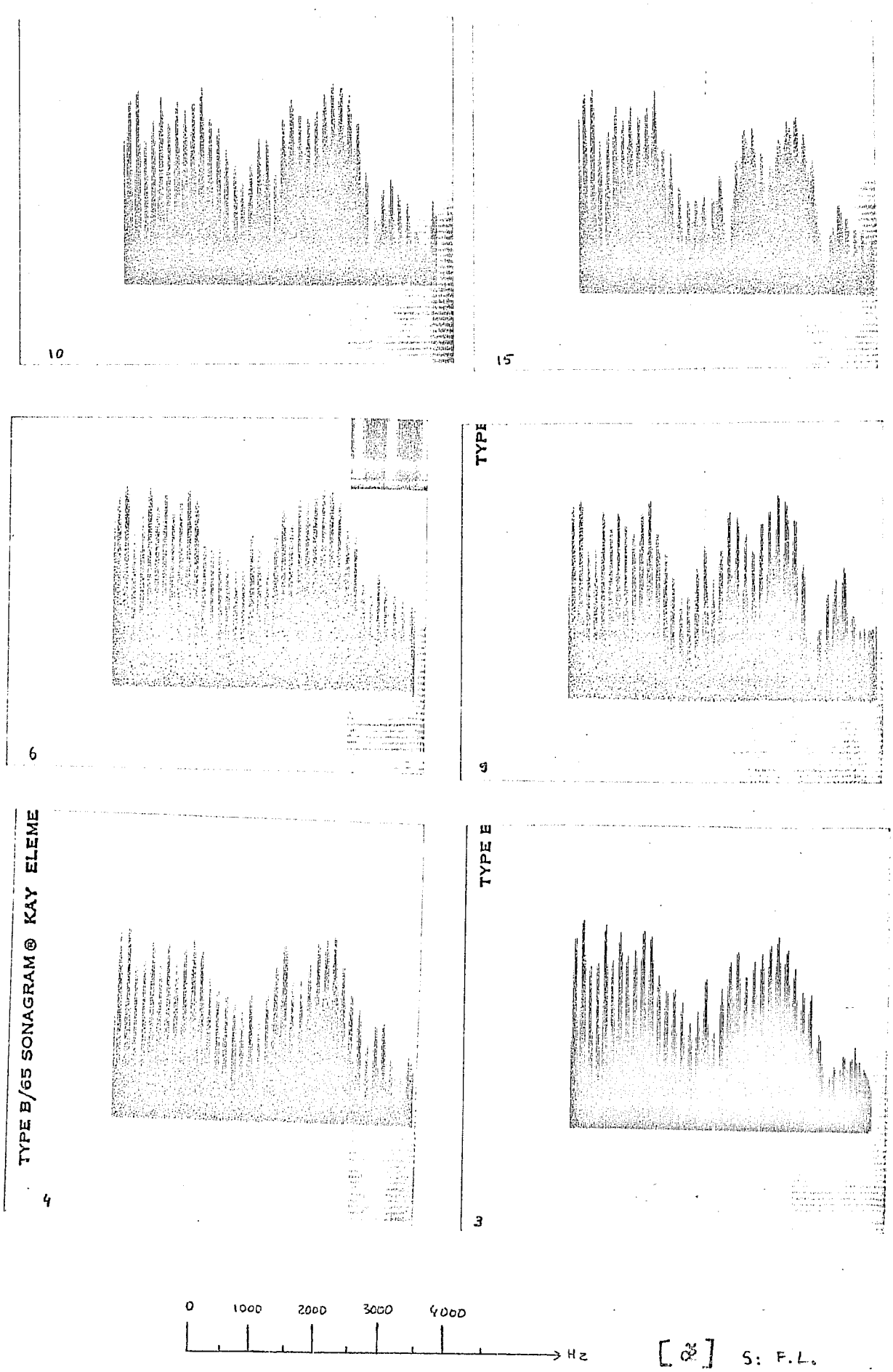

10
15
6
TYPE
9
TYPE B/65 SONAGRAM® KAY ELEME
4
TYPE E
3
0
1000
2000
3000
4000
Hz
S: F.L.

Fig. 27 - X

réalisées au même instant du temps. Pour cette voyelle, il faut admettre de plus que F'3 est à une fréquence élevée, très proche de F'4.

Toutes les études confirment que [ã] se caractérise par un premier formant (F'1) vers 700 Hz et un deuxième formant (F'2) proche de 1000 Hz.

VII - LA VOYELLE [œ̃]

A/ Description spectrale

Les spectrogrammes analysés sont reproduits aux fig. 27-X et 28-X. On constate globalement une très grande stabilité des caractéristiques spectrales.

1)- Pour les deux sujets, la première harmonique (P1) est plus intense que F_0 ou que la deuxième harmonique.

2)- Les sections du sujet JPZ possèdent un pic (P2) centré sur la 5ème harmonique vers 720 Hz, alors que pour cette même zone, les voyelles du sujet FL présentent deux harmoniques renforcées (P2, P2') au niveau des 4ème et 6ème harmoniques, séparées par une harmonique moins intense. Cette disposition différente est constante pour les deux sujets.

3)- Pour les deux sujets, la 9ème ou 10ème harmonique (1200-1350 Hz) est renforcée (P3).

4)- Chez le sujet FL, le pic suivant (P4), s'il est toujours visible, n'a qu'une intensité relativement faible. Par contre, pour le sujet JPZ, P4 est beaucoup plus apparent dans 4 sections sur 6.

5)- Les pics 5 et 6 font partie d'un groupe d'harmoniques renforcées particulièrement intenses pour le sujet FL. Le pic P5 est très stable en fréquence pour les deux sujets. Pour le sujet JPZ, l'intensité de P6 est largement supérieure à celle de P5, sauf pour la voyelle n° 11.

Les résultats des mesures fréquentielles sont présentés dans le tableau suivant :

Sujet JPZ	21	3	4	11	14	15	Moyenne
P1	265	260	240	300	260	280	270 Hz
P2	685	705	705 (935)	765	705	745	720
P3	1230	1275	1230	1125 1360	1275	1230	1250
P4	1800 (2005)	1870	(1550) 1865	1970	1910	1930	1900
P5	2575	2545	(2100) 2565	2585	2585	2540	2560
P6	3305 3685	3265 3605	3300 3660	3325 3685	3260 3725	3155 3410 3660	-

Sujet FL	4	3	9	6	15	10	Moyenne
P1	285	265	285	285	285	285	280 Hz
P2	640	620	620	620	640	640	630
P2'	875	830	850	850	875	875	860
P3	1250	1190	1315	1190	1230	1250	1240
P4	2110	2150	2175	2130	2240	2150	2160
P5	2615	2615	2510	2615	2660	2615	2600
P6	3370 4080	3245	3245 4210	3370	3330 4000	3245 4125	-

B/ Essai d'interprétation des caractéristiques spectrales

1)- Comparaison avec les travaux antérieurs

L'interprétation des données de DEBROCK est particulièrement difficile. On peut sans doute rapprocher le pic P3 de la série de valeurs 1210, 1340 et 1400 Hz. Par contre les valeurs de P2 sont nettement trop basses ; elles pourraient être des moyennes entre P1 et P2. Une seule valeur (pour le sujet 1) peut correspondre à P4.

	S1	S2	S3	Moyenne JPZ	Moyenne FL
P1	220	-	200	270	280 Hz
P2	-	510	420	720	630/860
P3	1210	1340	1400	1250	1240
P4	(1580) 2100	-	-	1900	2160
P5	2630	2800	-	2560	2600
P6	3300	3160	3130		

Il est possible d'identifier, dans les mesures publiées par GUTH, les pics P1 et P3. Une seule valeur pour le sujet S4 peut correspondre à P2. Il ne paraît pas possible d'identifier avec assurance les formants supérieurs.

	S1	S2	S3	S4	Moyenne JPZ	Moyenne FL
P1	350	300	500	350	270	280 Hz
P2					800	630/860
P3	1100 1750 2250 2800 3400	1350 1700 2500 3400	1200 2300 2800 3700	 1700 2500 3150 3600	1250	1240

Les valeurs publiées par DELATTRE correspondent bien à P1, P2 et P3 : 250, 600 et 1350 Hz.

On notera enfin le bon accord entre nos moyennes et les mesures de MRAYATI. Pour ces dernières, P3 est néanmoins de 200 Hz plus élevé.

	contexte ptk	contexte bdg	contexte mn ɲ	moyenne JPZ	moyenne FL
P1	400	415	400	270	280 Hz
P2	656	644	651	720	630/860
P3	1420	1388	1448	1250	1240
P6	3069	3036	2947	-	-

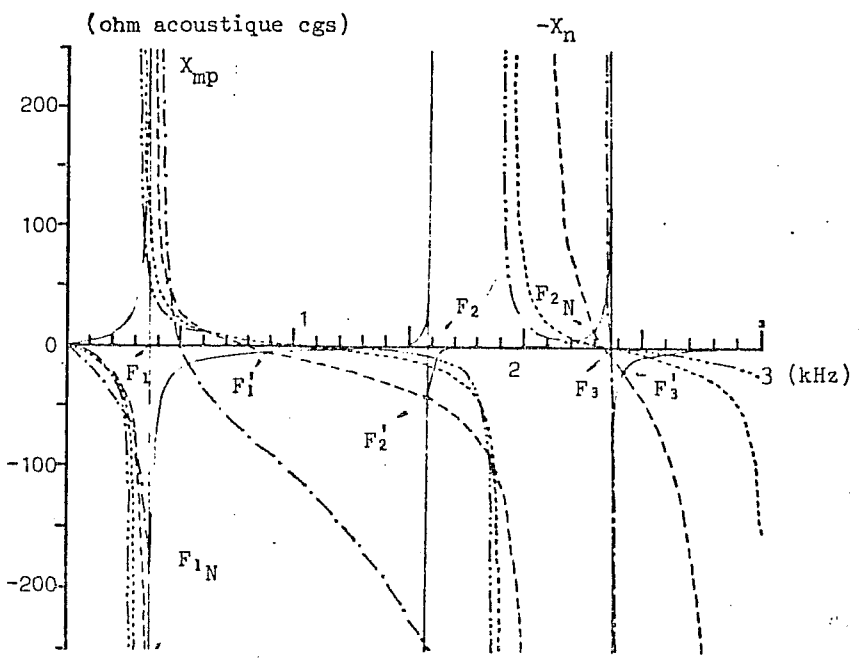
2)- Essai d'interprétation

L'analyse théorique de MRAYATI, reproduite à la fig. 29-X donne les résultats suivants en fonction du degré de couplage pour une voyelle dont les formants oraux s'établissent à 500, 1575 et 2375 Hz.

Degré de couplage	F1n	A1	F'1	F'2	F2n	A2	F'3
0.27 cm^2	350	500	525	1580	2350	2350	2375
1.66 "	350	800	850	1580	2300	2350	2400
3.50 "	350	1050	1200	1600	2200	2350	2525
5.00 "	350	1200	1300	1625	2200	2350	2600

La fréquence de F'2 est à nouveau franchement surestimée dans cet exemple théorique. Nous proposons l'interprétation suivante pour les différents pics relevés sur les sections :

P1 correspond à F1n. Si, pour le sujet JPZ, P2 correspond certainement à F'1, il est plus difficile d'associer P2 ou P'2 à F'1 pour le sujet FL. Les valeurs de MRAYATI plaident en faveur de P2. P'2 est sans doute un exemple supplémentaire de paire pôle/zéro du type déjà rencontré pour [ɛ̃] et [ɑ̃]. L'identification de P3 avec F'2 ne fait aucun doute. P4 correspond sans doute à F2n, alors que F'3 est représenté par P5. L'amplitude de P6 suggère à nouveau la présence de plusieurs pôles à des fréquences proches de F'4.



- Voyelle [œ̃]

Fig. 29 - X

3)- Essai de simulation

Les différents essais de simulation que nous avons effectués donnent des résultats satisfaisants. Dans tous les cas, le pic P2 (F'1) est unique. Nous présentons deux exemples aux fig. 30-X et 31-X. Dans les deux cas, l'amplitude de F'1 est inférieure à F'2. Par contre, l'amplitude de la zone d'harmoniques supérieure à 3000 Hz est trop faible. Notre programme de synthèse ne permettant de spécifier que 6 pôles, il ne nous a pas été possible de vérifier l'effet résultant de la présence de plusieurs pôles proches de F'4.

Les valeurs numériques utilisées pour les simulations sont les suivantes :

	F1n	A1	F'1	F'2	F2n	A2	F'3	F'4
Fig. 30-X	250	650	720	1250	2000	2100	2600	3500
Fig. 31-X	350	600	720	1250	2000	2150	2600	3500

C/ Conclusions

Nous présentons à la fig. 32-X une analyse des caractéristiques spectrales de la voyelle [œ̃] n° 3 (sujet F.L). On remarquera à nouveau le bon accord entre la section et l'analyse numérique. En particulier, l'analyse numérique révèle clairement la présence de deux pics aux environs de F'1 (P2 et P'2). La voyelle [œ̃] se caractérise par un premier formant vers 700 Hz et un deuxième formant vers 1250 Hz.

VIII - SYNTHESE DES VOYELLES NASALES : IMPORTANCE DES FREQUENCES FORMANTIQUES

Le corpus des voyelles naturelles étudié étant limité, nous ne présenterons pas de valeurs moyennes. L'établissement de celles-ci supposerait que l'on dispose de résultats provenant d'un plus grand nombre de locuteurs.

Quelques indices perceptifs intéressants sont fournis par l'étude d'une expérience d'identification que nous avons présentée au chapitre IV. 107 échantillons de voyelles

synthétiques ont en effet été réalisés pour nous par les laboratoires de la C.G.E. Un pré-test ayant démontré que certains échantillons paraissaient nasalisés, nous avons ajouté les 4 voyelles nasales aux catégories de la grille de réponse. Le test et les stimuli ont été présentés en détail au chapitre IV. Rappelons que ces voyelles synthétiques ne comprenaient aucun indice spécifique des voyelles nasales (affaiblissement de F1, présence de zéros, etc...). La source vocale comportait pourtant un zéro vers 2000 Hz, dû à l'emploi d'une onde glottale triangulaire asymétrique. Les largeurs de bande pour F1 (70 Hz) n'ont pu provoquer un affaiblissement de F1.

On a obtenu 209 réponses nasales sur 2259, c'est-à-dire 9,25 %. 150 réponses, soit près de 72 %, se trouvent concentrées dans la zone délimitée par F2 = 1800-800 Hz, F1 = 400-875 Hz. Les résultats, en pourcentage, sont présentés à la table I-X. Ces pourcentages sont calculés par rapport à l'ensemble des réponses nasales et non-nasales. Une deuxième zone, composée d'un F1 et F2 élevés, pour laquelle on observe 18 % du total des réponses nasales, est présentée à la table II-X.

Plusieurs constatations se dégagent de l'examen des résultats :

a- Pour certaines combinaisons de fréquences formantiques, le pourcentage de réponses nasales atteint 50 % (800 - 1700 - 2500 Hz). Ce résultat est obtenu sans l'aide d'indices de nasalité spécifiques. Ceci souligne l'importance des fréquences formantiques dans la perception de la nasalité. On s'aperçoit que ni l'affaiblissement de F1, ni la présence de formants de nasalité ou de zéros ne sont nécessaires pour l'obtention de réponses nasales. Le zéro glottal entre F2 et F3 (2000 Hz) n'a pu jouer un rôle important pour cet exemple.

b- Les combinaisons 700 - 900 - 2500 et 800 - 900 - 2500 Hz donnent une forte minorité de réponses [ã] . Ces chiffres correspondent aux valeurs de F'1 et F'2 que nous avons obtenues pour nos deux sujets :

F 2 Hz

F1 Hz \ F2 Hz	I800	I700	I600	I500	I400	I300	I200	II00	I000	900	800
400								ɔ̃ :I0 ɛ̃ : 5			ɔ̃ :5
500		ɔ̃ :5	ɔ̃ : I0	ɔ̃ :3	ɛ̃ :5	ɔ̃ : 6	œ :5	:I5 ɔ̃		œ̃ :5 ɔ̃ :5 ɑ̃ : 5	ɔ̃ :6
600	ɔ̃ :6	ɑ̃ :5 ɔ̃ : 5	ɔ̃ :I0 œ̃ :I0		ɔ̃ :5		ɔ̃ :3	œ̃ :6	ɑ̃ :I0		ɑ̃ :I2
700	œ̃ : 5	ɑ̃ :6 œ̃ :6 ɛ̃ :6	œ̃ :I2 ɛ̃ :4	ɛ̃ : I2 œ̃ : I2	œ̃ :I0 ɛ̃ : 5	œ̃ :I0 ɛ̃ :I0	ɛ̃ ;I0 œ̃ :5	œ̃ :3 ɛ̃ :3	ɑ̃ :I5 œ̃ :5 ɛ̃ :5	ɑ̃ :40 ɛ̃ :5 ɑ	ɑ̃ : 8 ɛ̃ :4 ɔ̃ : 8
800	ɑ̃ ;I0 ɛ̃ :I0 œ̃ :I0	ɛ̃ :20 œ̃ :25 ɑ̃ : 5 ɑ̃ - ɑ	ɑ̃ :6 œ̃ :I2 ɛ̃ :I2	ɛ̃ :35 œ̃ :6 ɑ̃ :6 ɑ̃	œ̃ :I0 ɛ̃ :20	œ̃ :4 ɛ̃ :8	ɛ̃ :6	ɑ̃ :6	ɑ̃ :6 ɛ̃ :6	ɑ̃ :35 ɛ̃ :3 œ̃ :3 ɑ	
875		ɔ̃ :6 œ̃ :I2 ɛ̃ :I8	ɑ̃ :5 œ̃ :I0	ɛ̃ :I0		ɑ̃ :3 œ̃ :3		ɑ̃ :I0	ɑ̃ :6		

NB: Les résultats sont en %

Table I - X a

	F3	2850	2750	2500	2500	2500	2500	
	F2	2500	2350	2200	2050	I900	I800	Hz
600		ɛ̃ :20 ɔ̃ :20	ɛ̃ :6	ɔ̃ :5	ɔ̃ :8	ɔ̃ :6	ɔ̃ :6	
FI Hz								
700		ɛ̃ :20 ɑ̃ :5	ɛ̃ :24	ɔ̃ :5 ɛ̃ :I0 ɑ̃ :5	ɑ̃ :I3 ɛ̃ :6	ɑ̃ :I2 ɔ̃ :4 ɛ̃ :I2	ɑ̃ :5	

NB: Les réponses sont en %

Table I - Xb

	F'1	F'2	F'3	
sujet JPZ	735	965	2475	Hz
sujet FL	630	905	2160	

En règle générale, les valeurs formantiques des voyelles synthétiques obtenant les scores nasaux les plus élevés sont proches des valeurs que nous avons relevées pour F'1 et F'2.

c- Les réponses [ɔ̃] sont éparpillées dans l'ensemble du tableau. Le pourcentage le plus élevé du tableau I-X n'est que de 15 %, pour la combinaison 500 - 1100 - 2500 Hz. Mais la combinaison (table II-X) 600 - 2500 - 2850 Hz, qui n'est pas physiquement réalisable, donne 20 % de réponses [ɔ̃]. Il est probable que, F1 est analysé comme la combinaison de F'1 et F'2, alors que F2 joue le rôle de F3. Dans ce cas, la correspondance avec les valeurs mesurées est bonne.

d- Aucune combinaison ne donne une majorité de réponses [œ̃]. Le pourcentage le plus élevé est obtenu pour la combinaison 800 - 1700 - 2500 Hz (25 %). Mais les réponses [ɛ̃] atteignent un pourcentage voisin (20 %). Ceci pourrait être le signe de la confusion [œ̃ - ɛ̃] que l'on observe en français contemporain. De plus la valeur de F2 est très nettement supérieure à la valeur mesurée (1250 Hz). Il paraît donc difficile de synthétiser une voyelle sans l'aide d'indices acoustiques spécifiques.

e- La combinaison 800 - 1500 - 2500 Hz donne 35 % de réponses [ɛ̃]. Ces valeurs formantiques sont en accord avec nos mesures. Mais une valeur supérieure de F2 (1700 Hz) donne également un nombre important de réponses [ɛ̃]. Ceci est en accord avec les variations observées dans la mesure des voyelles naturelles.

f- On pourrait faire l'hypothèse que la fréquence de F'2 ne joue qu'un rôle perceptif mineur au vu des résultats suivants :

600 - 2500 - 2850 Hz	ɛ̃ 20 %	ɔ̃ 20 %
700 - 2500 - 2850 Hz	ɛ̃ 20 %	ɑ̃ 5 %
700 - 2350 - 2750 Hz	ɛ̃ 24 %	

Dans tous les cas, l'identification de [ɛ̃] a été obtenue grâce à F1 seul. Mais on observe des hésitations entre [ɛ̃], [ɔ̃] et dans une moindre mesure [œ̃]. Le rôle de F'2 serait donc de préciser la nature phonétique de la voyelle nasale. F'1 serait l'indice le plus important de la nasalité.

Cette étude est incomplète dans la mesure où l'absence de caractéristiques spécifiques de nasalité ne permet un nombre important de réponses nasales que pour les stimuli qui ne produisent pas une bonne voyelle orale.

Mais elle montre sans ambiguïté que les fréquences formantiques jouent un rôle important dans la perception du timbre nasal. Ce résultat n'avait pas, à notre connaissance, été mis en évidence.

Nous espérons pouvoir bientôt étudier les indices des voyelles nasales à l'aide d'un synthétiseur à deux zéros. Il reste à préciser les plages de variations des fréquences formantiques en présence des indices spécifiques de nasalité. Il conviendrait également d'étudier en détail les mutations de timbre que les voyelles ont connues en se nasalisant (cf. OHALA 1974, ROCHET 1976). L'ouverture des voyelles nasalisées pourrait être la conséquence directe de l'élévation de F'1 par rapport à F1. Mais des contraintes articulatoires sont sans doute à l'oeuvre, qui expliquent la fermeture de certaines voyelles nasalisées au contact des consonnes nasales dans les langues germaniques. Il faudrait également préciser la différence entre voyelles nasales et nasalisées.

IX - L'ARTICULATION DES VOYELLES NASALES : caractérisation à partir des fréquences formantiques.

La théorie acoustique indique que le formant F'2 de la voyelle nasale est pratiquement insensible au degré de

nasalisation. Ce résultat est important pour l'étude de l'articulation des voyelles nasales. Il est permis de comparer la fréquence du deuxième formant d'une voyelle orale et d'une voyelle nasale. Ceci n'est pas possible pour le premier formant dont la fréquence est largement modifiée par le degré de nasalisation. On sait que le deuxième formant indique grossièrement la position de la voyelle sur l'axe articulatoire antérieur/postérieur. Il est donc possible de préciser l'un des paramètres articulatoires des voyelles nasales en comparant les fréquences de F2 pour les deux types de voyelles. Cette comparaison permet de résoudre partiellement la "querelle des symboles" portant sur les voyelles nasales. On vérifiera que ces indications articulatoires sont conformes aux conclusions de BRICHLER (1970) obtenues par cinéradiographie. Ces résultats devraient être incorporés dans un système de rééducation d'enfants mal-entendants par visualisation de la forme du conduit vocal. On sait, en effet, que l'un des défauts de ces enfants est la nasalisation intempestive frappant un certain nombre de voyelles. Cette nasalisation ne permet plus le calcul de la forme du conduit par la méthode de prédiction linéaire. L'utilisation de la valeur de F2 devrait permettre une estimation de la forme dans le cas difficile d'une nasalisation intempestive. Pour la comparaison avec les voyelles orales nous utiliserons les valeurs de DEBROCK et FORREZ (1976), que nous avons légèrement modifiées pour tenir compte de nos estimations personnelles.

Voyelles	F1	F2	F3	(Hz)
i	280	2300	2950	
e	350	1950	2550	
ɛ	450	1800	2470	
ə	660	1350	2380	
a	620	1150	2250	
ɔ	480	1050	2250	
o	360	780	2230	

Voyelles	F1	F2	F3	(Hz)
u	290	850	2270	
y	290	1800	2140	
ø	360	1450	2290	
œ	490	1380	2270	
ə	380	1400	2200	

a) - voyelle [ɔ̃]

F'2 est nettement inférieur à 1000 Hz. Chez le sujet JPZ, où la mesure semble la plus précise, F'2 s'établit à 735 Hz. F'2 est donc très proche de F2 de la voyelle [o]. BRICHLER note que la langue est plus retirée pour [ɔ̃] que pour la voyelle [ɔ] , avec un résonateur antérieur plus grand et un résonateur pharyngal plus petit pour [ɔ̃] que pour [ɔ] .
Mais la comparaison des divers tracés indique une grande variabilité de l'articulation, qui se trouve proche de [ɔ] dans le cas de la voyelle brève. Dans ce cas, F'2 devrait atteindre près de 1000 Hz. On peut s'attendre à des confusions avec [ɑ̃] . Pour les réalisations de nos deux sujets, le meilleur symbole est sans doute [õ] .

b) - voyelle [ɑ̃]

F'2 est proche de 1000 Hz. Cette valeur est large inférieure à celle de [a] , et est proche de celle de [ɔ]. BRICHLER note à nouveau que la langue est plus retirée en arrière pour [ɑ̃] que pour [a] , avec un angle maxillaire plus réduit. Le meilleur symbole pourrait donc être [ɔ̃] .

c) - voyelle [ɛ̃]

F'2 s'établit entre 1400 et 1550 Hz pour nos sujets. Des valeurs plus élevées (1600 - 1700 Hz) sont typiques des études antérieures. Il semble toutefois que cette voyelle nasale soit plus ouverte que [ɛ] , si l'on accepte une liaison linéaire entre F1 et F2 pour les voyelles antérieures. Le

symbole [æ̃] utilisé par MARTINET est le plus adéquat du point de vue phonétique. Les tracés de BRICHLER indiquent que [ɛ̃] est plus ouvert que [ɛ] au niveau palatal et alvéolaire. L'angle maxillaire est voisin pour les deux voyelles. La taille de la cavité pharyngale est égale ou inférieure selon les cas. Ces indications articulatoires sont tout à fait compatibles avec les conclusions que nous tirons des valeurs acoustiques.

d) - voyelle [œ̃]

Le deuxième formant (F'2) de la voyelle s'établit à 1250 Hz. Cette valeur est inférieure à la valeur moyenne pour [œ] (1380 Hz). Ceci indique une voyelle plus ouverte que l'on pourrait noter [ʌ̃] . Articulatoirement, la voyelle nasale est plus ouverte. On observe une diminution du volume pharyngal, et une augmentation de la distance langue-alvéole et langue-palais dur.

Il apparaît bien que la fréquence de F'2 permet une estimation grossière, mais néanmoins correcte, de l'articulation des voyelles nasales par comparaison avec les voyelles orales les plus proches. Si les voyelles nasales /ɔ̃, ɑ̃, ɛ̃, œ̃/ ont des articulations correspondant aux voyelles orales [o, ɔ, æ, ʌ] , l'hypothèse de DELATTRE (1968) selon laquelle les cavités pharyngales ont une taille identique pour les quatre voyelles nasales n'est pas vérifiée. La cavité pharyngale de la voyelle [o] est nettement plus étroite que celle de la voyelle [æ] .

X - NOTE SUR L'ANALYSE DES VOYELLES NASALES PAR PREDICTION LINEAIRE

Comme il a été démontré au chapitre VII, la méthode de prédiction linéaire exige que les voyelles puissent être décrites en termes de pôles, à l'exclusion de tout zéro. Il nous a paru intéressant de tester, sur quelques exemples, le résultat de la violation de cette contrainte. Cette étude nous a paru d'autant plus nécessaire que MRAYATI (1975) utilise la méthode de prédiction linéaire pour obtenir les

fréquences formantiques des voyelles nasales.

Nous avons simulé deux voyelles nasales correspondant approximativement à [ɑ̃] et [ɛ̃] . La table suivante précise les fréquences et les largeurs de bande utilisées :

	F1n	F'1	F'2	F2n	F'3	F'4	A1	A2
[ɑ̃]	250	700	950	1850	2300	3500	600	2000
	65	80	80	100	120	180	100	120
[ɛ̃]	250	700	1500	2200	2650	3500	600	2400
	65	80	80	100	120	180	100	120

Les spectres obtenus par transformée de FOURIER sont présentés aux fig. 33-X et 34-X.

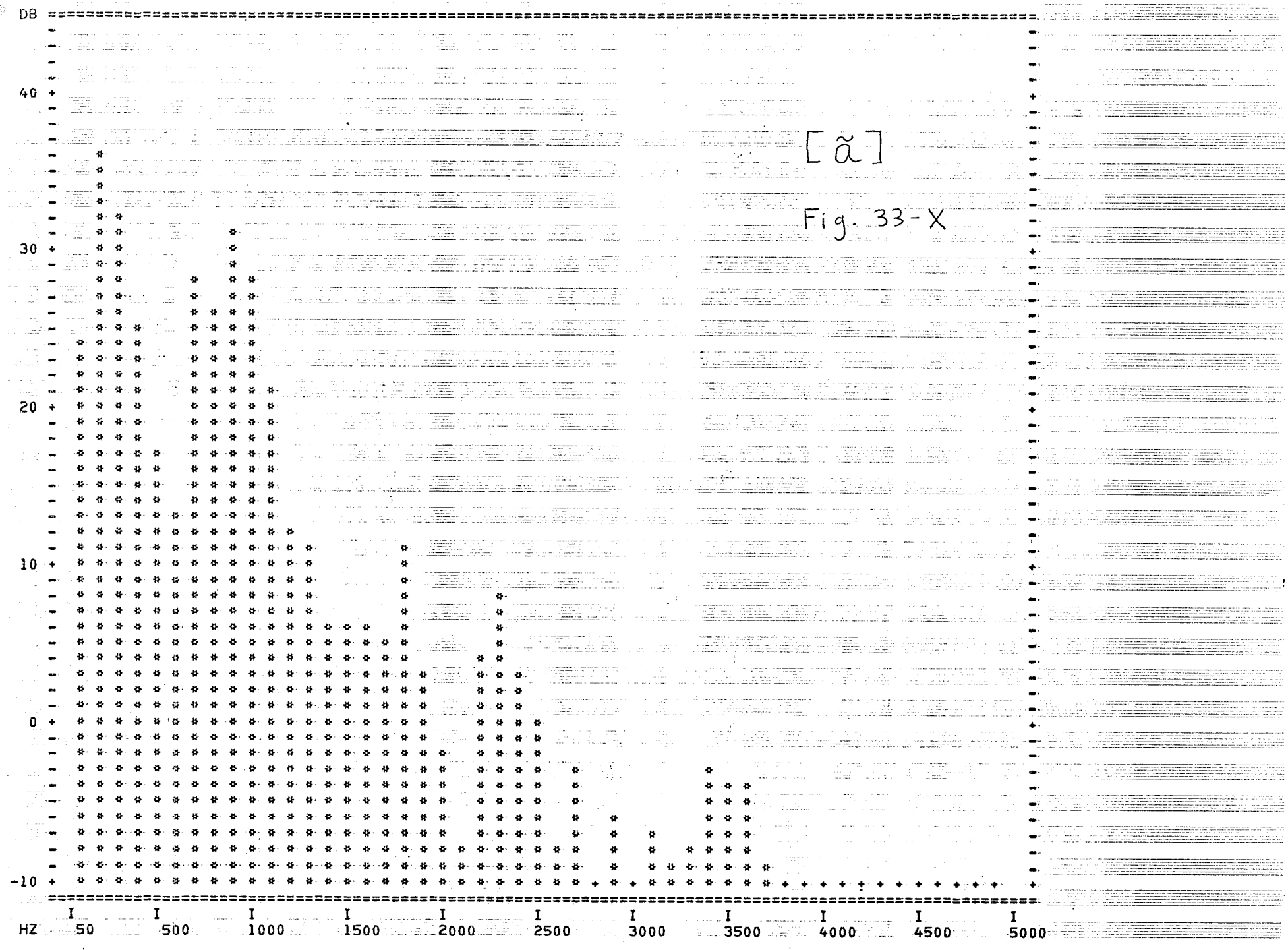
Ces mêmes voyelles ont ensuite été analysées par prédiction linéaire fig. 35-X et 36-X. Il apparaît clairement que la méthode de prédiction linéaire ne permet pas de calculer une bonne approximation de la fonction de transfert, bien que 12 coefficients aient été utilisés. Ce nombre de coefficients correspond aux 6 pôles ou formants utilisés à la synthèse.

Pour [ɑ̃] , on remarque que :

- la vallée correspondant au premier zéro (600 Hz) a été comblée.
- F'1 (700 Hz) n'apparaît plus.
- l'amplitude de F'2 est incorrecte.
- le deuxième formant nasal n'apparaît pas.
- la vallée correspondant au deuxième zéro est également comblée.

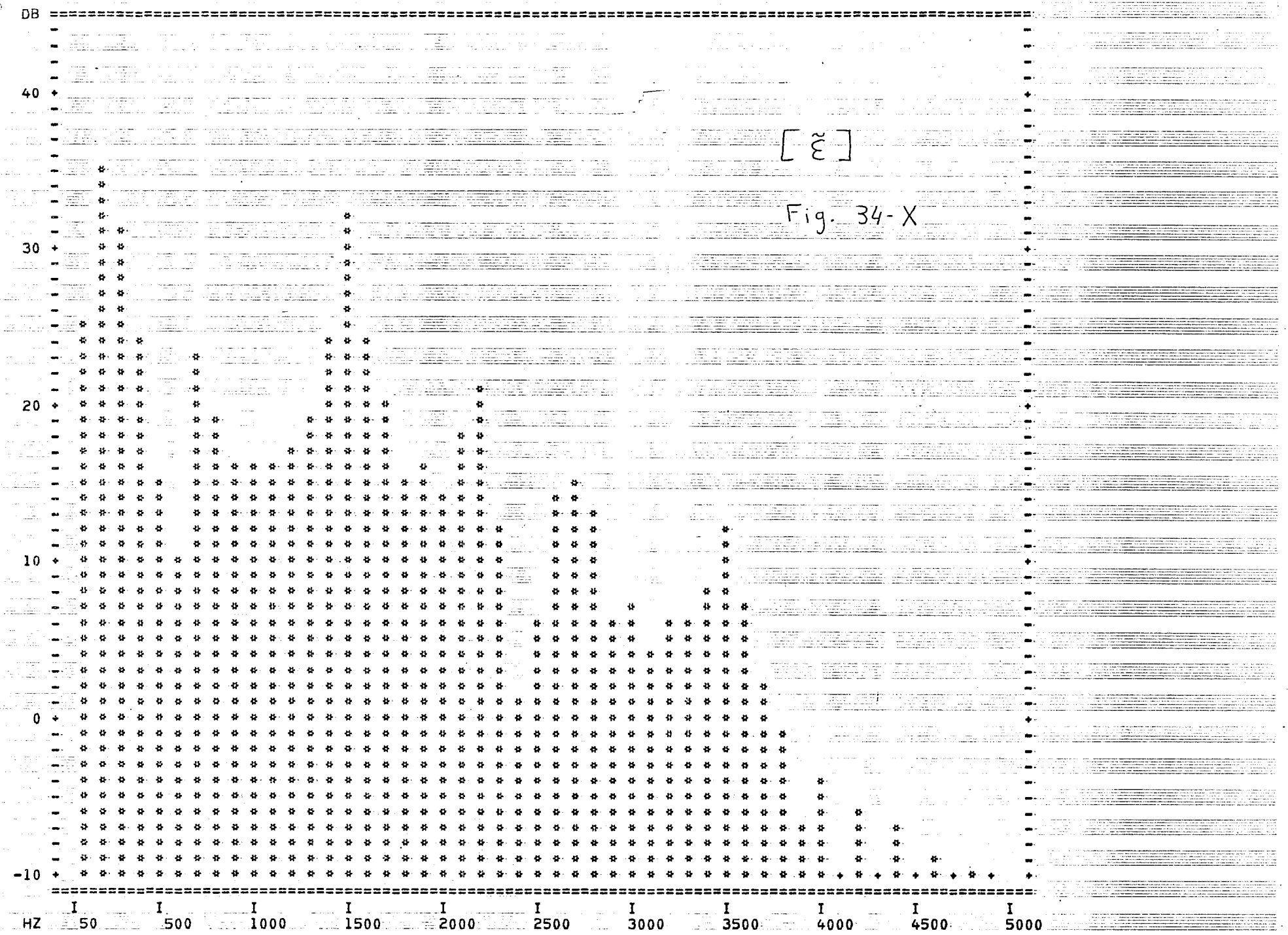
En résumé, toutes les caractéristiques de la voyelle nasale sont "gommées" par la méthode de prédiction linéaire. Les résultats de MRAYATI (1975) sont meilleurs que les nôtres. Une méthode de résolution numérique et un nombre plus important de coefficients permettent sans doute de détecter F'1.

Pour la voyelle [ɛ̃] simulée, les résultats sont identiques. F2n n'apparaît que comme une légère irrégularité vers 2200 Hz. Les deux zéros sont invisibles. F'1 et F'3 sont absents.


DB
40
30
20
10
0
-10
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

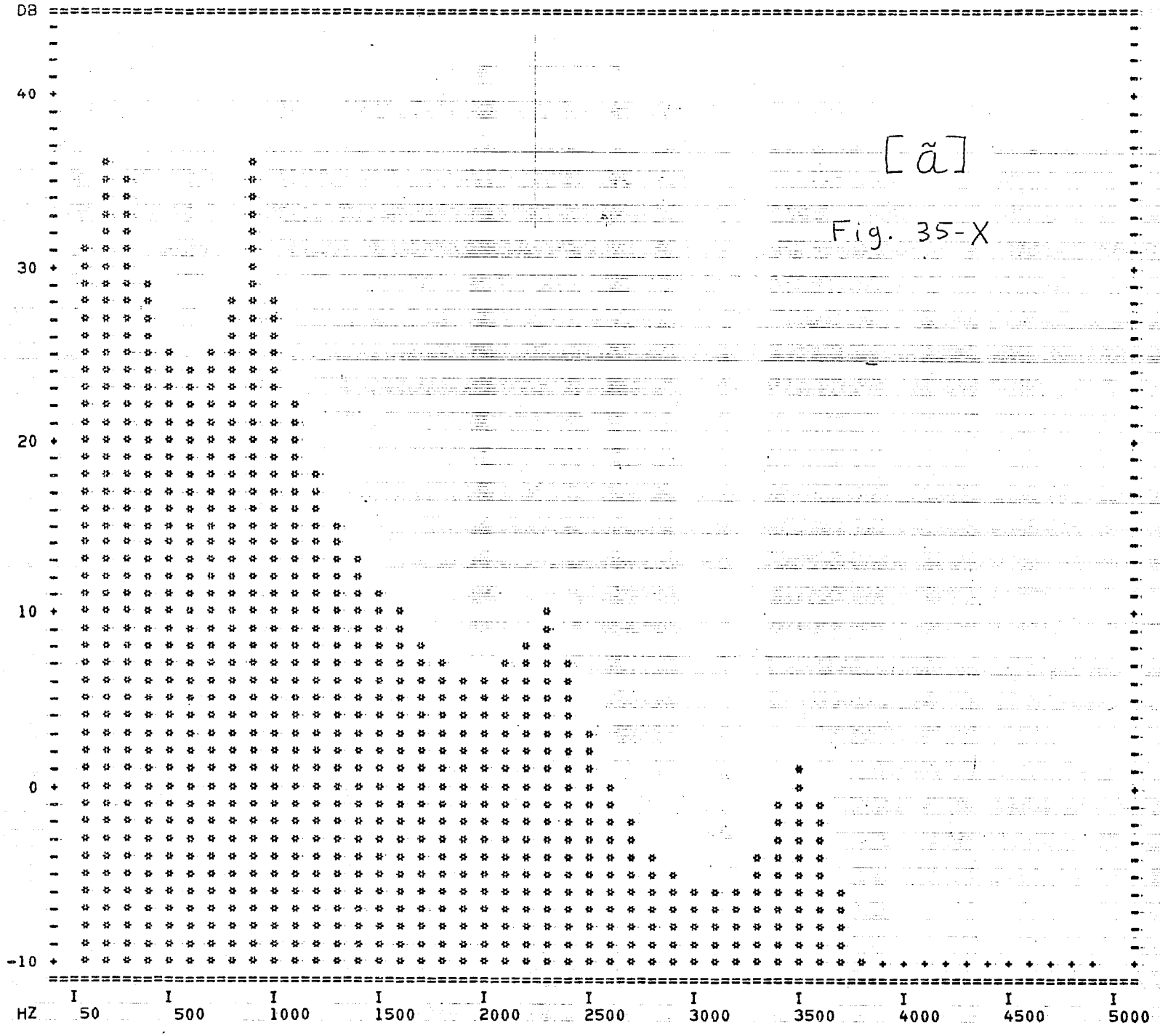
[ã]

Fig. 33-X



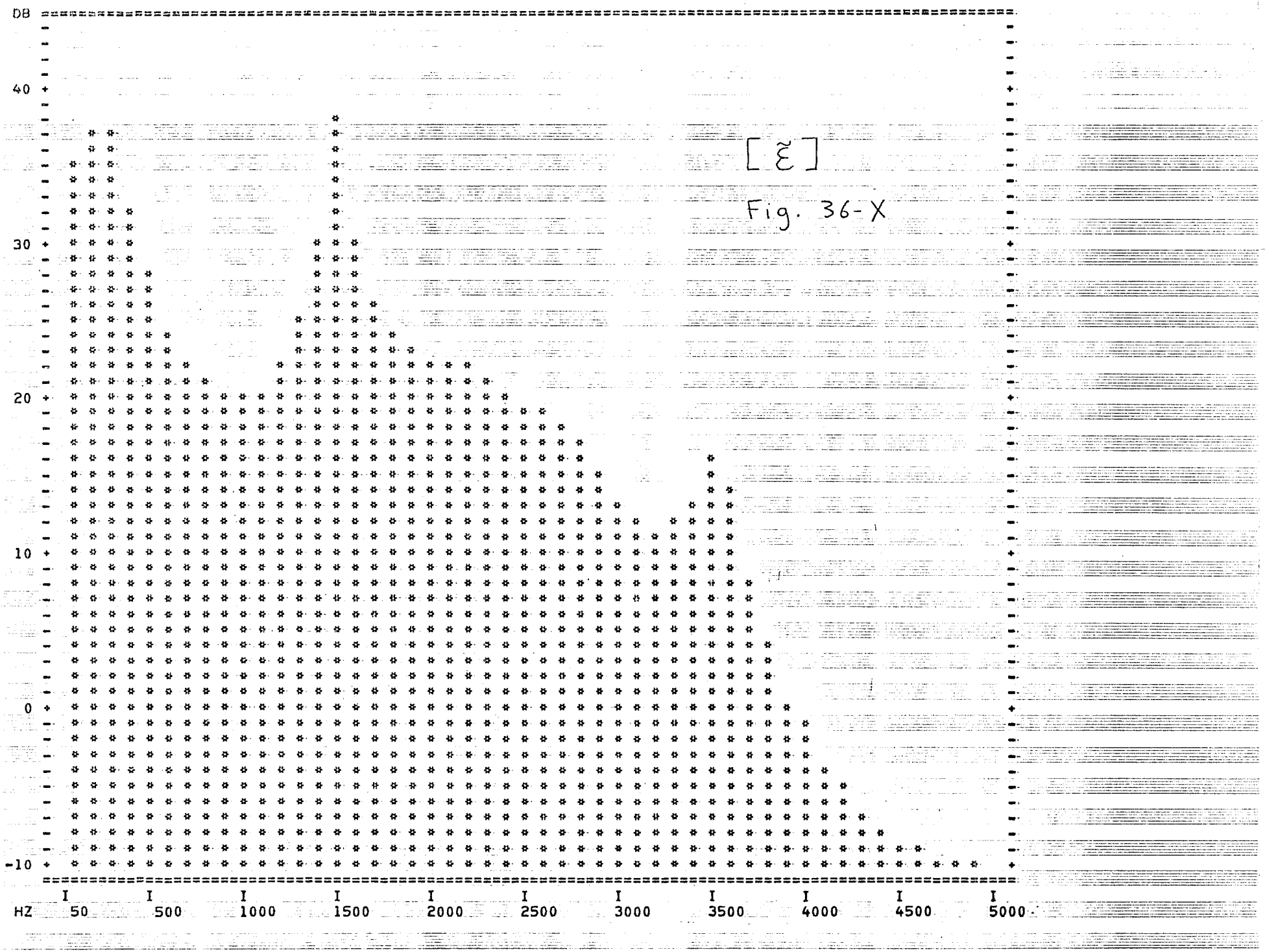
[$\tilde{\varepsilon}$]

Fig. 34-X


DB
40
30
20
10
0
-10
HZ
50
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

[ã]

Fig. 35-X



[$\tilde{\varepsilon}$]

Fig. 36-X

Il paraît difficile de recommander l'emploi de la méthode de prédiction linéaire pour l'étude des voyelles nasales si l'on désire des résultats détaillés. Mais des travaux importants sont en cours aux U.S.A. et au Japon pour mettre au point un modèle de prédiction permettant d'obtenir à la fois les fréquences des pôles et des zéros. Les voyelles nasales du français sont un matériau de choix pour tester ces nouvelles méthodes (MORIKAWA et FUJISAKI 1977, STEIGLITZ 1977). Nous disposerons sans doute bientôt de techniques permettant de faire une étude sérieuse sur un large échantillon de voyelles nasales.

XI - CONCLUSIONS

Ce chapitre doit être considéré comme le fondement d'une étude sur les indices perceptifs des voyelles nasales. Nous avons montré qu'il est possible d'identifier les caractéristiques acoustiques des voyelles nasales. On peut donc maintenant spécifier les fréquences des pôles et des zéros pour la synthèse. De plus, une expérience de synthèse sans zéros a montré que les fréquences de F1 et de F2 étaient des paramètres très importants. Ils sont capables, à eux seuls, d'induire un pourcentage non-négligeable de réponses nasales. Les résultats numériques de cette expérience de synthèse confirment les mesures effectuées sur des voyelles naturelles. Des études ultérieures devront préciser la hiérarchie des indices spéficiques de la nasalité (fréquences de F1n, et des zéros, etc...) et d'esquisser les limites des variations des formants F'1 et F'2. L'étude des mutations diachroniques devrait, également, bénéficier de ces études approfondies.

A N N E X E

S I M N A S
PROGRAMME DE SIMULATION DE VOYELLES ORALES ET NASALES

Ce programme FORTRAN a pour but:

- de synthétiser une voyelle définie par, au plus, 6 formants et 2 zéros;
- de calculer son spectre au moyen de deux méthodes différentes
  a) transformée de FOURIER (algorithme de GOERTZEL cf. LONCHAMP 1974 a,
  b) prédiction linéaire et F.F.T. (GOERTZEL) sur les coefficients de prédiction Ai;
- de faire l'affichage du spectre.

Le choix d'une méthode "longue" utilisant un programme de synthèse au lieu du calcul direct de la fonction de transfert à partir des positions des pôles et des zéros, permet

- de disposer, comme "sous-produit" du calcul, du tableau des échantillons temporels (sur 5 périodes): quelques instructions après la ligne 109 permettraient leur affichage;
- d'utiliser une onde glottale dont les caractéristiques se rapprochent de celles de l'onde réelle (courbe polynomiale de ROSENBERG 1971 a = 0.4, b = 0.16): la fonction de transfert globale est très proche de celle observée pour des voyelles naturelles;
- de tester la méthode de prédiction linéaire sur des voyelles orales et nasales;
- de simuler simplement la correction de haute-fréquence du spectrographe (par différentiation temporelle).

Ce programme comprend les modules suivants:

ligne	à ligne	
I2	25	: Lecture des données et initialisation
26	44	: Création de l'onde glottale (5 périodes)
45	60	: Simple ou double différenciation
6I	78	: Ecriture des paramètres initiaux
79	92	: Synthèse: filtre récursif du 2ème ordre simulant un pôle (formant)
93	II0	: Synthèse: filtre récursif simulant un zéro
II0	II5	: Fenêtre de HAMMING
II6	I54	: Calcul des coefficients de prédiction linéaire et calcul des aires.
I55	I9I	: F.F.T. (algorithme de GOERTZEL) sur les coefficients de prédiction ou sur les échantillons temporels.
I92	208	: Calcul de l'intensité des composantes spectrales en dB.
2I0	276	: Affichage numérique et graphique des résultats.

NB: La fréquence d'échantillonnage simulée est de I0 KHz. Cette valeur peut être changée en assignant une valeur différente à JFE (ligne 25). Cette valeur doit être vérifiée si l'on désire utiliser les résultats portant sur les aires.

PARAMETRES

Ligne

9 : NV: Nombre de voyelles dont on demande la simulation. Les paramètres suivants doivent être spécifiés pour chaque voyelle.

I2 : NF : Nombre de formants $\leqslant$ 6 .

IPL : si I, calcul par prédiction linéaire

si 0, calcul de la FFT sur les échantillons.

JZ : Nombre de zéros ($0 \leqslant JZ \leqslant 2$).

IDD : si I, double différenciation pour simuler la courbe de réponse du spectrographe; cette option est automatiquement appliquée lors de l'analyse par prédiction linéaire, où elle est obligatoire. Dans ce cas la pente moyenne du spectre est de 0 dB/ octave.
si 0, simple différenciation ; pente moyenne du spectre: - 6dB/octave.

IAFF: si I, affichage des valeurs numériques et affichage graphique;
si 0, affichage graphique seulement.

I5 : Liste des fréquences formantiques.

I7 : Liste des largeurs de bande.

20/22: Liste (le cas échéant) des fréquences et largeurs de bande des zéros.

SORTIE

ligne

6I : Fréquences formantiques.

66/69 : indication de la méthode utilisée: FFT directe ou prédiction linéaire.

7I : Largeurs de bande.

74/76 : Fréquences et largeurs de bande des zéros.

I53 : Aires du tube acoustique équivalent.

```
      DIMENSION F(401),AB(401),B(401),CC(201),PHI(201),ITAB(58,109),JFO(
     111),HDE(201),IZ(2),IZB(2)
      DIMENSION G(1000),IF(6),IB(6),R(15),A(15),E(15),AA(15),C(15)
      REAL K(16)
      DATA IBLANC,IMOINS,IEGAL,IVERT,IPLUS,IAST,IZER,ID,IV,IS4,IS3,IS2,
     1IS1/1H ,1H-,1H=,1HI,1H+,1H*,1H0,1HD,1HB,1H4,1H3,1H2,1H1/
      PI=3.141593
      E2PI=2*PI
      READ(105,10) NV
   10 FORMAT(I2)
      DO1000IS=1,NV
      READ(105,210)NF,IPL,JZ,IDD,IAFF
  210 FORMAT(I1,1X,I1,1X,I1,1X,I1,1X,I1)
      MP=NF*2
      READ(105,20)(IF(I),I=1,NF)
   20 FORMAT(6(I4,1X))
      READ(105,30)(IB(I),I=1,NF)
   30 FORMAT(6(I3,1X))
      IF(JZ.EQ.0)GOTO227
      READ(105,52)(IZ(I),I=1,JZ)
   52 FORMAT(2(I4,1X))
      READ(105,53)(IZB(I),I=1,JZ)
   53 FORMAT(2(I3,1X))
  227 CONTINUE
      JFE=10000
      DD=FLOAT(JFE)/10000
      XDD=DD*500
      YDD=DD*100
      D=1./FLOAT(JFE)
      JXDD=XDD
      JYDD=YDD
      JAX=40*DD
      JBX=16*DD
      XJAX=JAX
      XJBX=JBX
      DO200I=1,JAX
  200 G(I)=(3*((I/XJAX)**2)-2*((I/XJAX)**3))*100.
      DO201I=(JAX+1),(JAX+JBX)
  201 G(I)=(1.-(((I-XJAX)/XJBX)**2))*100.
      DO202I=(JAX+JBX+1),JYDD
  202 G(I)=0.
      DO203II=1,4
      DO203I=1,JYDD
  203 G(JYDD*(II)+I)=G(I)
      IF(IPL.EQ.1)GOTO226
      IF(IDD.EQ.0)GOTO211
  226 DO204I=5*JYDD,1,-1
  204 G(I+2)=G(I)
      G(1)=0.
      G(2)=0.
      DO205I=1,5*JYDD+1
  205 G(I)=G(I+1)-G(I)
      DO206I=1,5*JYDD
  206 G(I)=G(I+1)-G(I)
      GOTO212
  211 DO213I=5*JYDD,1,-1
  213 G(I+1)=G(I)
      G(1)=0.
      DO214I=1,5*JYDD
  214 G(I)=G(I+1)-G(I)
```

```
  212 WRITE(108,50)(IF(I),I=1,NF)
   50 FORMAT(5X,'F 1 = ',I4,5X,' F 2 = ',I4,5X,' F 3 = ',I4,5X,
     1' F 4 = ',I4,5X,' F 5 = ',I4,5X, ' F 6 = ',I4,5X/)
      IF(IPL.EQ.0)GOTO234
      WRITE(108,235)
  235 FORMAT(20X,'PREDICTION LINEAIRE : OUI'/)
  234 IF(IDD.EQ.0)GOTO236
      WRITE(108,237)
  237 FORMAT(20X,'DOUBLE DIFFERENTIATION : OUI',/)
  236 CONTINUE
      WRITE(108,51)(IB(I),I=1,NF)
   51 FORMAT(2X,' B(I) = ',6(3X,I3)/)
      IF(JZ.EQ.0)GOTO231
      WRITE(108,232)(IZ(I),I=1,JZ)
  232 FORMAT(5X,'Z(1) = ',I4,5X,'Z(2) = ',I4/)
      WRITE(108,233)(IZB(I),I=1,JZ)
  233 FORMAT(5X,'IZB(I) = ',2(I3,6X)/)
  231 CONTINUE
      DO102M=NF,1,-1
      FZ=IF(M)
      BX=IB(M)
      CS=COS(2*PI*FZ*D)
      EP1=EXP(-1*BX*PI*D)
      EP2=EP1**2
      J=1
      G(J)=G(J)+2*CS*EP1*(-G(J))-EP2*(-G(J))
      J=2
      G(J)=G(J)+2*CS*EP1*(G(J-1)-G(J))-EP2*(-G(J))
      DO103J=3,XDD
      G(J)=G(J)+2*CS*EP1*(G(J-1)-G(J))-EP2*(G(J-2)-G(J))
  103 CONTINUE
  102 CONTINUE
      IF(JZ.EQ.0)GOTO228
      DO54M=JZ,1,-1
      FZ=IZ(M)
      BX=IZB(M)
      CS=COS(2*PI*FZ*D)
      EP1=EXP(-1*BX*PI*D)
      EP2=EP1**2
      COF=1./(1.-2*EP1*CS +EP2)
      GM1=0
      GM2=0
      DO55J=1,XDD
      GA=G(J)
      G(J)=COF*(G(J)-GM1*2*EP1*CS+GM2*EP2)
      GM2=GM1
   55 GM1=GA
   54 CONTINUE
  228 CONTINUE
      IF(IPL.EQ.0)GOTO215
      ZDD=256*DD
      NZ=XDD-ZDD
      DO104J=(NZ+1),XDD
      G(J)=G(J)*(0.54-(0.46*COS(E2PI*(J-NZ-1)/(ZDD-1))))
  104 CONTINUE
      DO105KK=1,MP+1
      R(KK)=0.
      DO106I=(NZ+1),(XDD-(KK-1))
      R(KK)=R(KK)+G(I)*G(I+(KK-1))
  106 CONTINUE
```

```
      U2=U1
      U1=U0
  101 CONTINUE
      AB(IP)=(2/(2*N+1))*(F(1)+CY*U1-U2)
      B(IP)=(2/(2*N+1))*S*U1
      CC(IP)=SQRT(AB(IP)**2+B(IP)**2)
      PHI(IP)=ATAN(B(IP)/AB(IP))
      Q=C1*CY-S1*S
      S=C1*S+S1*CY
      CY=Q
  100 CONTINUE
      IF(IAFF.EQ.0)GOTO56
      WRITE(108,9)
    9 FORMAT(10X,'RANG',5X,'FREQUENCE (HZ)',8X,'AMPLITUDE (DB)',
     110X,'PHASE (RD)',/////)
   56 SUM=0.
      HMAX=-50
      DO116I=2,N
  116 SUM=SUM+(CC(I)**2)/2.
      SUM=SUM+(CC(1)**2)/4.
      PO2=SUM/10000.
      DO11I=1,N
      NFO=IFO*(I-1)
      II=I-1
      HDE(I)=10*ALOG10(((CC(I)**2)/2)/PO2)
      IF(IPL.EQ.0)GOTO225
      HDE(I)=34.-HDE(I)
      IF(HMAX.LT.HDE(I))HMAX=HDE(I)
  225 CONTINUE
      IF(IAFF.EQ.0)GOTO11
      WRITE(108,12)II,NFO,HDE(I),PHI(I)
   12 FORMAT(10X,I3,10X,I4,15X,F7.1,15X,F8.5)
   11 CONTINUE
      IF(IPL.EQ.0)GOTO229
      DO230I=1,N
  230 HDE(I)=HDE(I)+(38.-HMAX)
  229 DO15I=1,58
      DO15J=1,109
      ITAB(I,J)=IBLANC
   15 CONTINUE
      DO16I=2,56
      ITAB(1,6)=IMOINS
      ITAB(I,109)=IMOINS
   16 CONTINUE
      DO17J=6,109
      ITAB(1,J)=IEGAL
      ITAB(57,J)=IEGAL
   17 CONTINUE
      DO18J=17,108,10
      ITAB(58,J)=IVERT
   18 CONTINUE
      ITAB(58,8)=IVERT
      DO19I=6,56,10
      ITAB(I,6)=IPLUS
      ITAB(I,109)=IPLUS
   19 ITAB(I,4)=IZER
      ITAB(1,3)=ID
      ITAB(1,4)=IV
      ITAB(6,3)=IS4
      ITAB(16,3)=IS3
```

```
      ITAB(26,3)=IS2
      ITAB(36,3)=IS1
      ITAB(56,2)=IMOINS
      ITAB(56,3)=IS1
      IF(N.LT.103)GOTO220
      WRITE(108,31)
   31 FORMAT(1X,'SEUL FO ET LES 100 PREMIERES HARMONIQUES SONT
     1 REPRESENTEES'/)
      N=102
  220 CONTINUE
      WRITE(108,14)
   14 FORMAT(1H1)
      DO221J=3,N,2
      IHDE=HDE(J)
      IF(IHDE.LT.45)GOTO32
      ITAB(2,J+6)=IAST
      GOTO221
   32 CONTINUE
      IF((IHDE+10).LT.0)GOTO33
      DO21I=(46-IHDE),56
      ITAB(I,J+6)=IAST
   21 CONTINUE
      GOTO221
   33 ITAB(56,J+6)=IPLUS
  221 CONTINUE
      DO23I=1,58
      WRITE(108,34)(ITAB(I,J),J=1,109)
   34 FORMAT(2X,109A1)
   23 CONTINUE
      DO35I=1,10
      JFO(I)=IFO*I*10
   35 CONTINUE
      WRITE(108,36)IFO,(JFO(I),I=1,10)
   36 FORMAT(4X,'HZ',3X,I3,6X,10(I4,6X))
      WRITE(108,37)
   37 FORMAT(1H1)
 1000 CONTINUE
      STOP
      END
```

C H A P I T R E XI

OBSERVATIONS SUR LA THEORIE DES TRAITS DISTINCTIFS

C H A P I T R E XII

CONCLUSIONS ET PESPECTIVES

Pour clore ce travail portant sur les indices perceptifs des voyelles orales et nasales françaises, nous rappèlerons, dans une première partie les principales conclusions auxquelles nous avons abouti. Puis nous indiquerons brièvement les axes de recherche qui permettraient de poursuivre l'effort de recherche dans ce domaine particulièrement riche en interrogations.

I - ESPACE PERCEPTIF DES VOYELLES

L'étude de l'espace perceptif des voyelles repose sur l'hypothèse que le timbre peut être représenté comme un point dans un espace à plusieurs dimensions. Pour des raisons de simplicité mathématiques, on considère que les dimensions de l'espace sont indépendantes : la valeur prise par une voyelle sur une dimension ne dépend pas des valeurs sur les autres dimensions. Dans le cadre de ce modèle, deux questions doivent trouver réponse :

-Quel est le nombre de dimensions ?

-Comment s'effectue le calcul des coordonnées d'une voyelle dans l'espace ?

A ces questions fondamentales s'ajoute le problème de l'évaluation des techniques psychologiques permettant de construire cet espace à partir des distances observées.

A/ Analyse monodimensionnelle

Le chapitre II traite des techniques monodimensionnelles employant, sous une forme plus ou moins directe, la notion de "différentiel sémantique". Il n'existe à notre connaissance aucun exposé des fondements théoriques de ces techniques dans leur application au domaine de la perception des voyelles.

L'analyse des travaux publiés montre que les techniques monodimensionnelles sont inadéquates :

- Notre critique théorique majeure est que ces techniques ne permettent pas d'exprimer la notion d'indépendance des dimensions. Elles ne sont donc pas compatibles avec

le modèle de l'espace perceptif présenté plus haut. Beaucoup d'axes présentent les voyelles dans un ordre voisin, sans que l'on puisse décider si les légères différences révèlent une ou plusieurs dimensions.

- Une critique plus pratique concerne la présence d'artefacts.

  - Le premier est lié à la composition phonétique des termes du différentiel sémantique caractérisant les axes soumis aux jugements.

  - Un second artefact est l'apparition de dimensions "esthétiques" jugeant la qualité phonétique ou le naturel des stimuli, plutôt que leur timbre. L'utilisation de voyelles synthétiques semble renforcer cette tendance.

Nous pensons que ces défauts justifient l'abandon des techniques monodimensionnelles.

B/ Analyse multidimensionnelle

Le chapitre III analyse les quelques 15 analyses multidimensionnelles publiées jusqu'en 1977. Lorsque les données s'y prêtaient, nous avons procédé à une nouvelle analyse par la méthode INDSCAL. Nous décrivons également une expérience portant sur les voyelles françaises non-nasales.

a) Problèmes méthodologiques

Sans avoir voulu écrire un "manuel d'utilisation" de la méthode INDSCAL nous avons tenté de clarifier un certain nombre de difficultés :

- l'utilisation de la quantité de variance expliquée dépend de l'indépendance des distributions des coordonnées dans l'espace. Il est utile de vérifier la valeur de ce pourcentage en calculant le coefficient de corrélation entre les distances initiales et les distances dans l'espace perceptif.

- les avantages et les inconvénients des différentes méthodes d'estimation des distances sont passés en revue.

Aucune d'entre elles n'est exempte de problèmes.

- l'utilisation des matrices de confusions se heurte à la non-symétrie des confusions, qui ne permet pas une estimation fiable des distances. De plus, la présence de "cases vides" biaise les estimations des distances entre les stimuli les plus dissemblables

- la méthode triadique, qui exige un important support technologique, sous-estime systématiquement les distances importantes.

- la méthode des "échelles d'intervalle ", que nous avons adoptée, donne des estimations entachées de "bruit". Ceci nécessite l'utilisation des moyennes de groupe. Cette méthode ne permet donc pas l'analyse individuelle. Notre expérience révèle également un effet de "plafonnement" qui provoque une sous-estimation des distances les plus longues.

Toutes ces méthodes se heurtent donc au problème de l'estimation des distances entre les stimuli les plus dissemblables. L'utilisation future des analyses multidimensionnelles nous semble conditionnée dans une large mesure par des progrès à réaliser dans les méthodes d'estimation des distances.

- Notre expérience indique que la transformation des indices de dissimilarité en véritables "distances" améliore quelque peu les résultats.

- Les problèmes posés par l'interprétation des dimensions fait l'objet d'une discussion détaillée. Nous signalons plusieurs critères qui permettent d'estimer le nombre de dimensions, et donc de répondre à la première question. Mais il apparaît que l'analyse multidimensionnelle ne peut, à <u>elle seule</u>, résoudre le problème de l'interprétation détaillée de la transformation des indices acoustiques en indices perceptifs. Ceci est dû, non à un défaut de la méthode elle-même, mais à une caractéristique de notre domaine d'étude : le nombre de dimensions perceptives est inférieur au nombre d'indices acoustiques (formants) ; de plus, ces derniers ne sont pas indépendants

b) L'espace perceptif des voyelles

La conclusion essentielle du chapitre III est que l'espace perceptif des voyelles est bidimensionnel. Celle-ci repose sur les constatations suivantes :

- Dans tous les cas, l'hypothèse d'une troisième dimension est démentie par un ou plusieurs des faits suivants :

  - l'adjonction d'une troisième dimension n'améliore pas la concordance entre les distances subjectives initiales et les distances dans l'espace perceptif.

  - la configuration finale dépend de la position initiale des voyelles sur cette dimension, indiquant que le programme tente d'extraire une "information" dans ce qui n'est que "bruit".

  - les configurations obtenues sur la troisième dimension dépendent de la sous-estimation des distances les plus longues.

  - la "courbure" d'une dimension, résultant de la sous-estimation des points éloignés, crée une dimension supplémentaire artefactuelle.

- Notre expérience indique un coefficient de corrélation extrêmement élevé entre les distances subjectives moyennes et les distances dans l'espace perceptif. Ce résultat montre qu'il est possible de placer les voyelles de telle manière que les distances subjectives soient presque exactement proportionnelle aux distances sur un plan.

- La seule véritable exception est la troisième dimension qui sépare la voyelle rétroflexe [ɹ] de toutes les autres voyelles. A l'inverse des dimensions "artefactuelles", cette dimension se retrouve dans toutes les analyses portant sur l'anglais américain.

- Enfin, ce résultat est compatible avec la possibilité de synthétiser toutes les voyelles à l'aide de deux formants seulement.

Toute autre interprétation signifierait que les sujets ne sont pas capables de fournir une estimation de la distance subjective entre des couples de voyelles. Elle conduirait à rejeter la méthode d'analyse multidimensionnelle elle-même.

Cette conclusion implique que, ni les voyelles antérieures labialisées, ni les voyelles tendues ou relâchées ne se placent sur des dimensions spécifiques.

L'une de ces deux dimensions corrèle fortement avec F1, et il ne fait pas de doute qu'il s'agit du corrélat perceptif de cette caractéristique physique. L'autre dimension corrèle avec F2 en particulier, mais on sait que F2 et F3 ne sont pas indépendants. Il n'est donc pas possible de déterminer le rapport exact entre cette deuxième dimension et les formants F2, F3 et F4. L'analyse de cette deuxième dimension doit se faire à l'aide de techniques différentes, qui font l'objet du chapitre IV.

II - INTERPRETATION DE LA DEUXIEME DIMENSION PERCEPTIVE : F'2

On peut faire l'hypothèse que la deuxième dimension perceptive est le corrélat perceptif du "deuxième formant équivalent", ou F'2, mis en évidence par les expériences de synthèse à 2 formants. Cette interprétation a l'avantage d'unifier deux domaines de recherche distincts. F'2 est une combinaison linéaire des formants F2, F3 et F4. Pour les voyelles antérieures labialisées et les voyelles postérieures, F'2 est égal à F2. Mais pour les voyelles antérieures non-labialisées, F'2 est supérieur à F2. Le cas le plus typique est celui de la voyelle [i] où la différence entre F2 et F'2 atteint plusieurs centaines de Hz. Les conclusions du chapitre IV peuvent se résumer ainsi :

- L'étude directe de F'2 par comparaison du timbre de voyelles à 4 et à 2 formants, ou par test d'identification, donne des résultats trop variables. Ce point a été révélé par une expérience préliminaire de synthèse, qui a montré qu'il était difficile, sinon impossible, d'obtenir une voyelle

synthétique [i] à deux formants ayant un timbre acceptable. Ceci explique sans doute la grande variabilité des estimations directes de CARLSON et al. (1970).

- Un facteur, qui n'a pas été pris en considération jusqu'à présent, est la forte variabilité des fréquences formantiques de la voyelle [i] naturelle. Nous montrons, d'après des résultats publiés par JASSEM (1968) l'existence d'une corrélation inverse entre les quantités (F3-F2) et (F4-F3). Toute formule de calcul de F'2 doit être compatible avec ces variations. Il n'est donc pas permis de représenter la voyelle [i] par une seule combinaison de 4 formants.

- La nouvelle formule de calcul de F'2, proposée par CARLSON et al. (1975), est inadéquate. Nous illustrons ce point en analysant des données publiées qui ne sont pas conformes aux résultats calculés à l'aide de la formule.

- L'impossibilité d'une comparaison directe entre les voyelles à 4 et à 2 formants doit être attribuée au fait que F2, F3 et F4 forment un "bloc" fréquentiel relativement large, supérieur à une bande critique. La réalité perceptive de ce "bloc" fréquentiel est attestée par des données récentes sur le spectre de sonie et les seuils de masquage pour la voyelle [i].

- Nous présentons ensuite une tentative d'extraction directe de F'2 dans le domaine physique par comptage des passages par zéro. Les résultats ne sont pas entièrement satisfaisants. Mais les critiques adressées par CARLSON et al. (1970) ne paraissent pas justifiées.

- En revanche, une technique indirecte d'estimation de F'2 semble donner des résultats intéressants. L'originalité de cette approche est qu'elle intègre la notion de "bande critique". Elle ne nécessite que des tests d'identification portant sur des voyelles à 4 formants.

- Notre modèle d'identification des voyelles isolées dans un espace bidimensionnel montre qu'il est possible de retrouver les données d'une matrice de confusions à partir d'un

nombre limité de paramètres. L'utilisation d'un espace incorporant F'2 semble donner de meilleurs résultats que l'emploi d'un espace ayant une dimension fondée sur F2 seul. Les données expérimentales ont été fournies par un test de perception de voyelles synthétiques. En marge de ce chapitre, nous présentons quelques résultats quantitatifs sur la perception de voyelles synthétiques isolées par des sujets français.

Si ces premiers résultats sont prometteurs, l'établissement d'une formule de calcul de F'2 nécessitera encore des recherches expérimentales importantes.

III - PREDICTION DES SYSTEMES VOCALIQUES PAR UN CRITERE PERCEPTIF : UN ECHEC RELATIF

L'analyse et la prédiction de la stucture interne des systèmes vocaliques sont une tâche essentielle pour le phonéticien et le phonologue. Mais nous pensons que ces études mettent également à jour des renseignements précieux pour l'étude des contraintes articulatoires dans le processus de production ou de perception. Le chapitre V constate l'impossibilité de prédire la structure des systèmes vocaliques à partir de la seule notion de "différenciation maxima" dans l'espace bidimensionnel du timbre. Ce principe de "différenciation maxima", pour reprendre l'expression de MARTINET (1955), suppose que les voyelles tendent à se répartir de manière à assurer la plus grande distance entre elles. Notre conclusion négative est le résultat d'une analyse critique d'un travail de LILJENCRANTS et LINDBLOM (1972). Ces chercheurs tentent de démontrer que ce principe général est suffisant ; notre conclusion repose sur les constatations suivantes :

- L'espace perceptif défini par LILJENCRANTS et LINDBLOM n'est pas conforme aux données expérimentales présentées au chapitre IV. Bien qu'ils prétendent le contraire, nous montrons que leur modèle ne considère qu'un espace défini par F1 et F2 et non F1 et F'2. La distance séparant [i] et [u] est donc largement sous-estimée.

- Or le résultat de la simulation montre que le défaut principal de l'explication proposée réside dans la prédiction d'un trop grand nombre de voyelles "centrales" fermées entre [i] et [u]. Le terme "central" est un qualificatif commode pour toute voyelle qui n'est ni antérieure non-arrondie, ni postérieure arrondie. L'utilisation de la dimension F'2 aurait sans doute conduit à des résultats plus éloignés encore de la réalité.

- La simulation ne prévoit pas de voyelles "centrales" d'aperture moyenne.

- La prédiction d'une structure unique en fonction du nombre de voyelles s'oppose de façon flagrante à la diversité des systèmes vocaliques attestés. On rencontre des difficultés majeures à prédire au moyen d'un principe unique à la fois des systèmes à 4 voyelles comprenant la voyelle [ɨ] , et des systèmes à 7 voyelles ne comportant pas de voyelles centrales.

- Le nombre de degrés d'aperture prédit est incorrect. Ce résultat n'est d'ailleurs qu'une conséquence des mauvaises prédictions relevées précédemment.

Il est probable que des contraintes articulatoires jouent un rôle majeur dans la structuration des systèmes vocaliques. Ce chapitre tente de montrer que le niveau phonologique n'est pas le meilleur pour étudier la structure des systèmes vocaliques. Nous envisageons un niveau intermédiaire, où apparaissent des "allophones majeurs". Un allophone majeur est synchroniquement indépendant du contexte immédiat. Si un système à N phonèmes possède un allophone majeur, sa structure sera identique à l'une des structures possibles pour un système à N + 1 voyelles phonologiques. En d'autres termes, les allophones majeurs obéissent aux mêmes règles structurales que les phonèmes.

IV - ETUDE DE DEUX CONTRAINTES ARTICULATOIRES

L'analyse de la structure des systèmes vocaliques a montré que celle-ci dépend à la fois de critères perceptifs et articulatoires. Il est vraisemblable, comme nous l'indiquons

au chapitre VIII, que des informations articulatoires sont utilisées dans le mécanisme de perception des voyelles en contexte, par le biais d'un centre intégrateur "sensori-moteur". Mais il faut reconnaître que nous ne sommes pas encore en mesure de montrer expérimentalement le rôle des informations articulatoires dans le processus perceptif. Il est donc nécessaire d'étudier les contraintes articulatoires au niveau de la production, pour le moment du moins.

Une meilleure connaissance des contraintes de production permettrait de faire des hypothèses plus précises sur les informations qui pourraient être mises en jeu dans la perception des voyelles en contexte, et sur la forme du centre sensori-moteur. Cet aspect de notre recherche est donc un travail préliminaire à des études sur la perception des voyelles en contexte.

L'étude des contraintes articulatoires suppose que l'on puisse passer de façon précise de la forme du conduit vocal aux paramètres acoustiques, et vice versa. Les chapitres VI et VII présentent des techniques récentes permettant ces deux opérations.

A/ Calcul des fréquences formantiques à partir de la forme du conduit vocal

Nous avons poursuivi les recherches de ZERLING (1974) sur l'utilisation des équations de WEBSTER. Nous proposons un ensemble de corrections fréquentielles qui améliorent considérablement la précision de la méthode.

Ces corrections portent sur :

- l'augmentation de la fréquence du premier formant, due aux vibrations des parois buccales ;
- les variations fréquentielles dues à la radiation du son aux lèvres.

Une comparaison détaillée avec une méthode plus précise, mais plus complexe (MRAYATI et GUERIN 1976) a montré que nos estimations sont numériquement équivalentes aux leurs.

Une condition essentielle est que le conduit soit interpolé de manière à comporter un nombre de sections voisin de 150. Nous confirmons donc l'intérêt qu'il y a à utiliser une équation de WEBSTER discrétisée pour le calcul des fréquences formantiques à partir de la fonction d'aire du conduit vocal.

B/ Calcul de la forme du conduit vocal par prédiction linéaire

La méthode d'analyse par prédiction linéaire est une technique nouvelle qui a connu un succès foudroyant du fait de ses nombreux avantages. Il n'existe pas, à notre connaissance, d'introduction élémentaire satisfaisante. Aussi avons-nous tenu à dériver les équations fondamentales de façon explicite. Le modèle de production implicite est également présenté en détail. Si la méthode de prédiction linéaire sert principalement au calcul de la composition spectrale de la voyelle, WAKITA (1973) a montré qu'il était possible de calculer la forme d'un tube de section variable possédant la fonction de transfert voulue, si l'on dispose d'une estimation de la longueur. Nous avons tenté d'évaluer la précision de la méthode sur des fonctions d'aire connues. Ce calcul nécessite l'utilisation, en sens inverse, des résultats présentés au chapitre VI. Nous montrons qu'il est possible de retrouver les caractéristiques articulatoires essentielles d'une voyelle à partir de la connaissance des fréquences formantiques. Les programmes informatiques sont présentés en annexe. Ces deux "outils" permettent d'étudier deux contraintes articulatoires.

C/ La première contrainte articulatoire :

Pour une voyelles donnée, l'articulation retenue minimise l'écart avec le tube uniforme.

Les étapes de la vérification de cette hypothèse, proposée en premier par PAIGE et ZUE (1970), sont les suivantes :

- Pour une longueur de conduit donnée, les caractéristiques acoustiques définissent la forme du conduit de façon unique.

- On montre expérimentalement que la longueur réelle du conduit vocal, (mesurée par radiographie) est la longueur qui détermine le conduit le plus proche du tube de section uniforme. Ce résultat a été vérifié sur des données publiées (FANT 1960) et sur des données originales portant sur 30 mesures de la longueur du conduit vocal : 5 voyelles prononcées par deux locuteurs dans les trois contextes consonantiques [b-b, d-d, g-g].
L'expérimentation a consisté à calculer, par la méthode de WAKITA, un ensemble de formes articulatoires pour des longueurs variant de 15 à 20 cm. Les fréquences formantiques utilisées ont été mesurées sur les voyelles enregistrées lors de la séance de cinéradiographie. On a déterminé la forme la plus proche du tube uniforme et cette longueur "théorique" a été comparée à la longueur mesurée sur les clichés radiologiques.

- Pour les voyelles antérieures non-labialisées [i, ɛ], la longueur du conduit est faible, ce qui indique que ces voyelles doivent être délabialisées.

- Pour la voyelle [a], la longueur du conduit n'a qu'un effet très faible sur la forme du conduit. Ceci confirme expérimentalement un aspect de la "théorie quantale" de STEVENS (1972).

- Pour des voyelles postérieures labialisées, la forme du conduit est d'autant plus proche du tube uniforme que la longueur est plus importante. Ces voyelles seront donc "naturellement" labialisées. Ces résultats confirment partiellement les résultats de PAIGE et ZUE (1970) et WAKITA (1977).

On obtient donc le résultat fondamental :

LA NON-LABIALISATION DES VOYELLES ANTERIEURES, ET LA LABIALISATION DES VOYELLES POSTERIEURES, PEUVENT ETRE CONSIDEREES COMME LA CONSEQUENCE D'UNE CONTRAINTE ARTICULATOIRE TENDANT A MINIMISER L'ECART ENTRE LA FORME DU CONDUIT ET LE TUBE DE SECTION UNIFORME.

La liaison "naturelle" entre la labialisation et la postériorisation, reconnue par tous les phonologues, reçoit donc une explication "causale".

D/ Deuxième contrainte articulatoire :

Les voyelles se répartissent, de façon privilégiée, sur des "axes" de l'espace où l'effet acoustique est proportionnel à l'activité d'un muscle, ou d'un groupe fonctionnel de muscles.

Nous avons vérifié cette contrainte sur deux "axes" différents.

- Pour l'axe [i - ɛ] on montre que le passage de la voyelle fermée à la voyelle ouverte ne dépend que de l'action du muscle génioglosse.

- Pour l'axe [ɯ - i] on montre au contraire que l'avancée articulatoire n'est pas le résultat d'une activité plus ou moins importante d'un muscle ou d'un groupe de muscles. Mais elle est le résultat d'un réorganisation complète du mouvement articulatoire. L'axe [i - u] n'est pas un axe "naturel", au sens où le passage entre [i] et [ɛ] est déterminé "naturellement" par une diminution de l'activité du genioglossus. L'évolution diachronique [u > y] ou [u > ɨ] ne peut être le résultat que de l'action d'un phénomène extérieur, provoquant une "catastrophe" linguistique nécessitant une réorganisation de l'articulation.

- La voyelle [ɯ] pourrait être une exception à la première contrainte articulatoire, ce qui expliquerait sa relative rareté et son instabilité.

- Nous ne nous prononçons pas sur le cas de l'axe [a-ɯ] en raison de la complexité des processus musculaires mis en jeu.

V - LA DELABIALISATION COMME CAUSE INITIALE DE L'EVOLUTION u > y

Le chapitre VIII a montré que le processus d'évolution [u > y] ne pouvait provenir que d'une "catastrophe" impliquant une réorganisation motrice, et non d'une évolution naturelle. Le chapitre IX étudie les questions que pose l'origine de cette mutation. Nous avons montré tout d'abord que les deux hypothèses traditionnelles se heurtent à des difficultés sérieuses.

- L'explication physiologique de STRAKA n'est pas confirmée : la structure des muscles linguaux n'est pas telle qu'une diminution de la tension provoque automatiquement une avancée articulatoire.

- La "palatalisation spontanée" a eu lieu dans des cas oùle système ne comportait pas 4 voyelles postérieures, ce qui est contraire à la théorie articulo-structurale d'HAUDRICOUR et JUILLAND. L'hypothèse d'une chaîne de pression ne peut donc être retenue dans tous les cas.

Notre hypothèse est la suivante :

- La palatalisation spontanée a eu lieu alors que le système linguistique connaissait une phase de délabialisation.

- Acoustiquement, cette délabialisation conduirait à une confusion [u ~ o] , inacceptable pour des raisons de rendement linguistique.

- L'avancée articulatoire a donc pour but d'augmenter la distance acoustique entre les deux voyelles postérieures fermées.

- Cette délabialisation implique que la voyelle antériorisée est [ɯ] ou [ɨ] . Cette voyelle ne passera éventuellement à [y] , avec l'évolution concomitante [o > u] , que lorsque la langue connaîtra une phase de "relabialisation".

- Le système consonantique est également touché par la délabialisation. On constate souvent des évolutions [b > v, w > v] parallèles à la mutation vocalique.

- Cette hypothèse est parfaitement compatible avec la théorie de l'affaiblissement articulatoire proposée par STRAKA. Il est certain, également, que la présence de 4 voyelles postérieures rend indispensable une avancée articulatoire importante et précoce.

- L'évolution [o > ɸ] paraît liée à la "capture" de la voyelle postérieure délabialisée par une voyelle "centrale" préexistente.

- L'hypothèse d'une chaîne de pression semble pouvoir être retenue pour le scandinave central (suédois).

- L'évolution [ɑ>ɣ] est étudiée dans 10 langues ou dialectes différents.

Notre analyse est un exemple important de l'interaction entre des facteurs articulatoires et acoustico-perceptifs dans l'évolution diachronique des systèmes vocaliques.

VI - STRUCTURE ACOUSTIQUE ET PERCEPTION DES VOYELLES NASALES

Toute étude perceptive suppose que les problèmes de description acoustique ont été résolus. Dans le cas contraire, il est impossible d'identifier les paramètres physiques qui sous-tendent les dimensions perceptives. Or ceci n'est que partiellement vrai pour les voyelles nasales. La première partie du chapitre X avait pour but de vérifier si nous étions capables d'identifier les composantes spectrales des voyelles nasales.

L'analyse d'un corpus de 48 voyelles prononcées par deux locuteurs et la comparaison de nos résultats avec un ensemble d'études expérimentales et théoriques, permettent les constatations suivantes :

- Un premier formant nasal (F1n) est présent dans tous nos échantillons. Mais on note un écart important entre les prédictions théoriques et les valeurs mesurées. Ces dernières sont moins élevées que ne le prévoit la théorie.

- Deux formants oraux "décalés" (F'1 et F'2) sont visibles, sauf pour la voyelle [ɔ̃] où ils peuvent se confondre.

- La présence de 2 zéros n'est pas directement décelable sur les tracés spectrographiques. La simulation a montré que la présence de deux zéros était nécessaire pour expliquer l'amplitude réduite de F'1 par rapport à F'2, ainsi que la pente générale du spectre. Certaines difficultés d'interprétation nous ont conduit à nous demander si la fréquence du premier zéro ne pouvait être supérieure à la fréquence de F'1 dans le cas de la voyelle [ɔ̃] . Il n'existe à notre connaissance aucun moyen de déterminer la largeur de bande de ces zéros.

- On observe, dans certains cas, la présence d'irrégularités spectrales qui pourraient être dues à la présence d'une paire pôle - zéro supplémentaire liée à l'asymétrie des conduits nasaux et à la présence des cavités sinusales.

- La simulation indique que l'intensité des formants élevés (F3, F4) des voyelles naturelles est supérieure aux prévisions. Nous avons fait l'hypothèse que cette forte intensité pouvait être attribuée à une troisième paire pôle - zéro dans cette zone fréquentielle.

- Notre corpus est trop limité pour constituer une source de données normatives sur les fréquences moyennes des différents paramètres. Néanmoins, nos mesures de F'2 pour la voyelle [ɛ̃] révèlent des valeurs systématiquement plus faibles que celles relevées dans les études antérieures.

- Le deuxième formant nasal (F2n) est identifié dans la quasi totalité des voyelles de notre corpus.

- L'invariance de F'2, par rapport au deuxième formant de la voyelle orale ayant la même configuration articulatoire, permet de caractériser grossièrement l'articulation des voyelles nasales sur un axe antérieur-postérieur. Les symboles phonétiques les plus appropriés nous semblent être [æ̃, ɔ̃, õ, ʌ̃] pour les voyelles habituellement notées [ɛ̃, ɑ̃, ɔ̃, œ̃]

- La méthode de prédiction linéaire ne permet pas une étude fine de la structure des voyelles nasales. Mais on peut espérer disposer dans un proche avenir de modèles d'analyse plus adéquats.

Notre étude préliminaire de la perception des voyelles nasales au moyen de voyelles synthétiques sans zéros nous a permis de constater les points suivants :

- Des réponses "nasales" peuvent être obtenues sans que les stimulus ne comportent d'indices spécifiques de la nasalité.

- Les fréquences formantiques F1 et F2 qui induisent ces réponses nasales sont très proches des fréquences mesurées sur les voyelles de notre corpus.

- Ces résultats indiquent que les attributs spécifiques de la nasalité (présence de zéros et de formants nasaux modifiant l'amplitude de F1 et la pente générale du spectre) pourraient être d'une importance plus faible que ce que l'on pouvait imaginer. Il conviendrait pourtant de séparer, mieux que nous ne l'avons fait, la perception de la nasalité et la perception spécifique du timbre de la voyelle nasale. Ce résultat, s'il était confirmé, pourrait être important pour l'interprétation de la genèse des voyelles nasales en français.

VII - LE ROLE DES INDICES PERCEPTIFS ET DES CONTRAINTES ARTICULATOIRES : IMPLICATIONS POUR UNE THEORIE DES TRAITS DISTINCTIFS

Le chapitre XI envisage certains aspects de la théorie des traits distinctifs à la lumière des résultats obtenus.

La structure des systèmes vocaliques dépend à la fois de paramètres articulatoires et acoustico-perceptifs. Nous ne pensons pas que la définition des traits par rapport au seul domaine acoustico-perceptif permette de décrire de façon satisfaisante la structure des systèmes vocaliques. Si certaines caractéristiques phonologiques sont d'origine articulatoire, comme l'harmonie vocalique due à la position de la racine de la langue, d'autres ne peuvent être expliquées que par une interaction entre des paramètres perceptifs et articulatoires. La répartition quasi-équidistante des voyelles fermées ou la non-occurence des voyelles [ɯ] et [ɨ] dans un même système nous semblent liées au principe acoustico-perceptif de la différenciation maxima dans l'espace du timbre. Ce principe, ainsi que la première contrainte articulatoire, explique la labialisation "naturelle" des voyelles postérieures.

La phonologie "dynamique", qui étudie les règles de transformation donne des exemples encore plus probants. Tout système de traits lié trop étroitement à un seul domaine rencontrera de sérieuses difficultés dans la description des évolutions fondées en "substance" sur un domaine différent.

- la mutation [u > y] s'explique sans doute par une avancée articulatoire ayant pour but de faire échec à la confusion auditive entre [o] et [u] provoquée par la délabialisation.

- la deuxième contrainte articulatoire tente d'expliquer la répartition plus fréquente des voyelles sur certains axes de l'espace perceptif par une liaison simple entre l'activité d'un muscle et le résultat acoustique.

- l'ouverture des voyelles nasales en français pourrait être la conséquence de l'apparition d'un zéro qui "décale" le premier formant vers les fréquences élevées.

- dans le domaine consonantique on notera également les évolutions [p < k] et [f > ω]

Nous proposons un ensemble de 4 traits distinctifs pour décrire le sous-système des voyelles fermées. Cet ensemble dérive d'un graphe orienté permettant de générer tous les systèmes attestés et eux seuls. Ces traits définissent 4 positions sur un axe "antérieur-postérieur". L'hypothèse d'une marque globale (+ labialisé), appliquée au système entier, permet de considérer le trait de labialisation comm e redondant.

Nous donnons enfin une description explicite de la réalisation de ces traits au niveau articulatoire et acoustico-perceptif.

VIII - QUELQUES AXES DE RECHERCHE

Au fil des pages, nous avons eu l'occasion d'indiquer un certain nombre d'études complémentaires qu'il serait souhaitable d'entreprendre. Nous avons rassemblé, dans ce dernier paragraphe, la liste de celles qui nous paraissent les plus importantes.

A/ Analyses multidimensionnelles de l'espace perceptif

L'extension des recherches sur l'espace perceptif du timbre vocalique par analyse multidimensionnelle dépend dans une large mesure de l'élaboration de méthodes plus fiables

d'estimation des distances. Le problème de la sous-estimation des distances psychologiques les plus grandes devrait faire l'objet d'une étude menée conjointement par des psychologues et des phonéticiens. Le biais constaté est-il dû à un comportement du sujet ou est-il un artefact lié aux méthodes utilisées jusqu'à présent ?

On peut envisager, comme palliatif, d'étudier une portion seulement de l'espace vocalique. Mais un problème méthodologique doit être résolu: e st-il licite de faire estimer les distances psychologiques entre des stimuli qui n'appartiennent pas tous à la même classe phonologique ? L'utilisation d'une partie de l'espace perceptif exige en effet que l'on présente des voyelles acoustiquement distinctes mais appartenant à la même classe phonologique, car il faut un minimum de 8 voyelles pour effectuer une analyse multidimensionnelle. Une expérience limitée (KASUYA et al. 1971) indique que, pour un même écartement physique, la distance entre 2 voyelles proches d'une "frontière" phonologique est plus grande qu'au centre d'une région phonologique. La précision du placement des voyelles exige l'utilisation de stimuli synthétiques.

Une diversification des stimuli serait souhaitable. On peut envisager l'emploi de voyelles produites par des sujets différents. Mais il est probable que les variations de la fréquence fondamentale conduiront à l'extraction d'une dimension supplémentaire. La mise en évidence des dimensions propres au timbre sera plus difficile, car l'augmentation du nombre de dimensions accroît la sensibilité de celles-ci au "bruit". L'utilisation des voyelles en contexte (cf. ANGLIN 1971) ne semble pas modifier les résultats obtenus à l'aide des voyelles isolées. En revanche, une analyse multidimensionnelle mêlant des voyelles orales et nasales serait d'un grand intérêt. L'absence d'une dimension spécifique de "nasalité" jetterait un doute certain sur la validité de l'analyse multidimensionnelle.

S'il se révèle possible d'utiliser des réalisations d'un même phonème vocalique, il serait intéressant d'inclure, parmi les stimuli, des voyelles ayant subi une forte

coarticulation en contexte. Nous pensons en particulier à [u] dans un contexte [d - d] . Comme on peut le constater, le champ d'expérimentation est vaste, mais la diversification des stimuli ne va pas sans poser de sérieux problèmes méthodologiques.

B/ Le "deuxième formant équivalent"

La poursuite des recherches sur F'2 est un élément fondamental de toute recherche sur les indices de perception. Il est indispensable de diversifier nos méthodes d'estimation de F'2. On peut suggérer, en plus de la méthode indirecte que nous avons proposée au chapitre IV, une méthode de comparaison de voyelles à 4 et 3 formants. Pour ces dernières, F2 et F3 devraient être plus proches en fréquence qu'une bande critique. Ceci devrait permettre à la fois d'améliorer la qualité du timbre (en augmentant la largeur de bande perçue de F'2) et de conserver une estimation relativement précise du pic unique équivalent. Il faut également envisager des comparaisons systématiques entre les scores d'identification de voyelles naturelles ou synthétiques et la représentation de leurs spectres de sonie. L'utilisation de voix féminines et enfantines est également nécessaire pour évaluer l'effet, sur F'2, des variations acoustiques liées à des conduits vocaux différents (cf. FANT, CARLSON et GRANSTRÖM 1974). Ces expériences nous semblent être les plus intéressantes que l'on puisse envisager de réaliser à court terme.

C/ Les contraintes articulatoires

L'absence de paradigmes permettant de mettre directement en évidence le rôle des contraintes articulatoires exigera encore pour un temps que l'on se contente de les étudier au niveau de la production.

Ces études supposent que l'on affine les méthodes d'estimation de la forme du conduit vocal à partir des données acoustiques. On doit, pour cela, disposer de méthodes permettant d'extraire de la manière la plus précise possible les caractéristiques acoustiques (formants et largeurs de bande) des voyelles.

Il convient plus particulièrement d'étudier l'intégration des différents gestes articulatoires en fonction du contexte et de la rapidité d'élocution, afin d'expliquer la raison des écarts articulatoires non compensés. En particulier, le degré de coarticulation est-il fonction de l'effet acoustique produit ? Pour donner un exemple concret : la forte déviation du deuxième formant de [u] dans un contexte apico-alvéolaire, par rapport à la norme, est-il toléré parce qu'aucune confusion n'est possible en français, en raison de l'absence de la voyelle [ɯ] ?

D/ La perception des voyelles en contexte

Malgré le pessimisme de FLANAGAN (1971) que nous rappelions dans l'introduction, il existe au moins un axe de recherche qui pourrait se révéler fructueux. L'exigence d'une normalisation suppose que l'auditeur possède le moyen d'extraire une information concernant la longueur du conduit vocal du locuteur.

Il conviendrait d'étudier systématiquement la sensibilité des divers paramètres physiques des voyelles et des consonnes à des variations de longueur du conduit afin de déterminer les "candidats" possibles. Puis, par des expériences de synthèse, on pourrait étudier l'effet d'une variation de ces paramètres sur la probabilité de classification des voyelles. Un contrôle sur des échantillons naturels s'imposerait évidemment.

On nous permettra de proposer un critère qui pourrait permettre cette normalisation : une première simulation a montré que la fréquence de départ de la transition de l'occlusive [d] est une fonction quasi-linéaire de la longueur du conduit vocal. Cette simulation nous a été suggerée par les résultats de MILLAR et AINSWORTH (1972) que nous avons présentés dans l'introduction On peut également proposer la fréquence de la tache d'explosion (quand celle-ci existe). Mais le problème de la hiérarchie des indices reste posé. La fréquence fondamentale joue certainement un rôle, tout comme la fréquence de F'2, qui peut baisser lorsque F4 augmente (FANT et al. 1974).

Les problèmes de normalisation supposent au préalable que l'on tranche la question de savoir si l'articulation dépend de la longueur du conduit. NORSTRÖM (1977) a suggéré que tel était le cas, bien que les indices soient particulièrement ténus. Le besoin de données radiologiques sur des sujets féminins se fait cruellement sentir.

Enfin l'étude de la perception des transitions dans les cas d'"overshooting" et "d'undershooting" doit être poursuivie. Mais les problèmes paraissent formidables. On peut espérer néanmoins que l'utilisation d'un synthétiseur fournissant une parole de très haute qualité, ainsi que l'emploi de paradigmes expérimentaux réduisant les biais de réponses, pourront éliminer les artefacts qui dénaturent quelque peu les résultats publiés à ce jour (cf. par ex. LINDBLOM et STUDDERT-KENNEDY 1967).

E/ Les voyelles nasales

Nous pensons que les problèmes posés par la perception de la nasalité vocalique doivent être abordés à deux niveaux :

a) la perception de la nasalité elle-même ;
b) la perception du timbre des voyelles nasales.

L'étude détaillée de la hiérarchie des indices de nasalité reste à faire. L'utilisation d'un synthétiseur permettant de spécifier plusieurs zéros et autorisant de larges variations dans les largeurs de bande est une nécessité. L'étude du placement des voyelles dans l'espace multidimensionnel du timbre vocalique est l'une des études que nous espérons réaliser rapidement.
L'examen des rapports entre les indices de nasalité et les évolutions diachroniques nous semble un domaine où la phonétique peut contribuer de façon significative à la solution de problèmes phonologiques.

F/ Typologie des structures vocaliques et système des traits distinctifs

L'étude de la structure des systèmes vocaliques nous semble une source indirecte, mais féconde, d'hypothèses sur

l'utilisation des indices perceptifs et sur les contraintes articulatoires.

Le travail de SEDLACK (1969) sur la typologie des systèmes vocaliques devrait être amplifié.; ceci est d'autant plus souhaitable que les descriptions phonétiques et phonologiques apparaissent à un rythme soutenu. Il n'est plus guère de langues ou de dialectes pour lesquels nous ne possédons pas une étude détaillée. Nous manquons tout particulièrement de données sur les compatibilités et incompatibilités entre des voyelles mi-ouvertes et ouvertes :

L'étude d'un niveau infra-phonologique ("allophones majeurs") nous semble également un domaine plein de promesses. L'étude systématique des évolutions diachroniques et du fonctionnement de l'harmonie vocalique permettra sans doute d'affiner encore le système des traits distinctifs. La liaison entre l'utilisation des indices acoustiques, en rapport avec les contraintes articulatoires, et la structure des systèmes autorisera la phonologie "dynamique" à dépasser le stade notationnel pour atteindre à une véritable explication structurale des processus observés.

En arrêtant ici la description des axes de recherche, nous n'aurons pas la naïveté de nous étendre sur la longueur de la route qu'il reste à parcourir. Beaucoup de progrès ont été réalisés dans l'étude de la perception des sons du langage ; un long travail reste à faire.

Il ne nous appartient pas de juger si ce travail apporte une contribution significative à l'étude des indices de perception des sons vocaliques. Mais nous pouvons affirmer que les expériences faites et l'analyse des travaux de la communauté scientifique ont été pour nous un enrichissement intellectuel considérable. Nous espérons qu'il nous sera donné d'utiliser concrètement les connaissances que nous avons acquises dans des applications à caractère social, comme l'aide aux mal-entendants.

A N N E X E

LE SYNTHETISEUR A FORMANTS DE L'INSTITUT DE PHONETIQUE DE NANCY

LE SYNTHETISEUR A FORMANTS DE L'INSTITUT DE PHONETIQUE DE NANCY

I - INTRODUCTION

Afin d'être utilisable par un phonéticien, un système de synthèse de la parole doit satisfaire à de sévères exigences. Il doit permettre:

- de reproduire de la façon la plus naturelle toutes les caractéristiques acoustiques des sons de la parole (visibles sur un tracéspectrographique, par exemple);
- de spécifier tous les paramètres pertinents, pour les voyelles comme pour les consonnes;
- de produire une parole "naturelle", c'est-à-dire ne donnant pas l'impression de provenir d'une machine.

M. WAJSKOP déclarait à la suite d'un exposé de CARRE (1976) sur la synthèse de la parole:"...il serait important et urgent de veiller à fournir aux phonéticiens des appareillages sur lesquels ils puissent manipuler avec aisance les paramètres tels que la durée. A l'heure actuelle, il est difficile au phonéticien d'obtenir des stimuli de parole aussi simples qu'une syllabe CV dans laquelle la douzaine d'indices acoustiques qui manifestent l'opposition de voisement puissent être contrôlés avec précision et de manière indépendante".

C'est dans cet esprit que nous avons conçu le synthétiseur de l'institut de Phonétique de Nancy.

Cependant, il n'est pas sûr qu'il faille s'orienter vers un dispositif unique comportant l'ensemble des possibilités. La réalisation numérique sur mini-ordinateur permet d'envisager une gamme de synthétiseurs adaptés aux problèmes étudiés. Ceci suppose une construction modulaire dès la conception du projet. Chaque étape ou dispositif de synthèse

pourra être remplacé au besoin par un module plus performant ou plus complet sans désorganiser l'ensemble. Un exemple suffira: pour l'onde de pression glottique, on peut disposer de plusieurs programmes selon que l'on désire modifier ou non en cours de synthèse, les paramètres suivants:

- le type d'onde glottale; celle-ci est en effet différente pour les voyelles et les consonnes (KLATT 1973);
- le rapport phase ouverte/phase fermée ("duty ratio");
- le coefficient d'asymétrie de l'onde;
- la présence et l'importance des "micro-variations" temporelles de F_0 ("time jitter").

Il est bien sûr possible de réaliser un système complet, mais la taille du programme et la multitude des paramètres de commande (dont beaucoup ne seront pas utilisés à chaque expériences) rendront le système peu performant et peu pratique.

II - ARCHITECTURE DU SYSTEME: LA BRANCHE ORALE

Nous avons simulé un synthétiseur du type "à formants" sur un mini-ordinateur DATA GENERAL NOVA 2 (16kmots de 16 bits). Pour une typologie des systèmes de synthèse, on consultera CARRE (1976). L'absence de mémoire à accès lent (type disque ou bande) imposait de placer à la fois le programme et les échantillons de parole dans l'espace de 16 kmots (porté par la suite à 32k mots). La taille du programme devait être réduite au minimum pour obtenir une durée de synthèse maximale. Nous avons donc conçu le programme en langage assembleur. Sa durée est de 1,5 k mots pour la version minimale de la branche orale. Cette solution est également favorable du point de vue de la vitesse d'éxécution. Nous avons obtenu une synthèse en 5 à 10 fois le temps réel, selon les options. Les périphériques d'entrée se composent d'un téletype ASR 33 et d'un clavier-console de visualisation. La sortie des échantillons s'effectue sur un convertisseur digital/analogue de 10 bits, piloté par un générateur d'impulsions. La fréquence d'échantillonnage retenue

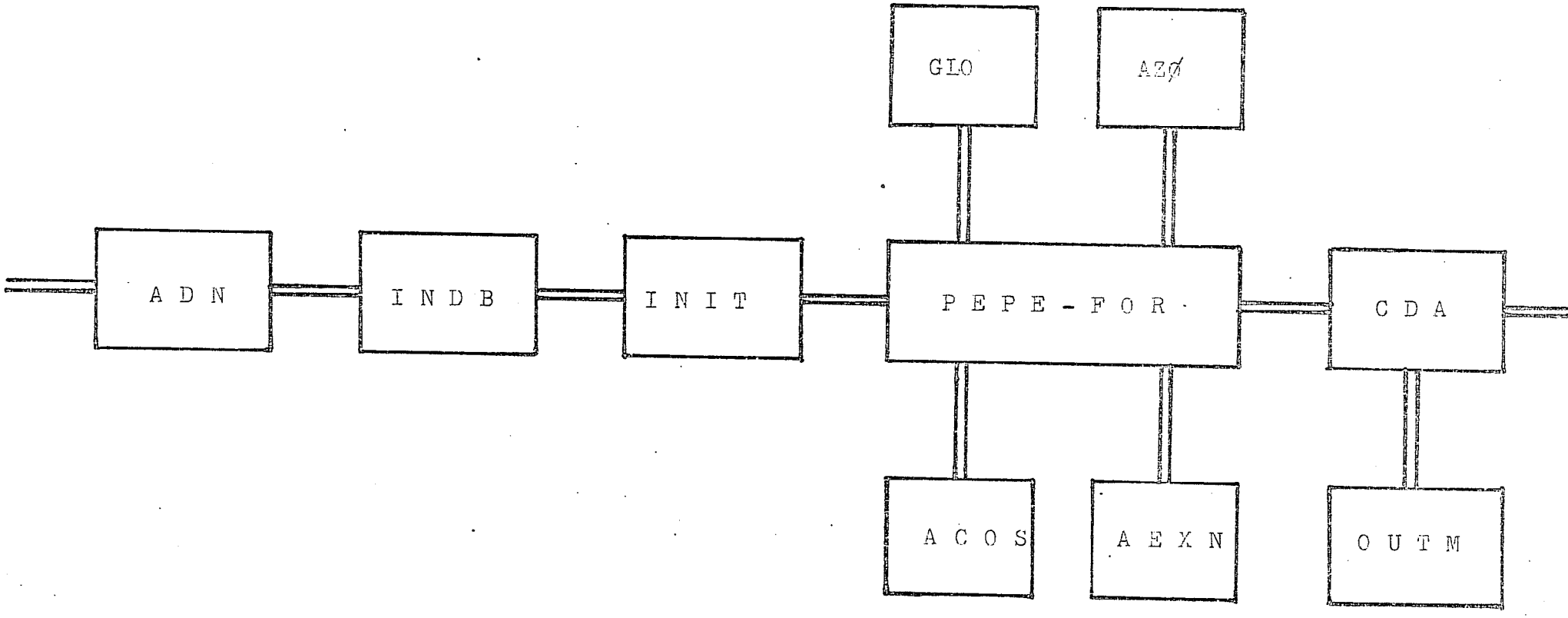


Fig. I - A

est de I0000 Hz, ce qui impose 5000 Hz comme limite supérieure de la synthèse. Un filtre passe-bande (75 - 4700 Hz) à pente raide (-36 dB/octave) élimine les bruits de basse fréquence et le bruit d'échantillonnage à 5 KHz. Le signal est ensuite amplifié et dirigé sur un haut-parleur.

Dans la configuration nécessaire pour synthétiser des sons vocaliques sans nasalité, le dispositif se compose de I0 modules (fig. I -A).

I - ADN: Acquisition des Données Numériques

2 - INDB: INput et conversion Décimale-Binaire

Ces deux modules réalisent l'entrée, le transfert et l'affichage des données (cf. paragraphe sur les commandes) numériques des paramètres de synthèse, ainsi que la conversion décimale-binaire. Les données sont séparées à ce stade en données numériques et alphanumériques. Les informations redondantes (espaces multiples etc.) sont supprimées.

3 - INIT: INITialisation de la table des paramètres.

Ce sous-programme réinitialise au début de chaque période la table des paramètres. Notre synthèse est donc du type "pitch-synchrone". Il interprète également les macro-instructions. Il gère enfin les appels au module PEPE-FOR.

4 - PEPE-FOR: Programme principal et filtrage récursif produisant les maxima du spectre ou FORmants.

Ces deux sous-programmes constituent le coeur du synthétiseur. Ils appellent 4 modules spécialisés que nous détaillerons plus bas. Pour le programme FOR, nous avons adopté la configuration exigeant le minimum de multiplications (fig. 2): cf. FLANAGAN I972 p. 264 fig. 6.36b. L'expression algébrique équivalente du filtre est

$$x(i) = x(i) + 2e^{(-\pi BiD)}.\cos(2\pi FiD)(x(i-I)-x(i)) - e^{(-2\pi Bid)}(x(i-2)-x(i))$$

D: période d'échantillonnage= 0,000Is; Fi: fréquence (en Hz) du ième formant; Bi: largeur de bande (en Hz) du ième formant.

Le calcul des échantillons se fait en "pseudo double précision" à l'intérieur du filtre récursif. Nous utilisons un mot (I5 bits + signe) pour la partie entière (quelque soit sa valeur), et un mot (I6 bits) pour la partie fractionnaire ou une somme de parties fractionnaires inférieure à 2^{I6} bits. A la sortie du filtre, les données sont converties en simple

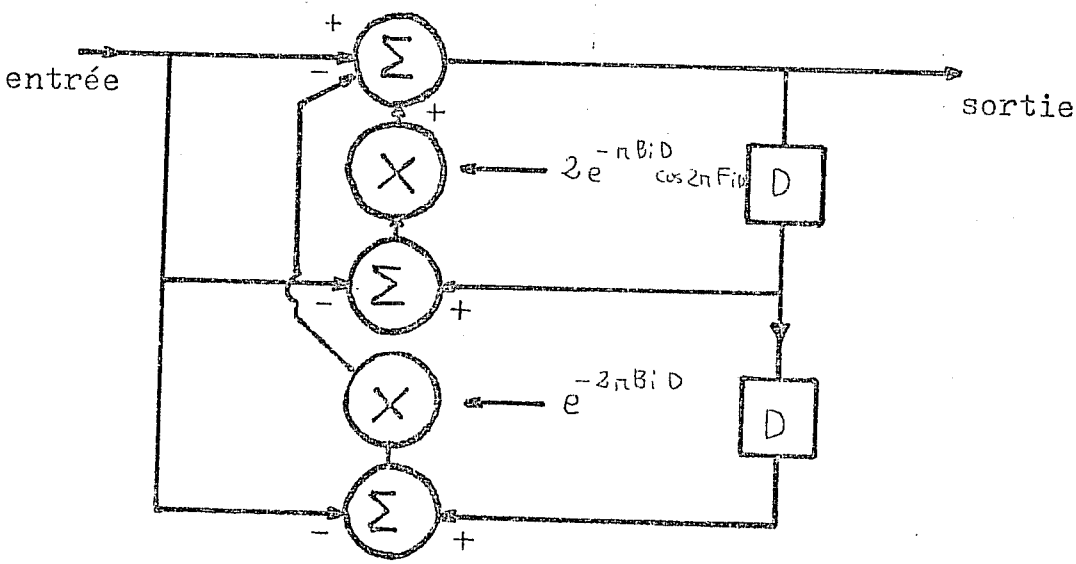


fig. 2 -A

précision avec arrondie. L'erreur maximale est donc de 0,5 bit par formant, c'est-à-dire au maximum 2,5 bits pour 5 formants. L'erreur quadratique moyenne (RMS) est beaucoup plus faible: I,5 bits. Le calcul des coefficients du filtre, la différenciation et la cadrage de l'amplitude de l'onde glottique ajoutent une certaine erreur qu'il est difficile de quantifier (I bit?). Cette précision est suffisante car le convertisseur ne prend que les IO ou I2 premiers bits. Une erreur de 4 bits est donc acceptable en théorie. L'utilisation d'un filtre récursif en simple précision peut provoquer une erreur de 2 bits par formant, ce qui est excessif. GOLD et RABINER (I968) ont d'ailleurs calculé que l'onde de pression devait être codée sur I8 bits pour assurer une précision suffisante de la synthèse.

Notre programme n'impose aucune restriction sur les valeurs de Fi et Bi. La précision est de l'ordre de I Hz.

Le programme PEPE gère les appels au programme FOR. L'ordre de calcul des formants est F4 - FI - F2 - F3. Cet ordre a été choisi pour éviter les débordements dus au gain des filtres. Le gain important de F4 s'applique à l'onde différenciée. Le gain faible, nettement inférieur à l'unité, de FI réduit la dynamique nécessaire. Les gains de F2 et F3 sont voisins de l'unité et ne posent guère de problèmes. L'utilisation de mots de I6 bits rend difficile le calcul de F5. Ce dernier, en raison de son gain très important, nécessite le recours à des décalages à droite (division par 2) dès que la fréquence fondamentale dépasse I80 Hz. Le filtre correspondant à F5 possède un gain supérieur à IO^2. Nous avons vérifié que la suppression de F5 ne détériore pas la qualité de la synthèse.

Quatre sous-programmes sont gérés par PEPE:

5 - <u>GLO</u>: Création de l'onde <u>GLO</u>ttale

La valeur de F_o est convertie en un nombre entier (arrondi) d'échantillons qui détermine la durée de la période. On introduit donc une erreur de I% environ sur la fréquence fondamentale. Il est possible (KLATT I973) de diminuer cette erreur en simulant une fréquence d'échantillonnage supérieure pour la création de l'onde glottale et en réalisant un "down-sampling" ultérieurement (réduction de la fréquence d'échantillonnage).

Nous n'avons pas jugé utile d'introduire cette procédure plus longue dans la partie préliminaire de notre étude.

Le programme GLO calcule une onde glottale qui sera décrite plus bas. Une des caractéristiques essentielles de notre synthétiseur est l'introduction de faibles variations aléatoires de F_o et A_o. Cet aspect sera examiné en détail plus bas. L'onde glottale est ensuite différenciée pour approximer l'effet spectral de la radiation (cf. chapitre VI annexe E).

6 - AZØ: Contrôle de l'amplitude A_o.

Ce module assure le calcul du coefficient multiplicatif correspondant à la valeur en dB. Il introduit également une légère variation aléatoire de l'onde glottale.

7 - ACOS: Calcul de la quantité cos(2πFiD) utilisé dans le filtre récursif.

Nous utilisons une expansion en série pour des raisons de précision et d'encombrement en mémoire. Celle-ci est différente selon la valeur de l'argument: pour des valeurs supérieures à $\pi/4$ et inférieures à $\pi-\pi/4$, le cosinus est utilisé; pour des valeurs hors de ces deux bornes nous utilisons l'expansion de sinus $(\pi/2 - x)$.

8 - AEXN: Calcul des deux EXponentielles Négatives du filtre récursif.

Une expansion en série est utilisée. Il faut noter que la deuxième exponentielle n'est que le carré de la première.

9 - CDA: Préparation de la Conversion Digitale/Analogue.

Ce module réalise le recadrage des échantillons afin d'utiliser au mieux les I0 ou I2 bits disponibles en sortie. Il gère la sortie des échantillons et réalise le cas échéant la répétition automatique de l'élément synthétisé toute les 2 ou 3 secondes.

I0 - OUTM: sortie numérique de valeurs positives ou négatives.

Ce programme est utilisé accessoirement pour obtenir le listing numérique des échantillons à tous les stades de la synthèse et pour des intervalles de temps variables. C'est un outil précieux pour la mise au point et la vérification des nouveaux modules. Sa commande se fait directement par les clés de l'ordinateur.

Notre objectif étant une parole de bonne qualité, nous présentons ci-dessous les solutions adoptées pour réduire le caractère "artificiel" de l'élocution de notre synthétiseur.

III - ONDE GLOTTALE

La forme retenue est la forme polynomiale testée par ROSENBERG (I97I). Elle a été jugée la plus naturelle pour la synthèse de syllabes et de phrases dans cette étude. On peut regretter que le dispositif de KLATT (I973), qui consiste en une impulsion modelée par passage dans un filtre à deux pôles, n'ait pas fait l'objet d'une recherche perceptive comparable. La fig. 3-A présente la forme de l'onde glottale pour F_o = I25 Hz. Le spectre calculé est présenté à la fig. 4-A et à la table I. On vérifie que la pente moyenne est très proche de -I2 dB par octave. On observera l'absence de zéro dans la zone fréquentielle utile.

Les deux polynomes décrivant l'évolution temporelle sont:

$$\begin{array}{ll} I \leqslant t \leqslant a & 3(t/a)^2 - 2(t/a)^3 \\ a+I \leqslant t \leqslant a+b & (I - ((t-a)/b)^2) \\ a+b+I \leqslant t \leqslant p & 0 \end{array}$$

Les paramètres a et b qui définissent la durée de la phase montante et descendante ont été fixés à 40 et I6% de la durée totale de la période(P). La forme de l'onde est donc invariante par rapport à F_o. Sur la fig 3, l'amplitude a été multipliée par 2^{I6}-I (65535) qui est la valeur maximale d'un mot de I6 bits.

ESPESSER (I977) utilise également cette onde glottale. Nous avons testé numériquement et auditivement la modification suggérée par RICE (I974). On constate une détérioration de la qualité. Celle-ci est peut-être à mettre en rapport avec l'apparition de zéros vers I200 et 2500 Hz. La pente moyenne devient également légèrement plus forte, ce qui réduit l'amplitude des formants élevés. Il ne semble donc pas souhaitable de modifier, comme le propose RICE, la manière "brutale" dont l'onde rejoint la phase fermée.

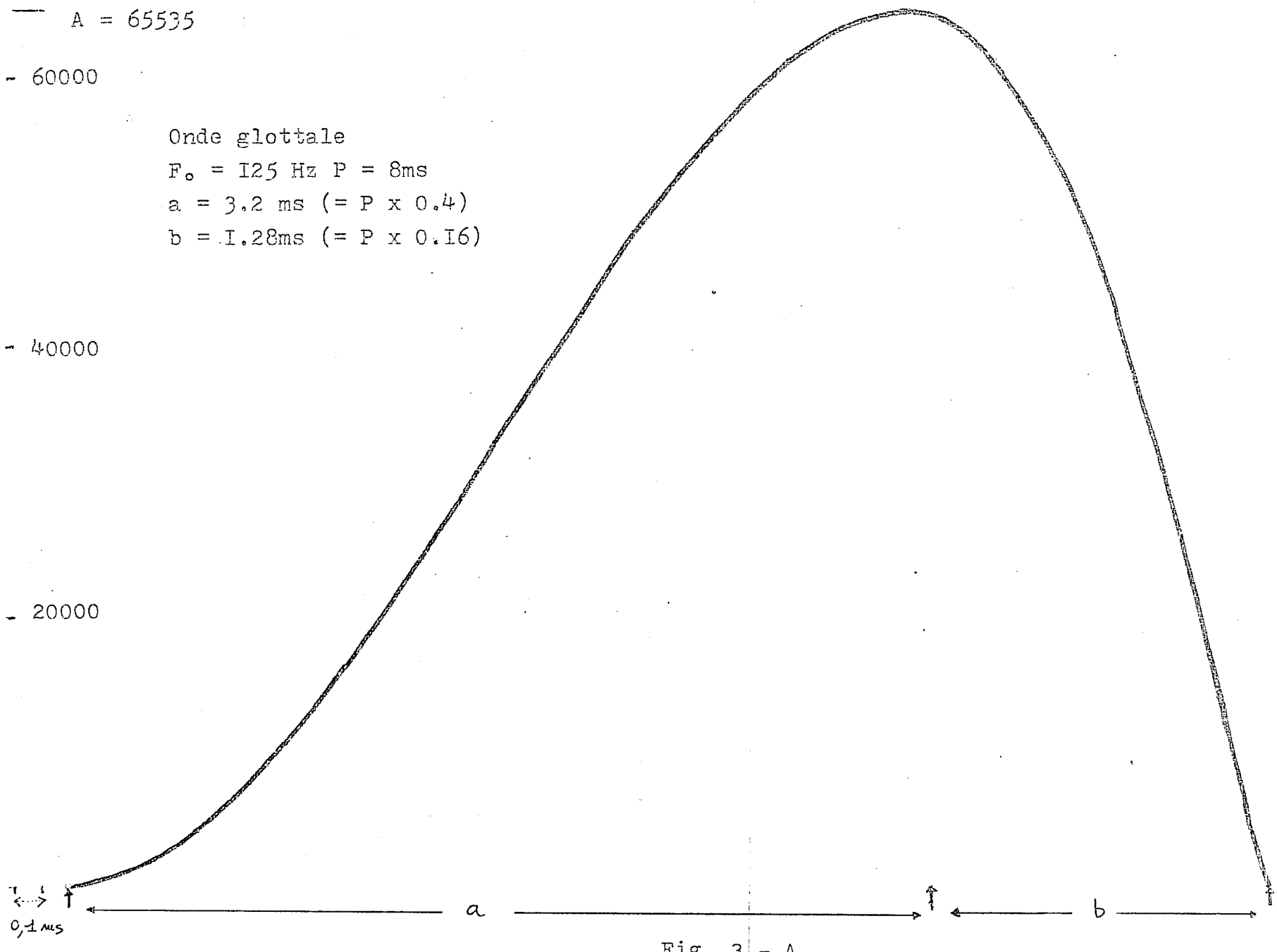

A = 65535
60000
40000
20000
Onde glottale
F₀ = I25 Hz P = 8ms
a = 3.2 ms (= P x 0.4)
b = I.28ms (= P x 0.I6)
0,1 ms
a
b

Fig. 3 - A

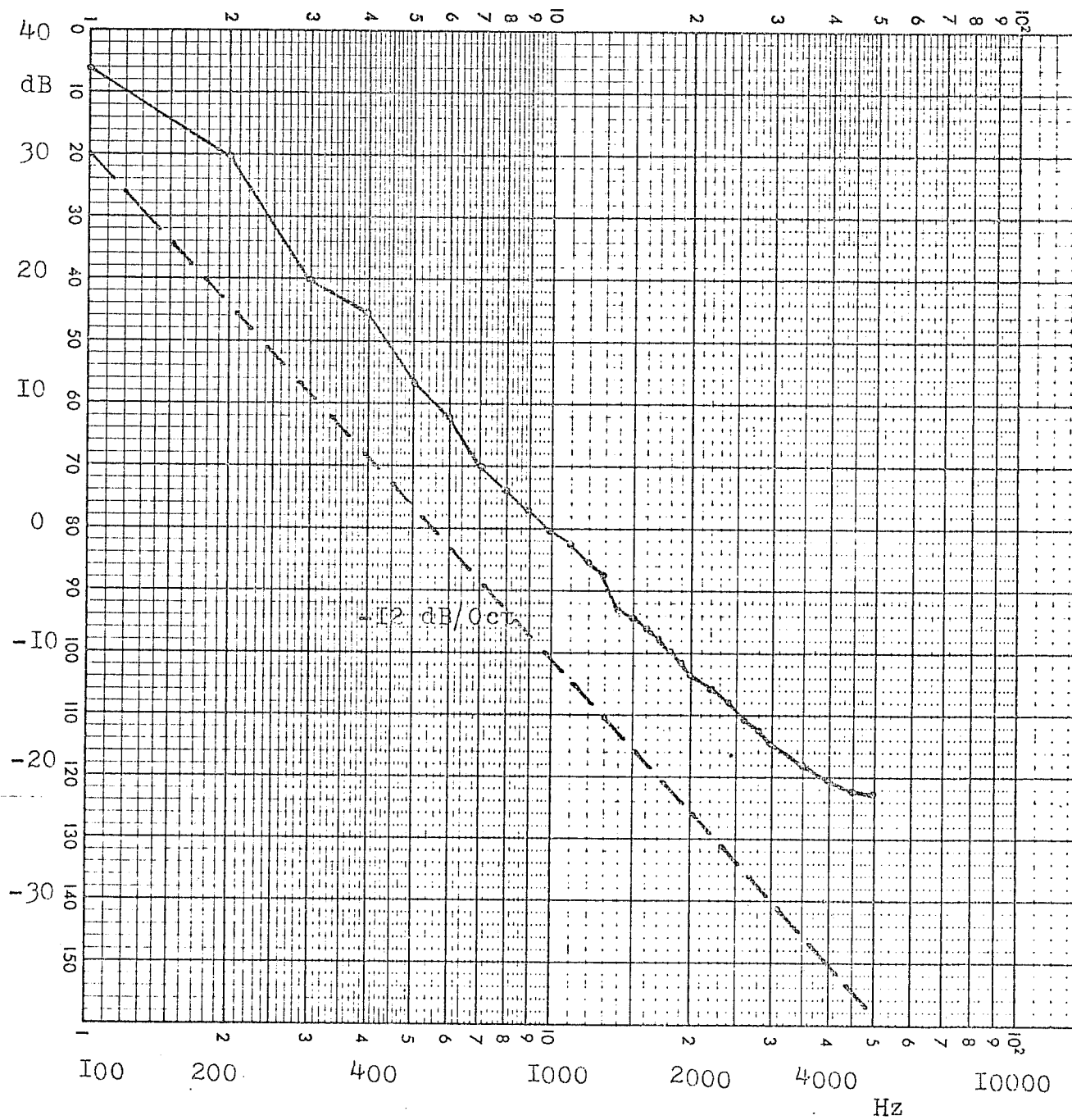

40
dB
30
20
I0
0
-I0
-20
-30
-I2 dB/Octv
I00
200
400
I000
2000
4000
I0000
Hz

Fig. 4 - A

RANG	FREQUENCE (HZ)	AMPLITUDE (DB)	PHASE (RD)
0	0	39.0	.00000
1	100	36.9	-1.05144
2	200	29.8	1.23423
3	300	20.2	1.33832
4	400	17.5	1.37275
5	500	11.7	-1.55859
6	600	9.1	-1.43048
7	700	5.1	-1.02419
8	800	3.1	-.81933
9	900	1.5	-.40747
10	1000	-.1	-.24831
11	1100	-1.2	.04843
12	1200	-3.4	.24291
13	1300	-4.5	.54233
14	1400	-6.4	.84829
15	1500	-6.9	1.12963
16	1600	-8.0	1.44483
17	1700	-8.6	-1.48704
18	1800	-9.7	-1.18770
19	1900	-10.7	-.96350
20	2000	-11.5	-.63385
21	2100	-12.5	-.36295
22	2200	-12.7	-.06246
23	2300	-13.5	.20355
24	2400	-13.8	.45623
25	2500	-14.7	.74196
26	2600	-15.2	1.00116
27	2700	-15.8	1.31930
28	2800	-16.2	-1.55955
29	2900	-16.5	-1.26130
30	3000	-17.0	-1.01060
31	3100	-17.3	-.73171
32	3200	-17.9	-.45585
33	3300	-18.1	-.17221
34	3400	-18.5	.12334
35	3500	-18.7	.39073
36	3600	-18.9	.67763
37	3700	-19.2	.93577
38	3800	-19.5	1.22455
39	3900	-19.8	1.49741
40	4000	-19.9	-1.35356
41	4100	-20.1	-1.07428
42	4200	-20.1	-.79919
43	4300	-20.3	-.52071
44	4400	-20.4	-.24826
45	4500	-20.6	.03491
46	4600	-20.7	.31466
47	4700	-20.7	.59689
48	4800	-20.8	.87289
49	4900	-20.8	1.15150

Table I- A

IV - VARIATIONS ALEATOIRES DE F_0 ET DE L'INTENSITE.

Une recherche bibliographique, présentée à la fin de cette annexe, a permis de trouver plus de IO études sur l'amélioration de la qualité de synthèse. Mais la plupart n'ont pas été publiées, ou concernent le problème du déphasage dans les vocoders à canaux. Une grande étude dans ce domaine reste à faire. On sait néanmoins que des "micro-variations" de la fréquence fondamentale ("F_0 random jitter"), et sans doute de l'amplitude, diminuent le caractère artificiel des sons de synthèse. Les résultats de ROSENBERG (I968) pour des voyelles naturelles modifiées ne confirment pas le rôle de ces micro-variations. FLANAGAN (I972 p. 245) résume ainsi cette expérience:"La structure fine de la forme de l'onde glottale et les variations à court-terme de la fréquence fondamentale ne sont pas perceptivement significatives, à l'inverse des variations moyennes de ces paramètres sur plusieurs périodes". Ceci est de toute évidence inexact et nous confirmons , après de nombreux phonéticiens, l'importance de ces micro-variations pour le naturel de la synthèse.

Pour l'implémentation de celles-ci, nous avons utilisé les deux seules sources de données expérimentales sur les variations de F_0, et de l'amplitude, à court-terme : (les données de LIBERMAN (I96I), obtenues sur des phrases dont la mélodie était variable, ne conviennent pas) LANDERCY (I973) et HOLLIEN et al. (I973).

Les principaux résultats quantitatifs de LANDERCY sont rappelés dans la table ci-dessous.

voyelle	a	i	u	o	e	
moyenne F_0	89.9	96.4	99.8	IOI.2	IOI.I	sujet MW
écart-type	0.95	0.58	0.49	0.49	0.43	
écart moyen absolu	0.65	0.46	0.36	0.35	0.26	
moyenne F_0	95.0	I02.2	I04.I	96.7	97.I	sujet AN
écart type	0.63	0.7I	0.83	0.74	0.33	
écart moyen absolu	0.44	0.59	0.58	0.50	0.29	

L'écart type moyen s'établit à 0.6I8, ce qui indique que 95% des variations sont comprises entre ± I.25 Hz. La moyenne des écarts absolus est de 0.45 Hz.

Les résultats de HOLLIEN et al. confirment ces ordres de grandeur.

Voyelle [a]	sujet	moyenne f_o	écart absolu	
	I	I00.2 Hz	0.55 Hz	
	2	I00.9	0.39	
	3	I03.I	0.40	
	4	IOI.6	0.40	moyenne 0.48 Hz

Pour les sujets testés, la moyenne des écarts absolus varie peu en pourcentage de I00 à 200 Hz.

Dans un premier temps, nous avons introduit des variations légèrement plus fortes que celles présentées plus haut.(moyenne nulle, écart type = I.22). On obtient une voix de mauvaise qualité que l'on pourrait décrire comme "rauque". Après quelques tentatives, nous avons retenu les valeurs suivantes: écart-type = 0.428, moyenne des écarts = 0.I8 Hz). Ces valeurs sont inférieures aux moyennes expérimentales. On peut se demander si ces dernières ne sont pas "gonflées" des variations dans le défilement de la bande magnérique ("tape flutter"). L'absence de tests portant sur de nombreux sujets ne permet pas d'affirmer que nos valeurs sont optimales. Elles sont des minima pour lesquels toute trace de "raucité" a disparu. Il est à noter que ces valeurs n'ont pas été testées en l'absence de variations de l'amplitude.

Techniquement, ces variations sont obtenues en faisant varier, en plus ou en moins, la valeur de la période d'un échantillon (0.I msec) selon le tableau suivant:

0	-I	0	0	0	0	I	0	0	0
0	0	0	-I	0	0	0	0	-I	0
0	0	-I	0	0	0	0	I	0	0
0	0	0	-I	0	0	0	0	0	0
I	0	0	0	-I	0	0	0	0	0

Les variations se répètent toutes les 50 périodes (0.5 s pour

F_o = I00 Hz). Le pointeur de la table n'étant pas réinitialisé au début de chaque synthèse, le point de départ dans la table est variable.

Les résultats expérimentaux de LANDERCY (I973) concernant les variations de l'amplitude sont plus succincts. Il indique une variation de $^{+}-$ I.5 % de l'intensité moyenne. Nous avons retenu les variations présentées dans la table ci-dessous. Le chiffre I indique que l'amplitude est diminuée de 0.28 dB. Cette valeur peut sembler très faible, mais les résultats du sujet A.N pour la voyelle [a] indiquent qu'une série de I0 périodes consécutives peut ne pas présenter des écarts de plus de 0,2dB entre chaque période.

0	0	I	0	0	I	0	0	0	I
0	0	0	I	I	0	0	0	0	I
0	0	0	0	I	0	0	0	I	0

La réduction de 0,28 dB est obtenue en diminuant la valeur de l'amplitude de I/32ème (5 décalages à droite). En effet si

$$20 \log x/y = z \text{ (dB)}, \quad 20 \log (x/y - x/32.y) =$$

$$20 \log (x/y.(3I/32) = 20\log x/y - 20\log 3I/32 = z - 0,28$$

V - LES COMMANDES

Un effort particulier a été fait pour simplifier la commande du synthétiseur. Le format d'introduction des paramètres est le moins restrictif possible(blancs redondants autorisés etc...).

- Le programme affiche sur l'écran de visualisation la liste des paramètres nécessaires. Pour la synthèse de voyelles orales, la liste est la suivante:

  F0 A0 F1 F2 F3 B1 B2 B3 F4 F5 B4 B5

  F_o, amplitude, fréquence de F1, F2 et F3, largeurs de bande de ces formants, fréquences, puis largeurs de bande de F4 et F5.

- L'utilisateur frappe au clavier les valeurs de ces paramètres en unités naturelles (Hz ou dB).

100 60 300 2200 2800 60 80 100 3400 4400 200 300

Pour des raisons pratiques, le "circuit" du 5ème formant n'est pas connecté. La ligne est terminée par un "retour chariot". La ligne précédente décrit la première période de la voyelle. 60 dB représente l'intensité maximale. La durée est implicitement spécifiée par F_o. La lignes suivantes décriront les périodes successives de la voyelle. Deux cas peuvent se produire:

a) - certains paramètres sont modifiés; il suffira de taper la liste des paramètres jusqu'au paramètre modifié. L'ordre adopté a donc pour but de placer en tête les paramètres qui sont le plus souvent modifiés.

b) - aucun paramètre ne doit être modifié; dans ce cas, on peut utiliser la "macro-commande" Ri (i nombre variant de I à 2^{I6}). Celle-ci indique au programme que les i périodes suivantes seront identiques à la ligne précédente.

- Une deuxième commande (F) signale la fin de la description et lance la synthèse. Le son synthétisé est répété automatiquement toutes les 2 secondes, pour faciliter l'évaluation auditive.

VI- DEVELOPPEMENTS FUTURS

Sur un plan pratique, de nombreuses améliorations peuvent être apportées; citons notamment

- corrections des erreurs à l'introduction des données;
- Contrôle du nombre de paramètres en entrée;
- Tabulation automatique des paramètres;
- Calcul automatique et affichage de la durée pour chaque ligne;
- vitesse d'échantillonnage variable.

Au plan du developpement, les étapes suivantes sont envisagée

- BRANCHE NASALE

Celle-ci pose un problème pratique d'encombrement mémoire et de temps de calcul. KLATT (I973) utilise une branche nasale consonantique (5 formants) parallèle à la branche orale. Pour les voyelles nasales, on peut adopter la même solution en ajout ant quelques zéros. Cette solution augmente assez sensiblement la taille du programme. Il est possible de n'utiliser qu'une seule branche, mais l'effet des zéros doit être annulé par un nombre identique de pôles pour les sons sans nasalité. Cette procédure double approximativement le temps de calcul. Nous envisageons une procédure "pseudo-parallèle" où les zéros sont introduits avant le "circuit" formantique . A la fin de chaque son nasalisé, on calcule sur quelques périodes la contribution temporelle de l'énergie résiduelle dans cette "pseudo-branche" nasale. Ce dispositif est en cours d'évaluation.

- BRANCHES CONSONANTIQUES

Si des travaux importants (RABINER I968, HOLMES I973, KLATT I973) ont proposé des solutions intéressantes pour les divers systèmes de synthèse des sons consonantiques, il reste de nombreux détails à préciser. A titre d'exemple, on peut citer les points suivants:

- Rôle de la forme de l'onde glottale dans les consonnes voisées;
- Spectre et contrôle du V.O.T;
- Création de la barre ou de la tache d'explosion des occlusives;
- Modélisation du mélange "bruit/son" dans le spectre des sonantes (l, R, etc..).

Comme l'écrit, à juste titre, CARRE (I976p. 4I)"...Il serait dangereux de laisser croire qu'il n'y a plus, ou presque plus, de problèmes dans ce domaine de recherche".

La synthèse nous semble être un outil indispensable pour les recherches sur la perception, à condition qu'elle soit de

de très haute qualité.

VII - ILLUSTRATIONS

La fig. 5 - A(a - g) présente des spectrogrammes illustrant certaines caractéristiques des voyelles produites par notre synthétiseur.

a) - Voyelle [a] synthétique; 4 formants; présence de micro-variations de la fréquence fondamentale et de l'intensité. La comparaison avec la fig. b montre l'effet de ces variations sur les harmoniques hautes.

b) - Voyelle [a] synthétique; 5 formants; absence de micro-variations de F_o et de A_o.

c) - Voyelle [i] synthétique; 4 formants; on notera l'attaque "douce".

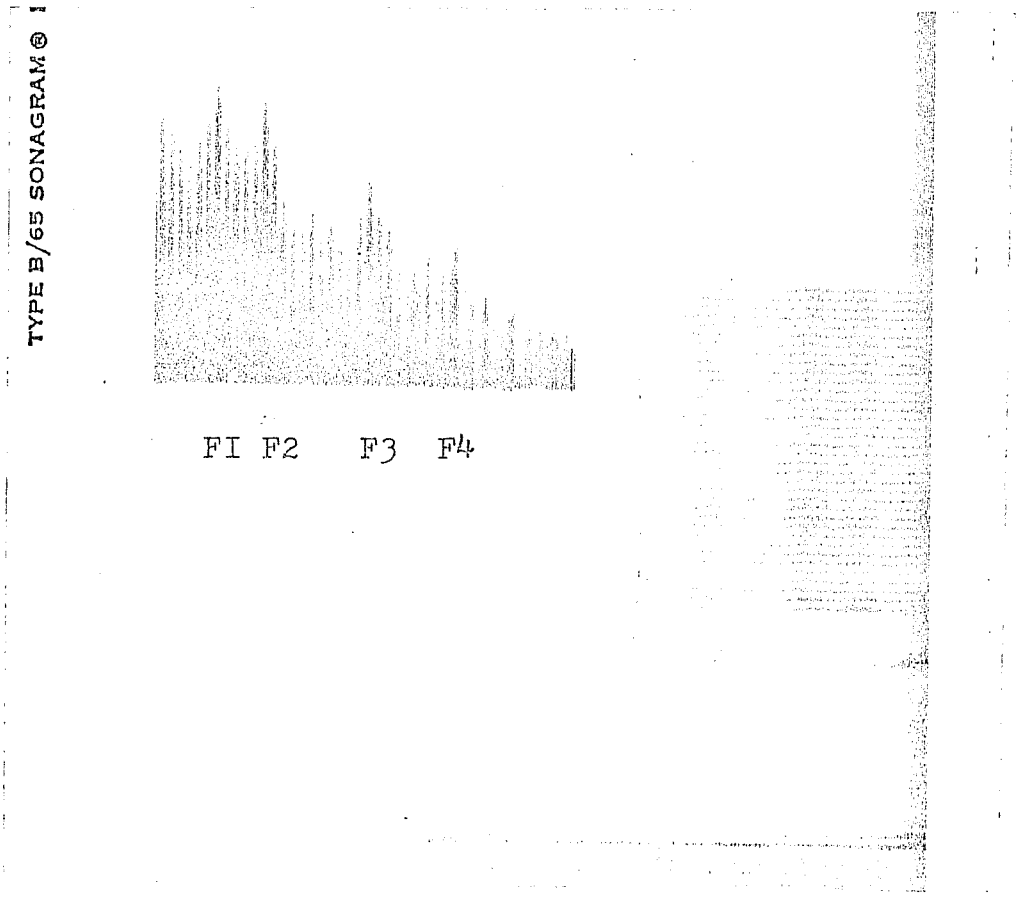
d) - Voyelle [i] synthétique; 5 formants; F_0 = 110 Hz, F1 = 250, F2 = 2200, F3 = 2900, F4 = 3400, F5 = 4400 Hz; absence de micro-variations.

e) - Voyelle [i] naturelle; on notera la très grande ressemblance entre les spectres synthétique et naturel.

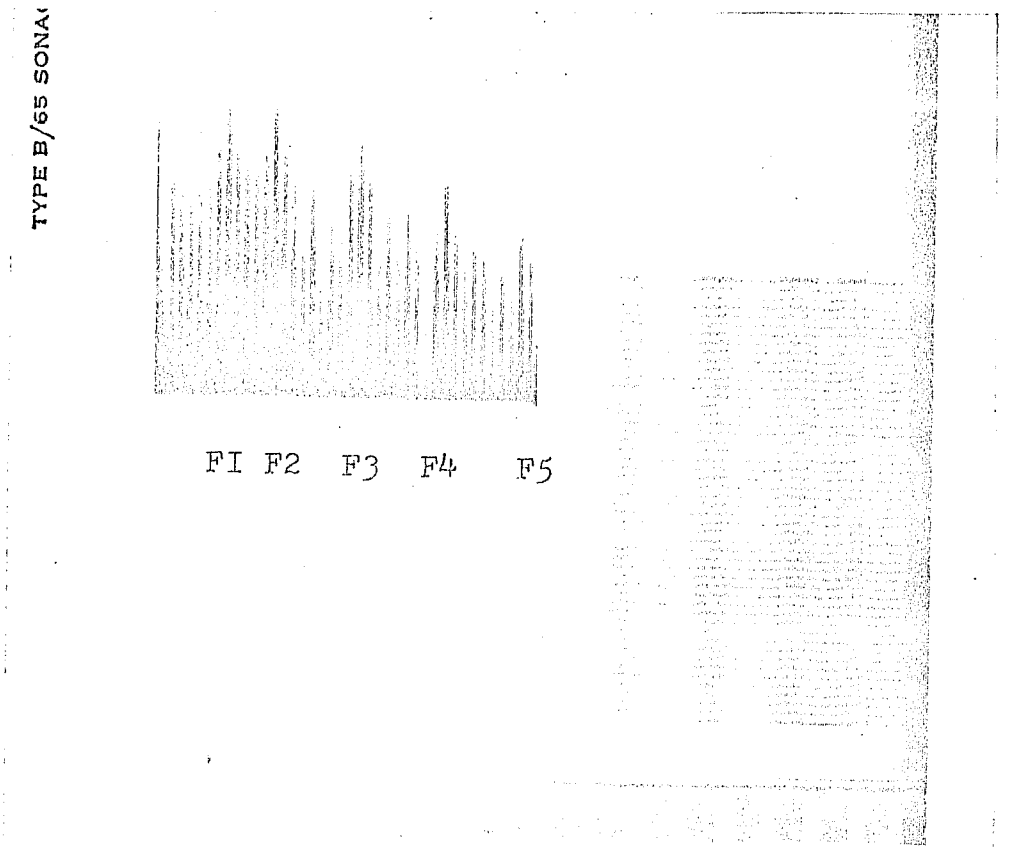
f) - Syllable [ga] synthétique; essai préliminaire de synthèse d'une tache d'explosion; absence de la barre de voisement (VOT = 0).

g) - Syllabe [da] synthétique; la dynamique du spectrographe est trop faible pour faire apparaître le 4ème formant.

La fig. 6 - A présente trois voyelles synthétiques [a] analysées au détecteur de mélodie et à l'intensimètre. On distingue nettement les variations aléatoires de la fréquence fondamentale et de l'intensité. La non-réinitialisation du pointeur de la table des variations assure que celles-ci sont différentes pour chaque voyelle.

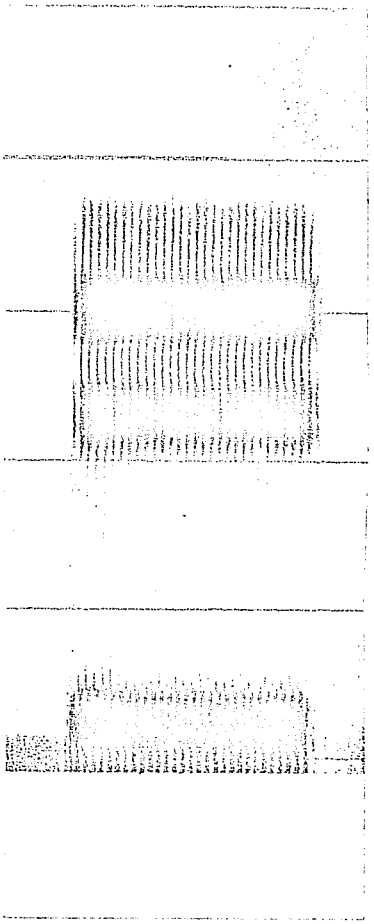


a)

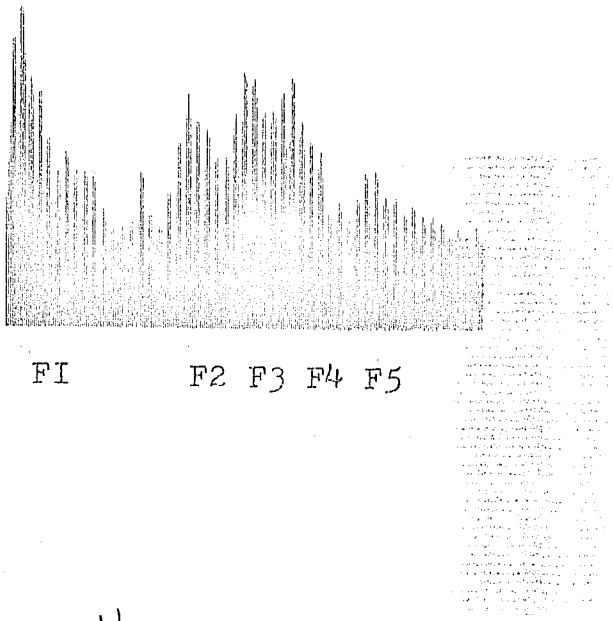


b)

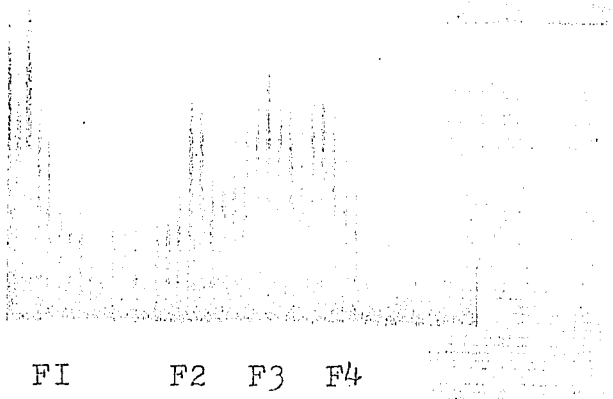
TYPE B/65



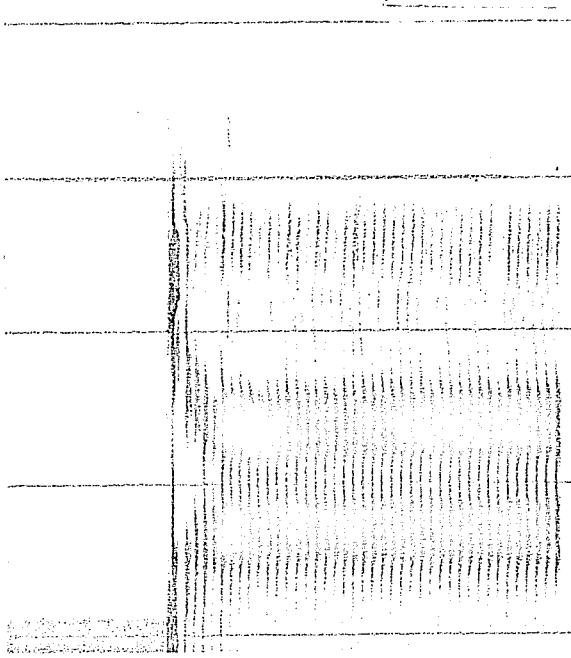
c)



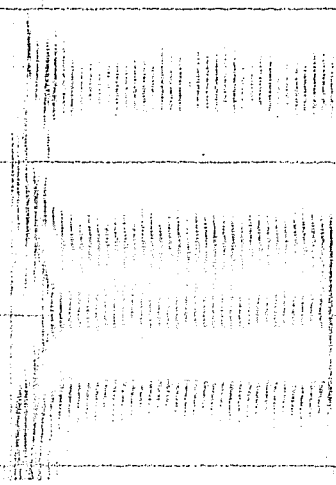
d)



e)



f)



g)

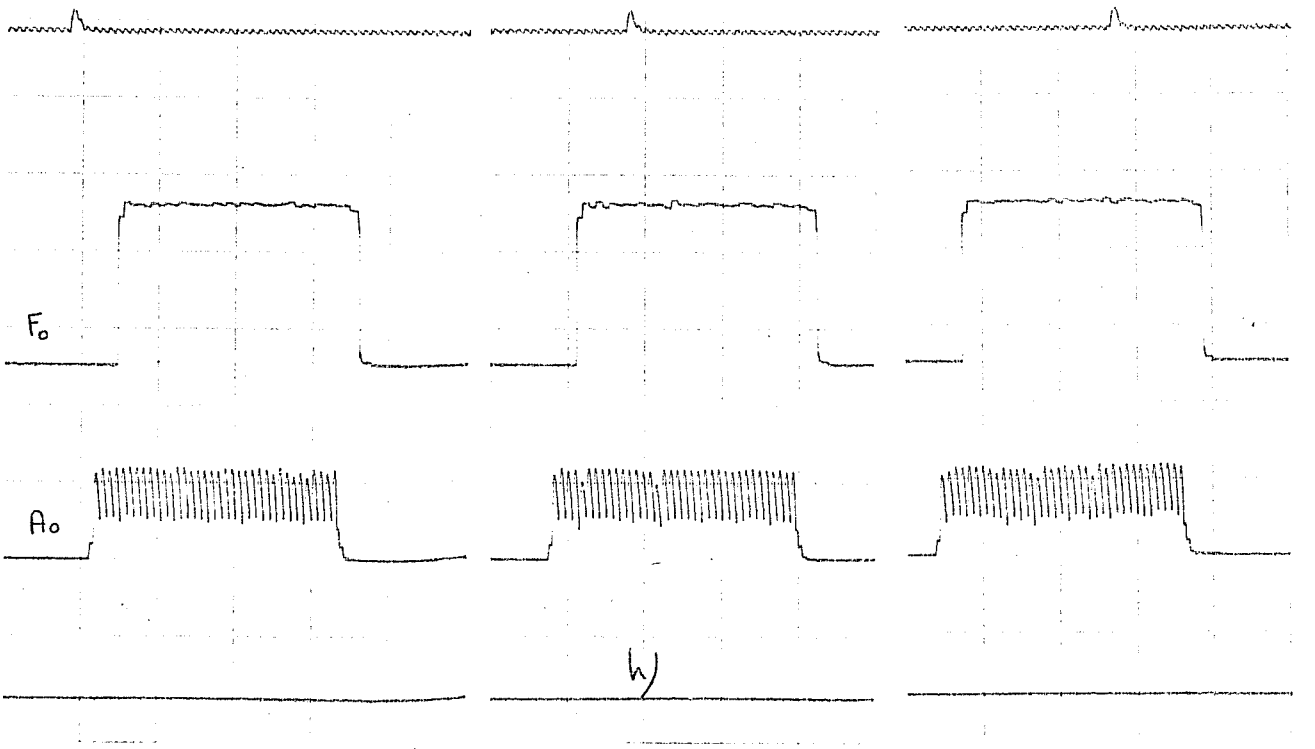


Fig 6 - A

R E F E R E N C E S B I B L I O G R A P H I Q U E S

CARRE R. (I976) "La synthèse de la parole", 7ème J.E.P. du GALF, NANCY, Vol. II, p. 4I - 58.

ESPESSER R. (I977) Contribution à la synthèse de la parole; étude d'une simulation numérique du conduit vocal, thèse de doctorat de spécialité (Acoustique), Université de Provence.

FLANAGAN J.L. (I972) Speech analysis, synthesis and perception, Springer Verlag, Berlin (2ème édition).

GOLD B. et RABINER L.R. (I968) "Analysis of digital and analog formant synthesizers", IEEE - AU, I6, p. 8I - 94.

HOLLIEN H., MICHEL I. et DOHERTY E.T. (I973) "A method for analysing vocal jitter in sustained phonation", J. of Ph., I, p. 85 - 9I.

HOLMES J.N. (I973) "Influence of glottal waveforms on the naturalness of speech from a parallel formant synthesizer", IEEE - AU, 2I, p. 298 - 305.

KLATT D.H. (I973) "Acoustic theory of terminal analog speech synthesis", Proceedings I972 Int. Conf. Speech Communication and Speech processing, paper D2, p. I3I - I35.

LANDERCY A. (I973) "Les variations temporelles à court-terme dans les vocoïdes", Rapport d'activité, Institut de Phonétique, Université Libre de Bruxelles, 7/I, p. 75 - 90

LIEBERMAN Ph. (I96I) "Perturbations in vocal pitch", JASA, 33, p. 597 - 603.

RABINER L.R. (I968) "Digital formant synthesizer for speech synthesis studies", JASA, 43, p. 822 - 828.

RICE D.L. (I974) "Articulatory tracking of the acoustic signal", SCS 74, vol. I, p. 2I - 26.

ROSENBERG A.E. (I968) "Effect of pitch averaging on the quality of natural vowels", JASA, 44, p. I592 - I595.

ROSENBERG A.E. (I97I) "Effect of glottal pulse shape on the quality of natural vowels", JASA, 49, p. 583 - 590.

LA QUALITE DE LA SYNTHESE : APERÇU BIBLIOGRAPHIQUE

ARKHIDOVA A.D. et SAPOZHKOV M.A. (I97I) "The quality of vocoder speech", Soviet Physics - Acoustics, I6, p. 292 - 298.

COOPER FS., PETERSON E. et FAHRINGER G.S. (I957) "Some sources of characteristic vocoder quality", JASA, 29, p. I83 (A).

FUJIMURA O. (I968) "An approximation to voice aperidiocity", IEEE - AU, I6, p. 68 - 72.

HOLMES J.N. (I963) "The effect of simulating natural larynx behavior on the quality of synthetic speech", Proceedings of the Speech Seminar, Stockholm, Vol. 2, paper F6.

HOLMES J.N. (I973) "Influence of glottal waveforms on the naturalness of speech from a parallel formant synthesizer", IEEE - AU, 23, p. 298 - 305.

KERSTA L.G., BRICKER P.D. et DAVID E.E. (I960) "Human or machine? JASA, 32, p. I502 (A).

KODAMA K., SATO T. et al. (I964) "Influence of pitch fluctuation on the quality of Japanese vowels", Research Institute in Electr. Communication, I6, p. 9 - 23. (TOHOKO University)

MAEDA S. et FUJIMURA O. (I968) "Factors of the glottal wave that contribute to the naturalness of speech", Annual Bull. RILP, 2, p. 46 - 5I.

MAEDA S., FUJIMURA O. et SAWASHIMA M. (I968) "Voice aperiodicity in terms of pulse shape interval", VIème I.C.A., vol. II-B, p. 43 - 46.

NAKAYAMA T. et al. (I968) "Generalization of control rule of buzz sources to improve the naturalness of synthetic speech", VIème ICA, Tokyo, Vol. 2, paper B5-5, p. I67 - I70

RICE L.D. (I97I) Glottal pulse waveform effect in line analog speech synthesis, W.P.P., n° I9.

SAPOZHKOV M.A. (I972) "Improving the quality of synthesized speech", Soviet Physics - Acoustics, I7, p. 5I0 - 5I3.

SLIS I.H. et CARDOZO B.L. (I968) "Experiment on the simulation of voice quality", I.P.O., 3, p. 45 - 47.

R E F E R E N C E S

B I B L I O G R A P H I Q U E S

INDEX DES ABREVIATIONS

Annual Bull. RILP	:	Annual bulletin of the Research Institute of Logopedics and phoniatrics - TOKYO
CTL	:	Current trends in Linguistics (ed T. A. SEBEOCK), Mouton, The Hague.
F. of L.	:	Foundations of Language, Dordrecht.
BSL	:	Bulletin de la Société de Linguistique de Paris.
Haskins S.R.	:	Haskins Laboratories, Status Report on Speech Research, New-York.
IEEE - AU	:	Institute of Electrical and Electronics Engineers - Transactions on Audio and Electro-acoustics, New-York.
IEEE - ASSP	:	Acoustics, Speech and Signal Processing (a remplacé le précédent)
IJAL	:	International Journal of American Linguistics, Baltimore.
JASA	:	Journal of the Acoustical Society of America, Lancaster, P.A.
JASJp	:	Journal of the Acoustical Society of Japan, Tokyo.
J. of Ph.	:	Journal of Phonetics, London.
JSHD	:	Journal of Speech and Hearing Disorders, Urbana.
JSHR	:	Journal of Speech and Hearing Research, Washington.
Lg.	:	Language, Baltimore.
Lg & S	:	Language and Speech, Teddington.
Math. Comp.	:	Mathematics of Computation, Providence R.I.
MIT-QPR-RLE	:	Massachussetts Institute of Technology, Quarterly Progress Report, Research Laboratory in Electronics.
P & P	:	Perception and Psychophysics, Austin.

Proc. IEE	:	Proceedings of the Institution of Electrical Engineers, London.
QPR-STL-RIT	:	Quarterly Progress Report - Speech Transmission Laboratory - Royal Institute of Technology, Stockholm.
R.A.	:	Revue d'Acoustique, Paris.
R. Ph. A.	:	Revue de Phonétique Appliquée, Mons.
SAS	:	Speech Analysis and Synthesis, Varsovie.
SCS 74	:	Speech Communication Seminar; Proceedings (1975) - Almqvist & Wiskell, Stockholm.
TCLP	:	Travaux du Cercle Linguistique de Prague.
TIP Nancy	:	Travaux de l'Institut de Phonétique de Nancy.
W.P.P.	:	Working Papers in Phonetics, University of California, Los Angeles.
Z.Ph.S.K.	:	Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, Berlin.
Xème I.C.A.	:	Proceedings of the Xth International Congress on Acoustics.
Xème J.E.P. du GALF	:	Actes des Xèmes Journées d'Etude sur la Parole du Groupe des Acousticiens de Langue Française.

ABBS J H. et EILENBERG G.R. (I976) "Peripheral mechanisms of speech motor control" in LASS N.J.(ed) Contemporary issues in experimental phonetics, Academic Press, New-York, p. I39 - I68.

ALLEN W.S.(I968) Vox Graeca, Cambridge University Press, Cambridge.

ANGLIN M. (I97I) Perceptual space of English vowels in word context, Master's thesis, Howard University, Washington (non publié; résumé in SINGH S. Distinctive features: Theory and validation, University Park Press, Baltimore, I976).

ATAL B.S. (I970) "Determination of the vocal tract shape directly from the speech wave", JASA, 47, 65(A).

ATAL B.S., CHANG J.J., MATHEWS W.V. et TUKEY J.W. (I977) " A method for inverting articulatory-to-acoustic transformation in the vocal tract based on a computer-sorting technique", 9ème I.C.A., paper I 59, p. 465.

ATAL B.S. et HANAUER S.L. (I97I) "Speech analysis and synthesis by linear prediction of the speech wave", JASA, 50(2), p. 637-655.

AUB-BÜSCHER G. (I962) Le parler rural de Ranrupt (Bas-Rhin), Klincksieck, Paris.

BACH E. et HARMS T. (I972) "How do languages get crazy rules?" in STOCKWELL R.D. et MACAULAY R.K.S. (eds) Linguistic change and generative theory, Indiana University Press, Bloomington, p. I - 2I.

BEECKMANS R. (I974) "Influence de la fréquence fondamentale sur l'espace perceptif des voyelles", Rapport d'activité, Inst. de Phonétique, Université Libre de Bruxelles, 8 - I/2, p. I09 - I28.

BEECKMANS R. (I977) " Une approche du processus d'identification des voyelles en français: influences comparées de plusieurs variables expérimentales sur l'espace perceptif", Rapport d'activité, Inst. de Phonétique, Université Libre de Bruxelles, II/I, p. 69 - 82.

BENNETT D.C. (I968) "Spectral form and duration as cues in the recognition of English and German", Lg & S, II, p. 65 - 85.

BERNARD J.R.L. (I970 a) "A cine-X-ray study of some sounds of Australian English", Phonetica, 2I, p. I38 - I50.

BERNARD J.R.L. (I970 b) " Towards the acoustic specification of Australian English", Z.Ph.S.K., 23(2/3), p. II3 - I28.

BERNSTEIN J.C. (I975) "Similarity structures among vowels", MIT-QPR-RLE, II4, p. I73 - I78.

BEYER E. (I964) La palatalisation vocalique spontanée de l'alsacien et du badois, Publications de la société savante d'Alsace, Strasbourg.

BJUGGREN B. et FANT G. (I964) "The nasal cavity structure", QPR-STL-RIT, 4, p. 5 - 7.

BLOCH B. (I950) "Colloquial Japanese", Lg., 26, p. 86 - I25.

BOURCIEZ E. (I967) Phonétique française, Klincksieck, Paris

BRICHLER - LABAEYE C. (I970) Les voyelles françaises, Klincksieck, Paris.

BRIGHAM E.O. (I974) The fast fourier transform, Prentice-Hall, Englewoods Cliffs N.J.

BRANDT J.F. (I976) "Perceptual psychophysics: Speech and Hearing", in LASS N.J. (ed) Contemporary issues in experimental phonetics, Academic press, New-York, p. 459 - 483.

BRANDT J.F. et THOMAS W.G. (I967) "Auditory analysis of vowel-shaped thermal noise", non publié, Communication Sciences Lab., University of Florida, Gainesville.

BROOKS S. et FALLSIDE F. (I976) "A technique for converting the linear prediction areas model of speech to a simple articulatory model", I.E.E.E. Conference Record, p. 7I - 74

BUTCHER A. (I974) "'Brightness - darkness' and the dimensionality of speech perception", J. of Ph., 2, p. I53 - I60.

CAILLIEZ F. et PAGES J.P. (I976) Introduction à l'analyse des données, Société de Mathématiques et Sciences Humaines, Paris.

CALBRIS G. (I974) "Test sur la perception des voyelles orales françaises par des français et des étrangers", Phonetica, 30, p. IOI - I22.

CARLSON R., FANT G. et GRANSTRÖM B. (I975) " Two-formant models, pitch and vowel perception" in FANT G. et TATHAM M.A.A.(eds) Auditory analysis and perception of speech, Academic Press, New-York, p. 55 - 82.

CARLSON R., GRANSTRÖM B. et FANT G. (I970) " Some studies concerning perception of isolated vowels", QPR-STL-RIT, 2-3, p. I9 - 35.

CARROLL J.D. (I972) "Individual differences and multidimensional scaling", in SHEPARD R.N., ROMNEY A.K. et NERLOVE S.B (eds) Multidimensional scaling: Vol. I - Theory, Seminar Press, New-York, p. IO5 - I55.

CARROLL J.D. et CHANG J.J. (I970) "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckhart-Young' decomposition", Psychometrika, 35, p. 283 - 3I9.

CARTON F. (I974) Introduction à la phonétique du français, Bordas, Paris.

CARTON F. et LONCHAMP F. (I977) "Validation d'indices perceptifs de l'intonation par analyse multidimensionnelle", VIIIème J.E.P. du GALF, Aix I977, p. 147 - 156

CHANDRA S. et LIN W.C. (I974) "Experimental comparison between stationary and non-stationary formulations of linear prediction applied to voiced speech", IEEE - ASSP, 22, p. 403 - 4I5.

CHIBA T. et KAJIYAMA M. (I94I) The vowel; its nature and structure, Phonetic society of Japan, Tokyo (2ème édition I958).

CHIPMAN S.F. et CAREY S. (I975) "Anatomy of a stimulus domain: the relation between multidimensional and unidimensional scaling of noise bands", P & P, I7(4), p. 4I7 - 424.

CHISTOVICH L.A. et KOZHEVNIKOV V.A. (I970) Theory and methods of research on perception of speech signals, JPRS - 50423, U.S. Dept. of Commerce, Washington D.C. (traduit du russe).

CHOMSKY N. (I964) Current issues in linguistic theory, Mouton, The Hague.

CHOMSKY N. (I968) Language and mind, Harcourt, Brace & World, New-York.

CHOMSKY N. et HALLE M. (I968) The sound pattern of English, Harper, New-York.

CHRICHTON R.G. et FALLSIDE F. (I974a) " A real-time articulatory display for deaf speech training", S.C.S. I974, Vol. IV, p. 35 - 40.

CHRICHTON R.G. et FALLSIDE F. (I974 b) "Linear prediction model of speech production with application to deaf speech training", Proc. I.E.E., I2I, p. 865 - 873.

CLARKE R.F. (I957) " Constant ratio rule for confusion matrices in speech communication", JASA, 29, p. 7I5 - 720.

COCKER C.H. (I968) "Speech synthesis with a parametric articulatory model", Proceedings of the speech symposium, Kyoto, Japon, p. A-4-I - A-4-6.

COLLINGE N.E.(I96I) "Phonetic shifts and phonetic asymmetries", IVème Cong. Int. des Sc. Phonétiques, Mouton, The Hague, p. 563 - 566.

COOLEY J.W. et TUCKEY J.N. (I965) "An algorithm for the machine calculation of complex fourier series", Math. Comp., I9, p. 297 - 30I.

CROWDER R.G. (I973) " Precategorical acoustic storage for vowels of short and long duration", P & P, I3(3), p. 502 - 506.

DARWIN C.J. (I976) "The perception of speech" in CARTERETTE E.C. et FRIEDMAN M.P. (eds) Handbook of perception; Vol. VII: Language and speech, Academic Press, New-York, p. I75 - 226.

DEBROCK M. (I974) "La structure spectrale des voyelles nasales", R. Ph. A., 29, p. I5 - 32.

DEBROCK M. et FORREZ G. (I976) " Analyses des voyelles orales du néerlandais et du français: méthode et résultats", R. Ph. A., 37, p. 27 - 72.

DELATTRE P. (I954) "Les attributs acoustiques de la nasalité vocalique et consonantique", Studia linguistica, 8, p. I03 - I09.

DELATTRE P. (I965a) "La nasalité vocalique en français et en anglais", French Review, 39, p. 92 - I09.

DELATTRE P. (I965b) Comparing the phonetic features of English, French, German and Spanish, J. Groos Verlag, Heidelberg.

DELATTRE P. (I968) "Divergences entre nasalités vocalique et consonantique en français, Word, 24, p. 64 - 72.

DELATTRE P. (I970) "Rapports entre la physiologie et la chronologie de la nasalité distinctive", Actes du Xème Congrès Int. de Linguistique (I967), Vol. IV, p. 22I - 227.

DELATTRE P., LIBERMAN A.M., COOPER F.S. et GERSTMAN L.J. (I952) "An experimental study of the acoustic determinant of vowel color: observations on one and two formant vowels synthesized from spectrographic patterns", Word, 8, p. I95 - 2I0.

DUMEZIL G. et NAMITOK A. (I954) "Le système des sons de l'oubykh", B.S.L., 50(I), p. I62 - I89.

DURAFFOUR A. (I927) "Trois phénomènes de nivellement phonétique en franco-provençal", B.S.L., 27, p. 68 - 80.

ESPESSER R. (I977) Contibution à la synthèse de la parole: étude d'une simulation numérique du conduit vocal, Thèse de doctorat de spécialité (Acoustique), Université de Provence.

EL-MALLAWANY I.I. (I975) Etude de vocoders à prédiction linéaire, Thèse de docteur-ingénieur, Grenoble.

EL-MALLAWANY I.I. (1976) "Approche de la détection de et à l'analyse sur l'intervalle de fermet ure de la glotte", VIIème J.E.P. du GALF, Nancy 1976, vol. I, p. 321 - 331.

EL-MALLAWANY I.I. (1977) "Analysis over the clossed glottis interval", 9ème I.C.A., paper I 65, p. 471.

FANT G. (1956) "On the predictability of formant levels and spectrum envelopes from formant frequencies" in For Roman Jakobson, Mouton, The Hague, p. 109 - 120

FANT G. (1959) "Acoustic analysis and synthesis with application to Swedish", Ericsson Technics, 15, p. 3 - 108.

FANT G. (1960) Acoustic theory of speech production, Mouton, The Hague (2ème édition:1970).

FANT G. (1971) "Notes on the Swedish vowel system", in HAMMERICH L.L. et al. (eds) Form and Substance, Akademisk Forlag, p. 259 - 268.

FANT G. (1972) "Vocal tract wall effect, losses and resonance bandwidths", QPR-STL-RIT, 2-3, p. 28 - 52.

FANT G. (1974) "Spatial characteristics of vocal tract resonance modes, S.C.S. 1974, vol. 2, p. 121 - 131.

FANT G. (1975) "Vowels, production and perception", VI ème J.E.P. du GALF, Toulouse 1975, vol. 2, p. 8 - 60.

FANT G., CARLSON R. et GRANSTRÖM B. (1974) "The [e - ø] ambiguity", S.C.S. 1974, vol. 3, p. 117 - 121.

FANT G., HENNINGSON G. et STÅLHAMMAR U. (1969) "Formant frequencies of Swedish vowels , QPR-STL-RIT, 4, p. 26 - 31.

FANT G. et RISBERG A. (1963) "Auditory matching of vowels with two formant synthetic sounds", QPR-STL-RIT, 4, p. 7 - 11.

FARNSWORTH P.R. (1937) "An approach to the study of vocal resonance", JASA, 9, p. 152 - 155.

FERGUSON T. (1976) A history of the Romance vowel systems through paradigmatic reconstruction, Mouton, The Hague.

FISCHER-JØRGENSEN E. (I967) "Perceptual dimensions of vowels" in To Honor R. JAKOBSON, Mouton, The Hague, vol. I, p. 667 - 67I.

FISCHER-JØRGENSEN E. (I975) Trends in phonological theory, Akademisk Forlag, Copenhagen.

FLANAGAN J.L. (I97I) "Focal points in speech communication research" 7ème I.C.A., vol. I, p. 57 - 76.

FLANAGAN J.L. (I972) Speech Analysis , Synthesis and Perception, Springer Verlag, Berlin.

FLETCHER H. et MUNSON W.A. (I937) "Relation between loudness and masking", JASA, 9, p. I - I0.

FOLEY J. (I977) Foundations of theoretical phonology, Cambridge studies in linguistics n° 20, Cambridge University press.

FONAGY I. (I963) Die Metaphern in der Phonetik, Mouton, The Hague.

FOUCHE P. (I952) Phonétique historique du français, vol. I, Klincksieck, Paris.

FOUCHE P. (I958) Phonétique historique du français, vol. II, Klincksieck, Paris.

FOURCIN A.J. (I968) " Speech source inference", IEEE - AU, I6, p. 65 - 67.

FRITZELL B. (I969) The velopharyngeal muscles in speech: an electromyographic and cineradiographic study, Acta Otolaryngologica Suppl. n° 250.

FROMKIN V. (I964) "Lip positions in American English vowels", Lg & S, 7, p. 2I5 - 225.

FRY D.B., ABRAMSON A.S., EIMAS P.D. et LIBERMAN A.M. (I962) "The identification and discrimination of synthetic vowels", Lg & S, 5, p. I7I - I89.

FUJIMURA O. "Spectra of nasalized vowels", MIT-QPR-RLE, 58, p. 2I4 - 2I8.

FUJIMURA O. (I967) " On the second spectral peak of front vowels: a perceptual study of the role of the second and third formants, Lg & S, I0, p. I8I - I93.

FUJIMURA O. et LINDQVIST J. (I97I) " Sweep-tone measurements of vocal tract characteristics", JASA, 49, p. 54I - 558.

FUJISAKI H. et KAWASHIMA T. (I970) "Some experiments on speech perception and a model for the perceptual mechanism", Annual Report of the Eng. Research Inst. Tokyo, 29, p. 207 - 2I4.

FUJISAKI H. et SEKIMOTO S. (I975) " Perception of time-varying resonance frequencies in speech and non-speech stimuli" in COHEN A. et NOOTEBOOM S.G. (eds) Structure and process in speech perception, Springer Verlag, Berlin, p. 269 - 282.

GALVAGNY M.H. (I973) Analyse acoustique et perceptive des voyelles de l'allemand, Thèse de doctorat de 3ème cycle, Paris

GRANT M. (I97I) INDSCAL analysis of Hindi and English vowels, Master's thesis, Howard University, Washington, non publié (Résumé in SINGH S. (I976)).

GREENBERG J.H. (I970) "Some generalizations concerning glottalic consonants, especially implosives", IJAL, 36, p. I23 - I46.

GREVIN J. (I974) Etude et réalisation d'un synthétiseur numérique de parole, Thèse de doctorat de 3ème cycle, Orsay.

GREVIN J., BERTIN F. et SAP J. (I975) "Synthétiseur de parole à formants à circuits numériques", VIème J.E.P. du GALF, Toulouse I975, vol. I, p. 356 - 363.

GUNNILSTAM O. (I974) "The theory of local linearity", J. of Ph., 2, p. 9I - I08.

GUTH G. (I975) Contribution expérimentale à l'étude des attributs acoustiques de la nasalité, S.N.E.D., Alger.

HALLE M. (I962) "Phonology in generative grammar", Word, I8, p. 54 - 72.

HALLE M. (1970) "Is Kabardian a vowel-less language", F. of L., 6, p. 95 - 103.

HANSON G. (1967) "Dimensions in speech perception", Ericsson Technics, I, p. I - 175.

HANSON G. (1968) "Distinctive features and response dimensions of vowel perception", Z.Ph.S.K., 21 (1/2), p. 99 - 101.

HARMAN H.H. (1960) Modern factor analysis, University of Chicago press, Chicago.

HARSHMAN R. (1970) "Foundations of the PARAFAC procedure: models and condition for an "explanatory" multi-modal factor analysis", W.P.P., 16, p. I - 84.

HATTORI S , YAMAMOTO K. et FUJIMURA O. (1958) "Nasalization of vowels in relation to nasals", JASA, 30, p. 267 - 274.

HAUDRICOURT A.G. (1947) "Problèmes de phonologie diachronique (français ei > oi)", Lingua, I, p. 209 - 218.

HAUDRICOURT A.G. et JUILLAND A.G. (1949) Essai pour une histoire structurale du phonétisme français, Klincksieck, Paris.

HAUGEN E. (1976) The Scandinavian Language, Faber & Faber, Londres.

HIROSE H., KIRITANI S., KAKITA K., USHIJIMA T., KJELLING O. (1977b) "Lingual electromyography related to tongue movement", 9ème I.C.A., paper I 15, p. 421.

HIROSE H., KIRITANI S., USHIJIMA T. et KJELLING O. (1977 a) "Lingual electromyography related to tongue mouvement in Swedish vowels production", Annual Bull. RILP, 11, p. 39 - 50.

HOCKETT C.F. (1955) A manual of phonology, IJAL, Memoir n° II.

HOOPER J.B. (1976) An introduction to natural generative phonology, Academic Press, New-York.

HOUSE A.S. et STEVENS K.N. (1956)"Analog studies of the nasalization of vowels", JSHD, 21, p. 218 - 232.

HOUTGAST T. (I974) "Auditory analysis of vowel-like sounds", Acustica, 3I, p. 320 - 324.

HOWARD J.H. Jr (I977) "Psychophysical structure of 8 complex underwater sounds", JASA, 62, p. I49 - I56.

HUSSON P. (I962) La physiologie de la phonation, Masson, Paris.

HYMAN L.M. (I975) Phonology: theory and analysis, Holt, Rinehart et Winston, New-York.

ISHIZAKA K. et FLANAGAN J.L. (I972) "Synthesis of voiced sounds from a two-mass model of the vocal cords", Bell System Tech. Journal, 5I, p. I233 - I268.

JAKOBSON R. (I939) "Les lois phonétique du langage enfantin et leur place dans la phonologie générale", communication prévue au Vème Congrès International de Linguistique, Bruxelles (I939): non publiée; repris et traduit in TROUBETZKOY N.S. (I949) Principes de phonologie, traduction J. CANTINEAU, Klincksieck, Paris, annexe V, p. 367 - 379.

JAKOBSON R. (I94I) Kindersprache, Aphasia und allgemeine Lautgesetze; traduction KEILER A.R. (I972) Child language, Aphasia and Phonological Universals, Mouton, The Hague.

JAKOBSON R. (I949) "On the identification of phonemic entities", Recherches structurales TCLP V, p. 205 - 2I3.

JABOBSON R., FANT G. et HALLE M. (I95I) Preliminaries to speech analysis, MIT Press, Cambridge (USA).

JAKOBSON R. et HALLE M. (I956) Fundamentals of language, Mouton, the Hague.

JANSON T. (I977) "Asymmetry in vowel confusion matrices", J. of Ph., 5, p. 93 - 96.

JASSEM W. (I968) "Vowel formant frequencies as cues to speaker discrimination", SAS, vol. I, p. 9 - 42.

JERRENTRUP A. et THÜRMANN E. (I974a) "Auditive Attribute und phonetische Ähnlichkeit: Untersuchungen mit dem Eindrucks-Differential an synthetischen Vokalklängen", Z.Ph.S.K., 27, p. 392 - 4I3.

JERRENTRUP A. et THÜRMANN E. (I974b) "Auditive Dimensionen der Vokalklangswahrnemung: Ein Vergleich des Urteilsverhaltenis von phonetischen Laien und Experten", Phonetica, 30, p. I75 - I93.

JOHANSSON K. (I975) "Perceptual characteristics of vowels", Working papers, Phonetics Lab., University of Lund, 10, p.83 - I04.

JOHNSON S.C. (I967) "Hierarchical clustering schemes", Psychometrika, 32, p. 24I - 254.

JONASSON J.(I97I) "Perceptual factors in phonology", 7ème Congrès Int. des Sciences Phonétiques, Mouton, The Hague, p. II27 - II30.

KADOKAWA Y. et SUZUKI J. (I968) "Simple calculation of the vocal tract configuration from three formant frequencies", J. of the Radio Research Lab., I5, 79, p. I47 - I64.

KANAMORI Y. (I974) "On the characteristics of formant frequencies of vowels in connected speech", JASJp, 30, p. 670 - 672.

KASUYA K., KANAMORI Y. et KIDO K. (I97I) "Psychological auditory space representing vowel quality", 7ème I.C.A., Budapest, paper 20C5,vol. 3, p. 4I - 44.

KENT R.D. (I976) "Models of speech production" in LASS N.J. (ed) Contemporary issues in experimental phonetics, Academic Press, New-York, p. 79 - I04.

KITZING K. (I97I) "Contrastive acoustic analysis of vowel phonemes pronounced by some North-German and South-Swedish high school pupils", Working papers, Phonetics lab., U. of Lund, 4 , p. 48 - 57.

KLEIN W., PLOMP R., et POLS L.C.W. (I970) "Vowel spectra, vowel space and vowel identification", JASA, 48, p. 999 - I009.

KNUDSEN M.J. (I975) "Real-time linear predictive coding of speech on the SPS-4I micro-programmed triple processor machine", IEEE-ASSP, 23, p. I40 - I45.

KÖHLER W. (I909) Akustische Untersuchungen, Diss. Leipzig.

KRUSKAL J.B. (I964 a) "Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis", Psychometrika, 29, p. I-27.

KRUSKAL J.B. (I964 b) "Nonmetric multidimensional scaling: a numerical method", Psychometrika, 29, p. II5 - I29.

LADEFOGED P. (I964) A phonetic study of West African Languages Cambridge University Press, Cambridge.

LADEFOGED P. (I97Ia) "The limits of phonology", in HAMMERICH L.L. et al. (eds) Form and substance, Akademisk Forlag, Copenhagen, p. 47 - 56.

LADEFOGED P. (I97Ib) Preliminaries to linguistic phonetics, University of Chicago Press, Chicago.

LADEFOGED P. (I972) "Phonetic prerequisites for a distinctive feature theory" in VALDMAN A. (ed) Papers in Linguistics and phonetics to the memory of Pierre Delattre, Mouton, The Hague, p. 273 - 285.

LADEFOGED P. (I975) A course in phonetics , Harcourt, Brace et Jovanovitch, New-York.

LADEFOGED P., DE CLERK J., LINDAU M., et PAPÇUN G. (I972) "An auditory-motor theory of speech production", W.P.P., 22, p. 48 - 75.

LAFON J.C. (I96I) Message et phonétique , P.U.F., Paris.

LANDERCY A (I973) "Les paramètres acoustiques de l'audition et de la perception des sons de la parole", R. Ph. A., 25, p. 3 - 24.

LEJEUNE M.(I946) Traité de phonétique grecque, Klincksieck, Paris.

LE ROUX J. (I976) "Optimisation du calcul des coefficients de corrélation partielle", VIIème J.E.P. du GALF, Nancy, vol. I, p. 359 - 373.

LIBERMAN A.M., COOPER F.S., SHANKWEILER D.S. et STUDDERT-KENNEDY M. (I967) "Perception of the speech code", Psychological Review, 74, p. 43I - 46I.

LILJENCRANTS J. et LINDBLOM B. (I972) "Numerical simulation of vowel quality systems: the role of perceptual contrast", Lg., 48, p. 839 - 862.

LINDAU M. (I975) Features for vowels, W.P.P., n° 30.

LINDBLOM B.E.F. (I974a) "Experiments in sound structure", 8ème Congrès Int. des sciences Phonétiques, Leeds, résumé.

LINDBLOM B.E.F. (I974b) "Les mécanismes des contrôles moteurs", Bull. de l'Institut de Phonétique de Grenoble, 3, p. I - 2I.

LINDBLOM B.E.F. et SUNDBERG J. (I969) " A quantitative theory of cardinal vowels and the teaching of pronunciation", QPR-STL-RIT, 2/3, p. I9 - 25.

LINDBLOM B.E.F. et SUNDBERG J. (I97Ia) "Neurophysiological representation of speech sounds", XVème Congrès Int. de Logopédie et de Phoniatrie, Buenos-Aires, résumé.

LINDBLOM B.E.F. et SUNDBERG J. (I97Ib) " Acoustical consequences of lip, tongue, jaw et larynx movement", JASA, 50, p. II66 - II79.

LINDBLOM B.E.F. et STUDDERT-KENNEDY M. (I967) "On the role of formant transition in vowel recognition", JASA, 42, p. 830 - 843.

LINDQVIST-GAUFFIN J. et SUNDBERG J. (I976) "Acoustic properties of the nasal tract", Phonetica, 33, p. I6I - I68.

ŁOBACZ P. (I973) "Non-unique phonemic interpretation of the Polish speech sounds", SAS, vol. 3, p. 53 - 74.

LONCHAMP F. (I97I) Un modèle probabiliste de perception de la parole, Mémoire de maîtrise, Nancy.

LONCHAMP F. (I973) "Modèle de perception de vocoïdes synthétiques isolés", IVème J.E.P. du GALF, Bruxelles, p. I25 - I46.

LONCHAMP F. (I974 a) "Notes sur les programmes d'ordinateurs mis au point au laboratoire de phonétique en I973 - I974", T.I.P. de Nancy, vol. I, p. 235 - 256.

LONCHAMP F. (I974b) "Recherches sur la validité des résultats de l'extraction automatique des formants: méthode de PINSON", T.I.P. de Nancy, vol. I, p. I93 - 234.

LUBKER J. (I968) "An electromyographic - cinefluorographic investigation of velar function during normal speech production", Cleft Palate J., 5, p. I - I8.

LUPAŞ L. Phonologie du grec attique, Mouton, the Hague.

LYONS J. (I962) "Phonemic and non-phonemic phonology: some typological reflections", IJAL, 28, p. I27 - I34.

LYONS J. (I970) Linguistique générale, Larousse, Paris.

MacNEILAGE P.F. (I970) "Motor control of serial ordering of speech", Psychological Review, 77, p. I82 - I96.

MAKHOUL J. (I975) "Linear prediction: a tutorial review", Proceedings of the IEEE, 63, p. 56I - 580.

MARKEL J.D. (I972) "Digital inverse filtering, a new tool for formant trajectory estimation", IEEE-AU, 20, p. I29 - I37.

MARKEL J.D. et GRAY A.H. (I976) Linear prediction of speech, Springer Verlag, Berlin.

MARTINET A. (I955) Economie des changements phonétiques, Francke, Berne.

MARTINET A. (I965) "De la morphonologie", La Linguistique, I, p. I5 - 30.

McALLISTER R.M., LUBKER J. et CARLSON J. (I974) "An EMG study of some characteristics of the Swedish rounded vowels", J. of Ph., 2, p. 267 - 278.

MERMELSTEIN P. (I967) "Determination of the vocal tract shape from measured formant frquencies", JASA, 4I, p. I283 - I294.

MERMELSTEIN P. (I975) "Vowel perception in consonantal context", JASA, 58, supplément I, S 56.

MILEWSKI T. (I967) "The evolution of the common Slavic system" in To Honor R. JAKOBSON, vol. 2, Mouton, The Hague, p. I362 - I372.

MILLAR J.B. et AINSWORTH W.A. (I972) "Identification of synthetic isolated vowels and vowels in H-D context", Acustica, 27, p. 278 - 283.

MILLER J.R et CARTERETTE E.C. (I975) "Perception space for musical structures", JASA, 58, p. 7II - 720.

MILLER R.L. (I953) "Auditory tests with synthetic vowels", JASA, 25, p. II4 - I2I.

MIYAWAKI K., HIROSE H., USHIJIMA T. et SAWASHIMA M. (I976) "An electromyographic study of the production of the Japanese vowels", JASJp, 32, p. 446 - 447.

MOHR B. et WANG W.S.Y. (I968) "Perceptual distance and the specification of phonological features", Phonetica, I8, p. 3I - 45.

MORIKAWA H. et FUJISAKI H. (I977) "A new method of speech analysis based on pole-zero model", 9ème ICA, Madrid, paper I 42, p. 448.

MRAYATI M. (I975) "Etude des voyelles nasales françaises", Bull. de l'Institut de Phonétique de Grenoble, 4, p. I - 26.

MRAYATI M. (I976) Contribution aux études sur la production de la parole: modèles électriques du conduit vocal avec pertes, du conduit nasal et de la source vocale , Thèse de doctorat (Sciences Physiques), Grenoble.

MRAYATI M. et CARRE R. (I976) "Relation entre la forme du conduit vocal et les caractéristiques acoustiques des voyelles françaises", Phonetica, 33, p. 285 - 306.

MRAYATI M. et GUERIN B. (I976) "Etude des caractéristiques acoustiques des voyelles orales françaises par simulation du conduit vocal avec pertes", R.A., 36, p. I8 - 32.

MUSHNIKOV V.N. et CHISTOVICH L.A. (I974) "Phoneme boundary fluctuations", Soviet Phys. Acoust., I9, p. 362 - 366.

NORDSTRÖM P.E. (I977) "Female and infant vocal tract simulated from male area functions", J. of Ph., 5, p. 8I - 92.

OSGOOD C.E., SUCI C.J. et TANNENBAUM P.H. (1957) The measurement of meaning, Urbana, Illinois.

OHALA J.J. (1974) "Phonetic explanation in phonology", Proceedings of the parasession on Natural Phonology, Chicago Ling uistic society, p. 251 - 274.

PAIGE A. et ZUE V.W. (1970 a) "Calculation of vocal tract length", IEEE-AU, 18, p. 268 - 270.

PAIGE A. et ZUE V.W. (1970 b) "computation of vocal tract area functions", IEEE - AU, 18, p. 7 - 18.

PERKELL J.S. (1969) Physiology of speech production, MIT Press, Cambridge (USA).

PETERS R.W. (1963) "Dimensions of quality for the vowel [æ] ", JSHR, 6, p. 239 - 248.

PETERSON G.E. et BARNEY H.L. (1952) "Control method used in a study of the vowels", JASA, 24, p. 175 - 184.

PHILIPP M. (1965) Le système phonologique du parler de Blaesheim, Annales de l'Est; publications de la Faculté des Lettres et Sciences Humaines de Nancy n° 27.

PHILIPP M. (1968) Phonologie des graphies et des rimes - L'alsacien de Th. Murner, C.N.R.S., Paris.

PIKE K. (1967) "Tongue root position in practical phonetics", Phonetica, 17, p. 130 - 140.

PILCH H. (1971) "Some phonemic pecularities of Australian English", in HAMMERICH L.L. et al. (eds) Form and substance, Akademisk Forlag, Copenhagen, p. 269 - 276.

PINSON E.N. "Pitch-synchronous time domain estimation of formant frequencies and bandwidths", JASA, 35, p. 1264 - 1273.

PISONI D.B. (1971) On the nature of categorical perception of speech sounds, Ph. D. Thesis, University of Michigan.

PISONI D.B. (1973) "Auditory and phonetic memory codes in the discrimination of consonants and vowels", P & P, 3, p. 253 - 260.

PISONI D.B. (I975) "Auditory STM and vowel perception", Memory & Cognition, 3, p. 7 - I8.

PLATT H. (I970) "A comparisive study of the phonetics of Australian English and German", Phonetica, 2I, p. I - 30.

PLOMP R. (I970) "Timbre as a multidimensional attribute of complex tones", in PLOMP R. et SMOORENDURG G. F. (eds) Frequency analysis aud periodicity detection in hearing, A.W. Sijthoff, Leiden, p. 397 - 4I4.

POLS L.C.W. (I970) "Perceptual space of vowel-like sounds and its correlation with frequency spectrum" in PLOMP R. et SMOORENBURG G.F. (eds) Frequency analysis and periodicity detection in hearing, A.W. Sijthoff, Leiden, p. 463 - 473.

POLS L.C.W., TROMP H.R.C. et PLOMP R. (I973)" Frequency analysis of Dutch vowels from 50 male speakers", JASA, 53, p. I093 - IIOI.

POLS L.C.W., Van der KAMP L.J.Th. et PLOMP R. (I969) "Perceptual and physical space of vowel sounds", JASA, 46, p. 458 - 467.

POSTAL P.M. (I968) Aspect of phonological theory, Harper & Row, New-York.

RAND T.G. (I97I) "Vocal tract size normalisation in the perception of stop consonants", Haskins Lab. S.R., 25/26, p. I4I - I46.

REID L.A. (I974) "Diachronic typology of Philippine vowel systems", in SEBEOCK T.A.(ed) Current trends in Linguistics, Vol. II, p. 485 - 506.

REMACLE L. (I948) Le problème de l'ancien wallon, Bib. de la Fac. de Philosophie et des Lettres, U. de Liège, fasc. n° LIX.

REMACLE L. (I953) Atlas linguistique de la Wallonie, tome I, Liège.

REMACLE L. (I972) "La géographie dialectale de la Belgique romane", in STRAKA G. (ed) Les dialectes de France au moyen-âge et aujourd'hui, Klincksieck, Actes et colloques n° 9, Paris, p. 3II - 336.

REUCHLIN P. (1976) Précis de statistique, coll. Le Psychologue, P.U.F., Paris.

RIORDAN C.J. (1977) "Control of vocal tract length in speech", JASA, 62, p. 998 - 1002.

ROCHET B.L. (1976) The formation and the evolution of the French nasal vowels, Max Niemeyer, Tübingen.

ROSENBERG A.E. (1971) "Effect of glottal pulse on the quality of natural vowels", JASA, 49, p. 583 - 590.

ROSSI M. (1971) "L'intensité spécifique des voyelles", Phonetica, 24, p. 129 - 163.

ROSSI M. (1977) "Les traits acoustiques", La Linguistique, 13, p. 63 - 82.

SACHS R.M. (1968) "Vowel identification and discrimination in isolation vs word context", MIT-QPR-RLE, 93, p. 220 - 229.

SAITO S. et ITAKURA F. (196) The theoretical consideration of statistically optimum methods for speech spectral density, Report n° 3107, Electrical Communication Lab., Tokyo.

SCHAFER R.W. et RABINER L.R (1970) "System for automatic formant analysis of voiced speech", JASA,47, p. 634 - 648.

SCHANE S.A. (1968) French phonology and morphology, MIT Press, Cambridge (USA).

SCHANE S.A. (1973) Generative phonology, Prentice-Hall, Englewoods Cliffs N.J.

SCHARF B. (1970) "Critical bands" in TOBIAS J.V. (ed) Foundations of modern auditory theory, Academic press, New-York, p. 159 - 202.

SCHÜRR F. (1954) "Zum Wandel u> ü im Französischen", in Estudios dedicadas a Menendez Pidal, vol. V, Madrid, p. 133 - 140.

SCOTT A.L. (1976) "Temporal factors in vowel perception", JASA, 60, p. 1354 - 1365.

SEDLAK P. (I969) "Typological considerations of vowel quality systems", Working papers on language universals, Stanford University, I, p. I - 40.

SHELDON D.R. (I973) "The identification of synthetic vowels in real vowel frames from different speakers", J. of Ph., I, p. I49 - I63.

SHEPARD R.N. (I957) " Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space", Psychometrica, 22, p. 325 - 345.

SHEPARD R.N. (I962) "The analysis of proximities: multidimensional scaling with an unknow distance function: part II", Psychometrika, 27, p. 2I9 - 246.

SHEPARD R.N. (I972) "Psychological representation of speech sounds" in DAVID E.D.Jr et DENES P.B. (eds) Human communication: a unified view, Mc Graw-Hill, New-York, p. 67 - II3.

SINGH S. (I976) Distinctive features: Theory and validation, University Park Press, Baltimore.

SINGH S. et WOODS D.R. (I97I) "Perceptual structure of I2 American English vowels", JASA, 49, p. I86I - I866.

SINGH S., WOODS D.R. et BECKER G.M. (I972) "Perceptual structure of 22 prevocalic English consonants", JASA, 52, p. I698 - I7I3

SNIDER J.G. et OSGOOD C.E. (eds) (I960) Semantic differential technique: a sourcebook, Chicago University Press, Chicago.

SOMMERFELT A. (I960) "The case of the norwegian [ɽ] ", Linguistics, 5, p. 48 - 50.

SONDHI M.M. (I974) " Model for wave propagation in a lossy vocal tract", JASA, 55, p. I070 - I075.

STEIGLITZ K. (I974) An introduction to dicrete systems, Wiley, New-York.

STEIGLITZ K. (I977) "On the simultaneous estimations of poles and zeros in speech analysis", IEEE - ASSP, 25, p. 229 - 234.

STEVENS K.N. (I968) "On the relations between speech movements and speech perception", Z.Ph.S.K., 2I, p. I02 - I06.

STEVENS K.N. (I972) "The quantal nature of speech: evidence from articulatory-acoustic data" in DAVID E.E.Jr et DENES P.B. (eds) Human communication: a unified view, Mc Graw-Hill, New-York, p. 5I - 66.

STEVENS K.N. et HOUSE A.S. (I963) "Perturbation of vowel articulations by consonantal context: an acoustical study", JSHR, 6, p. III - I28.

STEVENS K.N. et HOUSE A.S. (I972) " The perception of speech", in TOBIAS J. (ed) Foundations of modern auditory theory, vol. 2, Academic press, New-York, p. 3 - 62.

STEVENS S.S. (I956) "Calculation of the loudness of complex noise", JASA, 28, p. 807 - 832.

STEVENS S.S. et DAVIS H. (I938) Hearing: its psychology and physiology , Wiley, New-York.

STEVENS S.S et VOLKMANN J. (I940) "The relation of pitch to frequency", American Journal of Psychology, 53, p. 329 - 353.

STRAKA G. (I964) "L'évolution du latin au français sous l'effet de l'énergie et de la faiblesse articulatoire", Travaux de Linguistique et de Littérature, Université de Strasbourg, II, p. I7 - 98.

STRANGE W., VERBRUGGE R.R., SHANKWEILER D.P., et EDMAN T.R. (I976) "Consonant environment specifies vowel identity", JASA, 60, p. 2I3 - 224.

STUDDERT-KENNEDY M. (I974) "The perception of speech" in SEBEOCK T.A (ed) Current trends in linguistics, vol. XII, Mouton, The Hague, p, 2349 - 2385.

STUDDERT-KENNEDY M. (I976) "Speech perception" in LASS N.J. (ed) Contemporary Issues in experimental phonetics, Academic Press, New-York, p. 243 - 293

STUMP C. (1926) Die Sprachlaute, Berlin

SUMMERFIELD A.Q. et HAGGARD M.P. (1975) "Vocal tract normalisation as demonstrated by reaction time", in FANT G. et TATHAM M.A.A. (eds) Auditory analysis and perception of speech, Academic Press, New-York, p. 115 - 141.

TEACHER C.F., KELLETT H.G. et FOCHT L.R. (1967) "Experimental limited vocabulary speech recognizer", IEEE - AU, 17, p. 127 - 130.

TEODORSSON S.T. (1974) The phonemic system of the Attic dialect: 400 - 340 B.C., Acta Universitatis Gothoburgensis, Lund.

TERBEECK D. (1977) A cross-language multidimensional scaling study of vowel perception, W.P.P., 37.

TORGERSON W.S. (1958) Theory and Methods of scaling, Wiley, New-York.

TRUBETZKOY N.S. (1949) Principes de phonologie, traduction J. CANTINEAU, Klincksieck, Paris.

TUAILLON M.G. (1972) "Aspects géographique de la palatalisation u > ü en gallo-roman et notamment en franco-provencal" in STRAKA G. (ed) Les dialectes de France au moyen-âge et aujourd'hui, Actes et colloques n° 9, Klincksieck, Paris, p. 205 - 235.

TVERSKY A. et KRANTZ D.H. (1970) "dimensional representation and the metric structure of similarity data", Journal of Mathematical Psychology, 7, p. 572 - 596.

UMEDA N. et COKER C.H. (1975) "Subphonemic details in American English" in FANT G. et TATHAM M.M. (eds) Auditory analysis and perception of speech, Academic press, New-York, p. 539 - 564.

UNGEHEUER G. (1962) Elemente einer akustischer Theorie der Vokal-Artikulation, Springer Verlag, Berlin.

UNGEHEUER G. (I965) "Extensional-paradigmatische Bestimmung auditiver Qualitäten phonetischer Signale", Vème Congrès Int. des Sciences Phonétiques, Karger, Bâle, p. 556 - 560.

UNGEHEUER G. (I97I) "Zum Problem der auditiver Attribute, einige ältere Autoren", in HAMMERICH L.L et al. (eds) Form and substance, Akademisk Forlag, Copenhagen, p. I09 - I22.

Van der KAMP L.J. Th. et POLS L.C.W. (I97I) "Perceptual analysis from confusions between vowels", Acta Psychologica, 35, p. 64 - 77.

VON ESSEN O. (I964) "An acoustic explanation of the sound shift ɨ > u and l > i " in In Honor D. JONES, Longmans, Londres, p. 53 - 58.

WAGENAAR W.A. (I968) "Application of Luce's choice axiom to form discrimination", Ned. Tijds chr. Psychol., 23, p. 96 - I08.

WAKITA H. (I973) "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms", IEEE - AU, 2I, p. 4I7 - 427.

WAKITA H. (I974a) A theory of linear prediction:acoustic tube model for estimating vocal tract area functions, Speech Communication Research Lab., Santa Barbara, California, non publié.

WAKITA H. (I974b) "Estimation of the vocal tract length from acoustic data", JASA, 55, suppl. J7 (résumé)

WAKITA H. (I976) "Instrumentation for the study of speech acoustics", in LASS N.J. (ed) Contemporary Issues in experimental phonetics, Academic press, New-York, p. 3 - 40.

WAKITA H. (I977a) "Some considerations for estimating vocal tract area function from speech sounds, 9ème I.C.A., Madrid, paper I 60, p. 466.

WAKITA H. (I977b)" Normalization of vowels by vocal tract length and its application to vowel identification", IEEE - ASSP, 25, p. I83 - I92.

WALTER H. (I976) La dynamique des phonèmes dans le lexique français contemporain, France-expansion, Paris.

WARREN R.M. (I976) "Auditory illusions and perceptual processus", in LASS N.J. (ed) Contemporary issues in experimental phonetics, Academic Press, New-York, p. 389 - 4I7.

WEBSTER (I9I9) "Acoustical impedance and the theory of horns, Proceedings of the national Academy of Sciences (USA), 5, p. 275 - 282.

WELLS J.C. (I962) A study of the formants of the pure vowels of British English, M.A. Thesis, University of London.

WILSON H.F. et BOND Z.S. (I977) "An INDSCAL analysis of vowels excerpted from 4 phonetic context", J. of Ph., 5, p. 36I - 367

WISH M,, DEUTSCH M. et BIENER L. ((I972) "Differences in perceived similarity of nations", in ROMNEY A.K., SHEPARD R.N., et NERLOVE S.B. (eds) Multidimensional scaling: volume II Applications, Seminar Press, New-York, p. 289 - 3I3.

WOOD S. (I975) "Tense and lax vowels - Degree of constriction or pharyngeal volume?", Working papers, Phonetics Lab., U. of Lund, vol. II, p. I09 - I34.

WOOD S. (I977) "A radiographic analysis of constriction locations for vowels ", Working papers, Phonetics Lab., U. of Lund, vol. I5, p. IOI - I32.

ZERLING J.P. (I974a) "Etude d'un programme d'ordinateur permettant de déterminer les trois premiers formants à partir de la fonction d'aire", T.I.P. de Nancy, vol. I, p. 257 - 292.

ZERLING J.P. (I974 b) "Présentation d'un programme d'ordinateur permettant de déterminer une fonction d'aire à partir des trois premiers formants", T.I.P. de Nancy, vol. I, p. 293 - 324.

ZERLING J.P. (I979; à paraître) Paramétrisation de la forme du conduit vocal dans des énoncés dissyllabiques français , Thèse de 3ème cycle, Nancy.

ZWICKER E. et FELDKELLER R. (I967) Das Ohr als Nachrichtenempfänger, Hirzel Verlag, Stuttgart.

TABLE DES MATIERES

Page

C H A P I T R E I: INTRODUCTION

I - La perception des voyelles en contexte I

II - Problèmes de méthodologie 6

III- Motivation et plan de l'étude 7

C H A P I T R E II : ANALYSES MONODIMENSIONNELLES DE L'ESPACE PERCEPTIF DES VOYELLES

I - Introduction II

II - FISCHER-JØRGENSEN (I967) I2

III - JERRENTRUP et THÜRMANN (I974) I7

IV - JOHANSSON (I975) 22

V - CALBRIS (I974) 24

VI - Les expériences d'association de voyelles et de sons purs 26

VII - Conclusions 28

C H A P I T R E III : ANALYSES MULTIDIMENSIONNELLES DE L'ESPACE PERCEPTIF DES VOYELLES

I - Introduction 29

II - Fondements théoriques de l'analyse multidimensionnelle 29

III - La méthode INDSCAL de CARROLL et CHANG (I970) 33

IV - Revue critique et analyse des travaux antérieurs par la méthode INDSCAL 50

V - Analyse de IO voyelles françaises 89

VI - Conclusion générale: la bidimensionnalité de l'espace perceptif I06

C H A P I T R E IV : F'2, CORRELAT PHYSIQUE DE LA DEUXIEME DIMENSION PERCEPTIVE

I	-	Introduction	I09
II	-	F'2: le deuxième formant "équivalent"	II0
III	-	Echec de notre méthode expérimentale	III
IV	-	Estimation de F'2; les difficultés	II5
V	-	La réalité perceptive de F'2: spectres de sonie et expérience de masquage	II8
VI	-	Extraction de F'2 par une méthode fréquentielle	I20
VII	-	Estimation de F'2: une technique originale	I23
VIII	-	Vers un modèle probabiliste d'identification dans un espace bidimensionnel	I26
IX	-	Conclusions	I33
	-	Annexe A: score d'identification de voyelles synthétiques isolées	I35
	-	Annexe B: Intensité sonore: terminologie et méthodes	I39

C H A P I T R E V : ESPACE PERCEPTIF ET SYSTEMES VOCALIQUES: ECHEC D'UNE EXPLICATION PUREMENT PERCEPTIVE

I	-	Introduction	I43
II	-	Le principe de différenciation maxima	I43
III	-	Structures phonétiques ou structures phonologiques?	I44
IV	-	Formalisation du critère de différenciation maxima	I45
V	-	L'espace perceptif des timbres vocaliques	I46
VI	-	Les données empiriques: structures vocaliques attestées dans les langues	I48
VII	-	Discussion et conclusions	I52

C H A P I T R E VI : CALCUL DES FREQUENCES FORMANTIQUES A PARTIR DE LA FORME DU CONDUIT VOCAL

I	-	Introduction	I55
II	-	Equations de Webster	I56
III	-	Les conditions d'utilisation: problèmes d'interpolation	I57
IV	-	Les corrections	I60
V	-	Conclusions	I65
	-	Annexe: Programme FORTRAN	I66

C H A P I T R E VII : APPLICATION DE LA METHODE DE PREDICTION LINEAIRE AU CALCUL DE LA FORME DU CONDUIT VOCAL A PARTIR DES FREQUENCES FORMANTIQUES

I	-	Introduction	I67
II	-	Prédiction linéaire: les fondements théoriques	I68
III	-	Analyse spectrale	I7I
IV	-	Le modèle de production implicite	I74
V	-	Prédiction de la forme du conduit vocal	I78
VI	-	Test de la précision de l'estimation des fonctions d'aire	I8I
VII	-	Conclusions	I84
	-	Annexe A: Les nombres complexes	I85
	-	Annexe B: Exponentielle complexe	I86
	-	Annexe C: Notion de "phaseur"	I87
	-	Transformée en Z: définition et théorème fondamental	I88
	-	Annexe E: Exemple du calcul d'une fonction de transfert (spectre) à l'aide de la transformée en Z	I89
	-	Annexe F: Programme FORTRAN; prédiction linéaire et calcul des fonctions d'aire	I9I

C H A P I T R E VIII : LES CONTRAINTES ARTICULATOIRES

I	-	Introduction	I92
II	-	Théorie des maxima et théorie quantale	I94
III	-	Une contrainte articulatoire sur la forme du conduit vocal	I95
IV	-	Une contrainte sur les "axes" de l'espace acoustique	2II
V	-	Conclusions	2I8

C H A P I T R E IX : VERS UNE EXPLICATION ARTICULO - PERCEPTIVE DE LA MUTATION [u > y]

I	-	Introduction	220
II	-	La palatalisation [u > y] en ancien français	22I
III	-	La théorie physiologique de G. STRAKA	223
IV	-	L'explication structurale de HAUDRICOURT et JUILLAND	226
V	-	Pour une explication articulo - perceptive	228
VI	-	Ancien français	235
VII	-	Wallon	238
VIII	-	Lorrain	243
IX	-	Franco-provencal	245
X	-	Alsacien	247
XI	-	Grec ancien	25I
XII	-	Anglais australien	255
XIII	-	Slave commun	257
XIV	-	Japonais	258
XV	-	Suédois	259
XVI	-	Conclusions	260

C H A P I T R E X : LES VOYELLES NASALES

I	-	Introduction	26I
II	-	Modèle théorique de production et analyse descriptive des voyelles nasales	26I
III	-	Analyse descriptive d'un échantillon de 48 voyelles nasales françaises	268
IV	-	La voyelle [ɔ̃]	269
V	-	La voyelle [ɛ̃]	275
VI	-	La voyelle [ɑ̃]	279
VII	-	La voyelle [œ̃]	285
VIII	-	Synthèse des voyelles nasales: importance des fréquences formantiques	289
IX	-	L'articulation des voyelles nasales	292
X	-	Note sur l'analyse des voyelles nasales par prédiction linéaire	295
XI	-	Conclusions	297
	-	Annexe: Programme de simulation de voyelles orales et nasales (SIMNAS)	298

C H A P I T R E XI : OBSERVATIONS SUR LA THEORIE DES TRAITS DISTINCTIFS

I	-	Introduction	30I
II	-	Le rôle de la Phonologie	30I
III	-	L'inventaire des traits: le test de commutation	303
IV	-	Théorie des traits et modèle de perception	305
V	-	L'inventaire des traits et le problème des classes naturelles	3I0
VI	-	Evolutions diachroniques: implications pour un système de traits	3I6
VII	-	Vers un système de traits distinctifs pour les voyelles "fermées"	3I8

C H A P I T R E XII : CONCLUSIONS ET PERSPECTIVES

I	-	Espace perceptif des voyelles	324
II	-	Interprétation de la deuxième dimension perceptive: F'2	328
III	-	Prédiction des systèmes vocaliques par un critère perceptif: un échec partiel	330
IV	-	Etude de deux contraintes articulatoires	33I
V	-	La délabialisation comme cause initiale de l'évolution [u > y]	335
VI	-	Structure acoustique et perception des voyelles nasales	337
VII	-	Le rôle des indices perceptifs et des contraintes articulatoires: implications pour un système de traits distinctifs	339
VIII	-	Quelques axes de recherche	340

A N N E X E : Le synthétiseur à formants de l'Institut de Phonétique de Nancy 346

Références bibliographiques 36I

Table des matières 386

o o o o

UNIVERSITE NANCY BIBLIOTHEQUE