

UNIVERSITE DE NANCY II

INSTITUT DE PHONETIQUE

**ETUDES SUR LA PRODUCTION ET LA
PERCEPTION DE LA PAROLE**

**LES INDICES ACOUSTIQUES DE LA NASALITE VOCALIQUE
LA MODIFICATION DU TIMBRE PAR LA FREQUENCE
FONDAMENTALE**

**Thèse pour le Doctorat d'Etat
en Lettres et Sciences Humaines**

présentée par

François LONCHAMP

Sous la direction de M. le Professeur F. CARTON

Avril 1988

A ma mère

REMERCIEMENTS

Il est juste que les premières lignes de cette thèse aient pour objet de remercier Monsieur le Professeur F. CARTON, mon 'patron' depuis presque 16 années, de l'avoir rendue possible. Les mots me paraissent trop fades et faibles pour faire saisir l'ampleur de ma dette et de ma reconnaissance. Je ne peux que simplement dire " **merci**".

Je ne suis pas sûr de savoir exprimer aux membres du jury ma gratitude pour l'honneur qu'ils me font d'avoir accepté de juger cette thèse. C'est également une grande satisfaction d'avoir enfin l'occasion de remercier publiquement Madame le Professeur P. SIMON et Messieurs les Professeurs R. CARRE, J.P. HATON, M. ROSSI et S. MAEDA pour des raisons identiques : l'aide intellectuelle et matérielle qu'ils ont apportée à mes recherches à de très nombreuses reprises, l'intérêt qu'ils ont pris à leur développement, la confiance qu'ils m'ont témoignée en me permettant de collaborer à l'enseignement qu'ils dispensent et aux recherches qu'ils poursuivent et animent, et la réelle amitié qu'ils me témoignent. Je souhaite à tous les chercheurs d'avoir la chance exceptionnelle d'avoir été accepté, encouragé et aidé comme je l'ai été.

Une thèse, ce n'est pas la vie. C'est la vie qui la rend possible.
Merci à Julien et Delphine, pour les bisoux et la compréhension
à mon père, pour les rigueurs
à Nicole, "merci pour tout !"
à M.C.L., pour la durée
à Michèle, pour un premier tango parisien
à J.P.Z. et Françoise, pour l'accueil
à Jacqueline et Shinji, pour le gîte et l'amitié
à Noëlle C., pour nos travaux du lundi soir
à Christel S., pour la patience
à Isabelle, pour "saveurs du soir"
à Georges et Vicky, pour la porte ouverte
à Denis, pour le Raz de Sein à 4 h G.M.T.
à Michel, pour la musique
à Madame CARTON, pour deux petits pulls
et à tous ceux que j'ai oublié.

"Il ne m'a pas été possible de distinguer, par sa nature seule, une idée folle d'une idée raisonnable", Docteur LEVRET, Fragments psychologiques sur la folie, 1836;

"La recherche est une irrégularité sur laquelle nous fermons les yeux", Le Ministre de l'Instruction Publique, 1933.

$\pi \approx \sqrt{3} + \sqrt{2}$, à 0.15 % près.

PRESENTATION

La facture de cette thèse de Doctorat d'Etat n'est pas traditionnelle. En un mot, elle rassemble deux monographies dont le contenu est pour l'essentiel inédit, qui traitent, pour la première, des indices acoustiques de la nasalité vocalique, et pour l'autre, de la modification du timbre vocalique qu'entraîne un changement de la fréquence fondamentale et de la représentation perceptive du premier formant.

Cette thèse déroge donc au principe respectable de l'unicité du thème, qui est la règle dans les travaux soutenus dans les ex-Facultés des Lettres. Le choix de plusieurs thèmes, qui est plus fréquent dans le domaine scientifique, n'est pas dû à un désir d'imitation aiguillonné par un quelconque sentiment d'infériorité, mais la conséquence d'une pratique de la recherche et d'un goût personnel. De moins en moins nombreux sont les chercheurs, isolés, qui consacrent une vie de recherche entière à un sujet unique. Les laboratoires, et c'est le cas de l'Institut de Phonétique de Nancy, choisissent une palette de thèmes, qui se modifient progressivement sous les effets conjugués des difficultés rencontrées, des avancées réalisées dans le laboratoire ou ailleurs, des échanges et contacts personnels toujours plus nombreux et des contrats de financement proposés. L'éventail de cette thèse aurait été encore plus ouvert si elle avait rassemblé l'ensemble des travaux de son auteur depuis sa thèse de troisième cycle, de la tonologie du Bambara à la lecture de spectrogrammes dans une approche système-expert du décodage acoustico-phonétique.

Cette diversité peut faire craindre une excessive dispersion des efforts, et je ne prétends pas y avoir échappé. Mais je soulignerai que les deux parties de ce travail, pour disparates qu'elles soient, sont dans le droit fil des recherches entreprises pour le Doctorat de 3ème cycle, qui portait sur les indices perceptifs des voyelles orales et nasales. Cette thèse présente deux thèmes de recherche que j'ai choisis librement et développés seul. Et j'ai trouvé naturel d'être jugé sur l'état de mon pré carré. Je les ai trouvés ardues et passionnantes à la fois, par eux-mêmes et dans le contact qu'ils supposent avec des travaux de très haute qualité de chercheurs que j'admire. Je ne cacherai pas que le choix de deux thèmes est aussi affaire de prudence. Je diminuais ainsi la possibilité d'un blocage sur une interprétation ou d'un échec dans la réalisation d'une expérience. Ce choix est sans doute critiquable. Il m'a empêché de développer, notamment, l'étude expérimentale de la perception des indices de nasalité vocalique.

J'ai écrit que j'avais développé seul ces thèmes de recherche. Si j'assume seul les imperfections de ce travail, je veux affirmer à nouveau ma dette envers les nombreuses personnes qui m'ont formé et aidé, en premier lieu Monsieur le Professeur F. CARTON, et qui sont pour beaucoup dans ses éventuels mérites.

PREMIERE PARTIE



LES INDICES ACOUSTIQUES DE LA NASALITE VOCALIQUE

I - INTRODUCTION

La première partie de cette thèse est consacrée à l'étude des caractéristiques acoustiques de la nasalité vocalique (*). Comme il deviendra sous doute évident au fil des chapitres, ce dossier, ouvert dans la phase contemporaine de la phonétique expérimentale par la monographie de JOOS (1948) il y a 40 ans, n'est pas encore clos.

Deux raisons nous ont conduit à traiter, après beaucoup d'autres, de cette question. Il y a 10 ans, dans le chapitre de notre thèse (LONCHAMP 1978; cf aussi LONCHAMP 1979) consacré aux voyelles nasales, nous tirions certaines conclusions qui nous paraissent aujourd'hui douteuses ou erronées. Nous souhaitons les corriger. Plus important peut-être, des travaux récents et novateurs menés en France au CNET-LANNION et à l'Institut de la Communication Parlée de GRENOBLE nous ont donné envie de discuter de façon détaillée certaines de leurs conclusions.

Dès l'abord, il convient de préciser que ce chapitre n'abordera l'aspect articulatoire que pour tirer profit des excellents travaux de cinéradiographie conduits à l'Institut de Phonétique de STRASBOURG et pour préciser, à partir de documents publiés, certains détails anatomiques de la géométrie du conduit nasal.

L'important problème de la perception des indices de nasalité vocalique ne sera pas traité très longuement, et n'a pas fait l'objet d'expériences spécifiques. Nous pensons, et nous tâcherons de montrer, que l'étude de la perception des voyelles nasales est subordonnée à une meilleure connaissance de la réalité acoustique.

Un mot sur la méthode employée : caractériser les indices acoustiques de la nasalité consiste pour nous, comme pour un nombre croissant de chercheurs, à lier de façon étroite ce que nous savons de la forme du conduit vocal aux spectres acoustiques observés, au moyen de modèles de simulation numérique.

2 - ORGANISATION

Ce paragraphe précise le plan adopté, livrant dès l'abord quelques idées essentielles afin de permettre au lecteur de mieux saisir le fil de notre organisation des matériaux. Une partie importante de notre discussion portera sur l'une des conclusions tirées par les chercheurs de l'I.C.P. de GRENOBLE : seul un conduit nasal possédant une constriction relativement

* Le traitement de texte utilisé ne permet pas de placer le "tilde" nasal sur toutes les voyelles. Aussi serons nous contraints d'écrire [~e], par exemple, à la place de [ẽ]. Nous noterons [ã] pour le symbole [ǣ] de l'A.P.I., [õ] pour [õ], et [œe] pour [œ̃].

forte dans la zone des narines permet de rendre compte de la totalité des faits de nasalité vocalique. Cette constriction narinaire implique que la fréquence du formant lié à la cavité nasale (formant nasal) est relativement basse. Le rôle potentiel des sinus comme cavités résonantes auxillaires est alors moindre. Nous proposerons qu'un conduit nasal ne possédant pas cette propriété est également adéquat. L'étude des indices se concentrera sur ceux qui apparaissent à une fréquence inférieure à 1 KHz. Le peu que nous sachions sur la perception de la nasalité indique sans équivoque que les plus importants sont en basse fréquence, sous 1 kHz (HAWKINS & STEVENS 1985 par ex.). La caractérisation des indices inférieurs conditionne d'ailleurs celle des indices de fréquence plus élevée, qu'ils déterminent presque entièrement. Nous passerons d'abord en revue les éléments d'anatomie permettant d'évaluer la taille de la partie antérieure des fosses nasales, notamment des clichés de tomographie calculée ("scanner"). Nous préciserons ensuite les renseignements acoustiques qui peuvent être déduits de l'étude de la vue sagittale de l'articulation orale des voyelles nasales. Les chapitres suivants présentent tous les "outils" nécessaires aux simulations et à l'analyse acoustique : théorie du couplage des conduits; modèle de simulation numérique, choix raisonné de la F.F.T. comme méthode d'analyse, effets spectraux liés à la source vocale, valeurs des fonctions d'aire des conduits. Seront ensuite passées au crible les études antérieures dans leur ordre de parution en un essai de synthèse interprétative. Puis les résultats de nos simulations sont commentés et comparés à un échantillon de spectres réels, sur lesquels nous tenterons d'identifier les indices spécifiques de la nasalité vocalique. Enfin, le problème de la perception de ces indices fait l'objet d'un dernier chapitre, où nous analysons une série de travaux récents.

3 - ANATOMIE DES FOSSES NASALES

3.1 - INTRODUCTION

Cette présentation de l'anatomie des fosses nasales que nous proposons ci-dessous n'a pas simplement pour but de situer la nasalité dans son contexte articuloire, mais vise aussi à évaluer certaines critiques implicites émises sur les mesures réalisées par BJUGGREN & FANT (1964). La cavité nasale (ou fosses nasales) peut être divisée en trois parties de tailles inégales : d'avant en arrière, on trouve la zone du nez (≈ 2 cm), puis les fosses nasales proprement dites (6 - 7 cm), divisées verticalement par une cloison paramédiane, et enfin le conduit rhinopharyngal ou nasopharyngal (3 - 5 cm). La position générale de la cavité nasale dans la tête est trop connue pour mériter une description détaillée (figure 1.1).

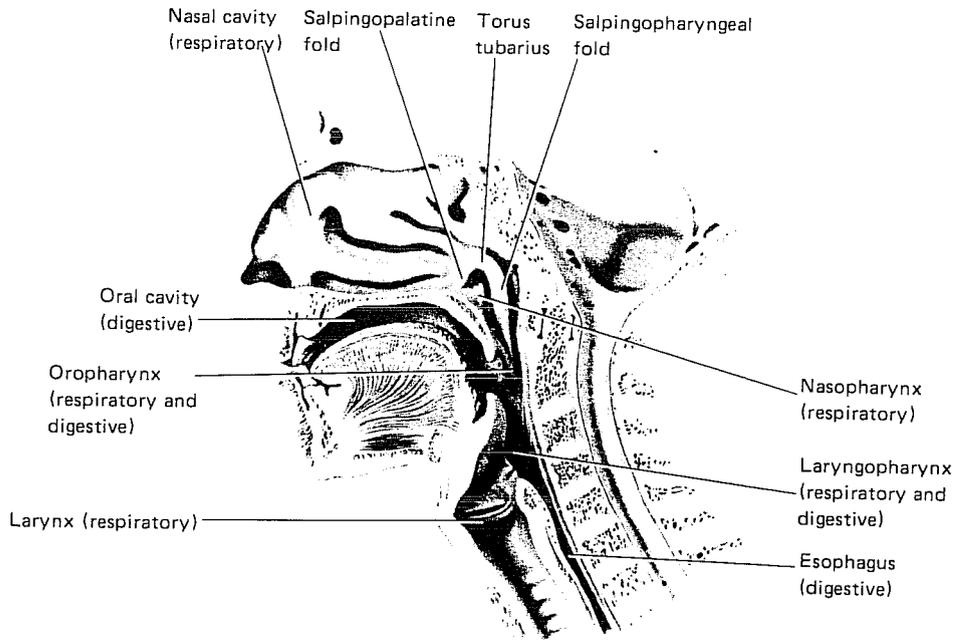


Figure 1.1 - Position anatomique des fosses nasales et des structures adjacentes. (in ZEMLIN 1981)

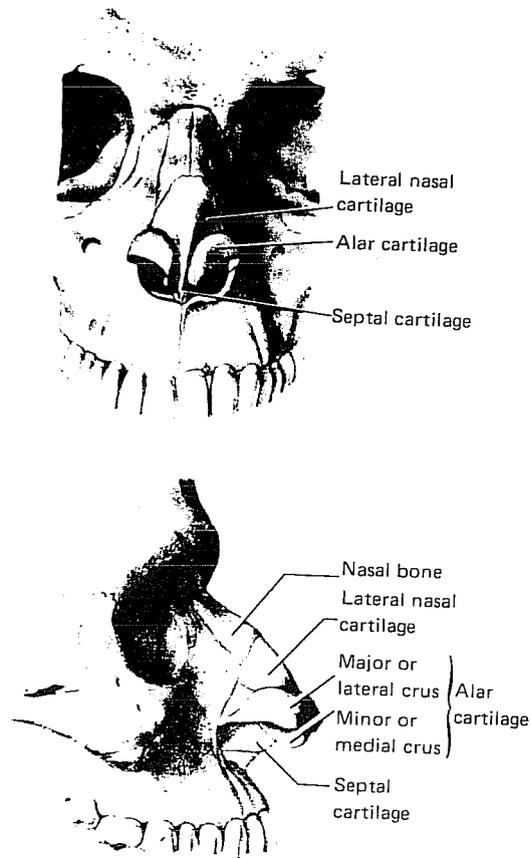


Figure 1.2 - Os et cartilages du nez. (in ZEMLIN 1981)

3.2 - LE NEZ

L'architecture osseuse et cartilagineuse du nez est bien connue car visible (figure 1.2). Notons simplement l'existence d'une lame cartilagineuse interne verticale (partie antérieure du septum nasal) qui le divise en deux parties plus ou moins égales.

3.3 - LES FOSSES NASALES et le RHINOPHARYNX

Les fosses nasales ont la forme d'un volume trapézoïdal allongé dont la base est plus large que le sommet. Leur longueur est d'environ 7 à 8 cm, la hauteur maximale de 5 cm et la largeur passe de 4 cm à la base à moins de 1 cm au sommet, compte non tenu des cavités accessoires (sinus). Elles sont divisées verticalement par une cloison cartilagineuse et osseuse. Le rhinopharynx a une forme grossièrement parallélépipédique de 3 à 4 cm de longueur lorsque le voile du palais est relevé.

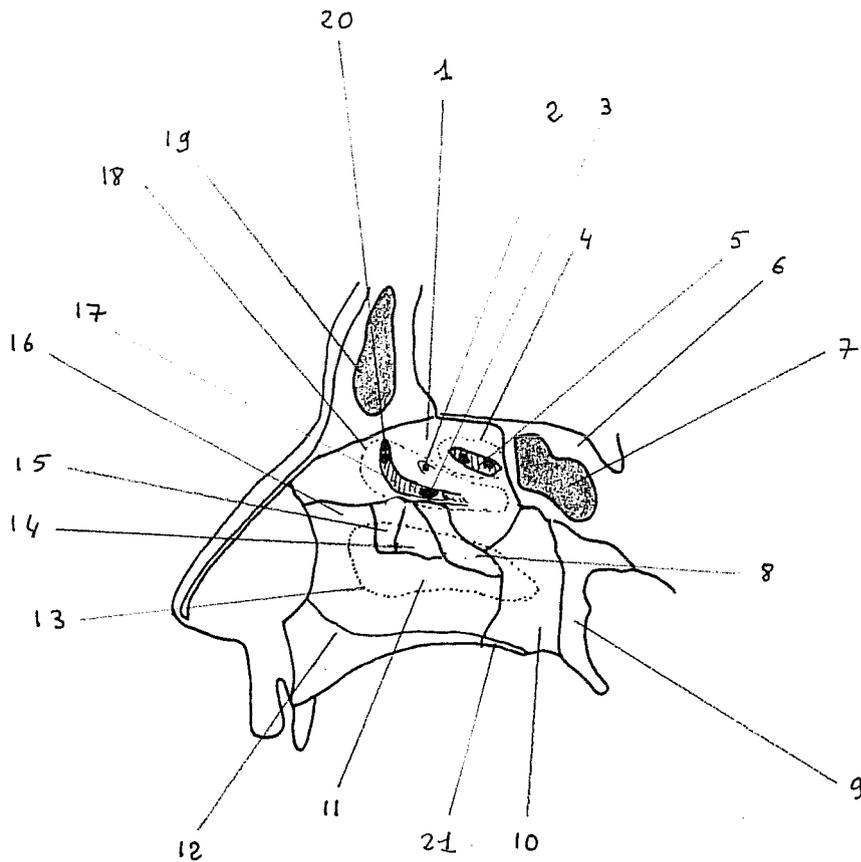
L'anatomie des fosses encombrées de replis cartilagineux est d'une très grande complexité d'autant que les parois ne sont qu'un puzzle d'os et de cartilages.

La paroi latérale (figures 1.3 et 1.4) est constituée dans sa partie antérieure et inférieure par le corps du maxillaire (apophyse antérieure et surface nasale), creusé du volumineux sinus maxillaire (figure 1.5) dont la béance est masquée par des plaques que nous allons décrire de gauche à droite : l'unguis ou os lacrymal, un petit triangle de muqueuse et l'apophyse unciforme de l'os ethmoïde. Au-dessus s'étend la masse latérale de l'os ethmoïde, notamment la face interne de l'os planum. Cette zone est si complexe qu'elle est impossible à décrire en vue latérale. Nous la retrouverons plus bas en coupe frontale. La partie postérieure de la paroi latérale est formée par la lame verticale de l'os palatin (figure 1.6), dont son apophyse sphénoïde dans la zone supérieure, puis plus loin dans le rhinopharynx, par l'aile ou apophyse ptérygoïde interne de l'os sphénoïde (figure 1.7).

Le plancher nasal se compose à l'avant de l'apophyse palatine du maxillaire (plaque horizontale) et à l'arrière de la lame horizontale du palatin. On sait que la masse musculaire du voile du palais prolonge celle-ci jusqu'à la paroi postérieure du (rhino)pharynx.

La partie postérieure des fosses est formée en haut par le corps du sphénoïde, creusé du sinus sphénoïde (figure 1.8), puis vers le bas et l'arrière de l'apophyse basilaire du même. La variabilité individuelle est bien illustrée sur cette figure. La paroi latérale est plissée dans la zone du débouché de la trompe d'Eustache (orifice tubaire).

Le toit des fosses sera présenté en coupe frontale plus loin. Notons dès à présent qu'à l'avant le toit est formé par l'os nasal (cf. figure 1.2) et l'apophyse nasale de l'os frontal (sous le sinus frontal; cf. figure 1.3).



- 1 - Os planum de l'ethmoïde (ou masse latérale)
- 2 - Orifice d'entrée des cellules ethmoïdes médianes dans la bulle ethmoïde.
- 3 - Orifice d'entrée (ostium) du sinus maxillaire
- 4 - Limite du cornet supérieur
- 5 - Orifice d'entrée des cellules ethmoïdes postérieures et du sinus sphénoïde dans la gouttière rétro-bullaire.
- 6 - Corps du sphénoïde.
- 7 - Sinus sphénoïde
- 8 - Apophyse unciforme du sphénoïde
- 9 - Apophyse (ou aile) ptérygoïde interne du sphénoïde
- 10 - Lambe verticale du palatin
- 11 - Corps du maxillaire (surface nasale: le centre du sinus maxillaire est à ce niveau)
- 12 - Maxillaire; apophyse palatine (lambe horizontale)
- 13 - Limite du cornet inférieur
- 14 - Muqueuse
- 15 - Unguis (os lacrymal)
- 16 - Apophyse frontale du maxillaire
- 17 - Gouttière unci-bullaire (hiatus semilunaris)
- 18 - Limite du cornet moyen
- 19 - Sinus frontal
- 20 - Orifice d'entrée du sinus frontal et des cellules ethmoïdes antérieures
- 21 - Lambe horizontale du palatin

Figure 1.3 - Paroi latérale des fosses nasales (cornets supprimés) : représentation schématique d'après plusieurs sources.

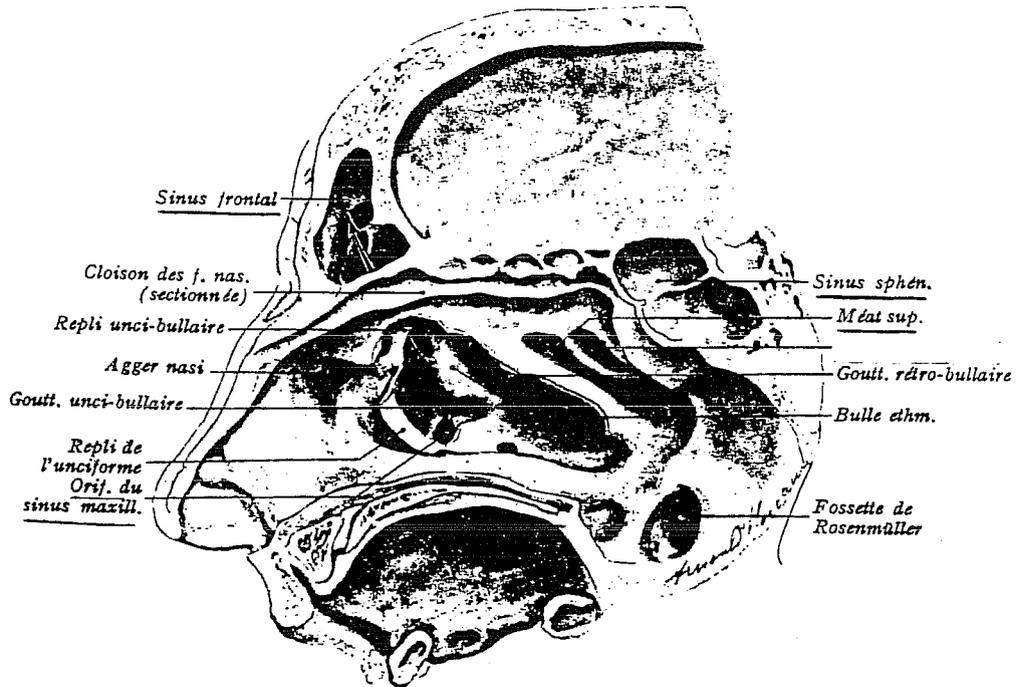


Figure 1.4 - Paroi latérale des fosses nasales (cornets supprimés) : in ROUVIERE (1948).

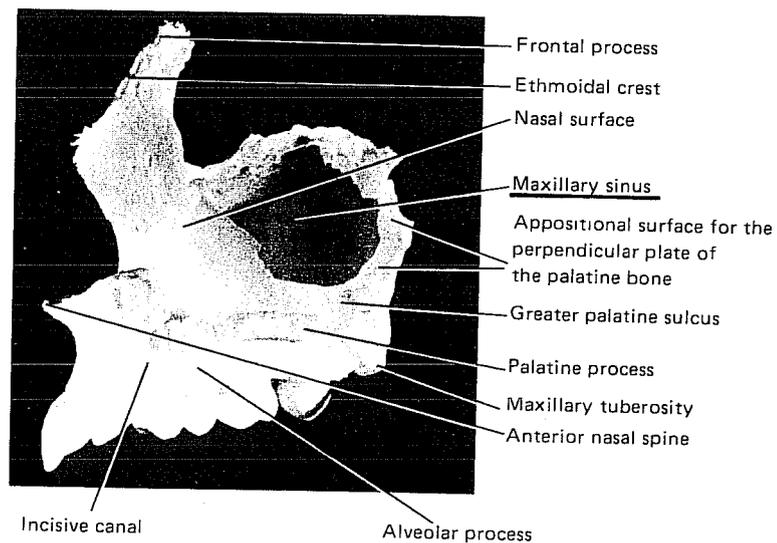


Figure 1.5 - Os maxillaire : vue interne (in ZELIN 1981).

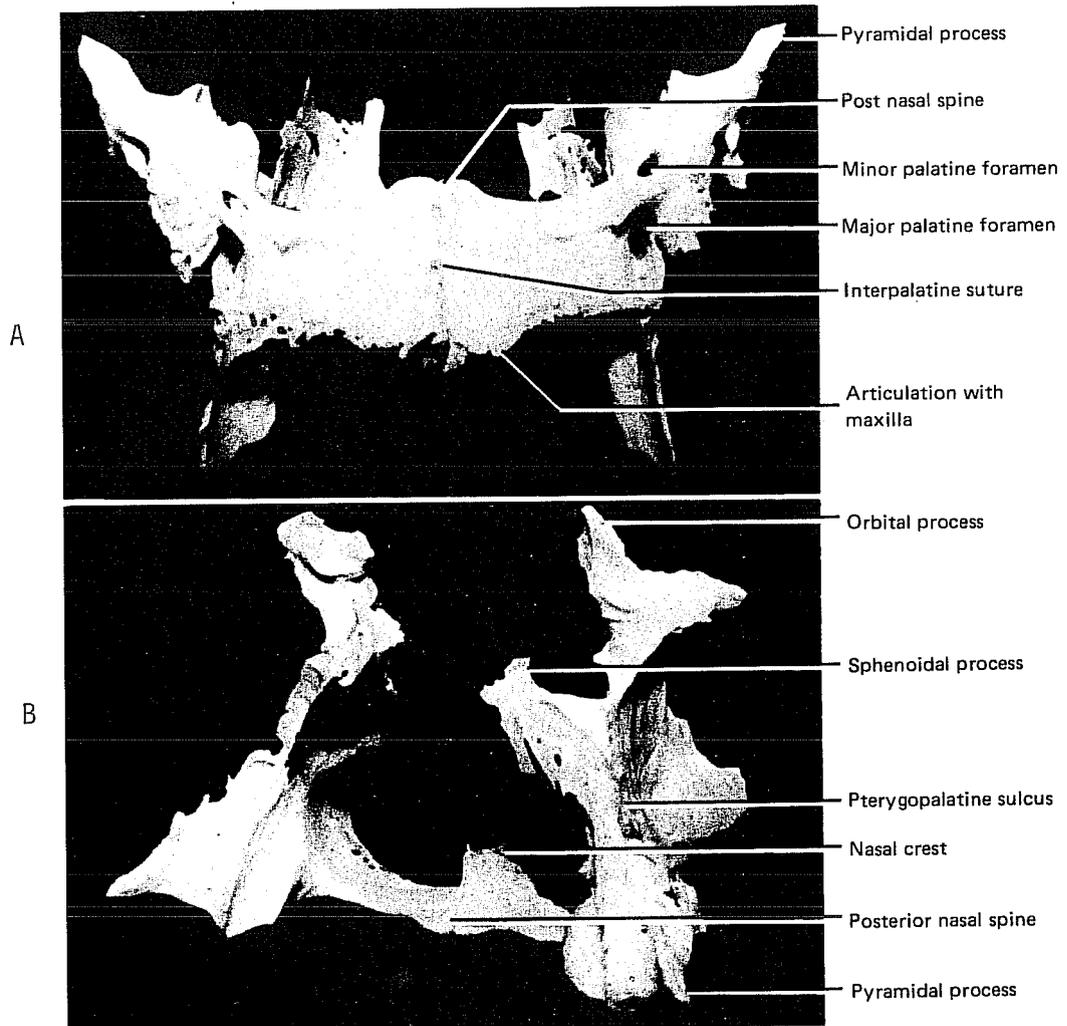


Figure 1.6 - Os palatin - A : vue de la face inférieure
 B : Vue par l'arrière. (in ZEMLIN 1981)

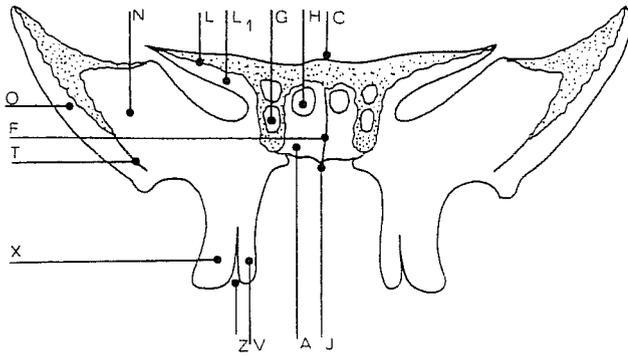


FIG 10 - SPENOIDE (VUE ANTERIEURE)

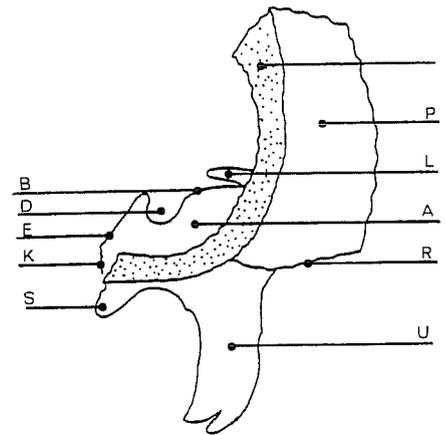


FIG 12 - SPENOIDE (VUE LATÉRALE)

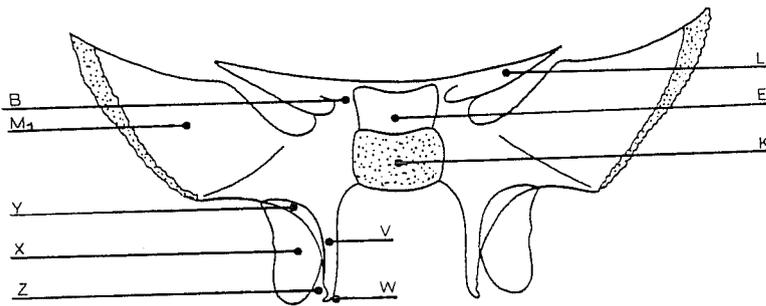


FIG 11 - SPENOIDE (VUE POSTÉRIEURE)

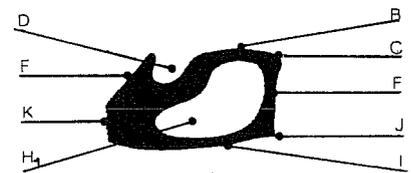


FIG 13 - SPENOIDE (COUPE SAGITTALE DU CORPS)

A corps. B jugum. C processus ethmoïdal. D fosse pituitaire. E lame quadrilatère. F crête sphénoïdale antérieure. G demi-cellule. H orifice d'entrée du sinus sphénoïdal. H1 sinus sphénoïdal. I crête sphénoïdale inférieure. J bec du sphénoïde. K Face postérieure du corps. L petite aile. L1 surface orbitaire. M grande aile. M1 face endocrânienne. M2 face exocrânienne. N face orbitaire. O face temporo-zygomatique. P partie temporale. Q partie zygomatique. R crête temporale. S épine du sphénoïde. T crête malaire. U apophyse ptérygoïde. V aile interne. W crochet. X aile externe. Y fossette scaphoïde. Z échancrure ptérygoïde.

Figure 1.7 - Os sphénoïde (in BOGNAR 1980).

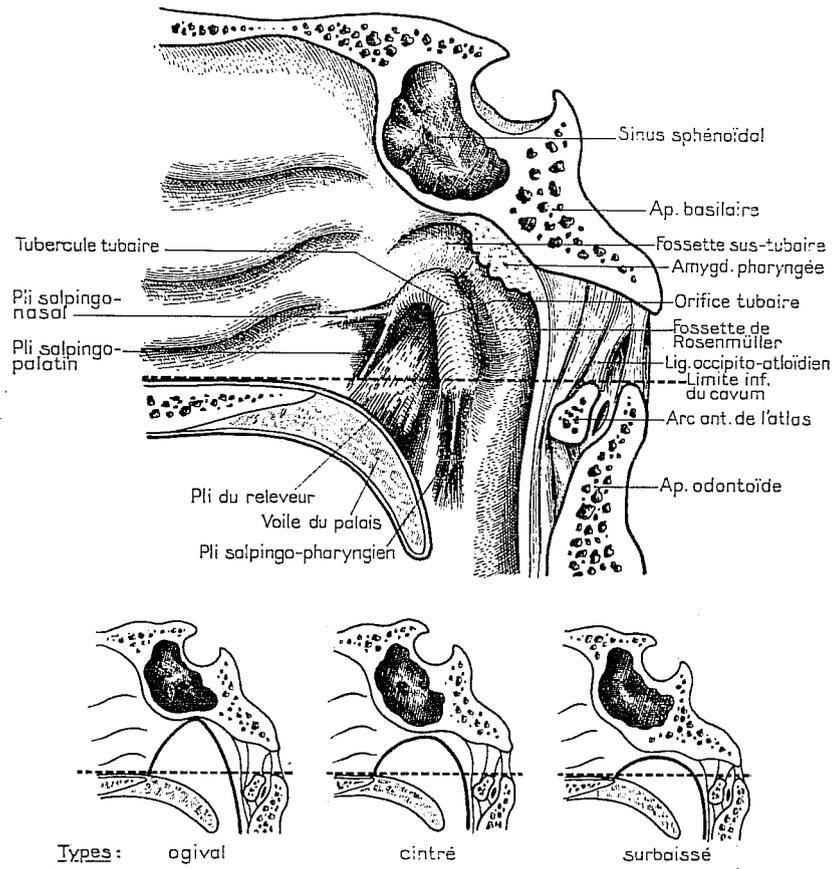


FIG. 15. — *Cavum naso-pharyngien.*
Parois latérale, supérieure et postérieure.

Figure 1.8 - Rhinopharynx (in PIALOUX 1975).

Figure 4-16 Schematic of medial wall of nasal cavity showing vomer bone, perpendicular plate of the ethmoid, and the articulation of the nasal bone with the septum.

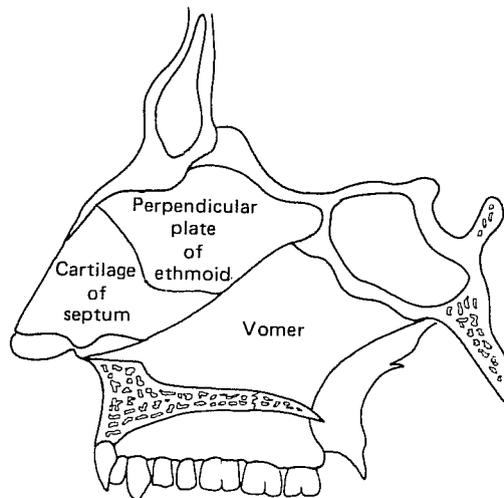


Figure 1.9 - Cloison nasale (in ZEMLIN 1981).

Une cloison verticale divise les fosses nasales en 2 parties plus ou (plutôt) moins égales. Elle est en effet presque toujours bombée latéralement, ce qui entraîne une dissymétrie des fosses droite et gauche. Elle comprend (figure 1.9) une plaque cartilagineuse antérieure se prolongeant dans le nez, qui s'articule sur la plaque perpendiculaire de l'ethmoïde ainsi que sur le vomer qui joint le sphénoïde au maxillaire et au palatin.

Les parois latérales sont partiellement masquées par 3 replis cartilagineux-osseux, ouverts vers le bas, appelés cornets supérieur, moyen et inférieur (figure 1.3 : tracés en pointillé et figure 1.10). L'espace entre les cornets et la paroi latérale, ou méat, ne mesure que quelques millimètres. Sous le cornet moyen se place la gouttière unci-bullaire (la bulle est une éminence arrondie de la paroi) où débouche, à l'arrière, l'ostium (orifice d'entrée ou foramen) du sinus maxillaire et, à l'avant, celui des sinus frontaux. Notez aussi sur la figure la bosse de l'agger nasi. La figure 1.13 précise la position du vestibule narinaire (volume des narines) et le resserrement appelé limen nasi (ou orifice nasal supérieur, à distinguer de l'orifice nasal inférieur ou narines).

Une coupe frontale est nécessaire pour préciser l'anatomie de la partie supérieure des fosses (figure 1.11 : coupe schématique ; figure 1.12 : coupe détaillée). On voit clairement les deux orbites sous lesquelles se placent les volumineux sinus maxillaires. Les fosses nasales s'insèrent entre ces structures. L'os ethmoïde, en forme de U renversé, ouvert donc vers le bas, est creusé de cellules (ou sinus) au sein des deux masses latérales, d'où sortent les cornets supérieur et inférieur. La partie supérieure plane de l'ethmoïde (lame horizontale ou cribriforme ou percée) est percée de minuscules passages pour les fibres du nerf olfactif. La lame perpendiculaire, surmontée à ce niveau de l'apophyse crista galli, s'articule à sa base, comme nous l'avons vu, sur le vomer pour diviser la cavité nasale. Les parois latérales sont donc formées à ce niveau par le bord interne de l'os planum de la masse latérale, par l'apophyse unciforme de l'ethmoïde et, plus bas, par l'apophyse maxillaire sur laquelle est pendu le cornet inférieur. Notez l'ostium du sinus maxillaire sur la figure 1.12. Les fosses nasales sont donc beaucoup plus étroites au sommet qu'à la base. Une vue par l'arrière du conduit nasal (figure 1.14) montre clairement la section rectangulaire de la partie antérieure du rhinopharynx, avec le septum vertical, les parois latérales (plaque ptérygoïde interne (ou médiale) du sphénoïde) et la plaque horizontale du palatin, au niveau de la crête nasale postérieure.

3.4 - LES SINUS PARANASaux

La figure 1.15 présente en vue cavalière la position des sinus et cellules (para)nasales.

- Les sinus frontaux sont deux cavités le plus souvent dissymétriques, creusées dans

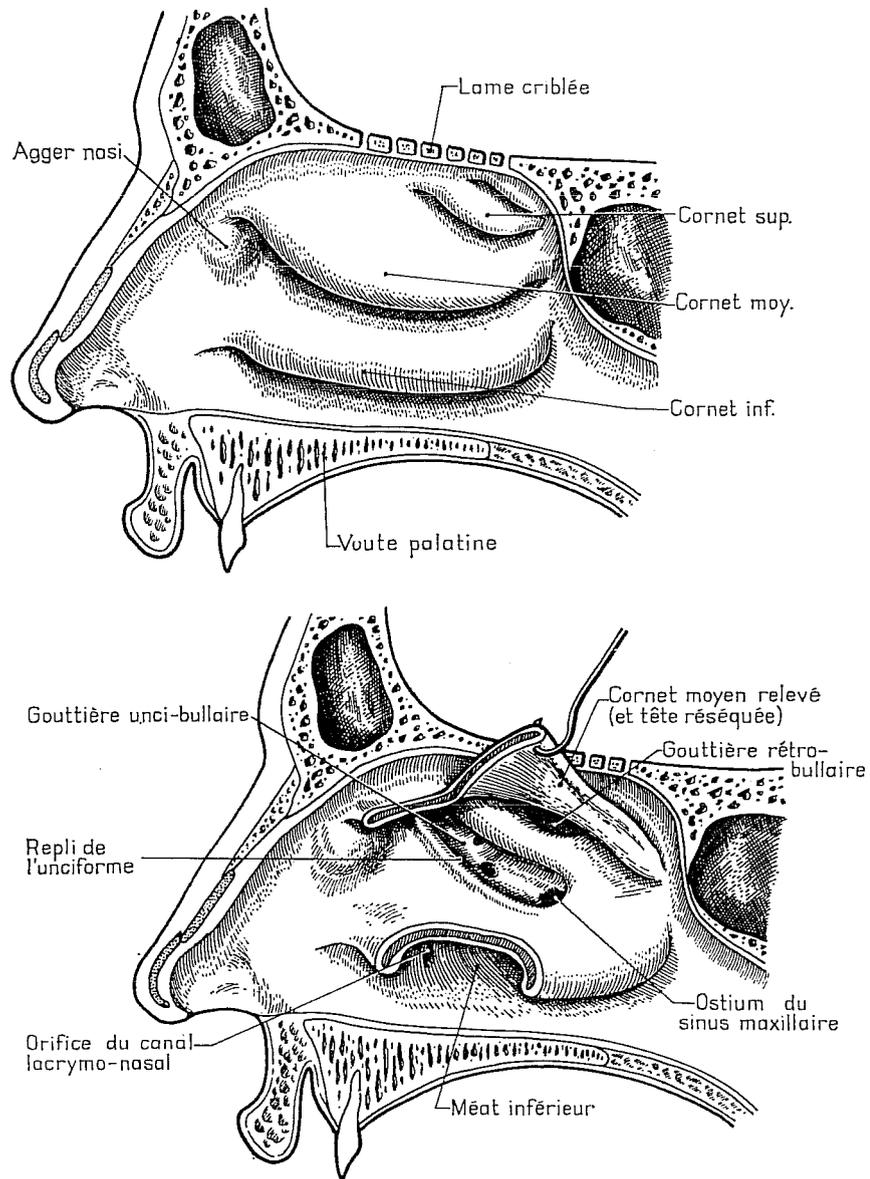


FIG. 13. — *Fosses nasales.* (Vue interne de la paroi externe.)

Figure 1.10 - Paroi latérale des fosses nasales : les cornets, et les structures sous-jacentes: méats, ostium du sinus maxillaire, gouttières, agger nasi.

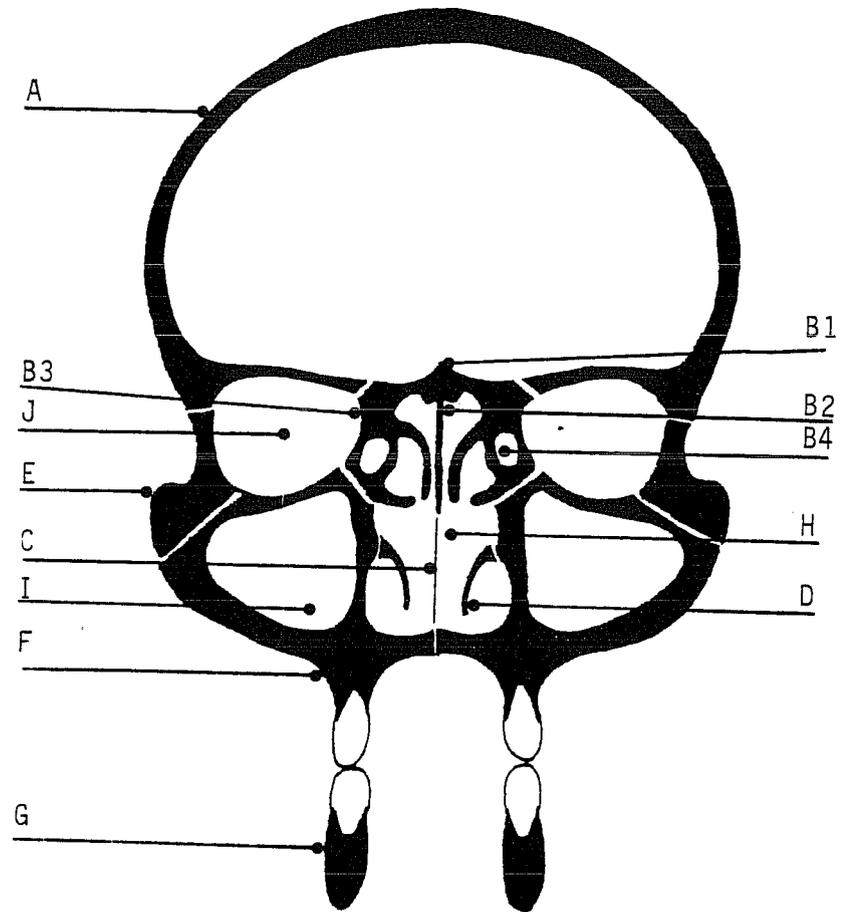


FIG 38 - TÊTE (COUPE FRONTALE)

A frontal. B ethmoïde. B1 apophyse crista-galli. B2 lame perpendiculaire. B3 os planum. B4 cellule. C cartilage de la cloison nasale. D cornet inférieur. E os malaire. F maxillaire. G mandibule. H fosses nasales. I sinus maxillaire. J orbite.

Figure 1.11 - Coupe frontale schématique de la tête
(in BOGNAR 1980).

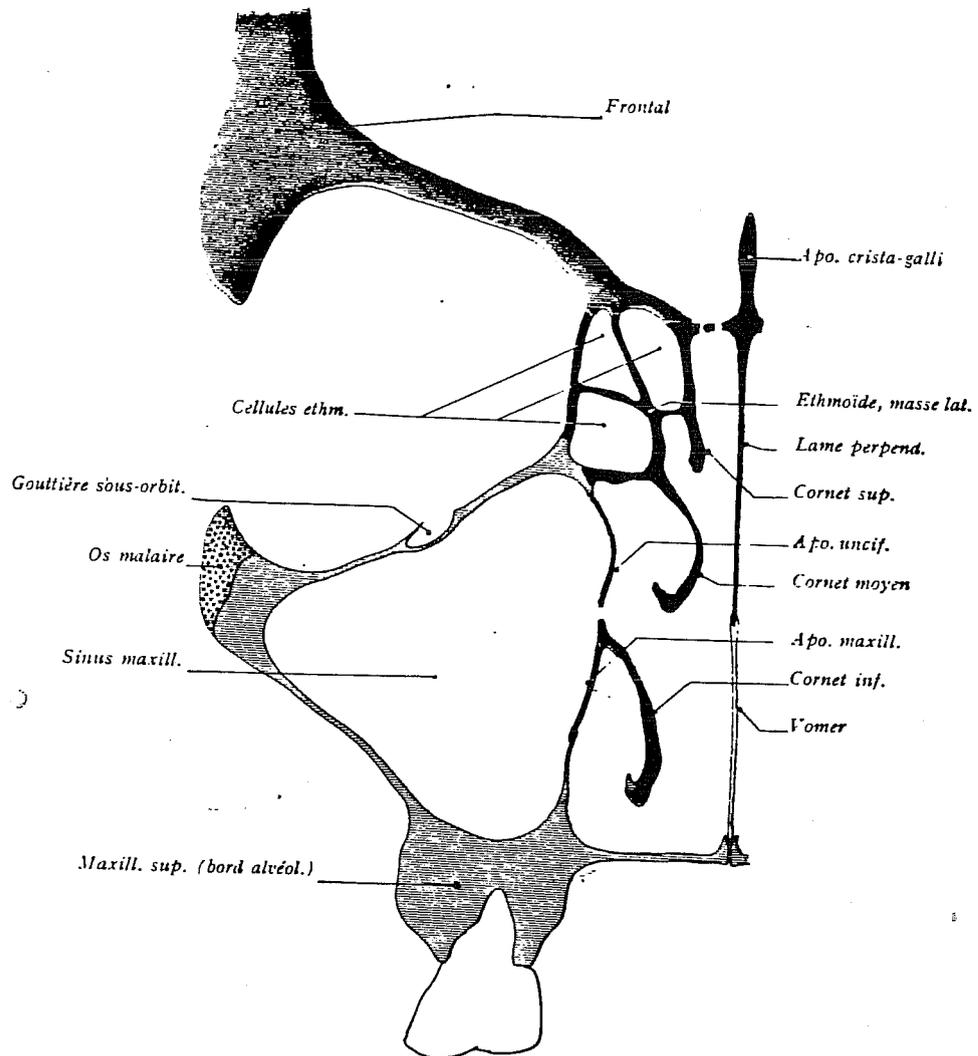


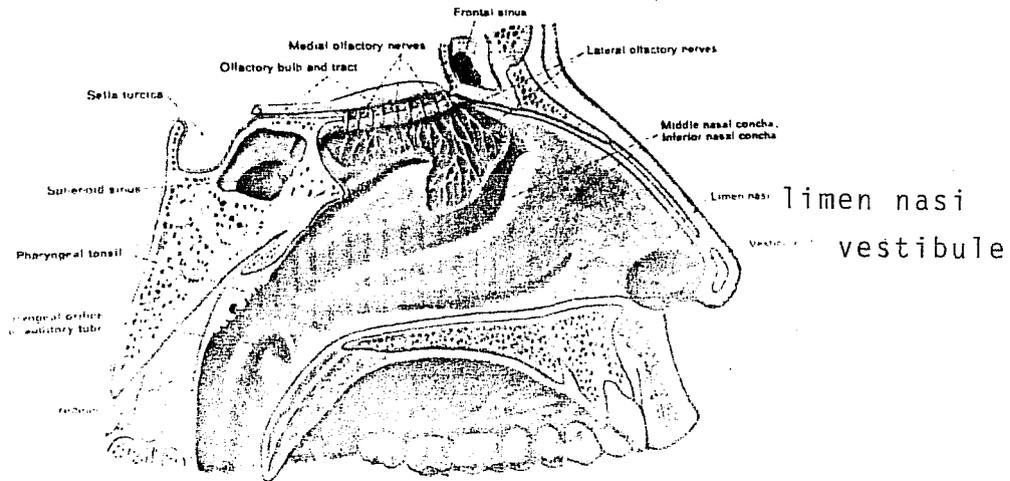
Figure 1.12 - Coupe frontale des fosses nasales et du sinus maxillaire (in ROUVIERE 1948).

Fig. 510: Arteries and Nerves of the Nasal Septum

Note that the mucous membrane has been removed from the nasal septum and nasopharynx revealing the septal vessels and nerves and the nasopharyngeal muscles.

Fig. 511: The Lateral Wall of the Left Nasal Cavity

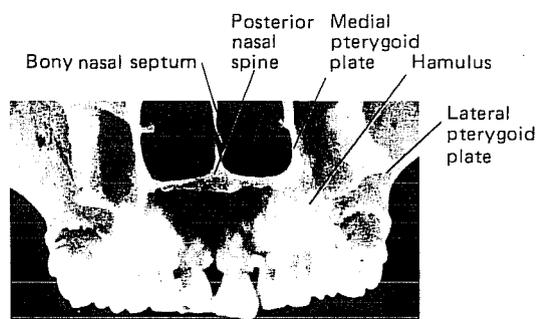
NOTE: the mucous membrane overlying the lateral olfactory nerves has been removed. The lateral wall of the nasal cavity is marked by the superior, middle and inferior nasal conchae. Beneath each concha courses its corresponding nasal passage or meatus.



· Figure 1.13 - Vue latérale des fosses nasales : position du limen nasi (seuil narinaire) et du vestibule.

(in CLEMENTE 1975)

Figure 4-78 The base of the skull as seen in detail, showing the hamulus of the medial pterygoid plate, in relation to the hard palate.



· Figure 1.14 - Les fosses nasales : vue postérieure (in ZEMLIN 1981)

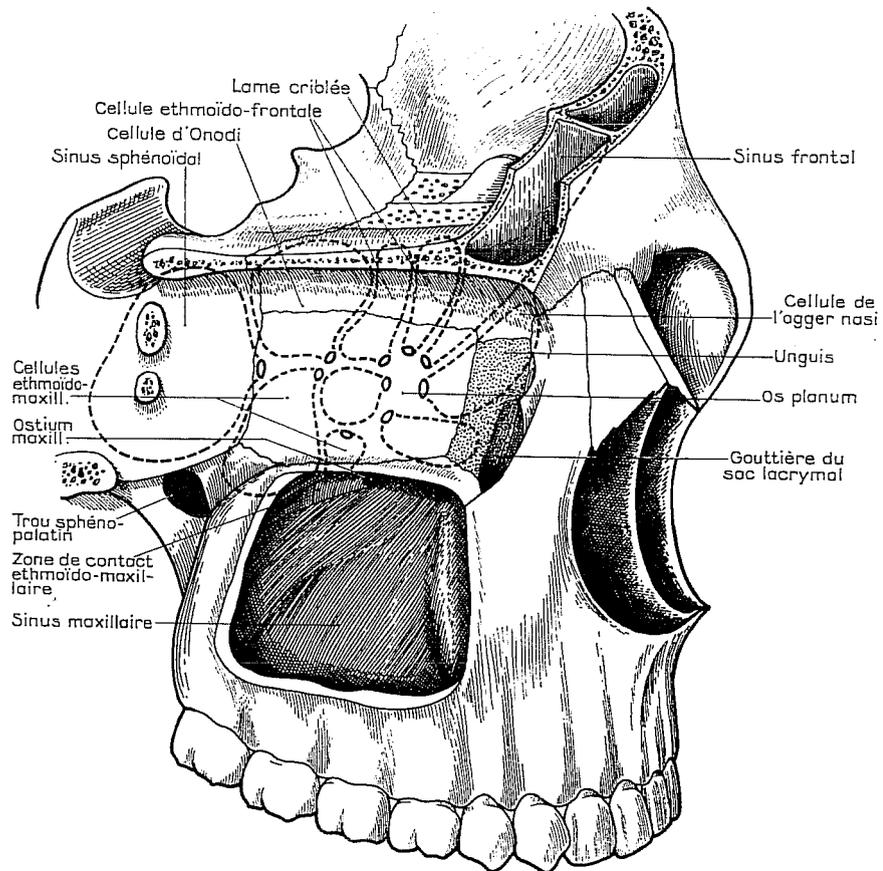


FIG. 14. — Les sinus de la face.

Figure 1.15 - Position anatomique des sinus paranasaux. (in PIALOUX 1975)

l'épaisseur de l'os frontal (cf. figure 1.3 par ex.) au-dessus des orbites. Des cloisons ou septae plus ou moins nombreuses les divisent. L'ouverture de drainage se fait au sommet de la gouttière unci-bullaire, directement ou par l'intermédiaire des cellules ethmoïdes antérieures qui y débouchent également. Les dimensions maximales à l'âge adulte sont : hauteur 3cm; largeur 2.5 cm; profondeur 2cm. Le volume théorique est donc au maximum de 15 cm³. TAKEUCHI & al. (1975) leur attribuent un volume de 5.7 cm³. Un cliché tomographique horizontal est présenté au bas de la figure 1.17. On distingue un septum non médian (flèche).

- Les sinus ethmoïdaux (ou cellules ethmoïdales) sont un ensemble de petites cavités aux parois très fines dans l'épaisseur de la masse latérale de l'éthmoïde. Leur nombre peut varier de 3 à 18 , en 3 ensembles non communicants (antérieur, médian et postérieur). Les orifices sont indiqués à la figure 1.3. Leur volume total est de 8.3 cm³ d'après TAKEUCHI & al.

- Le sinus sphénoïde est creusé dans le corps de l'os sphénoïde (figure 1.3 et 1.7). Il communique avec les fosses nasales au niveau de la gouttière rétro-bullaire sous le cornet supérieur. Nous n'avons pas trouvé de renseignements quantitatifs sur ses dimensions ou volume en dehors de la taille à la naissance (0.2 x 0.2 x 0.15 cm)!

- Les deux sinus maxillaires sont de loin les plus volumineux. Leur position sous les orbites, de part et d'autre des fosses nasales dans le corps du maxillaire, a été vue aux figures 10 et 11. Les dimensions moyennes s'établissent ainsi : hauteur 3.3 cm; largeur 2.3 cm; profondeur 3.4 cm, d'où un volume théorique de 25.8 cm³. AUST & DRETTNER (1975) rapportent les volumes mesurés sur des radiographies pour un dizaine de sujets masculins et féminins. La médiane des 8 valeurs masculines est de 24.9 cm³ (moyenne 23.7, et 24.5 en excluant les 2 extrêmes : 15.5 - 27.5 cm³). Les 3 valeurs féminines sont beaucoup plus faibles : 7.8, 10.1 et 13.8 cm³. SCHAEFFER (1920; d'après SOM (1984)) donne un volume moyen, mesuré sur des cadavres, bien inférieur : 14.75 cm³. Le volume théorique calculé plus haut sous l'hypothèse d'une forme parallélépipédique étant presque égal au volume moyen de AUST & al., ce dernier paraît trop élevé eu égard à la forme réelle, presque pyramidale, des sinus. Le volume de SCHAEFFER paraît plus plausible. TAKEUCHI & al. utilisent une valeur de 20.1 cm³. La section de l'ostium, mesurée par une technique indirecte par AUST & al. a une médiane de 4.75 mm², soit un diamètre de l'ordre de 2.5 mm. La variabilité est énorme : 1 à 24.6 mm²! D'une phrase peu explicite de ce travail, on peut déduire que la longueur de l'ostium est de 0.6 cm. ALBERTI (1976; cité par SOM) indique que le diamètre peut descendre jusqu'à 2.4 mm.

3.5 - COUPES TOMOGRAPHIQUES HORIZONTALES ET FRONTALES - LE PROBLEME DE LA TAILLE DU LIMEN NASI.

Comme nous le verrons plus loin, la taille exacte de la zone narinaire est un paramètre important de la fonction d'aire utilisée pour la modélisation acoustique de la nasalité. FENG (1986; cf. aussi FENG & al. 1985), s'appuyant, semble-t-il, sur une photographie d'une dissection où apparaissent ailleurs des distortions (PALACIOS & al. 1980, figure SA-12-A reproduite ici à la figure 1.16) et une citation, propose une aire aux narines inférieure de plus de la moitié à la valeur publiée par BJUGGREN & FANT (1964). Nous avons consulté environ 75 clichés tomographiques en coupe horizontale et frontale dans les atlas de BERGERON & al. (1984) et LATCHAW (1985). Beaucoup d'entre eux se trouvent dans des sections concernant d'autres structures que les fosses nasales. Concernant pour la plupart des sujets différents, ils offrent un échantillonnage raisonnable de niveaux de coupe. Après un tour d'horizon visuel insistant surtout sur la partie antérieure des fosses, nous donnerons quelques estimations numériques en les comparant aux données de BJUGGREN & FANT. Les figures 1.17 à 1.19 sont une sélection des clichés les plus contrastés en séquence anatomique. Ils n'ont pas été sélectionnés pour défendre une thèse particulière! L'identification du niveau a souvent été faite par l'auteur de cette thèse.

La première ligne de la figure 1.17 présente une coupe de la base des fosses nasales et des sinus maxillaires. Le vestibule nasal (narines) n'est pas étroit et si la cloison épaisse bouche l'extrême base des fosses (à gauche), ce n'est plus vrai quelques mm plus haut (à droite). Les fosses sont mêmes relativement larges dans leur partie postérieure. Notez la belle dissymétrie de la cloison à droite. La seconde ligne illustre le niveau de la base du cornet inférieur. Le vestibule peut varier de taille (N.B. il n'est pas exclu que certains sujets soient de race noire) et la cloison d'épaisseur. Le cliché de gauche est le seul exemple du seuil osseux entre le maxillaire et le nez figuré par certains anatomistes. Il y a peu d'espace entre les cornets et la cloison, mais les méats sont bien dégagés. La partie supérieure du cornet inférieur est illustrée à la troisième ligne. La taille du vestibule nasal décroît rapidement. La cloison est fine et si l'espace entre les cornets et la cloison est faible, il n'y a pas de resserrement entre le nez et les fosses. La forme est plutôt en entonnoir. Le cliché horizontal inférieur (sinus maxillaire gauche comblé par une tumeur) présente la section la plus étroite de nos données.

Les 4 clichés supérieurs de la figure 1.18 sont au niveau des cornets moyens. Le vestibule nasal a presque disparu. La cloison présente un soudain épaissement chez certains sujets qui correspond, sauf erreur de notre part, à la suture entre la partie cartilagineuse du septum et la plaque perpendiculaire de l'ethmoïde. Dans le cas où cet épaissement n'est

SA-12-A

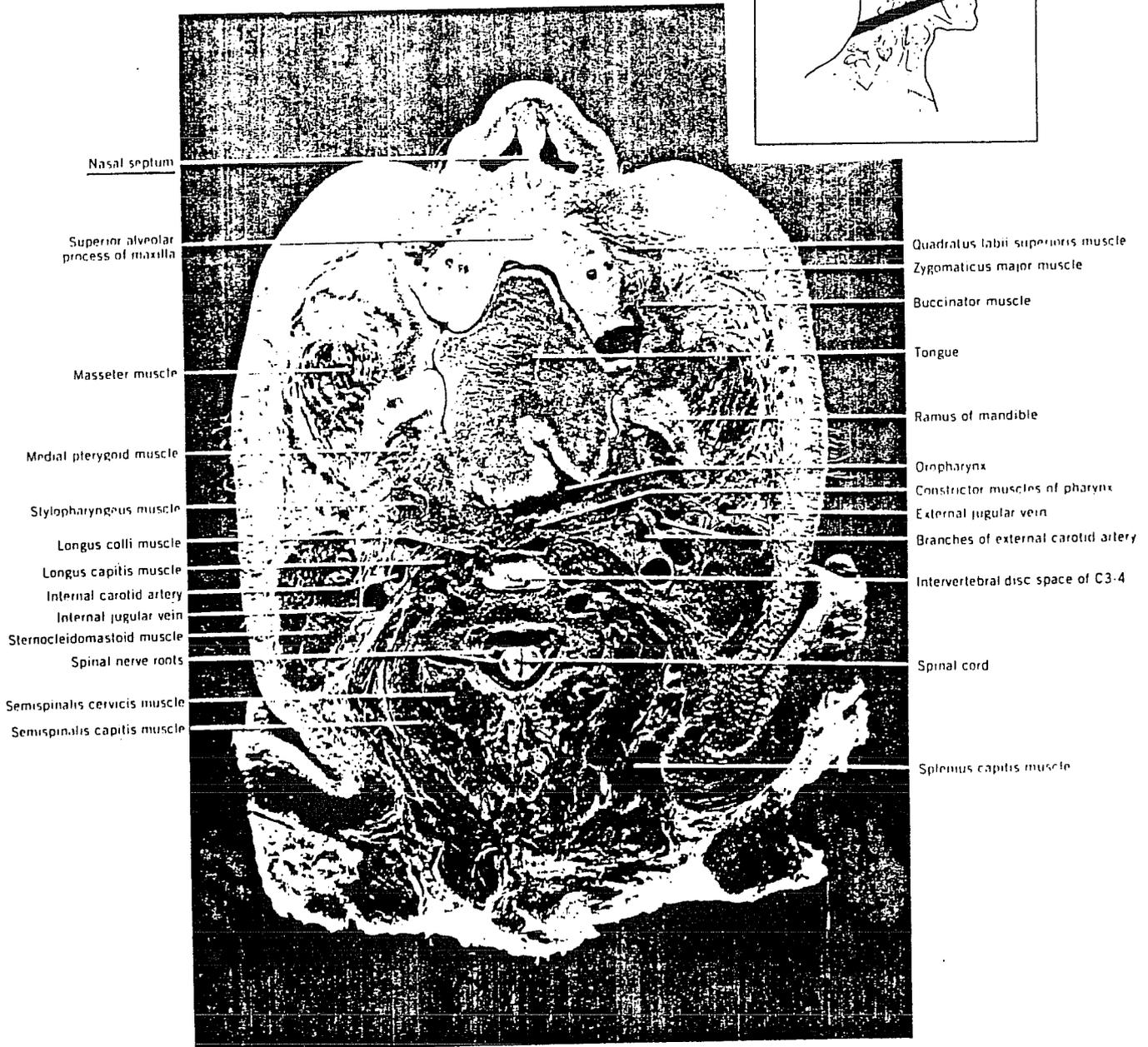
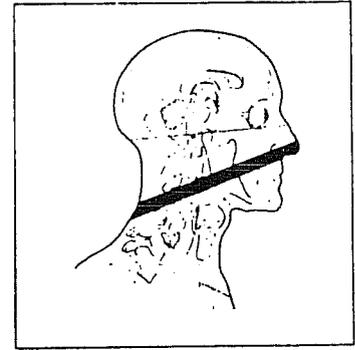
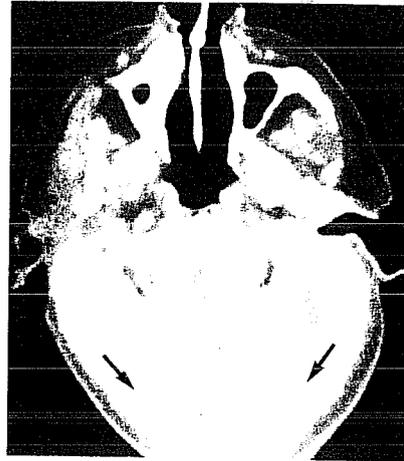


Figure 1.16 - Coupe para-horizontale de la tête :
 "forme et taille d'une section à la sortie du conduit nasal,
 aux environs du limen nasi" (FENG 1986; d'après PALACIOS
 & al. 1980).

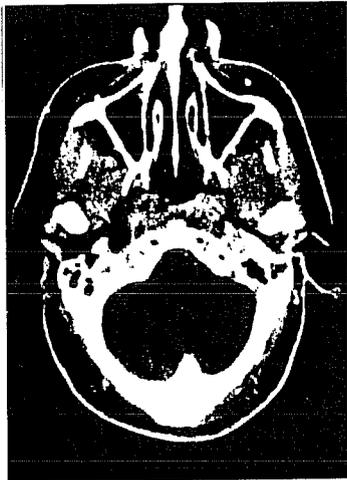


A

10 - 3E

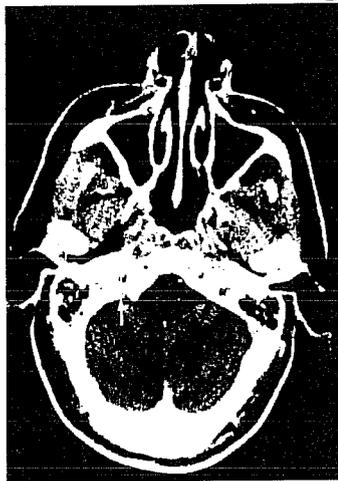


13 - 10C

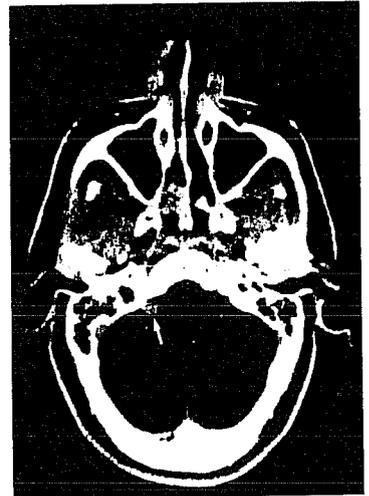


B

13 - 13C

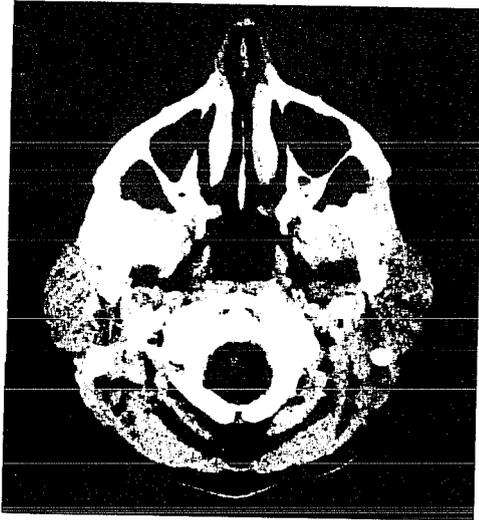


13 - 13B

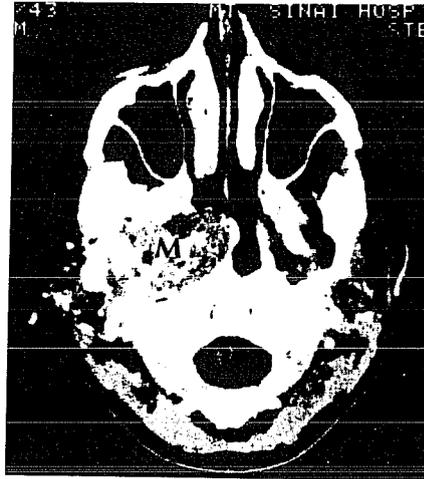


16 - 145C

Figure 1.17 - Coupes tomographiques horizontales
niveau A : sous les cornets inférieurs ; niveau B : partie
inférieure des cornets inférieurs. (in BERGERON & al. 1984)



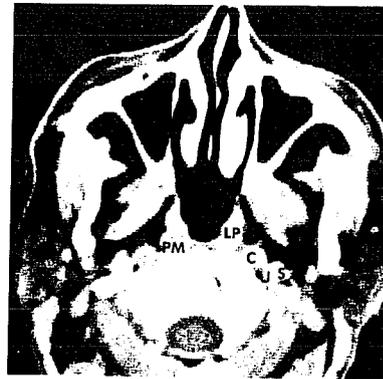
4 - 35B



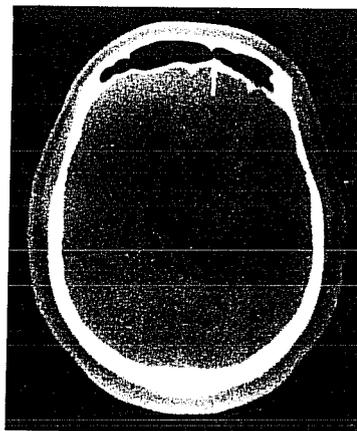
5 - 5B



10 - 4B



10 - 3C

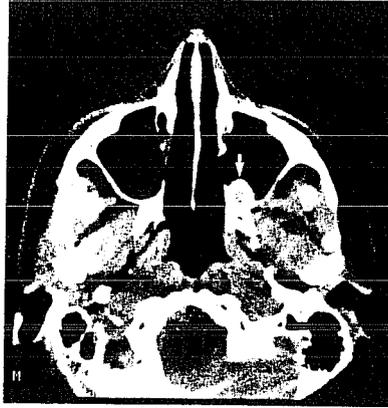


1 - 67

figure 1.17 (suite) - Niveau médian des cornets inférieurs
1-67 : niveau du sinus frontal (flèche = septum)



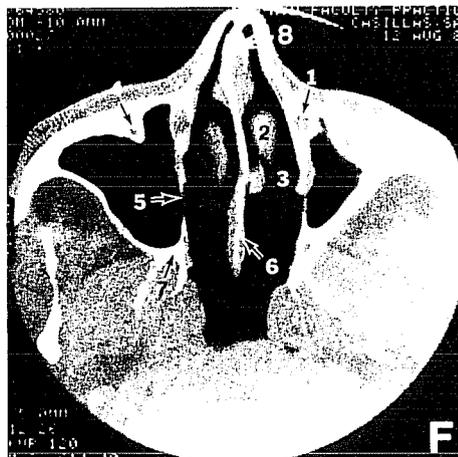
Fig. 1-80. Axial CT scan. Open arrow lies on inferior orbital rim. Apparent bone destruction or soft tissue mass results from curved orbital rim and periorbital soft tissues. Curved white arrow rests on midline meatus and points to middle turbinate. Black arrow points to rostrum of sphenoid and posterior nasal septum.



1 - 110



1 - 255 B



2 - 7F

Figure 1.18 - Coupes tomographiques horizontales : niveau des cornets moyens (in BERGERON & al. 1984)

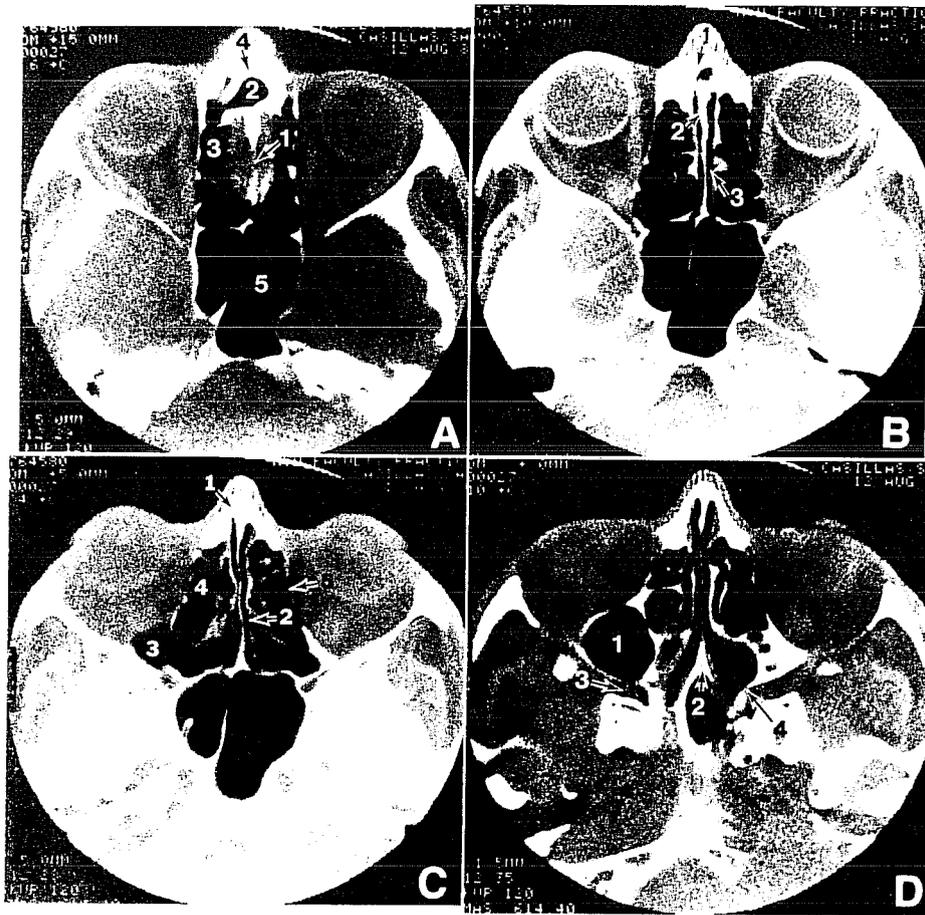


Fig. 2-7. A, Axial CT scan without contrast enhancement at midorbital level. 1, Superior attachment of perpendicular plate of ethmoid; 2, inferior portion of frontal sinus; 3, ethmoid labyrinth; 4, nasal process of frontal bone (glabella); 5, sphenoid sinus. B, Axial CT scan 5 mm below A. 1, Nasal bone; 2, osseous nasal septum (perpendicular plate of ethmoid); 3, sphenoid-ethmoid recess. C, Axial CT scan 5 mm below B. 1, Nasal bone; 2, osseous nasal septum; 3, apex of right maxillary sinus; 4, ethmoid labyrinth; 5, lamina papyracea. D, Axial CT scan 5 mm below C. 1, Maxillary sinus; 2, base of sphenoid with attachment of vomer; 3, pterygopalatine fossa; 4, sphenopalatine foramen. (From Osborn, A.G., and Meff, E.B.: *Head Neck Surg.* 4:182-192, 1982.)

2 - 7 A, B, C, D.

Figure 1.18 (suite) - Niveau des cornets supérieurs et du toit des fosses nasales.

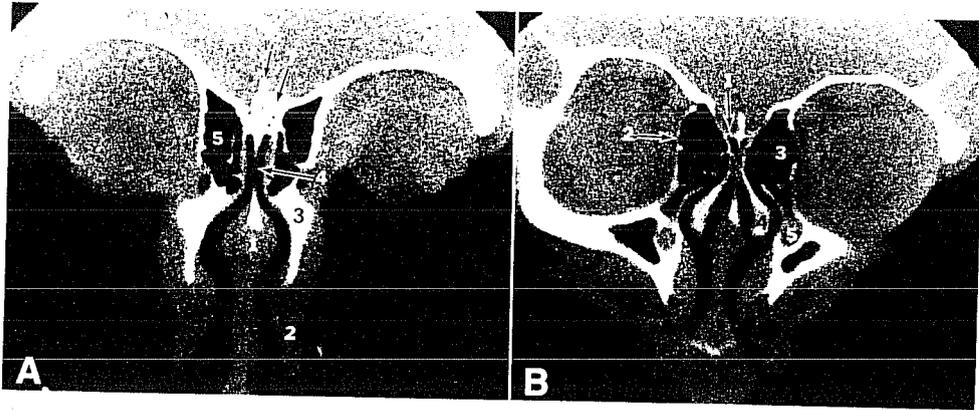


Fig. 2-8. A, Coronal CT scan without contrast enhancement. 1, Cartilaginous nasal septum; 2, lateral crus, greater alar cartilage; 3, frontal process of maxilla; 4, perpendicular lamina of ethmoid; 5, ethmoid sinus; 6, crista galli; 7, olfactory fossa. B, Coronal CT scan 8 mm posterior to A. 1, Cribriform plate; 2, lamina papyracea; 3, ethmoid sinus; 4, middle concha; 5, nasolacrimal duct (obliquely sectioned). (From Osborn, A.G., and McElf, E.B.: *Head Neck Surg.* 4:182-192, 1982.)

Continued

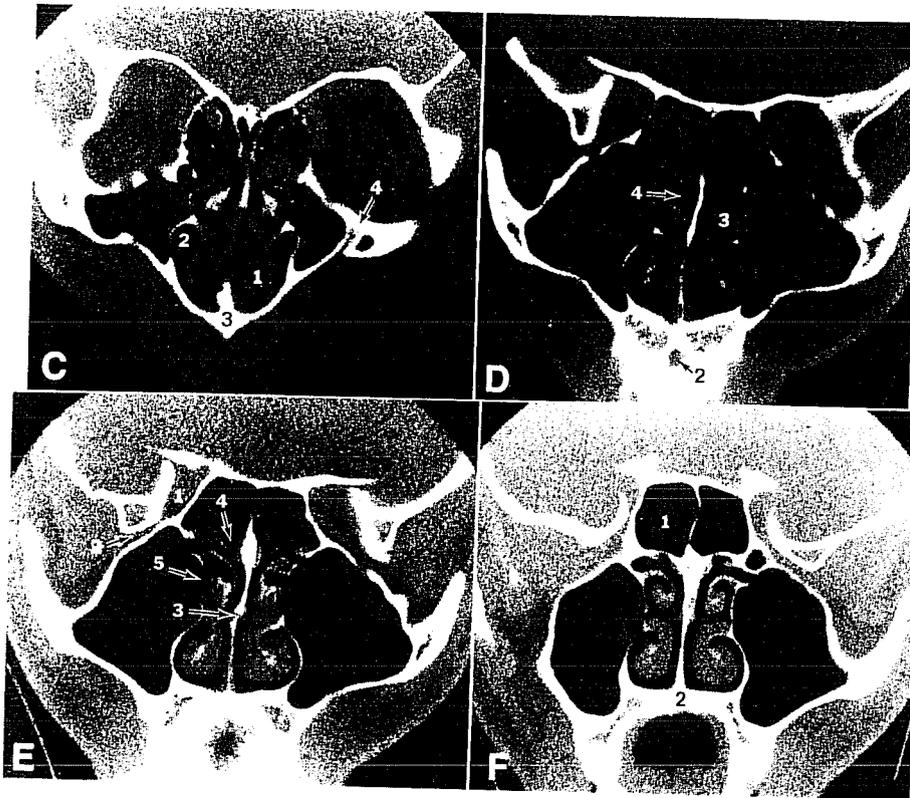


Fig. 2-8, cont'd. C, Coronal CT scan 8 mm posterior to B. 1, Inferior turbinate; 2, inferior meatus; 3, nasal spine of maxilla; 4, inferior orbital fissure (obliquely sectioned). D, Coronal CT scan 10 mm posterior to C. 1, Vomer; 2, incisive canal; 3, middle turbinate; 4, middle meatus. E, Coronal CT scan 6 mm posterior to D. 1, Superior orbital fissure; 2, inferior orbital fissure; 3, suture between ethmoid and vomer; 4, sphenoidal recess; 5, hiatus semilunaris. F, Coronal CT scan 6 mm posterior to E. 1, Sphenoid sinus; 2, hard palate.

Figure 1.19 - Coupes tomographiques frontales : distance des sections par rapport à A : B = + 8mm, C = 16 mm, D = + 26 mm, E = + 32 mm, F = 38 mm. (in BERGERON & al. 1984).

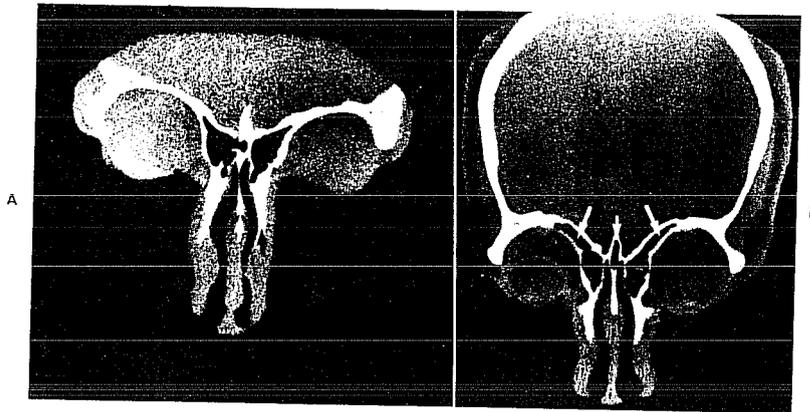


Fig. 1-86. A, Coronal CT scan. Small black arrow points to crista galli. Long black arrow indicates frontal process of maxilla. Anterior ethmoid cells are clearly seen, as is cartilaginous nasal septum and perpendicular plate of ethmoid bone (*white arrow*). B, Coronal CT scan. Long white arrows indicate supraorbital ethmoid cells as they extend up into orbital roofs. Short white arrow points to pneumatization of crista galli.

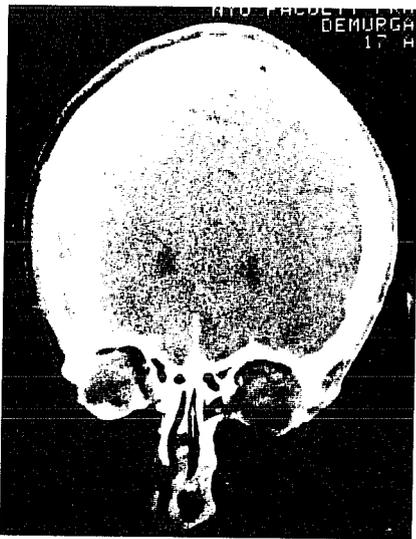


Fig. 1-87. Coronal CT scan. Curved black arrow indicates frontal process of maxilla. Long white arrow indicates crista galli. Thin white arrow lies in olfactory recess of nasal cavity.



Figure 1.19 (suite) - Coupes tomographiques frontales à la jonction du nez et des fosses nasales.

pas important, la section antérieure est relativement plus large que le passage entre la cloison et les cornets moyens. Le groupe de 4 clichés à droite en séquence descendante montre l'étroit sommet des fosses nasales entre les cellules ethmoïdes (B : flèche 3, C : flèche 2). le chiffre 2 de D se trouve dans le sinus sphénoïde. Le niveau est sous doute celui des cornets supérieurs.

Les clichés de la figure 1.19 sont en coupe frontale. Pour le même sujet (A à F), on parcourt les fosses à partir du nez. en A, on note l'épaississement de la jonction du septum cartilagineux et de la plaque perpendiculaire de l'éthmoïde, et la bosse interne de la paroi latérale des cartilages du nez. Il n'est pas certain que l'inclinaison soit la même que le plan de coupe de BJUGGREN & FANT. Elle semble légèrement inclinée d'arrière en avant. En B, notez l'élargissement de la cloison à sa jonction avec le maxillaire. De C à F, on distingue avec beaucoup de clarté les cornets inférieurs et moyens, et les sinus maxillaires. Remarquons surtout l'étroitesse apparente des passages et méats. Les deux clichés inférieurs et ceux de droite montrent la jonction entre nez et fosses. Le plan semble presque parallèle à la pente de l'arête du nez. On voit une constriction nette pour certains sujets au tiers inférieur du vestibule.

Les coupes de BJUGGREN et FANT (1964) sont reproduites à la figure 1.20. La fonction d'aire est donnée plus loin , à la figure 1.65. La disposition, qui alterne la partie à droite et à gauche de la coupe, n'est pas très parlante. Nous avons grossièrement reproduit la séquence anatomique en regard. Il n'est pas possible de comparer quantitativement les clichés tomographiques frontaux aux sections de FANT en l'absence d'une échelle. Une simple comparaison visuelle révèle cependant que les coupes 11 ou 12 ne montrent guère la constriction notée pour certains sujets au tiers inférieur du vestibule. De plus, la largeur des méats et des passages entre la cloison et les cornets semble trop forte. En revanche on peut tenter une estimation plus quantitative sur les tomographies horizontales. La procédure est la suivante : de la mesure de la surface d'une section (64 mm^2 pour la somme des aires de la coupe 6 par exemple), nous déduisons le facteur de réduction linéaire (2.7 : racine carrée du rapport aire mesurée / aire réelle) de la figure de FANT, et donc l'échelle absolue. Sur les tomographies horizontales, nous pouvons presque toujours mesurer soit la distance entre les condyles ou les apophyses coronoides de la mandibule, soit entre la partie supérieure du corps (ramus). Nous avons mesuré cette distance sur plusieurs sujets et sur des crânes. Cette mesure (14.5 cm et 13 cm en moyenne) nous permet une approximation grossière de l'échelle de la tomographie. Le problème étant celui de la taille de la zone antérieure (coupes 15 à 11), nous avons mesuré les tomographies dans cette zone à plusieurs niveaux. La largeur maximale chez FANT pour la section 12 (position juste antérieure aux cornets inférieurs) est de 14 mm en faisant la somme de la partie gauche et droite. Voici un échantillon de largeurs mesurées sur les tomographies

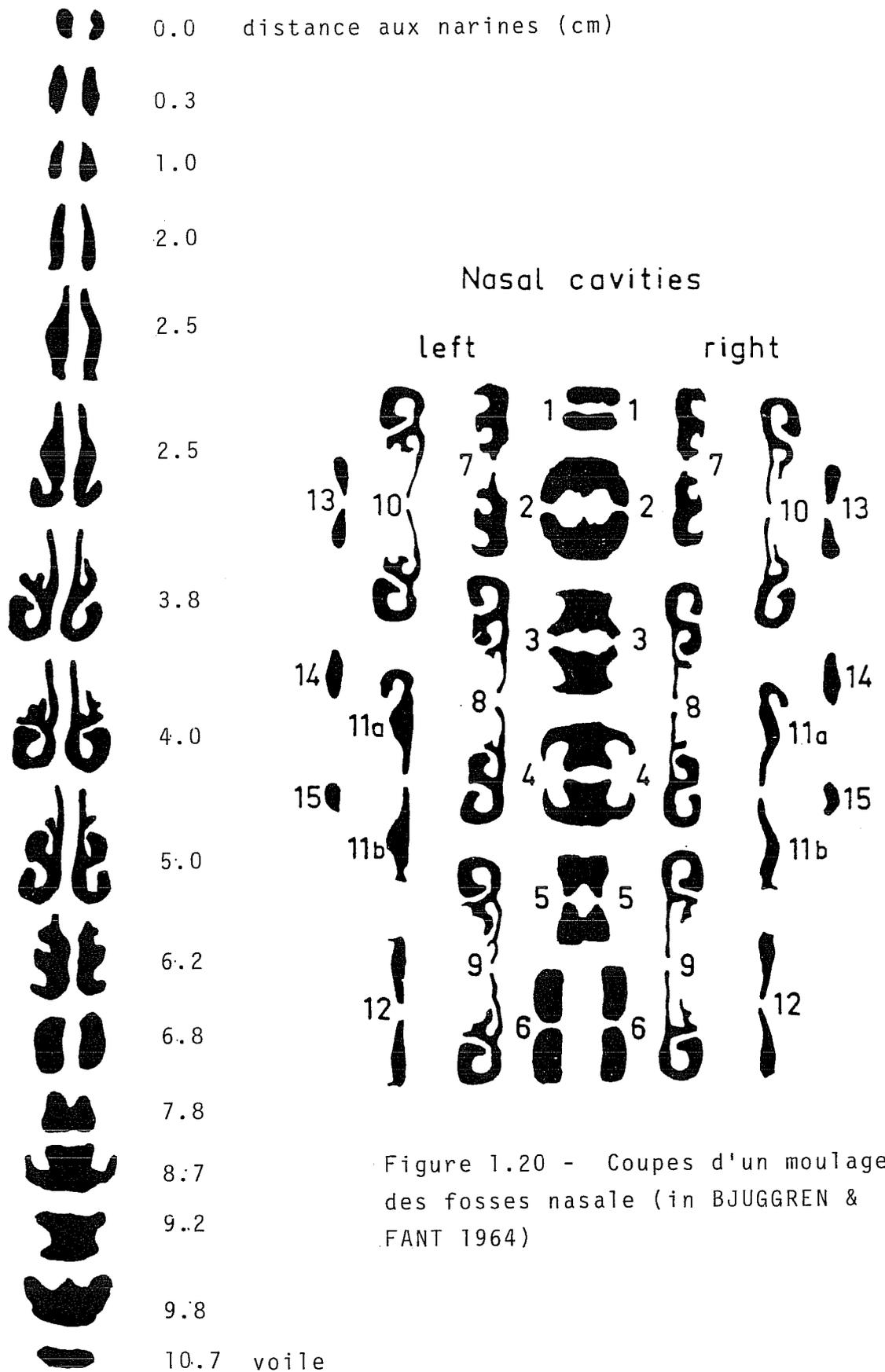


Figure 1.20 - Coupes d'un moulage des fosses nasale (in BJUGGREN & FANT 1964)

(BERGERON & al.) à l'endroit le plus étroit :

figure	largeur gauche	largeur droite	somme (mm)
1.82	4.5	4.5	9
1.101B	9	4.5	13.5
1.110	5.5	6.5	12
1.124B	7.5	6	13.5
13.13B	8.5	4	12.5
16.145C	7	4	11
10.4B	4	4	8

Il apparaît clairement que l'ordre de grandeur est bien celui de FANT et que la suggestion de FENG & al. (1985) d'une réduction des 2/3 est peu vraisemblable. De plus, il est évident que l'aire du vestibule nasal, bien visible de l'extérieur, est supérieure à 1cm^2 . Une erreur importante sur la hauteur du passage est improbable. Mais il n'est pas exclu que les valeurs moyennes du limen nasi soient légèrement inférieures. Une aire trop faible (0.3cm^2 pour chacun des conduits, d'où un diamètre de 0.6cm) ne permettrait pas d'insérer un fibroscope dans le nez, ce qui n'est que rarement le cas pour des appareils de taille supérieure. Mais il faut insister sur la possibilité d'une forte diminution par gonflement des muqueuses dans les cas d'infection (coryza). C'est l'expérience commune qu'une ou les deux narines peuvent être bouchées par le gonflement et l'excès de mucus.

Notons enfin que la longueur du conduit nasal ne peut être déduite des tomographies, celle-ci devant être mesurée le long des "centres de gravité" des aires et non comme une corde.

Nous retiendrons de ce long paragraphe que la taille de la zone narinaire est égale ou un peu inférieure aux valeurs de BJUGGREN & FANT (1964). Dans les paragraphes suivants, nous nous tournons vers les renseignements acoustiques que l'on peut tirer des vues sagittales du conduit oral.

4 - ARTICULATION DES VOYELLES NASALES

Ce paragraphe est consacré à une revue de nos connaissances sur l'articulation linguale des voyelles nasales du français. Il s'appuie principalement sur l'excellent travail de J.P. ZERLING (1984) et le riche recueil récent de BOTHOREL & al. (1986). Il a également pour objet de prévoir qualitativement la position des formants des voyelles nasales par

rapport à ceux de la voyelle orale d'articulation la plus proche. ZERLING (1984) compare, pour deux sujets, dont l'auteur de cette thèse, l'articulation linguale des voyelles [a - ā] , [oe - õe], [ɔ - õ] et [ɛ - ẽ] dans des contextes consonantiques identiques quant au lieu d'articulation. Les faits saillants, visibles aux figures 1.21 et 1.22, peuvent se résumer ainsi :

- [ɛ] / [ẽ] On note un recul horizontal de la langue pour les deux sujets et un accroissement de la labialisation pour un sujet.
- [oe] / [õe] Pour un sujet, la position linguale est celle de [ẽ]. Il y a donc neutralisation articulatoire de l'opposition. Pour l'autre sujet, [õe] est plus labialisé que [oe].
- [a] / [ā] Après avoir vérifié, sur d'autres tracés, l'absence de différence entre [a] et [ɑ] pour les deux sujets, ZERLING note une forte labialisation pour [ā] , et un recul de la langue pour l'un d'entre eux.
- [ɔ] / [õ] Un recul de la langue est évident, avec un abaissement de la partie antérieure de la langue pour l'un, ce qui suggère une articulation proche de [o] . Il y a de plus une surlabialisation de [õ] par rapport à [ɔ] et [ā] .

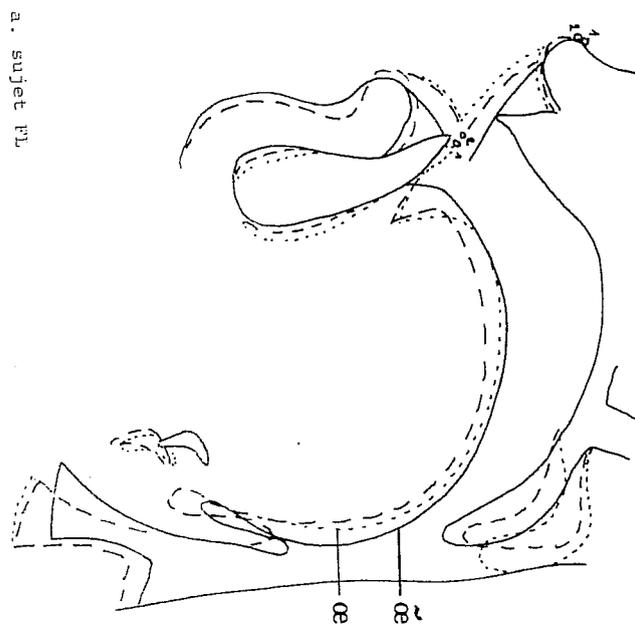
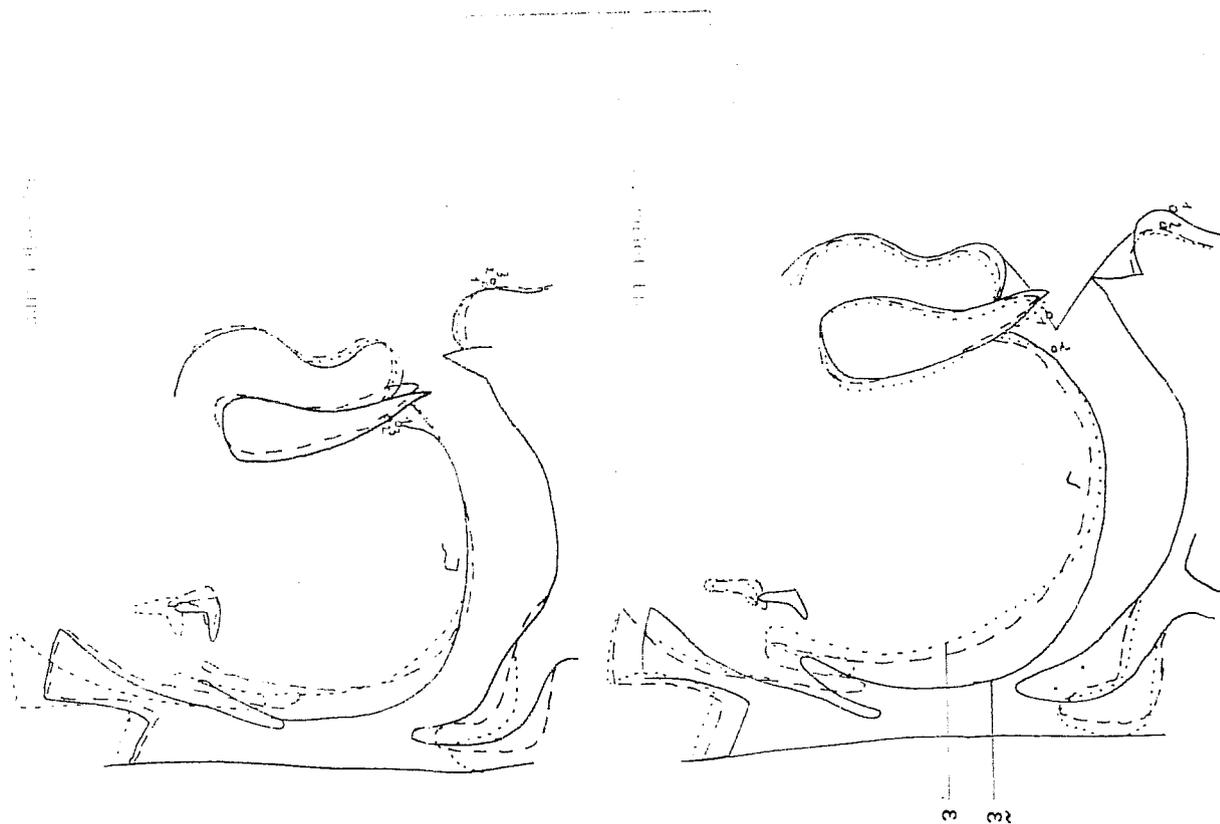
Au plan de la labialisation, les deux sujets emploient des stratégies différentes : F.L. réalise une opposition ternaire : [ɛ] : -labial ; [ẽ] - [õe] : 0 labial ; [oe] : + labial

J.P.Z. conserve l'opposition binaire traditionnelle : [ɛ] - [ẽ] : - labial ; [oe] - [õe] : + labial.

Il est clair également que la conclusion de DELATTRE (1968 par ex.) quant à l'identité de la forme du pharynx pour les voyelles nasales est infirmée. Enfin, ZERLING met en évidence (figure 1.23) que les articulations linguales de [ā] et [õ] sont très voisines, ne différant que par la surlabialisation de [õ] , tout comme celles de [ẽ] et [õe]. Les deux couples s'opposent par le caractère / antérieur - postérieur /. On trouve une bonne confirmation de ce fait chez les deux sujets de BOTHOREL & al. (1986) pour lesquels les contextes permettent une comparaison (figure 1.24).

Connaissant les différences articulatoires entre les voyelles orales et nasales proches, on peut prévoir qualitativement la variation des formants dépendant du conduit oral. Un recul de la langue implique un abaissement de F2, et un accroissement de la labialisation provoque une chute de F1. Cette analyse néglige le rôle de l'abaissement du voile qui crée un resserrement dans la zone vélaire du conduit. Ce point sera développé en détail au paragraphe 9.2. Ces modifications acoustiques sont présentées à la figure 1.25. Elles sont conformes aux résultats des analyses acoustiques. Nous disposons par exemple de statistiques d'analyse par prédiction linéaire pour 40 [ɛ] et [ẽ] prononcées par l'auteur dans deux conditions : voix "forte" et "faible". Le but de cette expérience était d'estimer l'influence du "formant" glottal (cf. § 7) sur les résultats L.P.C. Les détails ne nous concernent pas ici, d'autant que les résultats ne sont pas clairs de ce point de vue. Les

FIG. 2 : COMPARAISON DES VOYELLES [ɛ], [ɛ̃], [œ], [œ̃]



a. sujet PL



b. sujet JZ

FIG. 2 : COMPARAISON DES VOYELLES [ɛ], [ɛ̃], [œ], [œ̃]

Figure 1.21 - Positions articulatoires des voyelles [ɛ], [ɛ̃], [œ] et [œ̃] (in ZERLING 1984)

FIG. 3 : COMPARAISON DES VOYELLES [a], [ã], [ã]

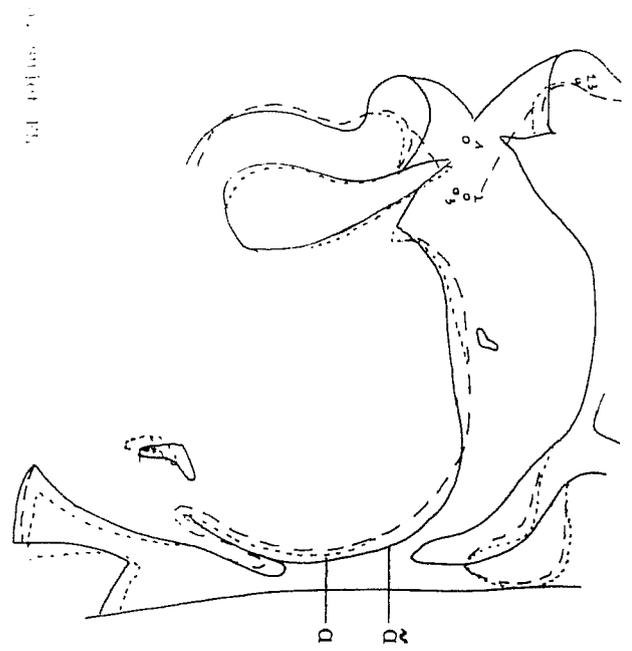
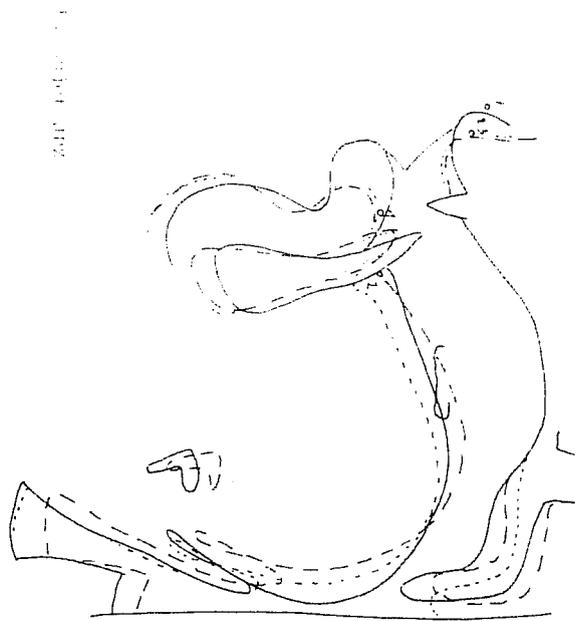


FIG. 4 : COMPARAISON DES VOYELLES [ã], [ã], [ã]

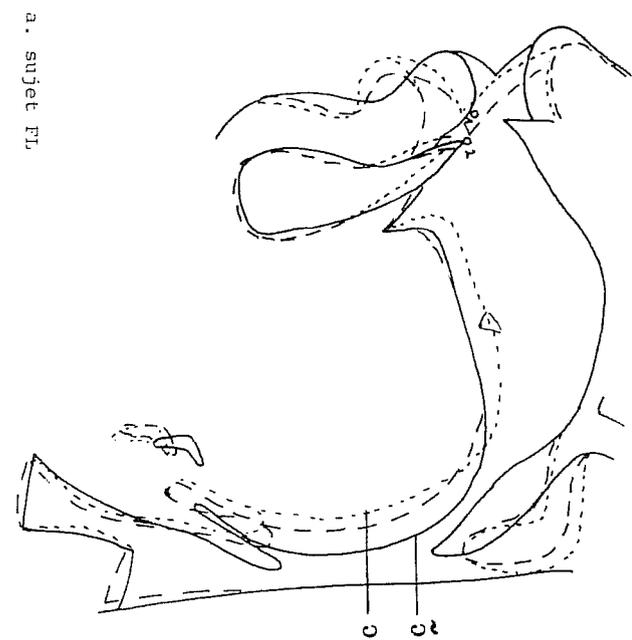
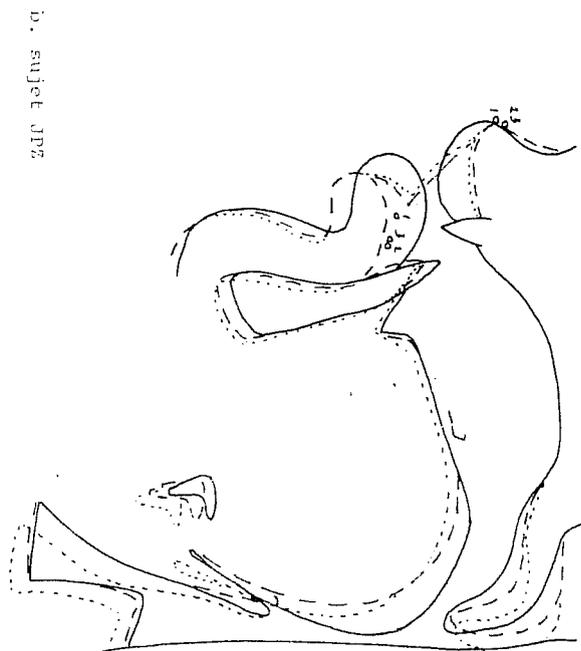


Figure 1.22 - Positions articulatoires des voyelles [a], [ã] [o] et [õ] (in ZERLING 1984).

FIG. 5 : COMPARAISON DES VOYELLES NASALES [ɛ̃] ET [œ̃]

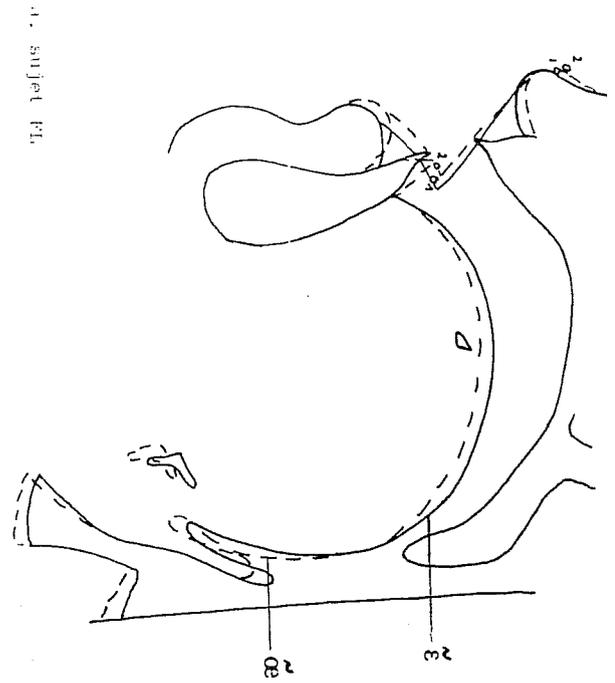


FIG. 5 : COMPARAISON DES VOYELLES NASALES [ɛ̃] et [œ̃]

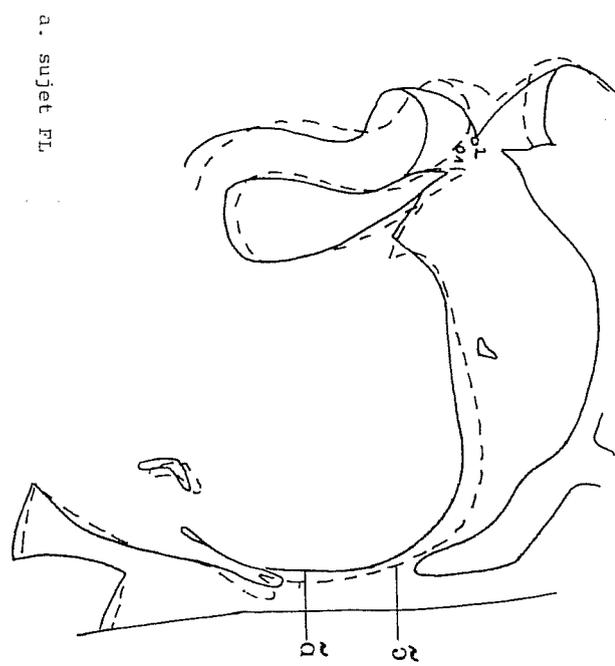


Figure 1.23 - Comparaisons des positions articulatoires des voyelles [ɛ̃] - [œ̃], et [ã] - [õ] (in ZERLING 1984).

VOYELLES ã ET õ

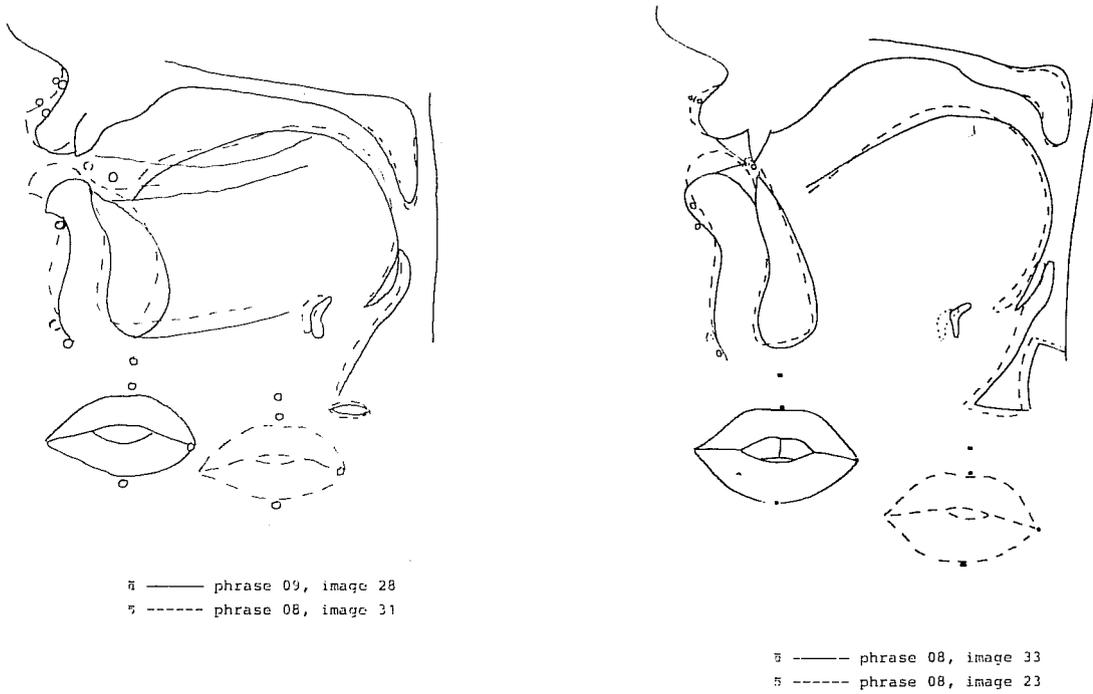


Figure 1.24 - Comparaisons des positions articuloires des voyelles |ã| et |õ| (in BOTHOREL & al. 1986).

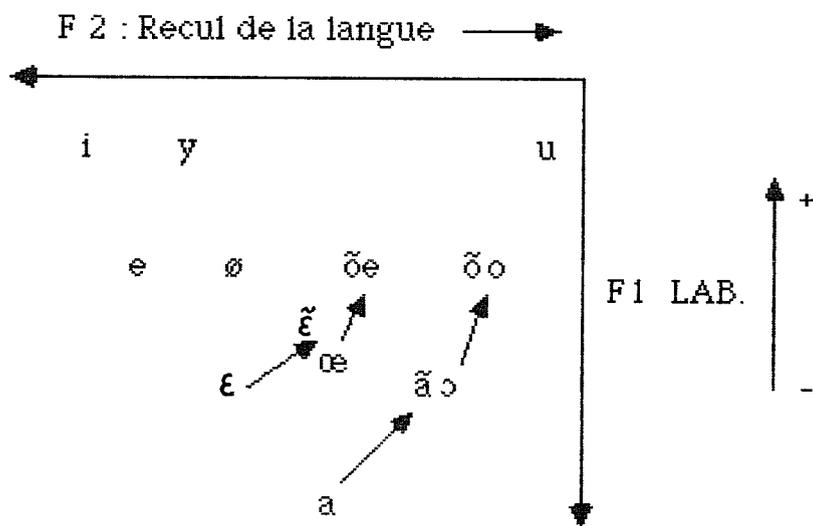


Figure 1.25 - Evolution schématique des deux premiers formants en fonction des différences articuloires entre voyelles orale et nasale homologues

moyennes (et écarts-type pour les orales) des fréquences formantiques sont les suivantes :

[ɛ] faible	495 (13)	1676 (21)	2532 (31)
[ɛ] fort	526 (19)	1706 (26)	2531 (32)
[ɛ̃] faible	520	1525	(2600 ?)
[ɛ̃] fort	604	1505	2607

Les prédictions tirées de l'articulation sont confirmées pour F2, où la différence est d'environ 150 - 200 Hz. La mesure de F1 pose des problèmes qui seront longuement débattus au paragraphe 10. Les spectrogrammes permettent une illustration visuelle parlante (figure 1.26). Pour [ɛ̃] en hiatus (a), on voit clairement la chute de F2, et la présence de deux formants en dessous qui interdit "la" mesure de F1. La comparaison entre [aã] et [ã] (b, c) est instructive. Les formants de [ã] sont bien plus proches de [ɔ] que de [a]. On peut montrer de la même manière que [õ] est plus proche de [o] que de [ɔ]. Comme l'ont noté LONCHAMP (1978) et ZERLING (1984) après d'autres, les symboles "classiques" de l' A.P.I. ne correspondent pas à la réalité articulatoire ou acoustique. Mais à l'inverse de ce que nous avons fait auparavant (LONCHAMP 1978), nous ne proposerons pas de nouveaux symboles. Il est impossible de choisir pour chaque nasale un symbole reflétant à la fois ce que nous savons de son acoustique et de son articulation. Pour [ɔ̃] et [õ] par exemple, l'articulation suggère un symbole unique, avec un diacritique précisant pour [ɔ̃] la surlabialisation, alors que l'acoustique s'accommoderait mieux des symboles [ɔ̃] et [õ].

5 - ACOUSTIQUE THEORIQUE DE LA NASALITE ET MODELES DE SIMULATION

5.1 - ANALYSE GRAPHIQUE PAR LE SCHEMA DE REACTANCE

L'analyse du spectre des sons nasalisés, et en particulier des voyelles nasales, est facilitée par une construction graphique fondée sur l'évolution de la réactance, en fonction de la fréquence, au point de couplage des conduits oral et nasal. Cette construction a été proposée par FANT (1960) et utilisée également par FUJIMURA (1960).

On sait que les fréquences de résonance d'un système physique, appelées formants dans notre domaine, se produisent lorsque l'impédance d'entrée (Z_i) est maximale, c'est-à-dire infinie quand on suppose les pertes nulles. Dans ce cas, l'impédance se réduit à la

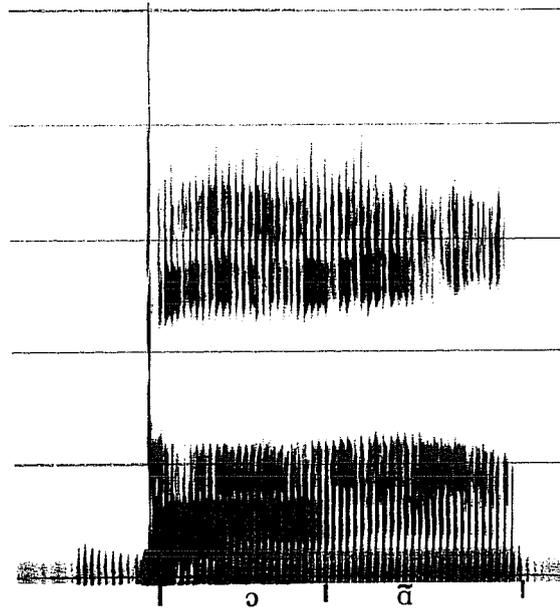
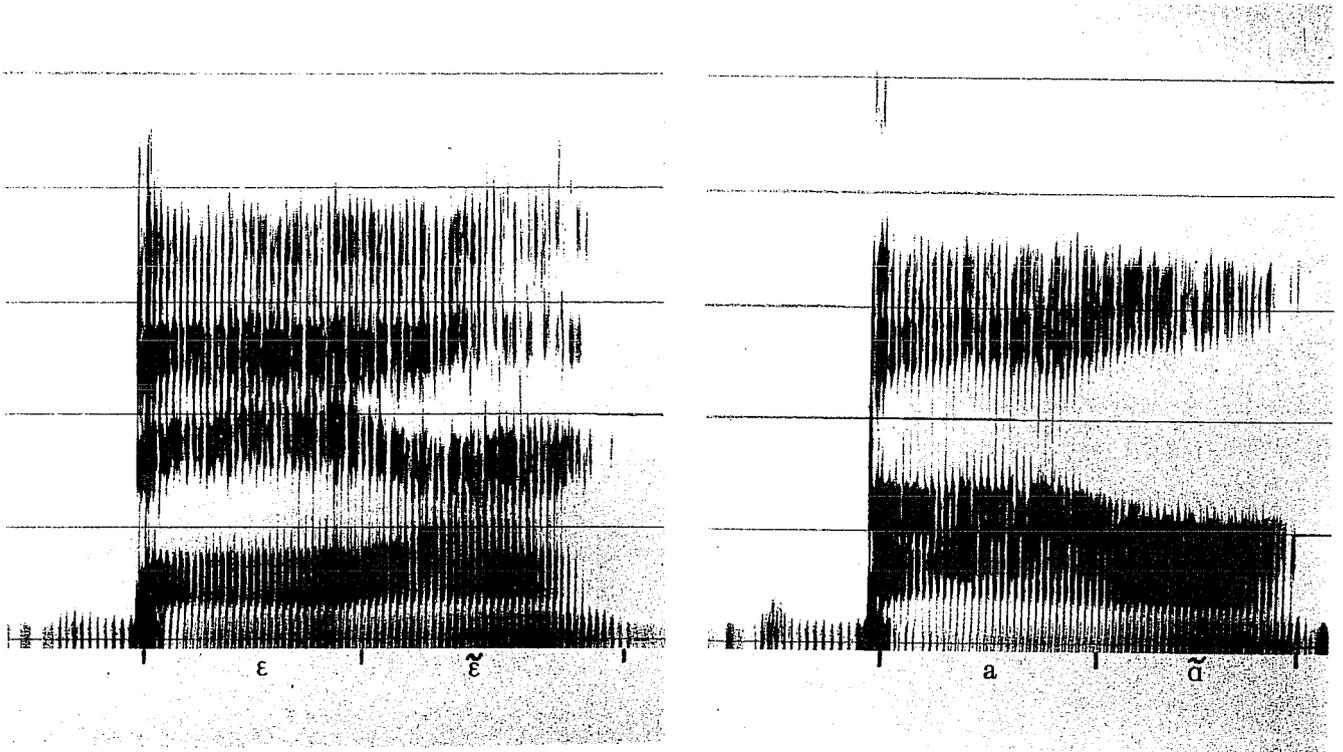


Figure 1.26 - Spectrogrammes des voyelles $[\epsilon\tilde{\epsilon}]$, $[a\tilde{a}]$ et $[o\tilde{o}]$ en hiatus.

réactance (notée ici $X(w)$). Les fréquences de résonances se produiront donc pour les fréquences (appelées pôles) où $X(w) = \infty$ ou $1/x(w) = 1/\infty = 0$. Pôles et fréquences de résonance coïncident par définition quand les pertes (largeurs de bande) sont nulles. Pour les pertes faibles présentes en parole, les valeurs numériques ne diffèrent que d'une dizaine de Hz au plus. Si l'on considère tout d'abord le système cavité buccale et cavité pharyngale en se plaçant au point de couplage (zone du voile), les deux cavités sont branchées en parallèle. L'impédance du conduit oral, notée $X_o(w)$, est donc donnée par :

$$1/X_o(w) = 1/X_b(w) + 1/X_p(w)$$

$X_b(w)$ et $X_p(w)$ notant la réactance des conduits buccal et pharyngal. La théorie des circuits électriques appliquée à la propagation des ondes dans un tuyau montre que l'évolution de $X_b(w)$ a la forme d'une fonction tangente (tuyau ouvert à l'extrémité) et celle de $X_p(w)$ une fonction cotangente (conduit fermé possédant une première résonance à 0 Hz). La valeur de la réactance passe de $+\infty$ à $-\infty$ aux fréquences de résonance (figure 1.27). La réactance du conduit oral $X_o(w)$ complet a la forme indiquée à la figure 1.27.

Le conduit nasal est connecté en parallèle au système oral au même point de couplage. La réactance du conduit nasal seul (c'est-à-dire sans couplage) est noté $X_n(w)$ et a la même forme que celle du conduit buccal (tuyau ouvert à une extrémité). Dans nos figures, l'allure de l'évolution de la réactance pour un couplage moyen a été "empruntée" à MRAYATI (1975) (elle n'a pas fait l'objet de calcul dans ce travail). Il y aura donc résonance lorsque la réactance totale (orale + nasale) sera nulle, ce qui correspond aux cas où

$$1/X_o(w) + 1/X_n(w) = 0, \text{ c'est-à-dire } 1/X_o(w) = -1/X_n(w) \text{ ou encore } X_o(w) = -X_n(w).$$

L'astuce graphique consiste à tracer sur le schéma de la figure 1. 27 l'évolution de l'inverse de $X_n(w)$. Graphiquement, les résonances se produiront pour les fréquences où les courbes $X_o(w)$ et $X_n(w)$ se croisent, donc sont égales et de signe opposé (figure 1.28). Notons que l'on peut également représenter graphiquement les susceptances, c'est-à dire la partie imaginaire de l'inverse de l'impédance, ou admittance (FUJIMURA & LINDQVIST 1971, HAMADA 1983). Cette représentation est plus claire quand on veut reporter plusieurs degrés de nasalisation sur un même schéma.

Les fréquences pour lesquelles $X_o(w)$ est infini correspondent aux fréquences des formants (F_1, F_2, \dots) du conduit vocal sans couplage. Les fréquences pour lesquelles $X_n(w)$ est infini sont les pôles (formants) propres au conduit nasal seul. Aux points d'intersection (marqués d'un cercle noir à la figure 1.28) on obtient les fréquences de résonance (formants) du conduit couplé. Des antirésonances ou zéros se produisent lorsque la réactance est nulle, dans le cas sans pertes. Nous les noterons généralement A_i , et quelquefois Z_i dans la discussion de certains travaux, pour conserver la notation de l'auteur. La pression aux lèvres, produit du débit et d'une impédance nulle, est nulle. Le

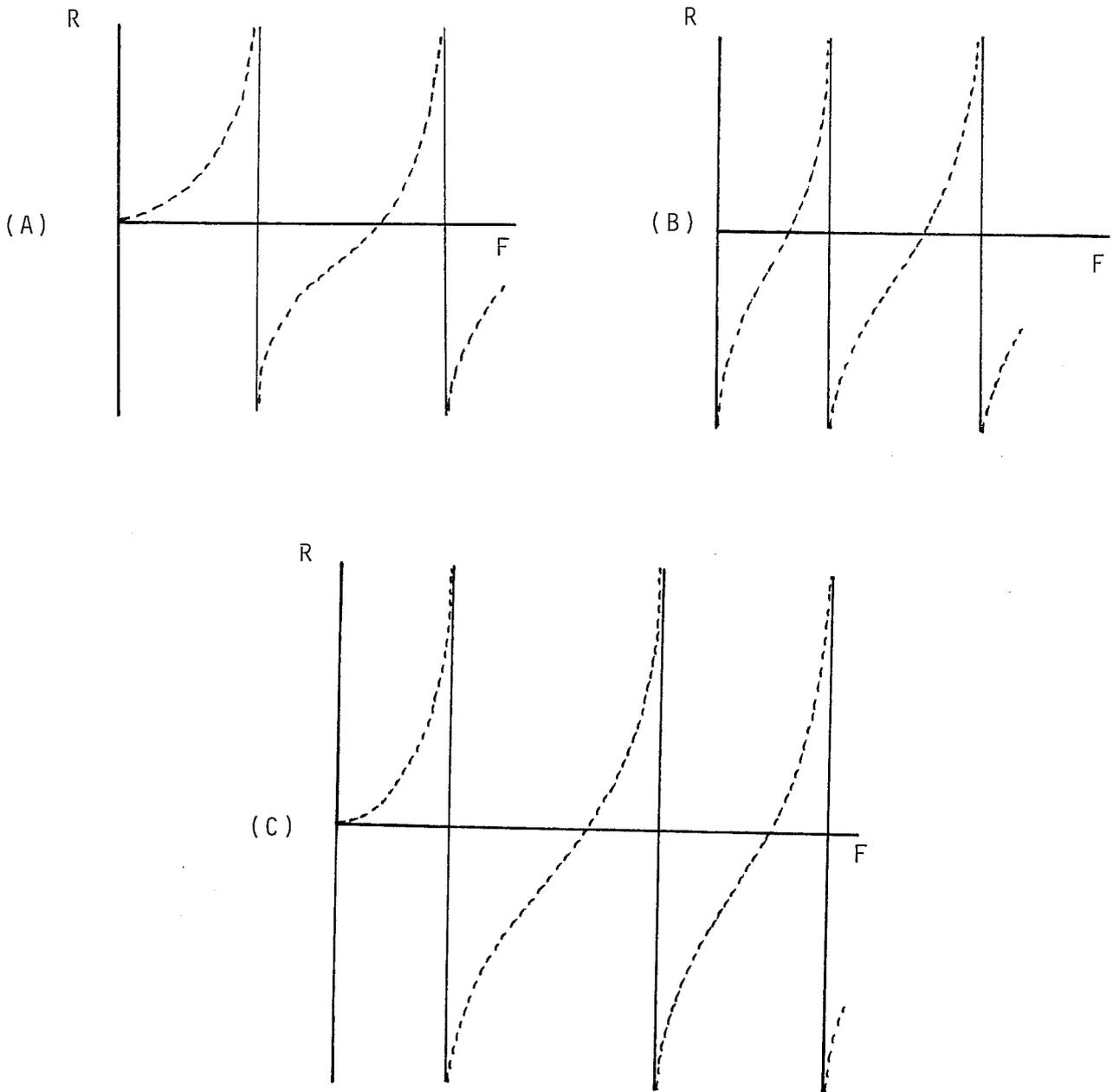


Figure 1.27 - Evolution schématique de la réactance R pour le conduit buccal ou le conduit nasal (A), le conduit pharyngal (B), et le conduit vocal complet (C).

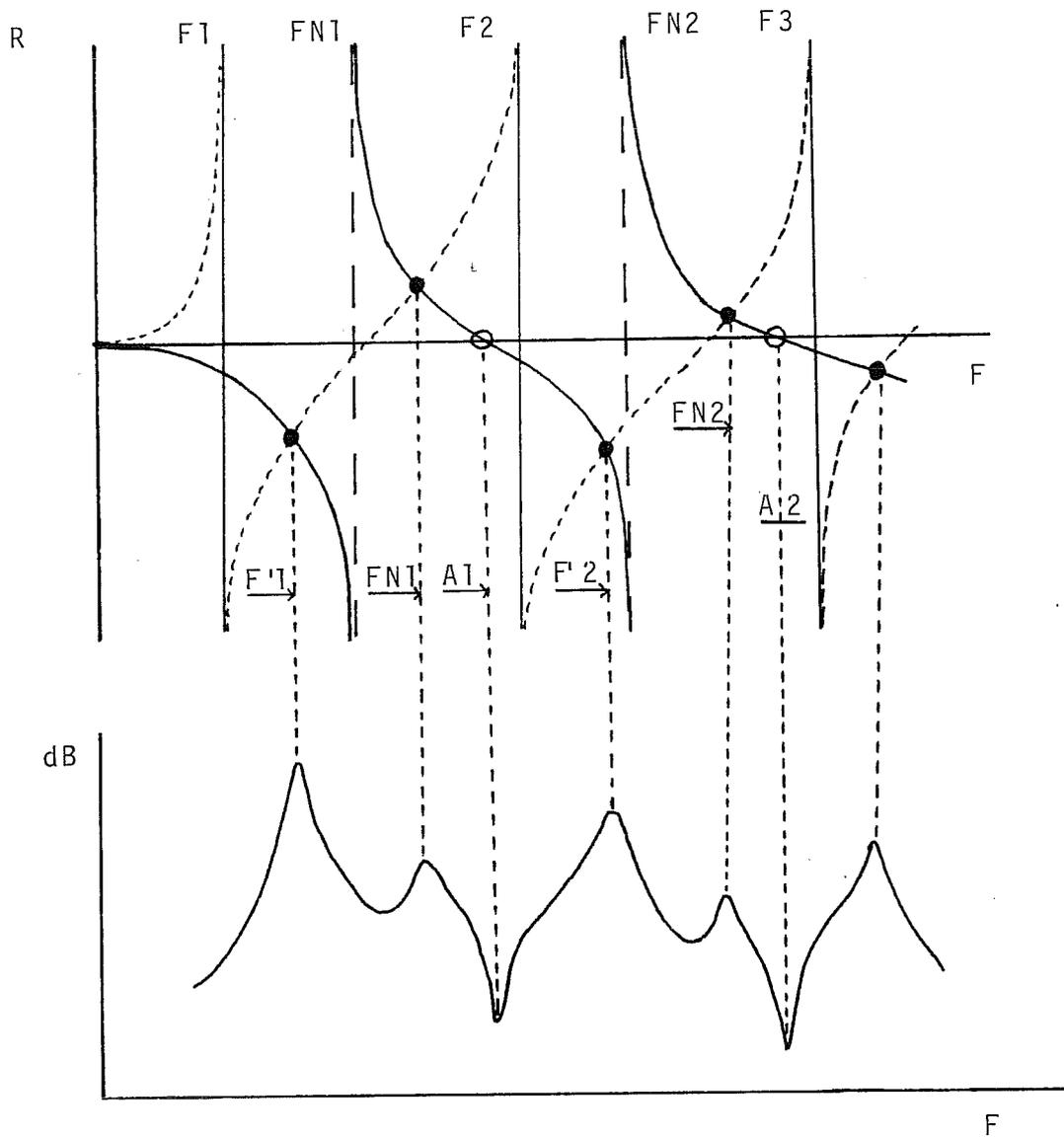


Figure 1.28 - Prédiction (schématique) du spectre d'une voyelle nasale à partir du schéma de réactance.

couplage provoque deux ensembles de zéros à des fréquences différentes, l'un présent dans le spectre issu de l'extrémité du conduit oral, et l'autre à l'extrémité du conduit nasal. Le son recueilli par un microphone en position normale est un mélange des sources orales et nasales (MERLIER 1984). Si les maxima d'énergie s'additionnent en première approximation, il n'en va pas de même pour les zéros. L'énergie rayonnée par le conduit buccal étant supérieure, un zéro nasal, qui est un déficit d'énergie, est souvent comblé par l'énergie rayonnée par le conduit buccal. En revanche, un zéro buccal a moins tendance à être masqué par de l'énergie nasale. En pratique donc, on ne s'intéressera qu'aux zéros provoqués par le couplage nasal dans le spectre émis par la bouche, qui sont marqués d'un cercle blanc à la figure 1.28.

Sous le schéma de réactance, on a placé l'allure du spectre rayonné en identifiant les formants liés à la cavité orale (notés F_i) et les formants de nasalité (notés FN_i). Notons enfin que dans une paire pôle / zéro, le zéro peut se placer avant ou après le pôle selon la fréquence des autres maxima de réactance. Les modifications ("shifts") fréquentielles dues au couplage et l'ordre d'apparition des pôles et zéros feront l'objet des paragraphes suivants. La notation F_i souligne l'affiliation avec le formant homologue F_i de la voyelle non nasalisée, ainsi que le déplacement fréquentiel.

5.2 - DEPLACEMENTS FREQUENTIELS DUS AU COUPLAGE NASAL.

On remarque sur la figure 1.28 que les fréquences des formants de nasalité et des formants oraux sont décalés vers le haut. L'ampleur du décalage dépend à la fois de la séparation fréquentielle entre deux résonances successives et de la pente de la réactance. On ne peut donner qu'une règle très générale. Plus un formant oral et un formant sont lointains, plus le décalage fréquentiel du premier est important. La forme exacte de l'évolution de la réactance dépend de la valeur du couplage, et sera discutée plus loin au paragraphe 8.5. Ce décalage est, avec la présence de pics supplémentaires et l'introduction de zéros, un indice essentiel de la présence d'un couplage.

5.3 - ORDRE DES POLES ET ZEROS

FUJIMURA & LINDQVIST (1971) présentent un résumé particulièrement clair du comportement des pôles et zéros, déjà décrit par FANT (1960) et FUJIMURA (1960). Pour un couplage très faible, le pôle nasal et le zéro coïncident presque en fréquence. Le cas limite de l'absence de couplage peut donc être considéré comme l'annihilation des formants de nasalité par leur zéro. Quant le couplage augmente, la fréquence des zéros croît

plus vite que les fréquences des formants oraux et nasaux. Un zéro peut donc "rattraper" un formant oral ou nasal placé au-dessus de lui et le détruire (figure 1.29 d'après FUJIMURA 1960). Un formant nasal peut "dépasser" un formant oral et réciproquement, mais cela ne peut se produire pour deux formants du même type (oral ou nasal). La règle la plus importante concerne l'ordre d'apparition des formants en basse fréquence : le pic le plus bas peut être un formant oral décalé ($F'i$) ou un formant nasal (FNi). Si la fréquence du premier formant du conduit oral seul est supérieure à un certain seuil, le pic le plus bas sera le premier formant nasal. Ce seuil est la fréquence de la première résonance du conduit nasal seul, fermé au voile (fréquence critique ou caractéristique : p. 552). La fréquence du premier pic des conduits couplés est toujours comprise entre les fréquences des premiers formants des conduits oral et nasal non couplés. Sa plage de variation est donc limitée. Cette discussion ne tient pas compte de l'existence potentielle de résonances de basse fréquence liées aux cavités sinusales. Un dernier point, qui à ma connaissance, n'a jamais fait l'objet de remarques, vaut d'être mentionné. Dans un paire pôle / zéro, l'ordre d'apparition des deux éléments n'est pas fixe. Deux résultats de mesure sur maquette de MERLIER (1984) illustrent parfaitement ce point (figure 1.30). Pour le couplage faible, on note un zéro suivi d'un pôle entre $F1$ et $F2$ (a), alors qu'à un degré de couplage plus élevé, l'ordre est inversé : le pôle précède le zéro (b). Le schéma de réactance de la figure 1.31 présente l'analyse théorique : le zéro précèdera le pôle quand le formant nasal est placé juste au-dessus d'un formant oral. Dans cette position, la pente des réactances est forte, et la séparation pôle / zéro ne sera jamais importante. Si le formant nasal est à mi-chemin de deux formants oraux, la pente faible des réactances provoque un plus large écart du pôle et du zéro, qui seront plus visibles sur le spectre.

5.4 - LE SINUS COMME RESONATEUR DE HELMHOLTZ

En raison de sa forme de cavité reliée au conduit principal par une ouverture de faible diamètre, un sinus satisfait aux conditions de calcul de sa fréquence de résonance par l'équation de HELMHOLTZ. L'équation qui s'applique dans notre cas est celle d'un résonateur fermé en dérivation sur un conduit, et non comme le fait FENG (1986) celle d'un résonateur ouvert à l'air libre :

$$F = C / 2 \pi * \sqrt{(S / (L + 1.7 A) * V)} \quad (\text{KINSLER \& FREY (1962) p. 205 par ex.})$$

Les paramètres sont introduits à la figure 99.

Pour le sinus utilisé par MAEDA (1982 a) où $V = 20.8 \text{ cm}^3$, $S = 0.1 \text{ cm}$, $L = 0.5 \text{ cm}$, la fréquence de résonance F vaut 431 Hz, corrigée à 467 Hz si l'on admet que la correction classique de vibration des parois s'applique dans ce cas, et non 507 (corrigée à 538) Hz

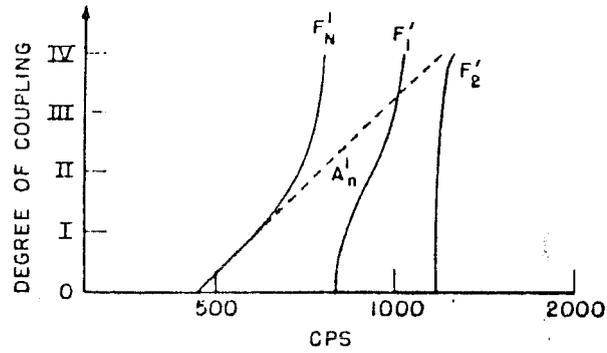
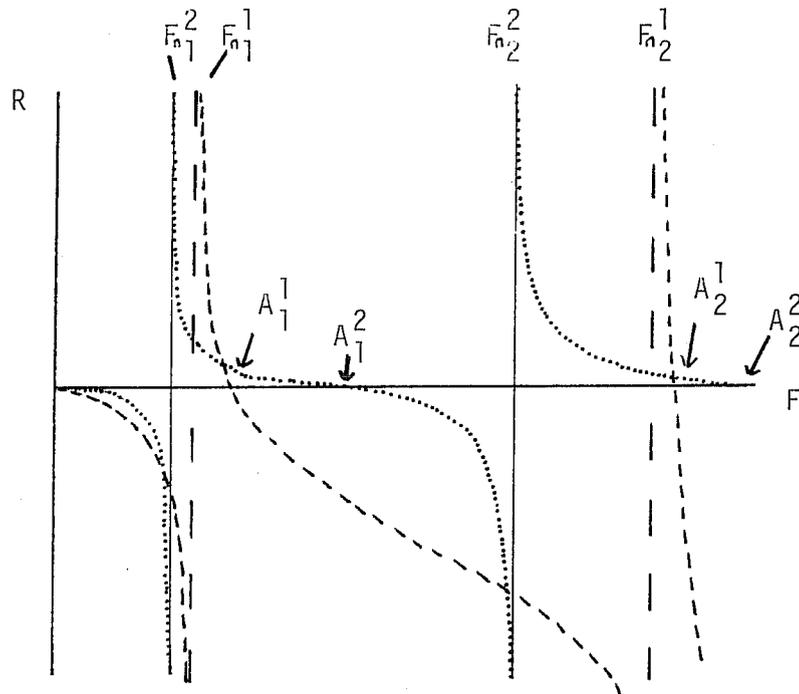


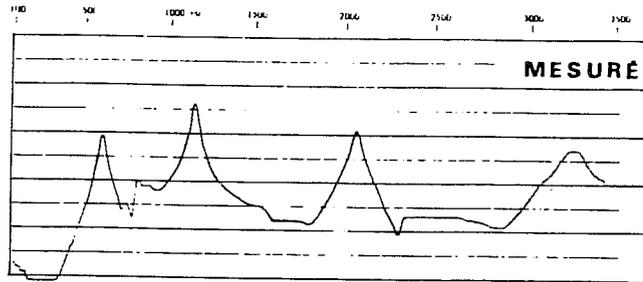
Figure 1.29 A - Evolution de la fréquence de F_1' , F_2' , F_{N1} et du premier zéro (A_{n1}) en fonction du degré de couplage. (in FUJIMURA 1960)



1 : Couplage faible (-----)

2 : Couplage fort (.....)

Figure 1.29 B - Evolution schématique de l'inverse de la réactance du conduit nasal seul en fonction du degré de couplage (adapté de MRAYATI 1975).



PRESSION ORALE POUR LA MAQUETTE DU CONDUIT VOCAL (A)
avec couplage nasal 0,5 cm²

Figure 1.30 A - in MERLIER (1984) : Mesure sur une maquette de la pression orale pour une voyelle orale.

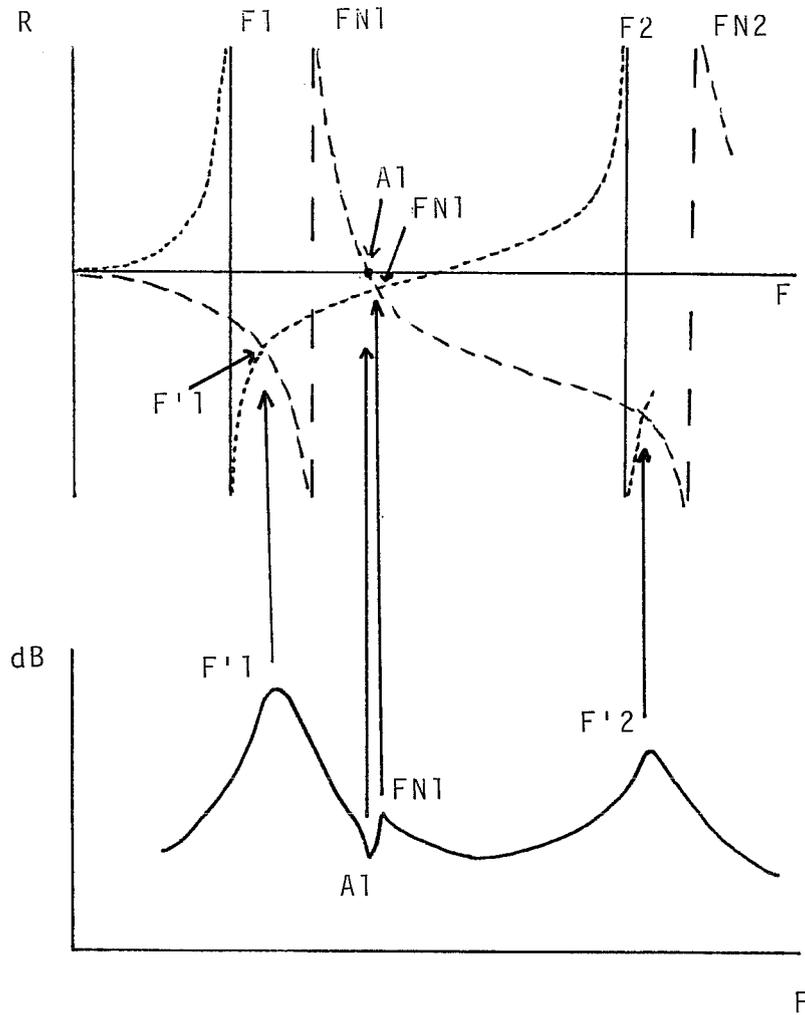
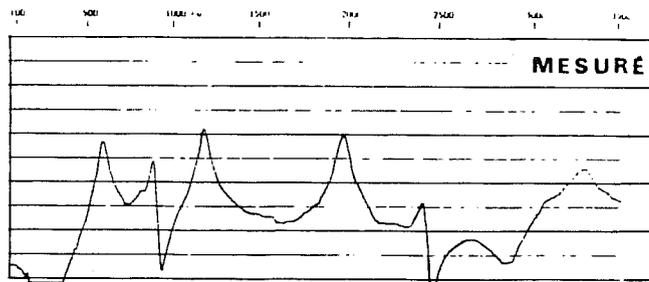


Figure 1. 31 A - Schéma de réactance théorique du spectre mesuré (fig. 1.30 A : A1 précède FN1).



PRESSION ORALE POUR LA MAQUETTE DU CONDUIT VOCAL (A)
avec couplage nasal 1 cm²

Figure 1.30 B - in MERLIER (1984) - Mesure sur une maquette de la pression orale pour une voyelle nasale.

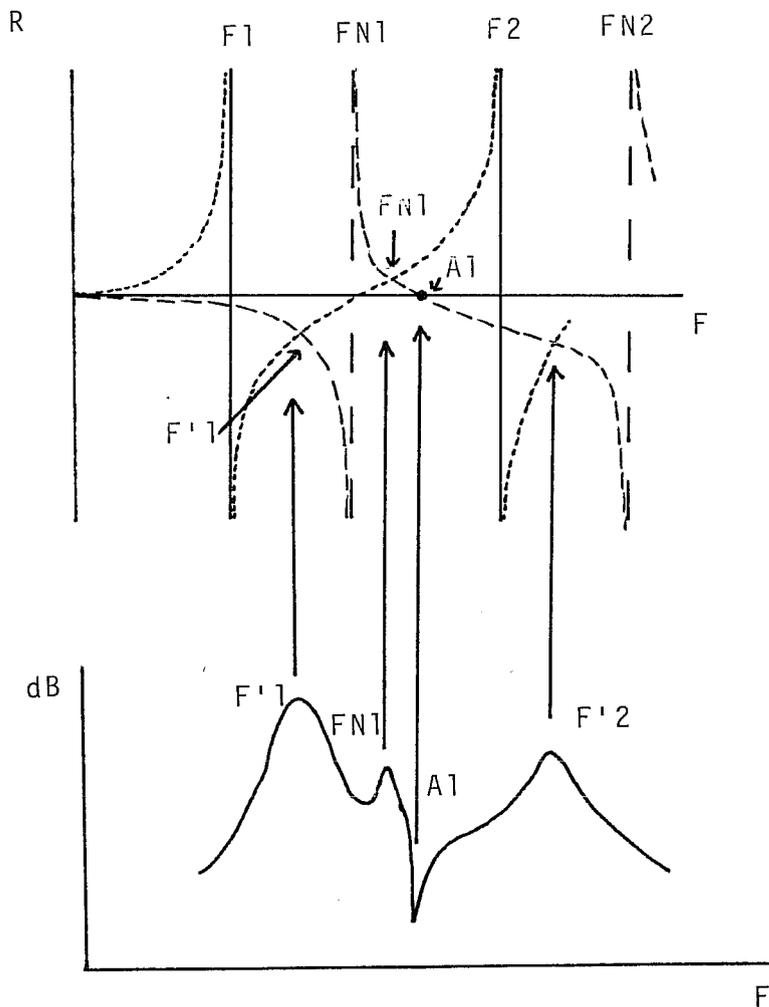


Figure 1.31 B - Schéma de réactance théorique du spectre mesuré (fig. 1.30 B : FN1 précède A1).

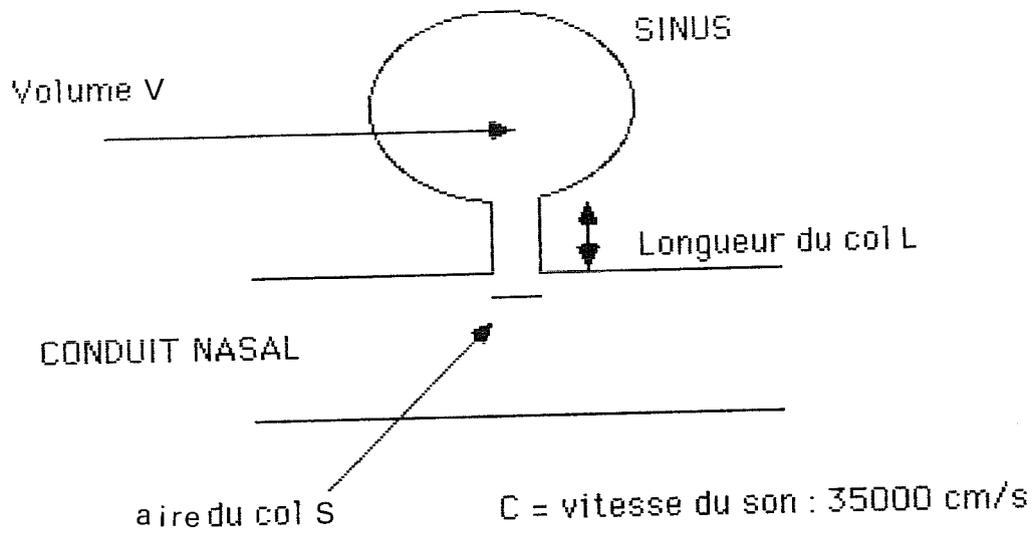


Figure 1.99 - Le sinus comme résonateur de Helmholtz : les paramètres.

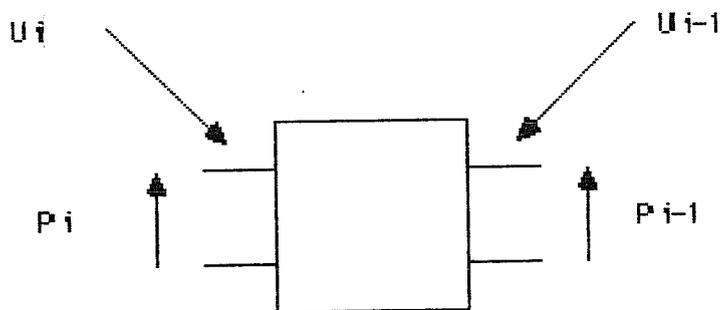


Figure 1.98 - Paramètres d'entrée - sortie d'une cellule de simulation du conduit vocal.

pour FENG. Comme nous l'avons vu, les dimensions des sinus sont variables et mal connues. Le tableau suivant, obtenu en faisant varier les différents paramètres dans des limites "raisonnables", montre que de faibles variations provoquent des modifications importantes de la fréquence de résonance : 280 à 590 Hz ($V = 25 \text{ cm}^3$, $a = 0.1 \text{ cm}$, $L = 0.7 \text{ cm}$; $V = 15$, $a = 0.2$, $L = 0.5$).

Rayon = 0.125 cm	V cm ³	F Hz	F corrigée Hz
Ø du col = 0.25	15	354	397
aire col = 0.049 cm ²	20	306	355
longueur col = 0.6 cm	25	273	327

Volume V = 25 cm ³	Rayon col	Diamètre col	Aire col	F	F corrigé
Longueur col = 0.6 cm	0.1 cm	0.2 cm	0.031 cm ²	225 Hz	288 Hz
	0.12	0.24	0.045	264	319
	0.15	0.30	0.071	320	367
	0.2	0.40	0.126	407	445

Volume = 25 cm ³	longueur col	F	F corrigé
rayon col = 0.125 cm	0.5 cm	292 Hz	345 Hz
diamètre = 0.25 cm	0.6	273	327
aire col = 0.049 cm ²	0.7	257	316

FANT (1985 p. 60, eq. 23) donne une expression simple de la valeur de la partie imaginaire de l'impédance d'un résonateur de Helmholtz. Cette expression n'est pas très utile en simulation où la partie réelle (largeur de bande) est également importante. Au plan pratique, on retiendra que le pôle associé aux sinus maxillaire se situe entre 300 et 500 Hz.

5.5 - LE MODELE DE SIMULATION

Dans cette partie nous décrirons dans ses grandes lignes le modèle de simulation utilisé dans ce travail. Après avoir réalisé un modèle de simulation identique au modèle harmonique de LHERM (1984), lui-même très proche de celui d'ATAL & al. (1978), nous avons tourné notre choix vers la structure proposée par FANT (1985). L'identité des résultats garantissait l'absence d'erreurs graves. La comparaison avec les résultats détaillés de BADIN & FANT (1984) offrait une garantie supplémentaire. La taille des programmes et le temps de calcul sont voisins. La totalité du logiciel a été mis au point par l'auteur.

Le point de départ est constitué par les équations classiques d'une ligne de transmission électrique avec pertes qui sont formellement identiques à celles de la transmission d'un son dans un tuyau avec pertes par viscosité et chaleur, si on fait correspondre la tension à la pression P et le débit (volumique) V ("Volume velocity" ou "flow") au courant.

Dans une section de tuyau de longueur l on a (cf FANT 1985 eq. 4(a) et (b) par ex.)

$$(1) \quad P_i = U_i Z_i \coth \Theta_i - U_{i-1} / \sinh \Theta_i \quad (\text{numérotation de l'article de FANT 1985})$$

$$(2) \quad P_{i-1} = U_i Z_i / \sinh \Theta_i - U_{i-1} Z_i \coth \Theta_i$$

avec P_i : pression à la sortie P_{i-1} : pression à l'entrée Z_i : impédance de la section

U_i : débit à la source U_{i-1} : Débit à l'entrée

$\Theta_i = \gamma l$: constante de propagation $\gamma = a + j \omega / c$ l : longueur

a : facteur de perte ω : pulsation = $2\pi f$ f : fréquence c : vitesse du son

Les paramètres d'entrée et de sortie d'une cellule sont rappelés graphiquement à la figure 98. On note $Z_{i,i}$ l'impédance du conduit vue de la sortie et Z_b l'impédance vue de l'entrée. Une simple manipulation de (2) permet d'exprimer le rapport débit à l'entrée / débit à la sortie pour la i ème section. Cette expression servira à calculer la fonction de transfert.

$$(10) \quad U_i / U_{i-1} = \cosh \Theta_i + Z_b / Z_i \sinh \Theta_i$$

$$(11) \quad Z_{i,i} = Z_i (\cosh \Theta_i - U_{i-1} / U_i) * 1 / \sinh \Theta_i$$

Un conduit sera considéré comme une concaténation de tuyaux de faible longueur (0.5 à 1 cm). La fonction de transfert est prise par définition comme égale au rapport du débit à la sortie par le débit à l'entrée du conduit : U_o / U_i ($o = \text{output}$, $i = \text{input}$). Si on numérote les sections de façon décroissante du larynx vers les lèvres, U_i est le débit à l'entrée de la i ème section et U_{i-1} le débit à la sortie de la i ème section, qui est égal au débit à l'entrée de la $i-1$ ème section.

Au delà d'un branchement en dérivation d'impédance Z_s , la règle de calcul des impédances en parallèle permet d'écrire

$$(12) \quad U_{i,s} = U_i (1 + Z_{i,i} / Z_s)$$

$$(13) \quad Z_{i,s} = Z_{i,i} / (1 + Z_{i,i} / Z_s)$$

Le calcul commence par la spécification de l'impédance de radiation Z_b à l'extrémité du conduit :

$$(9) \quad Z_b = R_o + j \omega L_o \quad \text{ce qui conduit en utilisant (10) à}$$

$$(16) \quad U_1 / U_o = \cosh \Theta_1 + (R_o A_o / \rho c + j \omega L_o / c) \sinh \Theta_1$$

si on exprime de façon classique l'inductance de rayonnement par

$$(14) \quad L_0 = \rho \cdot 0.8 \sqrt{(A_0 / \pi) / A_0} \quad A_0 : \text{Aire de l'extrémité} \quad \rho : \text{densité de l'air}$$

Il se poursuit par le calcul de U_2 / U_1 à l'aide de (10) avec l'impédance propre de la section 2 prise égale $\rho c / A_2$ puis de $Z_{2,2}$ par (11). On obtient finalement le rapport

U_2 / U_0 en remarquant qu'il est égal à $(U_2 / U_1)(U_1 / U_0)$. Le cas échéant, on applique (13) dans le cas d'un branchement en parallèle. Dans le cas d'une cavité fermée, comme un sinus, A_0 tend vers 0, et l'équation (16) se réduit à son premier terme. Après les calculs portant sur la dernière section (larynx) on dispose de l'impédance d'entrée du conduit (par l'équation 11), et on calculera la valeur de la fonction de transfert à la fréquence donnée par 20 Log_{10} du module de l'inverse de U_i / U_0 (eq. 10).

Dans le cas du couplage d'une cavité fermée, nous ne calculons pas son impédance par l'équation (23) car il n'est pas possible d'estimer a priori la partie résistive. Nous appliquons la méthode plus lourde du conduit fermé en dérivation.

L'article de BADIN & FANT (1984) contient une discussion extrêmement complète de tous les paramètres physiques de la modélisation. Les constantes (densité de l'air, vitesse du son, constante adiabatique ...) de notre modèle sont ceux de la page 107 de cet article, sauf la vitesse du son fixée à 35000 cm / s. L'impédance de radiation est calculée par le modèle WF (p. 68) y compris la correction $K_S(w)$. Pour les pertes par vibration des parois, la formulation par deux cellules ("lumped element representation" : p. 80) a été retenue, avec la correction d'aire. Cette structure réduit notablement le temps de calcul par rapport à la formulation des pertes pour chaque élément du conduit, sans sacrifier la précision. Ne sont pas inclus dans notre simulation l'effet de l'impédance glottale et du système sous-glottique. Le conduit nasal est traité de la même manière que le conduit oral, y compris les pertes par vibration des parois. La contribution des sinus, en revanche, ne comportent pas ces pertes. Les facteurs de forme oral et nasal sont fixés à 2 et 3.5 respectivement. Le pas d'incrémentation fréquentiel est de 20 Hz et nous balayons l'intervalle 100 - 4000 Hz ou 4500 Hz. Pour l'affichage d'une fonction de transfert plus réaliste, nous incorporons l'une de deux corrections suivantes : Une pente de -6 dB / octave à partir de 250 Hz (introduisant une petite discontinuité à 250 Hz visible sur certains tracés) pour simuler l'effet combiné de l'impédance de rayonnement et de la pente moyenne de l'onde glottale, ou la modélisation du spectre glottal d'après FANT (1983), qui est décrite plus loin. Les pics de résonance et les largeurs de bande sont estimés par interpolation parabolique sur les valeurs adjacentes à un maximum spectral. Les largeurs de bande calculées ainsi ne sont qu'approximatives.

La configuration des conduits est présentée à la figure 1.32. Le conduit nasal se branche à l'extrémité droite de la section la plus proche de 53% (9 / 17) de la distance bouche - larynx. Il comprend, du voile vers les narines, une partie unique, représentant le rhinopharynx, dont la longueur nominale (conduit fermé) est de 3.3 cm. Ce conduit se

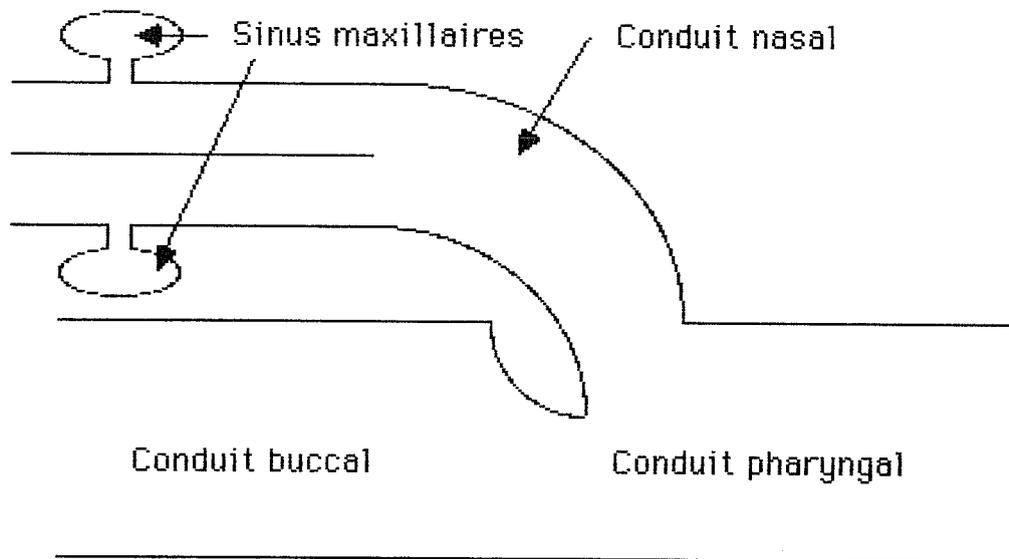


Figure 1.32 - Configuration des conduits utilisés dans la simulation.

divise en deux pour représenter les fosses nasales et permettre l'étude de l'effet de la dissymétrie. Sa longueur nominale est de 7.5 cm. A chaque branche est connecté un sinus à environ 4 cm des narines. Le rétrécissement du conduit vocal causé par l'abaissement du voile est réparti sur 3 cm environ. Les formes et dimensions exactes seront précisées dans les paragraphes traitant des résultats des simulations.

6 - ANALYSE ACOUSTIQUE : LE CHOIX D'UNE METHODE

Ce paragraphe précise la méthode et les conditions d'analyse des voyelles naturelles. Nous recherchons une analyse fine, capable de représenter des spectres possédant un nombre élevé de pics et de vallées. Cette condition élimine en pratique l'analyse par prédiction linéaire fondée sur un modèle ne comportant que des pôles alors que la théorie acoustique prévoit l'existence de zéros. L'analyse LPC suppose un choix délicat de l'ordre du filtre de prédiction pour "suivre" le détail de l'évolution spectrale (CARRE & TRIBOLET 1974). De plus, ce modèle utilise un nombre variable et non contrôlable de pôles pour supprimer l'effet spectral de la source et de la radiation, d'où découle qu'un nombre variable de pôles seront utilisés pour rendre compte du spectre nasal proprement dit. On trouvera dans l'analyse des travaux de MRAYATI (1975) un exemple d'une difficulté inhérente à cette méthode.

Il ne reste en compétition que l'analyse cepstrale et la F.F.T. L'expérience accumulée par les chercheurs du Centre de Recherche en Informatique de Nancy (C.R.I.N.), en particulier Y. LAPRIE de l'équipe R.F.I.A., montre que les résultats d'analyse sont très voisins quand le nombre de coefficients cepstraux est bien choisi. Pour des raisons de simplicité, nous avons retenu la F.F.T. Pour éliminer l'effet éventuel d'une fenêtre d'analyse, nous utilisons en fait une transformée de Fourier classique sur la durée exacte d'une période. Le temps de calcul est réduit par l'utilisation de l'algorithme de GOERTZEL (1960). Au plan pratique, l'utilisateur détermine la zone d'analyse sur l'évolution de la courbe d'énergie et le programme détermine la période. Nous échantillons donc le spectre aux seules fréquences où il est défini, sans pertes d'informations. La fréquence d'échantillonnage du signal temporel est de 10 KHz, avec une précision de 12 bits. Dans le cas d'une fréquence fondamentale grave, la définition spectrale est excellente. L'inconvénient d'une F.F.T. est que la forme exacte de l'onde glottale pour une période particulière est présente dans le spectre, à l'inverse de la sorte de moyenne réalisée par les méthodes analysant plusieurs périodes consécutives.

Le micro 1/2 pouce et l'amplificateur Bruel & Kjaer sont connectés à un filtre anti-repliement de -36 dB/octave dont la fréquence de coupure est à 4 KHz.

7 - LA CONTRIBUTION SPECTRALE DE LA SOURCE GLOTTALE ET DES COUPLAGES NON-VELAIRES

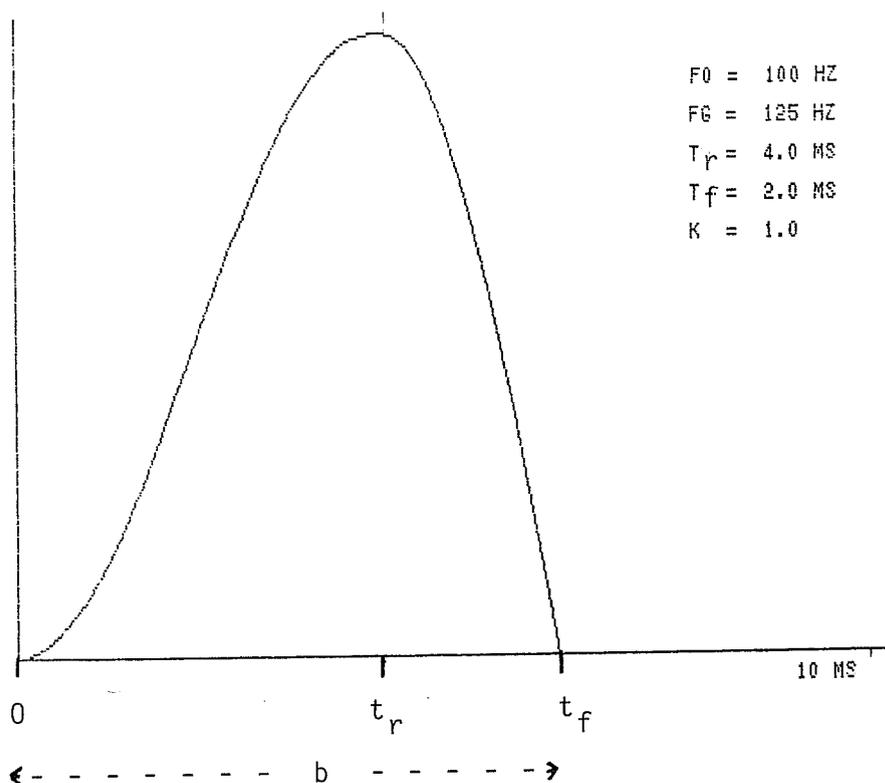
L'analyse fine des spectres des voyelles naturelles suppose que les effets spectraux liés à la source glottale soient identifiés pour être séparés de la contribution propre de la forme des cavités. Il faut donc disposer d'un modèle précis de l'onde glottale. Nous nous appuyons principalement sur les travaux que mène G. FANT et ses collaborateurs depuis 1979. Les paragraphes suivants résument, en visant les conclusions utilisables concrètement, les connaissances actuelles sur le spectre glottal, mais également la contribution potentielle d'autres facteurs comme le couplage sous-glottique et nasal externe.

7.1 - LA SOURCE GLOTTALE

La forme de l'onde de débit glottal est en première approximation proportionnelle à la variation de l'aire de l'aperture glottale. Les simulations numériques de la vibration du système mécanique que constituent les ligaments vocaux et les muscles associés quand ils sont soumis à une pression sous-glottique (cf. GUERIN 1978 par ex.), tout comme l'observation directe par cinématographie ultra-rapide, révèlent que l'aire glottale au cours d'une période présente une phase croissante (ouverture ou abduction des cordes vocales) suivie d'une phase décroissante (fermeture ou adduction) plus brève, et d'une phase fermée (accolement des cordes). L'onde de débit présente grossièrement cette forme triangulaire mais la pente de fermeture est plus raide en raison de l'effet des inductances glottale et vocale (figure 1.33). FLANAGAN (1958) a montré qu'une onde triangulaire symétrique de durée ouverte b présentait des minima spectraux à des fréquences égales à des multiples de $2/b$: pour une fréquence fondamentale de 100 Hz par exemple, et une durée de phase ouverte de 6 ms, les zéros tombent aux multiples de 330 Hz. L'ordre de grandeur reste le même pour des triangles dissymétriques.

FANT (1979, 1983 par ex.) a proposé un modèle relativement simple d'onde de débit glottal dont le spectre possède l'avantage de pouvoir être mis sous une forme analytique, bien utile pour les simulations dans le domaine fréquentiel. Dans le domaine temporel, la branche montante de cette onde (figure 1.33) est décrite entre 0 et t_r (rise time) par une portion de (co)sinusoïde de valeur $0.5 U_{gm} (1 - \cos 2\pi F_g t)$ où U_{gm} est la valeur du débit maximum, et où F_g est égal à $1/(2 t_r)$. La branche descendante, entre t_r et t_f (fall time) est décrite par $U_{gm} (K \cos 2\pi F_g (t - t_r) - K + 1)$, et rejoint la phase de débit nul à $t = t_r + (1 / 2\pi F_g) \arccos ((K - 1) / K)$. Le paramètre K détermine la pente de la partie

Débit



$$x = f/F_g$$

$$\varphi_3 = \arccos[(K-1)/K]$$

$$A = \cos[x(\pi + \varphi_3)]$$

$$B = (1-2K)\cos(x\varphi_3)$$

$$C = 2K-2$$

$$D = 2x(2K-1)^{1/2}$$

$$E = \sin[x(\pi + \varphi_3)]$$

$$F = (1-2K)\sin(x\varphi_3)$$

$$H(f) = -\sum_{n=1}^4 10 \log[(1-f^2/F_n^2)^2 + (B_n f/F_n^2)^2]$$

$$H_{k4} = 0.54x^2 + 0.00143x^4 \quad (\text{dB})$$

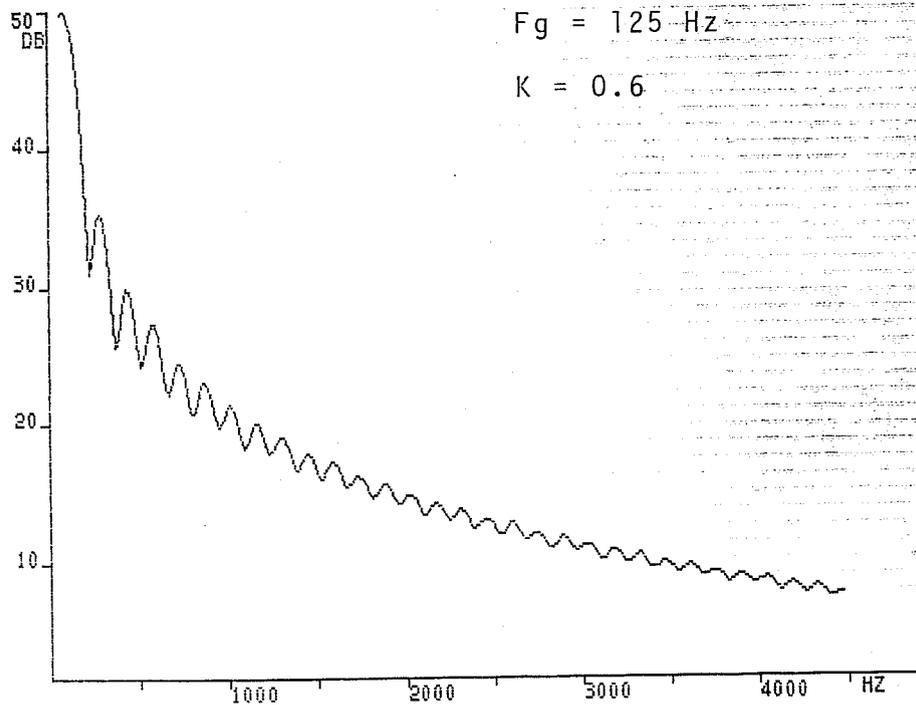
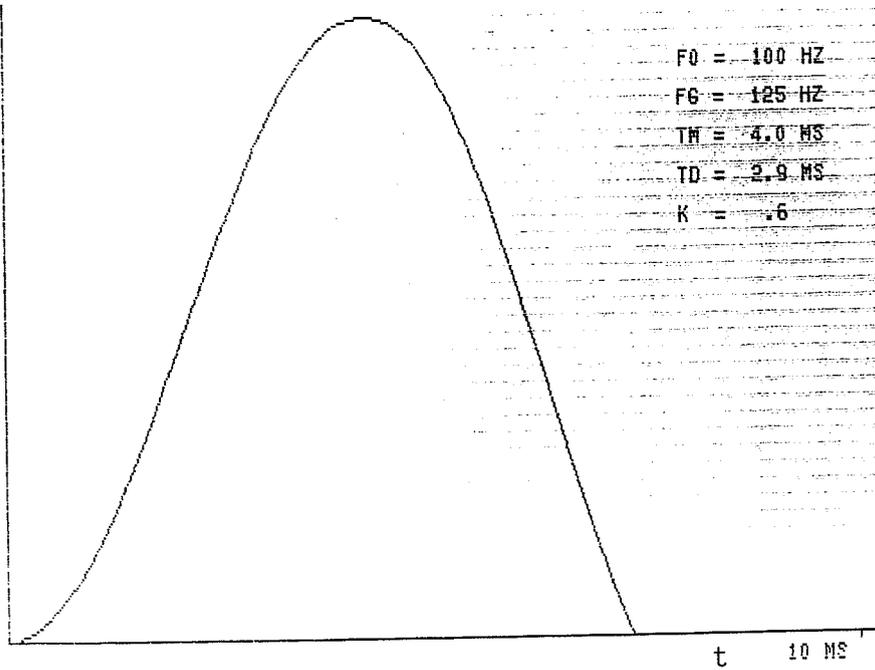
$$x = 17.5f / 500 T_t$$

$$S(f) = 20 \log \{ [(A+B+C)^2 + (D+E+F)^2]^{1/2} / 4\pi |x^2 - 1| \} \cdot \{ F_0 / 100 \} \quad (\text{dB})$$

Figure 1.33 - Onde de débit glottal et équations du calcul de son spectre et de la fonction de transfert d'un voyelle complète. (FANT 1983)

descendante et peut varier théoriquement entre 0.5 (onde à flancs symétriques et de phase fermée nulle) et $+\infty$ (branche descendante de durée nulle). L'expression analytique est donnée à la figure 1.33 ainsi que les équations permettant le calcul de la fonction de transfert complète d'une voyelle dont on connaît les fréquences formantiques et les largeurs de bande (F_n, B_n). On fixe la valeur d'une fréquence fondamentale théorique (F_0). Les figures 1.34 à 1.36 illustrent les spectres obtenus en faisant varier les paramètres K et F_g . On voit à la figure 1.34 la forme de l'onde de débit glottal, le spectre glottal avec sa pente caractéristique de -12 dB / octave corrigé de l'effet de radiation de $+6$ dB / octave, le spectre après l'application d'une préemphasis de $+6$ dB / octave mettant en évidence les oscillations par rapport à une pente nulle en moyenne et enfin le spectre d'une voyelle [a] complète : F_g est ici fixé à 125 Hz, F_0 à 100 Hz et K à une valeur faible (0.6) caractéristique d'une voix "faible" à phase ouverte longue. La figure 1.35 montre les mêmes simulations pour $K = 1$, correspondant à une voix plus "forte". Enfin, la figure 1.36 montre la voyelle [a] pour une phase ouverte plus courte ($F_g = 175$ Hz) correspondant à un effort vocal modéré ($K = 0.8$). On remarque tout d'abord que le spectre présente les oscillations prévues par FLANAGAN. On voit également sur les spectres avec préemphasis une concentration d'énergie à la fréquence F_g qui domine le spectre d'une hauteur variable selon K . La comparaison des voyelles complètes révèle que cette zone peut varier de 8 dB relativement à l'amplitude de F_1 . "Formant" glottal est le terme maintenant consacré, bien qu'impropre, pour désigner cette concentration d'énergie qui domine le spectre en très basse fréquence. Sa fréquence est approximativement égale à l'inverse du double de la durée de la branche montante de l'onde de débit, c'est à dire approximativement la fréquence fondamentale multipliée par un facteur 1.7 pour une voix masculine. Cette concentration d'énergie à (ou juste au-dessus de) la fréquence F_g reste en fait à un niveau constant quand le facteur K varie. C'est le reste du spectre qui subit un déplacement vertical tout en conservant une pente moyenne de -12 dB / octave. Plus la pente est forte (K croissant), plus les hautes fréquences seront proéminentes. Auditivement, ceci explique pourquoi on ne confond pas une voix faible amplifiée avec une voix forte. Le facteur K étant en moyenne plus faible chez les femmes que chez les hommes pour des raisons aéro-mécaniques, les deux premières harmoniques domineront plus le spectre pour les voix féminines. C'est, avec un plus grand niveau de bruit turbulent glottal et la valeur de F_0 , un indice puissant de féminité vocale. FANT (1980) remarque que le "formant" glottal est souvent plus visible pour les voyelles nasales, car la présence d'un zéro proche réalise une sorte de filtrage inverse de F_1 rendant F_g plus proéminent. Le formant nasal à 250 Hz de DELATTRE (et LONCHAMP 1978 !) est vraisemblablement en réalité un formant glottal. Lorsque F_g et F_1 sont proches, il se produit un renforcement

Débit



, Figure 1. 34 - Onde de débit glottal : $F_g = 125 \text{ Hz}$, $K = 0.6$;
spectre corrigé de l'effet de radiation.

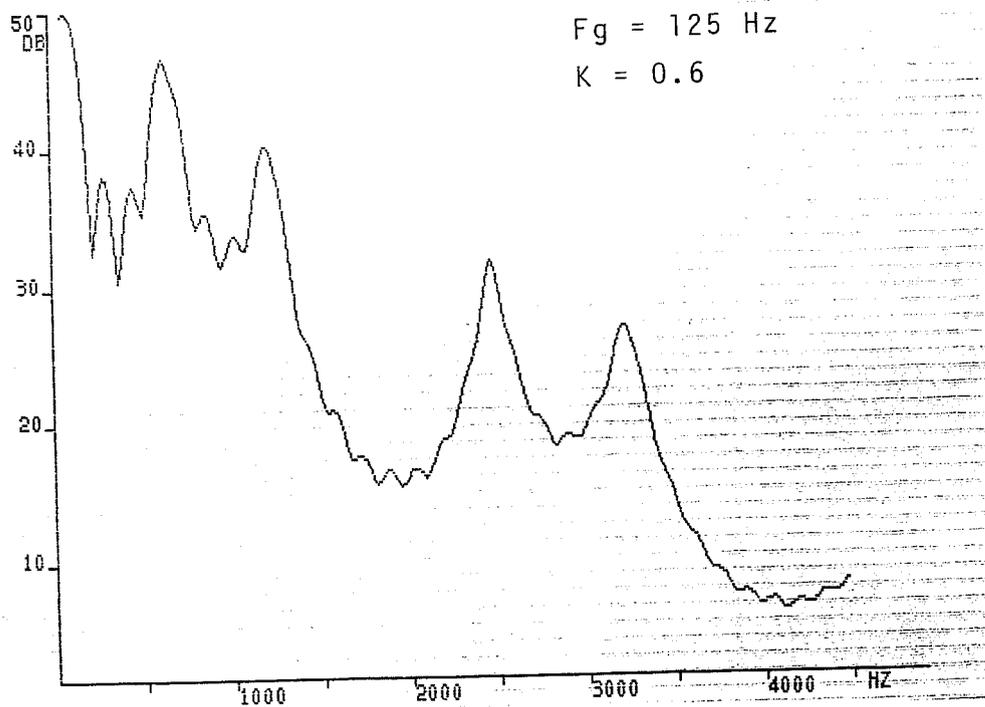
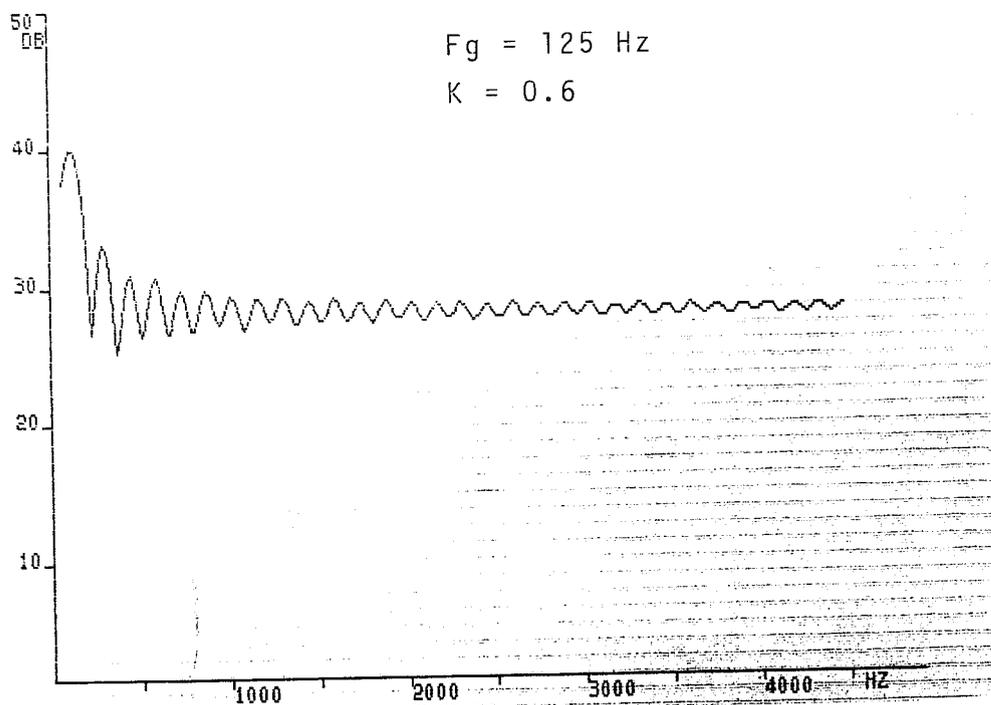


Figure 1. 34 (suite) - Spectre avec préemphasis de + 6 dB / octave
spectre d'une voyelle [a] : $F_1 = 684$, $F_2 = 1256$, $F_3 = 2503$
 $F_4 = 3262$ Hz ; $l = 16.5$ cm. $B_d = 80, 100, 120$ et 150 Hz.

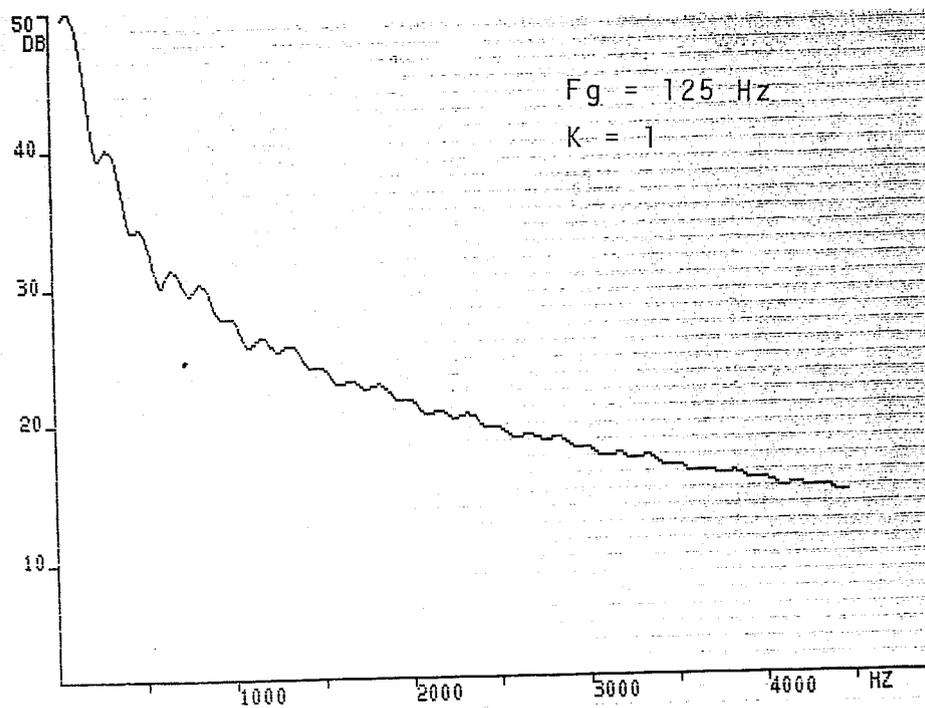
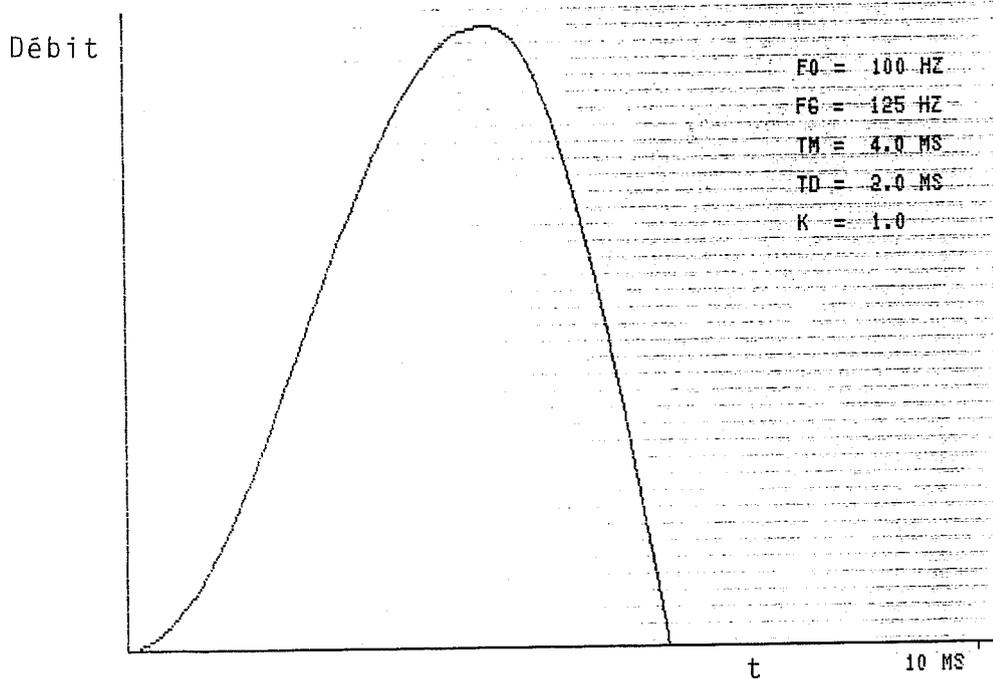
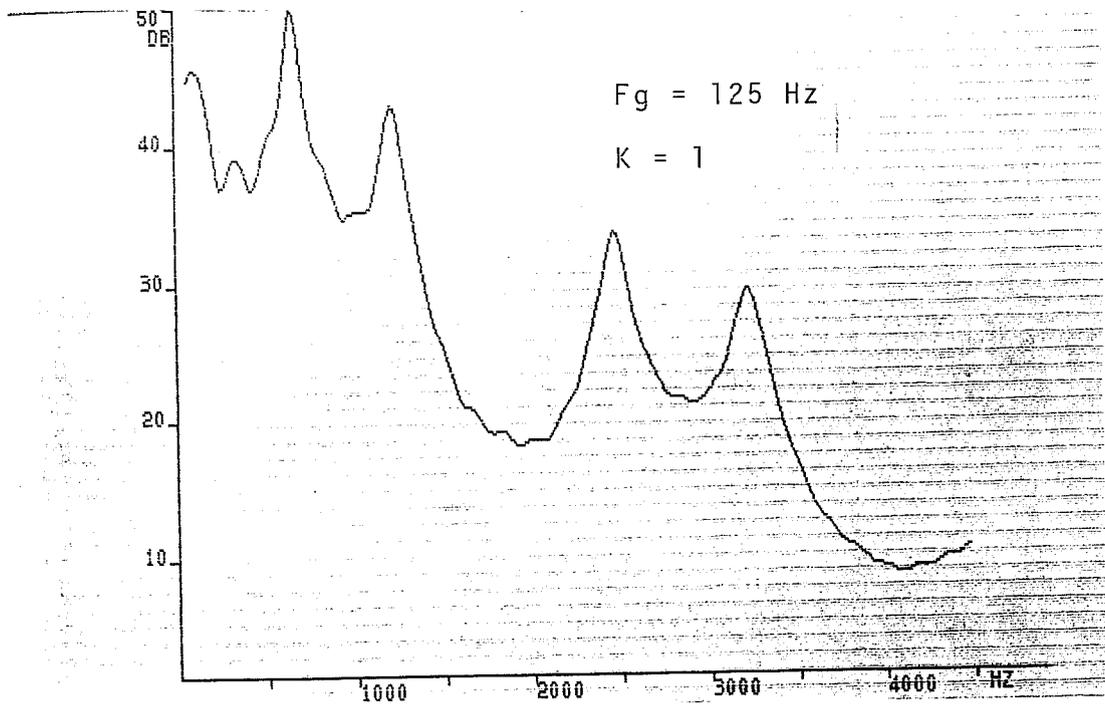
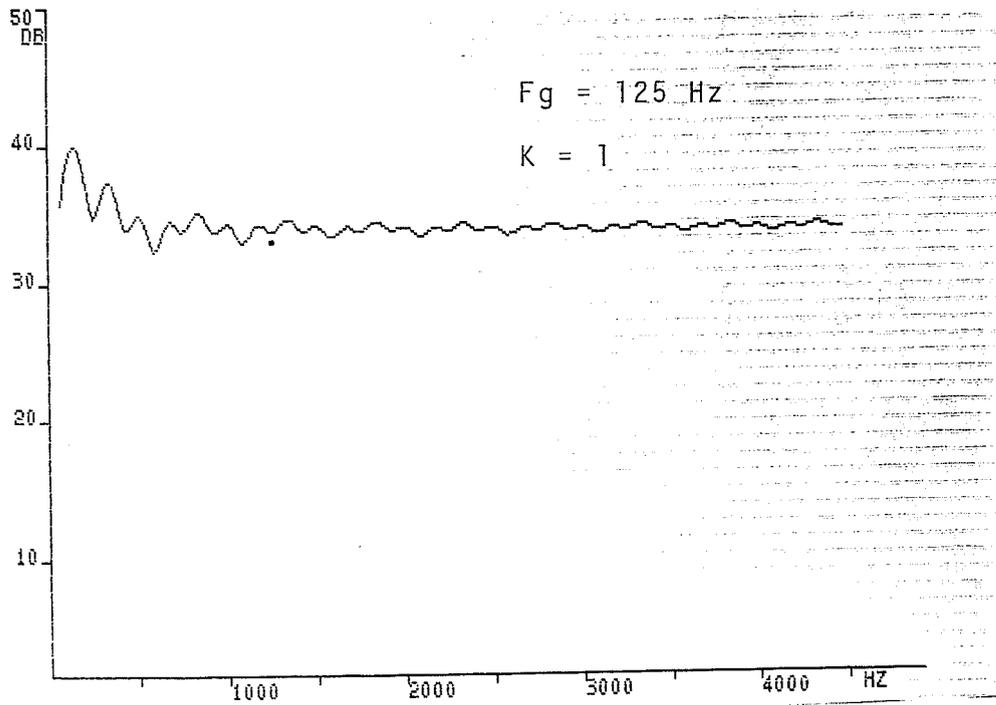


Figure 1.35 - Onde de débit glottal : $F_g = 125 \text{ Hz}$, $K = 1$;
 spectre corrigé de l'effet de radiation.



. Figure 1.35 (suite) - Spectre avec préemphasis de + 6 dB/octave spectre d'une voyelle [a] (cf. fig. 1.34).

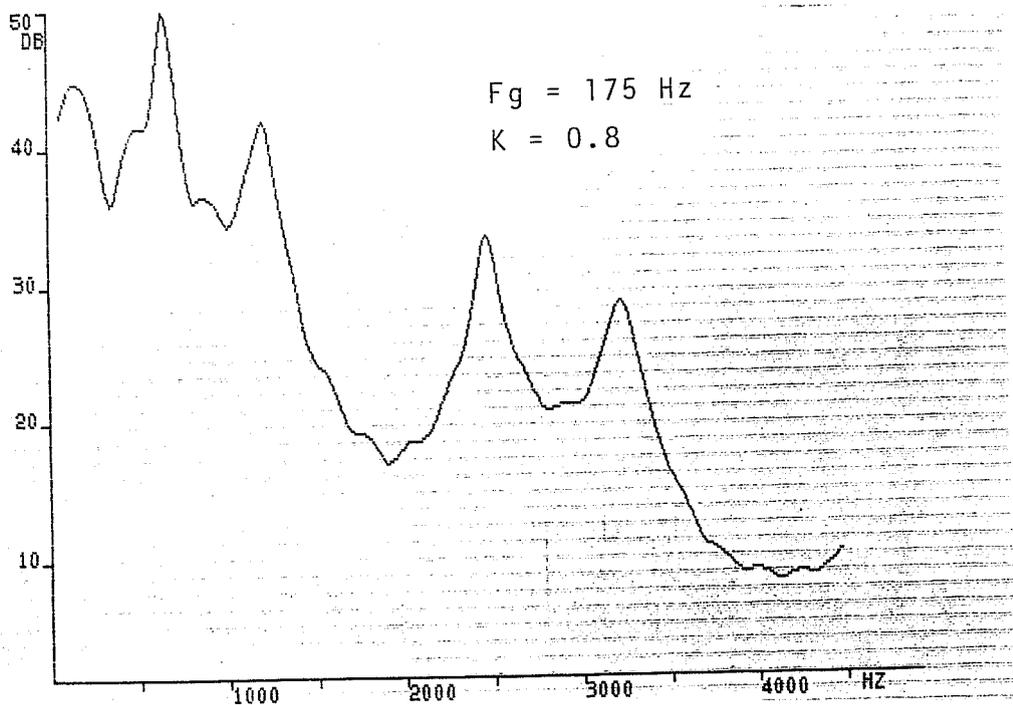
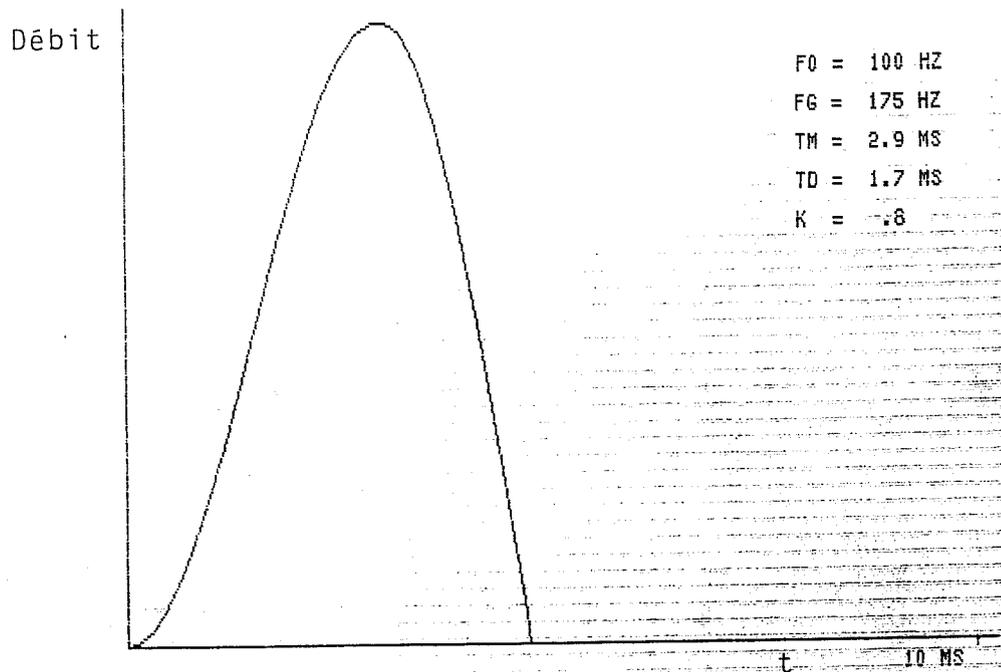


Figure 1.36 - Onde de débit glottal : $F_g = 175 \text{ Hz}$; $K = 0.8$
 spectre d'une voyelle [a] (cf. Fig. 1.34).

similaire à celui observé pour deux formants ordinaires : une réduction de moitié de la distance provoque un renforcement de + 6 dB. Des oscillations de forme $\sin x / x$ se surajoute à la pente moyenne. Il est probable que les minima sont moins creusés dans la réalité en raison de la variation de l'amortissement glottal. Ces oscillations ont une période de $1.33 \times F_g$, et leur amplitude diminue quand K augmente. La raideur de la pente de la branche descendante détermine presque complètement l'intensité sonore émise. FANT a montré dès 1959 (cf. FANT 1979) que les voyelles accentuées doivent leur plus grande énergie non à un accroissement de la pression sous-glottique et du débit U_g , mais à un accroissement du facteur de pente K , qui passe typiquement de 0.8 pour une voyelle non accentuée à 1.5 ou plus.

Les largeurs de bande des formants varient d'une manière appréciable pendant une période. La valeur instantanée peut dépasser 500 Hz pendant la phase ouverte d'un [a], car l'inductance glottale devient très grande, alors que la valeur moyenne reste inférieure à 150 Hz. Il se produit une troncation des oscillations dans le domaine temporel (FANT & ANANTHAPADMANABHA 1982), qui se traduit spectralement par un affaiblissement de l'amplitude du premier formant par accroissement de la largeur de bande, et par des oscillations de forme $\sin x / x$ sur les flancs des formants ("formant ripple") : l'ordre de grandeur de l'espacement des minima est de 150 Hz pour une voix masculine. Cette phase d'amortissement élevé empêche tout transfert d'énergie important d'une période à l'autre. C'est la situation typique pour les voix masculines que nous analyserons. La fréquence du formant sera également modulée par la variation de l'inductance glottale, et la largeur de bande observée dans une analyse sur une période sera artificiellement élevée. On trouvera une superbe illustration dans CHENG & GUERIN (1987a ; figure 1.3).

On peut certainement critiquer notre choix de l'onde glottale de FANT (1979). FUJISAKI & LJUNGQVIST (1986) ont montré que des modèles comportant un plus grand nombre de paramètres ajustables permettent une meilleure modélisation de voyelles naturelles par analyse-par-synthèse. L'aspect le plus criticable est la terminaison abrupte de la branche descendante. ANANTHAPADMANABHA (1984) par exemple décrit en détail un modèle plus réaliste. Une terminaison plus douce, toute chose égale par ailleurs, augmente la proéminence du formant glottal et la pente moyenne, et renforce légèrement les oscillations spectrales.

Un autre phénomène remarquable est l'effet du couplage supra-glottal sur l'onde de débit. Phénoménologiquement, le mécanisme est simple. La forme de l'onde dépend de la valeur de la pression transglottique, c'est à dire de la différence de pression au-dessus et en dessous du larynx. Pendant l'émission d'une voyelle, si la pression sous-glottique reste pratiquement constante, la pression supraglottique est variable, car le conduit, notamment dans sa partie laryngo-pharyngale; est le siège d'une oscillation de pression à la fréquence

de F1. On peut donc s'attendre à une modulation du débit transglottal à la fréquence de F1 ("vocal tract interaction ripple"). GUERIN (1978) a montré avec netteté (figure 1.37), en raisonnant sur l'impédance d'entrée du conduit vocal qui présente un maximum aux fréquences formantiques, que cette oscillation de l'onde de débit provoque un minimum d'énergie à la fréquence de F1. Mais FANT & al. (1985) prévoient et illustrent par des simulations un minimum spectral (bordé de deux petits rebonds) pouvant atteindre 2 fois la fréquence de F1. Comme l'analyse des voyelles nasales nécessite l'étude détaillée du spectre, y compris des petites irrégularités, nous avons cherché à comprendre les raisons de cette divergence. Les simulations de GUERIN (1978) portent sur des ondes de débit dont la phase ouverte, sur laquelle peut s'inscrire les oscillations dues à F1, est longue par rapport à la période de F1. En revanche, toutes les simulations de FANT & al. (sauf une !) portent sur des ondes dont la phase ouverte est courte par rapport à la fréquence de F1. Notre figure 1.38 reproduit la figure 6 de FANT & al. (1985). On note sur la dérivée de l'onde de débit (UG') une oscillation de période 1 ms (1000 Hz), créée par un formant de 500 Hz. La phase ouverte (cf. l'onde de débit UG) est très brève (3 ms), pour une période de 8 ms. Ce n'est pas la situation courante pour des voix masculines ou féminines. L'oscillation due à F1 est donc tronquée. Ce "filtrage non linéaire" provoque un déplacement de la fréquence apparente comme on peut le voir à la figure 1.39. Si on analyse une sinusoïde de 500 Hz modifiée par une fenêtre triangulaire symétrique plus large qu'une période (a), la fréquence du maximum reste à 500 Hz, et le spectre présente les rebonds en $\sin x / x$ classiques. Le filtrage est linéaire. Si la fenêtre triangulaire est égale à la moitié de la période, ce filtrage non-linéaire provoque un déplacement du maximum (b). Dans le seul cas où FANT & al. emploient une phase fermée longue (6.5 ms pour une période de 8 ms : fig 14), le minimum spectral principal est bien à la fréquence de F1 (500 Hz). Retenons qu'en pratique, le zéro d'interaction perturbera le premier formant, en élargissant sa largeur de bande apparente, mais ne doit pas créer un couple pic / vallée ailleurs dans le spectre.

Les mesures de CARRE & GUERIN (1985) révèlent enfin que le couplage source-conduit peut provoquer une variation de l'amplitude des harmoniques pouvant atteindre 20 dB quand on compare les spectres à plusieurs fréquences fondamentales.

7.2 LE COUPLAGE SOUS-GLOTTIQUE

Pendant la phase ouverte de l'oscillation glottale, la cavité pharyngale est couplée à la trachée respiratoire et aux poumons. A partir des dimensions du système respiratoire, FANT & al. (1972) ont calculé que ses fréquences de résonance propres étaient de l'ordre

FIGURE 32. Onde de débit calculée pour trois voyelles à partir de la même fonction d'aire d'ouverture des cordes vocales.

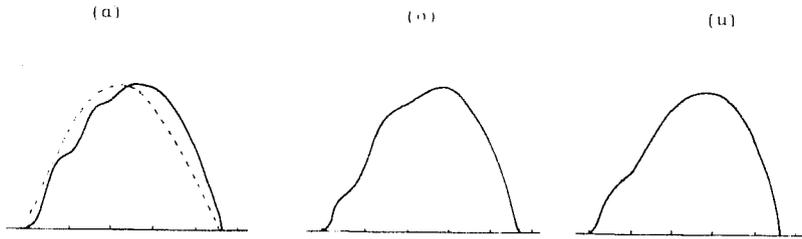


FIGURE 33. Influence de la résistance équivalente de l'impédance d'entrée du conduit vocal sur le "zéro" du spectre de l'onde de débit à la fréquence du premier formant (en b/, la résistance est trois fois plus grande qu'en a/).

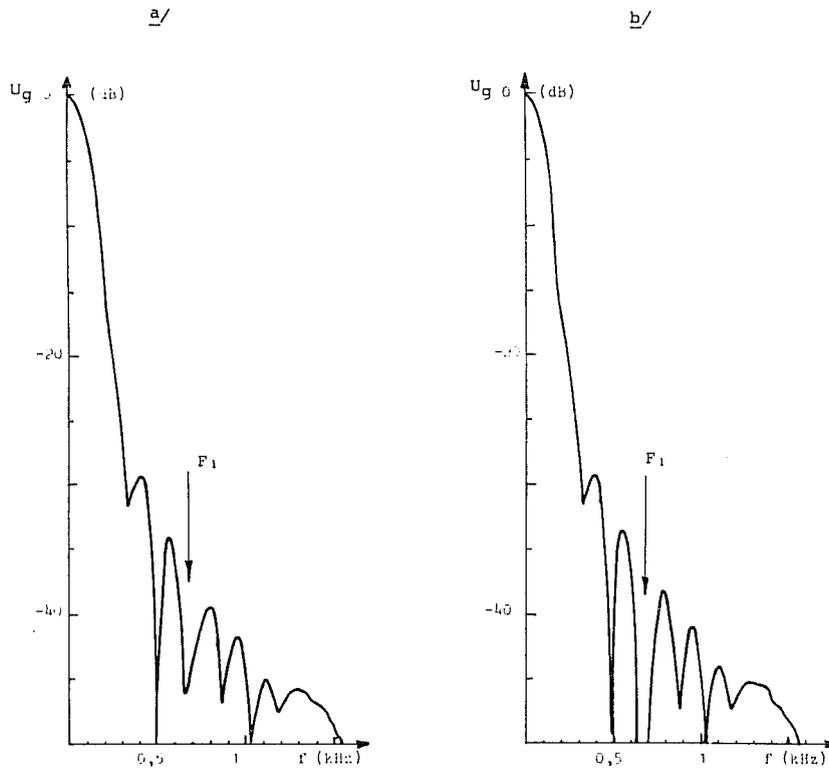
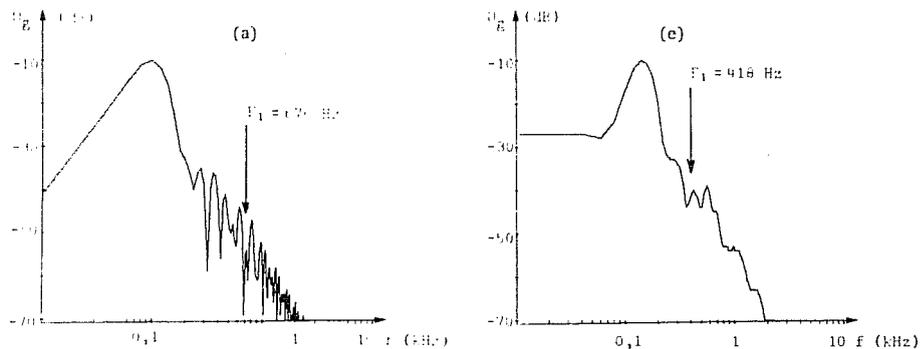


FIGURE 34. Spectre de l'onde de débit pour deux voyelles avec le modèle de la source vocale à deux masses.



. Figure 1.37 - Influence du couplage sous-glottal sur le spectre de l'onde de débit glottal : présence d'un zéro à la fréquence de F_1 (in GUERIN 1978).

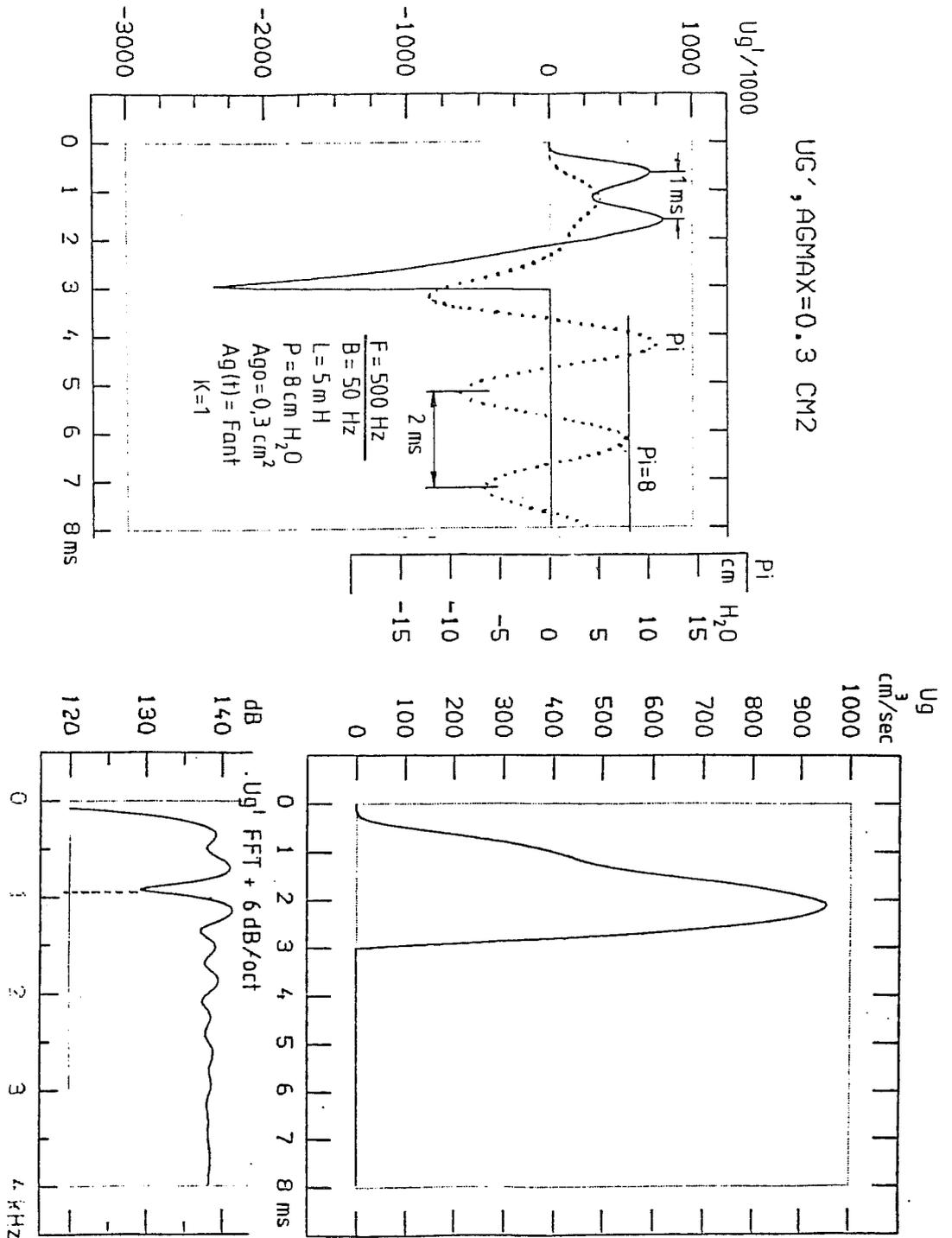


Fig. 6. Solid lines - flow and flow derivatives. Dotted line - supra-glottal pressure assuming single resonant circuit load. First pulse only.

- Figure 1.38 - Influence du couplage sous-glottique sur l'onde de débit glottal : présence d'un zéro à 2 fois la fréquence de F_1 ; notez la phase ouverte très brève (in FANT & al. 1985).

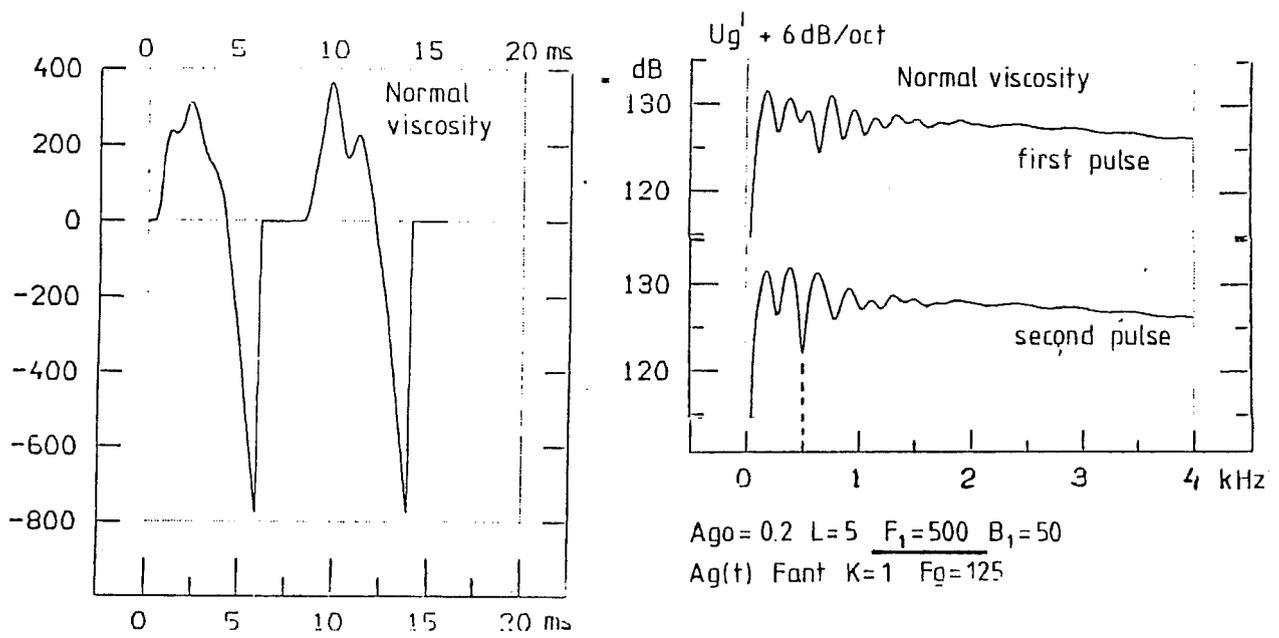


Figure 1.38 (suite) - présence d'un zéro à la fréquence de F_1 ; notez la phase ouverte longue (in FANT & al. 1985).

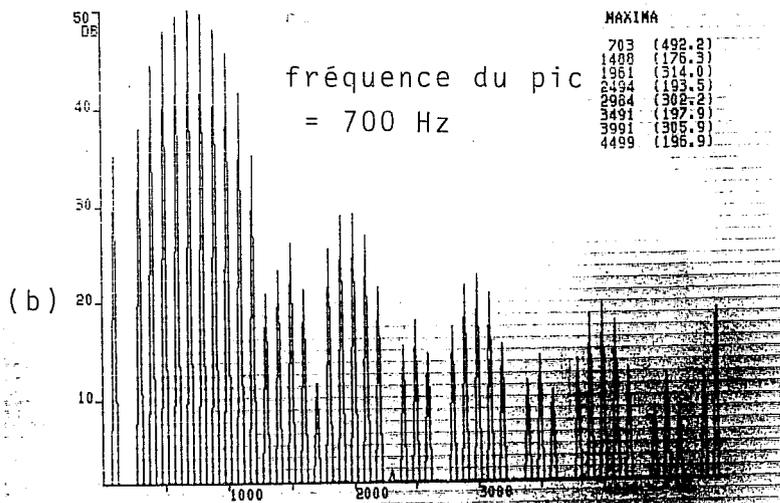
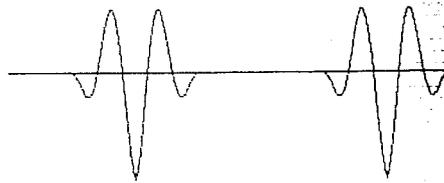
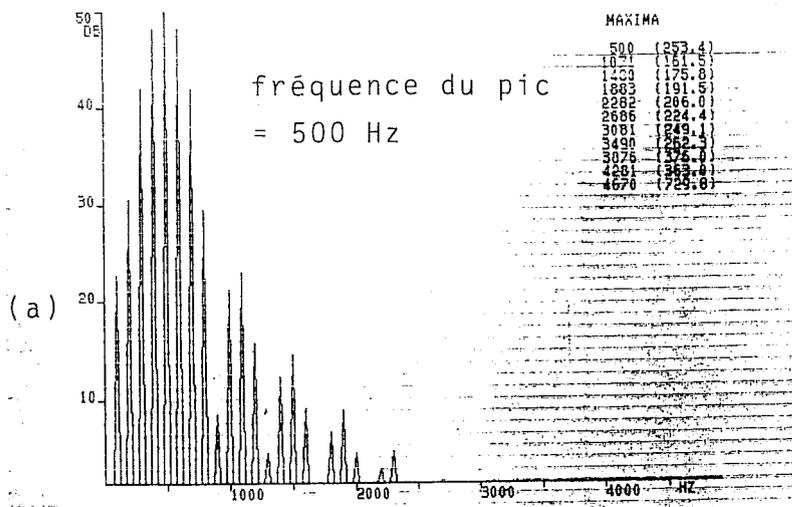


Figure 1.39 - Rôle de la durée d'une fenêtre triangulaire sur le spectre d'une sinusoïde: (a) durée supérieure à une période (b) durée = moitié de la période: Notez le déplacement du pic dans ce cas.

de 650, 1400, 2150 et 2850 Hz. Les mesures de CRANEN & BOVES (1987) fournissent des résultats voisins : 510, 1355 et 2290 Hz. La fréquence de la première résonance varierait légèrement en fonction de la fréquence du premier formant. Mais "... Il n'existe pas encore d'indice clair d'une influence significative des résonances sous-glottiques dans le spectre supraglottal des sons voisés. En dehors des segments aspirés, l'impédance glottale est suffisamment grande pour assurer un découplage efficace (des conduits sous et supraglottal), mais elle est suffisamment faible pour accroître de manière appréciable l'amortissement (des formants) " écrit FANT en 1979 (p. 88). Il ajoute: "En théorie, on pourrait obtenir une interaction quand le premier formant de la voyelle coïncide avec la première résonance du système sous-glottique vers 600 Hz ou quand la glotte est très ouverte". De fait, les formants sous-glottiques ne sont clairement visibles que dans l'aspiration des consonnes occlusives voisées (FANT & al. 1972). KLATT (1987) a pourtant donné une belle illustration de la présence d'une résonance sous-glottique dans une voyelle [a] féminine (figure 1.40). On sait qu'une légère aspiration est une caractéristique essentielle des voix féminines. L'identification est rendue certaine par le fait que ce formant sous-glottique continue un formant visible dans la consonne [h] qui la précède. Il s'agit dans ce cas de la troisième résonance, les deux premières étant faiblement visibles vers 600 et 1600 Hz. La zone du premier formant est perturbée, et la mesure de F1 rendue difficile. Les voix masculines, qui feront seules l'objet d'analyses dans ce travail, ne semblent donc pas devoir présenter cette caractéristique. Néanmoins, beaucoup de chercheurs, dont G. FANT ont mentionné la présence d'un minimum spectral vers 800 Hz, accompagné d'un pic à une fréquence légèrement plus basse, en faisant l'hypothèse d'une interaction sous-glottique. CRANEN & BOVES (1987) émettent l'hypothèse d'un effet du déplacement vertical des cordes vocales. La figure 1.41 présente le spectre d'un [e] présentant une telle paire pôle / zéro. FANT & al (1972) rejettent cette hypothèse et suggèrent, pour le zéro, un effet de la forme de l'onde glottale. FANT et ANANTHAPADMANABHA (1982) (cf. aussi CARRE (1981)) l'attribuent à l'effet du couplage supraglottal. Nous ne sommes pas totalement convaincu par cette explication pour l'exemple de la figure 1.41. Comme il a été vu plus haut, le zéro de couplage doit se produire à, ou juste au-dessus de la fréquence du premier formant pour une voix masculine typique dont la phase d'ouverture glottale est longue par rapport à la période de F1 (cf. GUERIN 1978). Pour la voyelle [e] dont le premier formant est bas, le minimum et le pic associé ne peuvent pas être à une fréquence double. Il revient à MERLIER (1984) d'avoir suggéré une cause possible qui est décrite dans le paragraphe suivant.

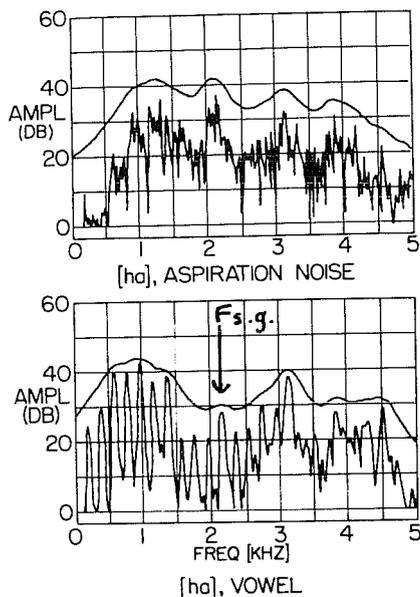
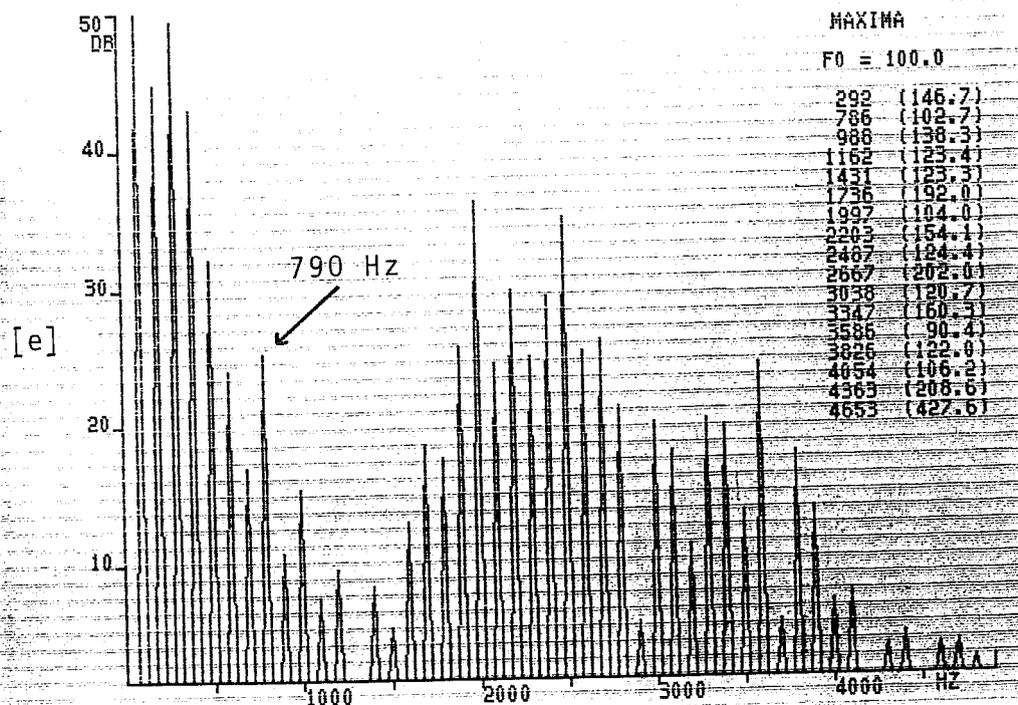


FIG. 13. The magnitude spectrum of 50 ms of aspiration noise in the syllable [ha] (top) reveals a strong subglottal formant at 2150 Hz, between $F_2 = 1300$ Hz and $F_3 = 3200$ Hz. Additional subglottal resonances are faintly evident at about 600 and 1600 Hz. The effect of the subglottal system on the vowel spectrum, measured about 40 ms after voicing onset (bottom) is to create a spurious peak at 2150 Hz, and to modify harmonic amplitudes in the F_1 - F_2 region so as to make it difficult to tell whether two or three formants are present between 600 and 1400 Hz. These changes are typical of normal breathy vowels of women (Klatt, 1986b).

• Figure 1.40 - Formant sous-glottal à 2150 Hz (KLATT 1987).



• Figure 1.41 - Couplage nasal externe : composante proche de la fréquence de Helmholtz du conduit nasal pour une voyelle orale.

7.3 - LE COUPLAGE NASAL EXTERNE (MERLIER 1984)

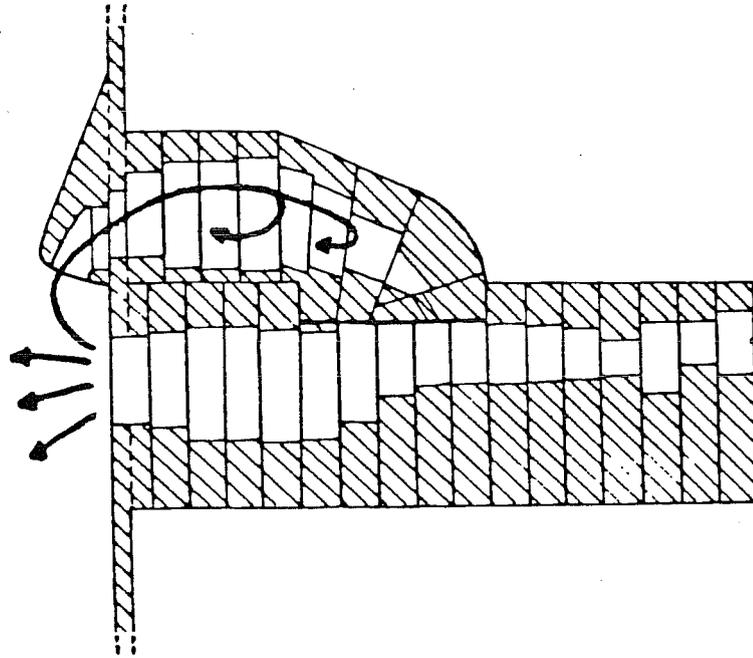
A la suite de mesures sur un modèle mécanique des conduits oral et nasal, MERLIER (1984 p. 143 - 146) a suggéré que l'intensité de la source buccale et la proximité de l'extrémité nasale permettaient l'existence d'un couplage nasal externe (figure 1.42). Ce couplage, présent que le conduit nasal soit ou non couplé au conduit oral, est visible à la figure 1.43. L'impédance d'entrée du conduit vocal est modifiée par le couplage nasal : le premier formant du [a], à une fréquence très proche de la résonance naturelle de la cavité nasale, est affaibli d'environ 4 dB. La courbe présente d'ailleurs une épaulement sur le flanc de basse fréquence qui est la trace d'une seconde résonance. L'intensité d'une résonance de couplage externe sera faible en dehors des fréquences formantiques car l'intensité rayonnée à la bouche est faible. Nous proposons donc d'identifier le pic de 780 Hz de la figure 1.41 comme une résonance de couplage externe. D'autres analyses montrent une résonance à une fréquence un peu plus élevée, jusqu'à 900 Hz. Ces fréquences sont compatibles avec le conduit nasal vu comme un résonateur de HELMHOLTZ, le col étant aux narines. Pour un volume de 40 cm³, un col de 1.5 cm de long et de 2 cm² d'aire, la fréquence corrigée de résonance est de 760 Hz. On peut également imaginer un renforcement de cette résonance nasale par transmission d'énergie à travers le voile du palais en position haute (couplage transvélaire). Nous ignorons si l'énergie transmise est suffisante pour rendre cette hypothèse plausible. Ces deux mécanismes ne sont pas mutuellement exclusifs. Les paragraphes précédents ont présenté la plupart des "outils" qui nous serviront à étudier la nasalité vocalique. Mais il convient tout d'abord de faire notre miel des nombreux travaux déjà consacrés à ce problème.

8 - LES INDICES ACOUSTIQUES DE LA NASALITE VOCALIQUE : revue interprétative des études antérieures.

Dans ce long paragraphe, nous allons passer en revue les études antérieures les plus importantes portant sur la caractérisation des indices acoustiques de la nasalité vocalique. Il s'agit, tout en présentant avec scrupules les conclusions des auteurs, de cerner certaines difficultés et de proposer des solutions alternatives.

8.1 - FANT (1960)

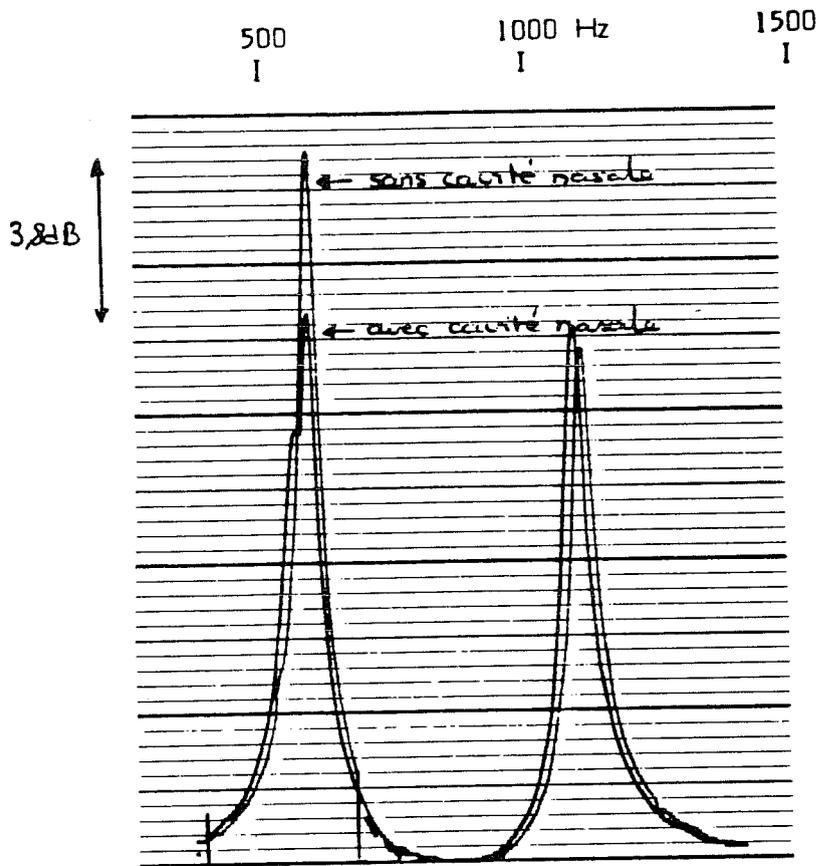
"The Acoustic Theory of Speech Production" de FANT (1960) reste un monument incontournable pour tout de ce qui concerne le domaine. Saluons par exemple sa



Conduit vocal avec fosses nasales.

Figure 9.7 : Mise en évidence du couplage externe avec les fosses nasales.

Figure 1.42 - Couplage nasal externe (in MERLIER 1984).



Comparaison des deux premières résonances
du conduit vocal $|A_0|$,
avec ou sans couplage externe avec les fosses nasales.

Figure 9.8 : Impédance d'entrée du conduit vocal ;
Mise en évidence de l'effet du couplage nasal externe
en l'absence de couplage nasal interne.

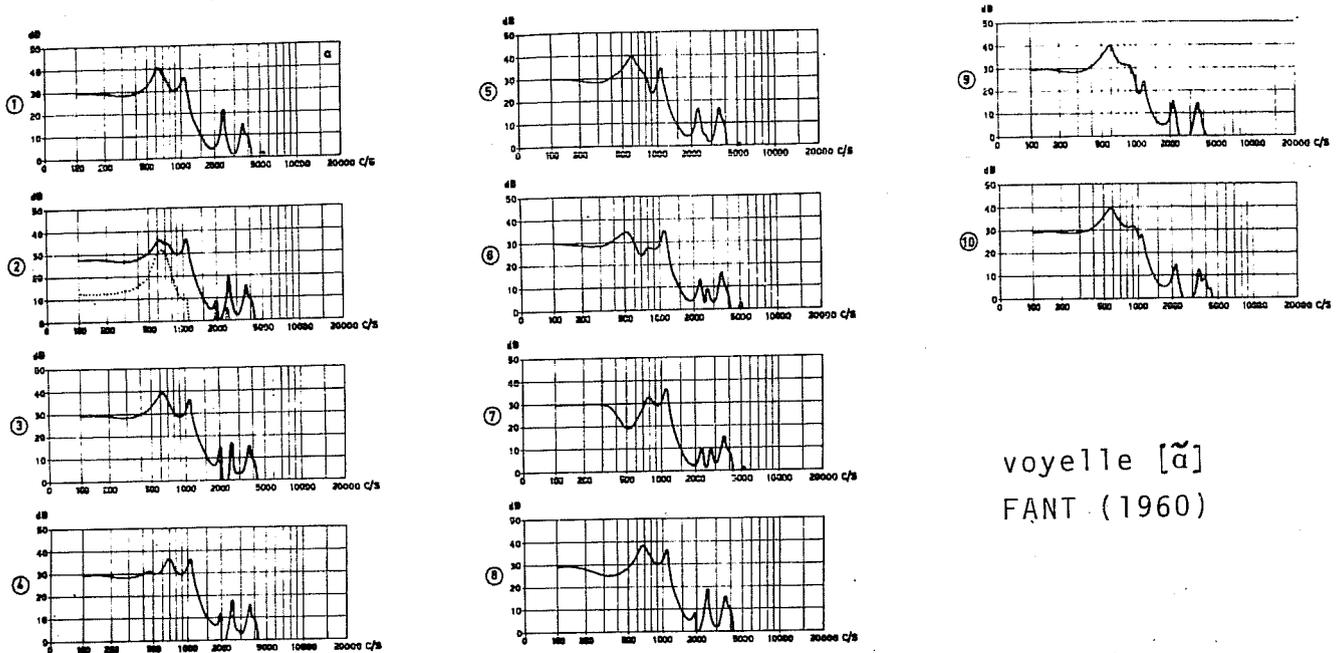
- Figure 1.42 - Impédance d'entrée du conduit vocal seul : mise en évidence d'un effet spectral du couplage nasal externe.
(in MERLIER 1984)

fantastique culture phonétique qui lui fait affirmer comme une évidence dès cette époque que les voyelles nasales françaises "sont réalisées avec une position linguale spécifique en plus de la nasalisation "(p. 138). Pour les nasales et la nasalisation (p. 139-161), il offre tout d'abord (p. 148-149) une revue compacte des travaux antérieurs sur la nasalité vocalique : JOOS (1948), SMITH (1951), DELATTRE (1954) , HOUSE & STEVENS (1956) et HATTORI & al. (1956; cf 1958). Les indices suivants (sous F3 seulement !) se dégagent :

- 1 - Un formant supplémentaire au-dessus de F1 vers 1000 Hz (JOOS, SMITH), plus d'autres plus haut (HATTORI, HOUSE & STEVENS), entre les formants vocaux (2000 Hz pour SMITH, DELATTRE);
- 2 - Un zéro vers 900 Hz, souvent peu visible (JOOS, HOUSE & STEVENS, DELATTRE d'après FANT), ou vers 500 Hz (HATTORI), et vers 1800 Hz (HOUSE & STEVENS);
- 3 - Affaiblissement de F1 (SMITH, DELATTRE (indice principal) , HOUSE & STEVENS);
- 4 - Un formant nasal vers 250 Hz (DELATTRE (indice secondaire), HATTORI)
- 5 - F2 affaibli et rehaussé (SMITH) ou fixe (DELATTRE).

La discussion de Fant s'ouvre sur une présentation des fonctions d'aire. Celle du conduit nasal est reproduite ici à la figure 1.44. Bien que tracée à une petite échelle, elle n'apparaît pas très différente de celle décrite plus tard par BJUGGREN & FANT (1964 : cf. § 8.3 pour une analyse détaillée), sinon que sa longueur est supérieure de 1.7 cm. Ses fréquences propres, vraisemblablement en conservant la forte section postérieure illustrée, sont de 500, 2000 et 3200 Hz , d'après la figure 1.2.4-3 (non reproduite). Les simulations des consonnes nasales ne nous concernent pas directement à l'exception de celles de [ŋ]. Avec la partie pharyngale de la figure 1.44 (celle d'un [ŋ] palatalisé, dont on notera la grande taille), le premier formant s'établit à 300 Hz, F2 à 1000 et F3 à 2200 Hz. F1 est la fréquence de résonance du conduit entier, alors que F2 et F3 semblent dépendre du volume nasal et pharyngal respectivement. Il est clair qu'un conduit pharyngal de taille plus raisonnable aurait élevé la fréquence de F1 au-dessus de 300 Hz, comme dans nos propres simulations.

La discussion de la simulation d'un [a] nasalisé ([ã] p. 155) est très complexe. Il est utile de se reporter au schéma de réactance que nous avons déduit de l'exposé (figure 1.45) ainsi qu'à la reproduction de la figure 1.2.4 - 5 (figure 1.46). F1 et F2 se placent vers 630 et 1070 Hz. Pour un faible couplage (courbe 3 : 0.16 cm²), le formant nasal apparaît vers 660 Hz, ce qui indique que la fréquence caractéristique du conduit nasal est entre 630 et



voyelle [ã]
FANT (1960)

Fig. 2.4-5. The effect of nasal coupling on the calculated spectrum envelope of the vowel [ã]. Glottis resistance $5 \rho c$.

- (1) No nasal coupling;
- (2) Mouth output (solid curve), nasal output (dotted curve), assuming a coupling area of 0.16 cm^2 at the naso-pharynx;
- (3) The sum of the mouth output and the nasal output of (2);
- (4) Same as (3) but the area at the narrowest region of the nasal tract outlet reduced to 0.32 cm^2 ;
- (5) Same as (3) but for increase of the naso-pharyngeal coupling area to 0.65 cm^2 ;
- (6) Same as (5) but for a reduction of the nasal tract outlet as in (4);
- (7) Same as (6) except for a complete obstruction at the nostrils;
- (8) Same as (7) but the naso-pharyngeal area reduced to 0.16 cm^2 ;
- (9) Normal nasal tract outlet area. Naso pharyngeal coupling area 2.6 cm^2 ;
- (10) Same as (9) except for a reduction of the area at the entrance to the mouth cavity to 0.65 cm^2 caused by an approach of the velum toward the back of tongue.

Figure 1.46 - Simulations de la fonction de transfert de la voyelle [ã] (in FANT 1960).

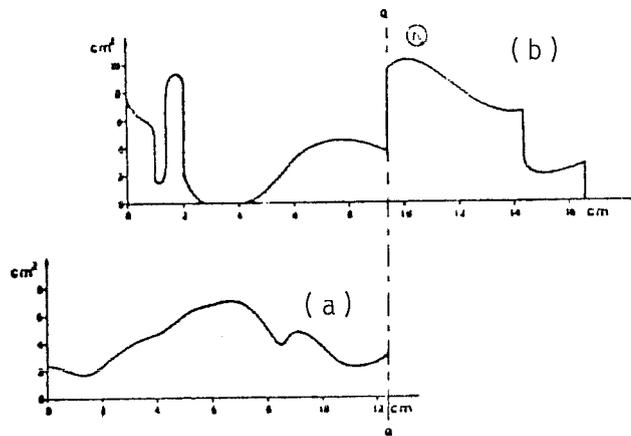


Figure 1.44 - Fonction d'aire du conduit nasal de FANT (1960): a fonction d'aire du conduit pharyngal de la consonne nasale vélaire | | : b (in FANT 1960).

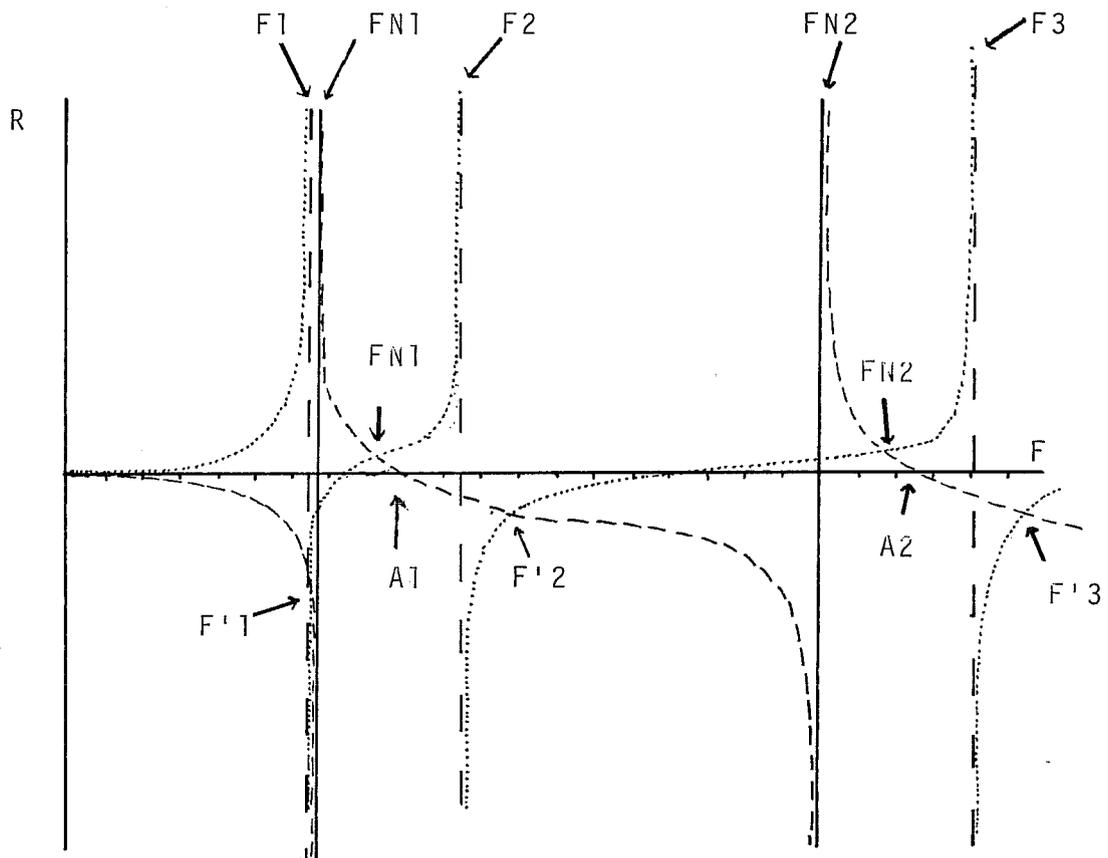


figure 1. 45 - Schéma de réactance déduit de FANT (1960) : voyelle | |.

660Hz, très proche donc de F1. Au vu de la valeur de FN1, on doit supposer une forte réduction de l'aire de la partie postérieure du conduit nasal dont la première fréquence de résonance était de 500 Hz seulement à la figure 1.2.4-3. Le faible degré de couplage implique que l'élévation des formants est très modérée. F'1 doit donc être très voisin de 630 Hz, et FN1, à 660 Hz, très proche de A1 (noté Fzn1 par FANT à 700 Hz). Deux formants et un zéro en 70 Hz conduisent nécessairement à un pic unique vers 650 Hz, dont la largeur de bande est un peu augmentée, d'où une réduction faible de l'intensité (-1dB). F'2 est proche de F2. L'indice le plus clair, au plan acoustique, de cette faible nasalisation est la zone vers 2 - 3 KHz, où l'on trouve dans l'ordre FN2, A2 et F'3, avec F4 visible vers 3500 Hz.

Une réduction de l'aire aux narines (courbe 4) abaisse le formant nasal sous F'1, vers 450 Hz. Si le couplage est plus important (0.5 cm^2 : courbe 5), le zéro atteint 900 Hz. FN1 est alors également un peu plus haut et forme une "épaule" sur le flanc descendant du premier pic. Le groupe supérieur ne forme plus qu'un seul pic. Une réduction de l'aire aux narines (courbe 6) montre à nouveau FN1 vers 500 Hz et F'1 presque annihilé par A1 (-8 dB). Une fermeture complète aux narines (courbe 7) pousse FN1 sous 300 Hz. Si, dans ce dernier cas, on réduit le couplage à 0.16 cm^2 (courbe 8), A1 s'abaisse et détruit toute trace de FN1. En revanche, comme A1 s'éloigne alors de F'1, ce dernier regagne en intensité. Cette voyelle est presque orale, en dehors de la zone vers 2 KHz. Si on augmente fortement le couplage (2.6 cm^2 : courbe 9), avec une aire aux narines normales, A1 s'approche de F'2 qui disparaît presque. Si on simule enfin la constriction vélaire du conduit oral (courbe 10) provoqué par l'abaissement du voile, F'2 s'abaisse un peu et est alors complètement masqué par A1. Nous retiendrons principalement de cette simulation que le premier formant nasal peut être si proche de F'1 qu'ils se confondent en un pic unique, et que dans ce cas la baisse d'énergie est peu marquée.

La simulation de [~e], avec une aire aux narines normale, est plus simple à interpréter. Les figures de FANT sont reproduites à la figure 1.47. F1 à 420 Hz est plus bas que FN1. Un degré important de nasalisation (courbe 3) et le relatif écartement de FN1 et F1 conduit à une augmentation de la fréquence de F'1 (490 Hz). La réduction de l'amplitude est notable (-4dB), ce qui montre que l'amplitude de F'1 peut varier en dehors de la présence d'un zéro proche par le simple jeu du couplage : A1 est en effet à 1350 Hz environ, 900 Hz au-dessus de F'1, FN1 passe de 650 Hz à 1100 Hz, et ils deviennent visibles comme une légère irrégularité entre F'1 et F'2. Cette voyelle ne possède néanmoins pratiquement aucune trace de nasalité en dehors de l'affaiblissement somme toute modéré de F'1. Une réduction du couplage à 0.65 cm^2 , puis à 0.16 cm^2 , abaisse le zéro (A1) à 1000, puis 800 Hz (courbe 4 et 5). Le spectre est celui d'une voyelle orale.

On retiendra surtout la réduction d'amplitude de F'1 qui peut être également amenée par le

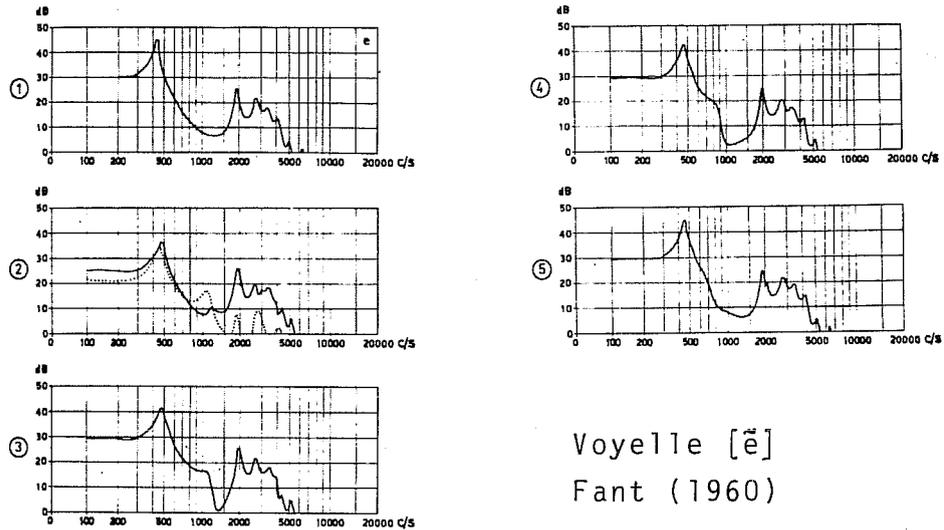


Fig. 2.4-6. The effect of nasal coupling on the spectrum envelope of the vowel [ẽ]:
 (1) No nasal coupling;
 (2) 2.6 cm^2 naso-pharyngeal coupling area. The solid line represents the mouth output and the dotted line the nose output;
 (3) The sum of the nasal and the oral outputs of (2);
 (4) Same as (3), except for a naso-pharyngeal coupling area of 0.65 cm^2 ;
 (5) The naso-pharyngeal coupling area reduced to 0.16 cm^2 .

Figure 1.47 A - Simulations de la fonction de transfert de la voyelle [ẽ] (in FANT 1960).

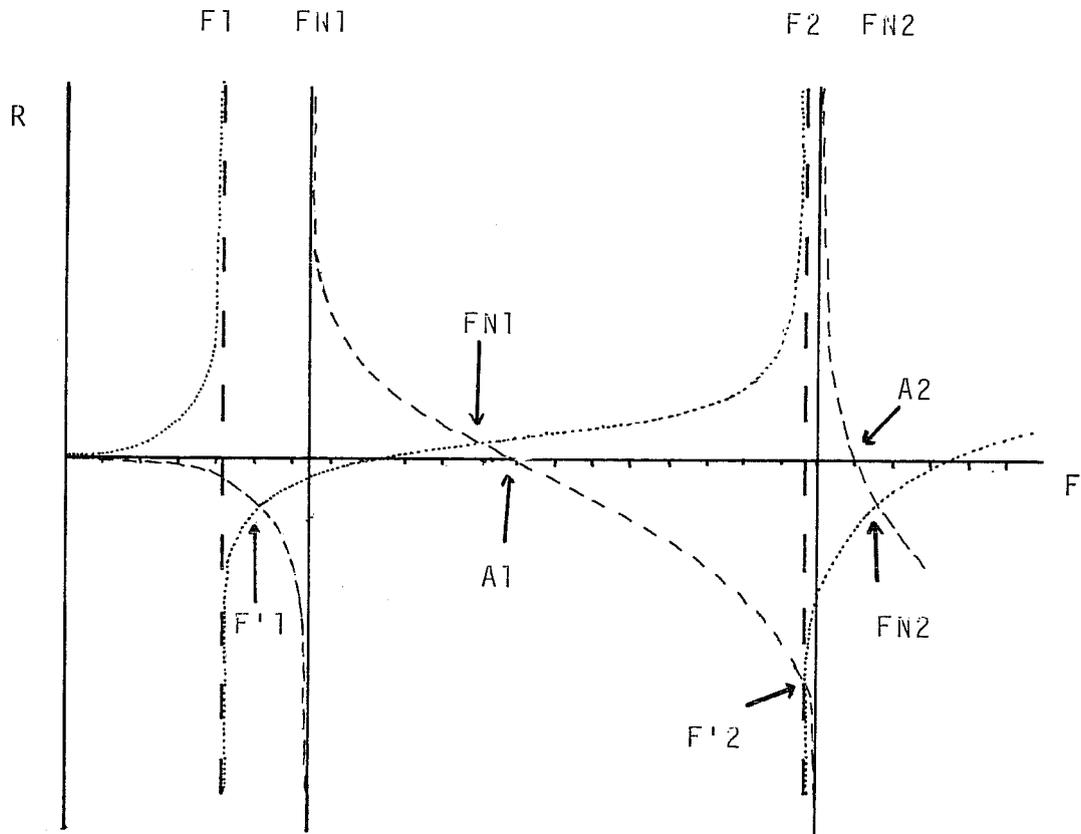


Figure 1.47 B - Schéma de réactance déduit de FANT (1960) :
 voyelle [ẽ]

simple jeu du couplage, sans "l'aide" d'un zéro de couplage.

Les simulations des voyelles fermées [\sim i et \sim u] n'apportent que peu d'informations nouvelles .

FANT note également que la présence d'une paire pôle / zéro provoque au delà de sa zone d'influence propre un affaiblissement constant du spectre égal à $40 \cdot \text{Log}(A_i / F_i)$. - 6 dB est donc une valeur typique. Mais il faudrait également tenir compte de l'effet des pôles supplémentaires sur le niveau des formants supérieurs (cf. § 12.1 pour une discussion). De l'analyse de deux voyelles [a] naturelles nasalisées par une consonne nasale précédente, il conclut que le premier zéro se trouve vers 750-800 Hz, et F1N vers 1000 Hz, comme dans un exemple de DELATTRE (1954). Il attribue l'affaiblissement de F1 (-2 et -5 dB) à la proximité du premier zéro. Il mentionne sans explication la baisse de fréquence de F1, alors que la théorie prévoit une augmentation. On peut supposer que le zéro proche "creuse" le flanc droit de F'1 et décale ainsi visuellement son sommet vers les basses fréquences. Pour expliquer le formant nasal très bas visible à la figure 1 de HATTORI & al. (1956), il accepte l'hypothèse d'une forte constriction aux narines.

8.2 - HOUSE & STEVENS (1956)

Nous présentons ce travail après celui de FANT parce que la comparaison nous a semblé plus fructueuse dans ce sens. Il est entièrement consacré à une modélisation acoustique. La forme du conduit nasal retenue (cf. § 9.1), bien que fondée sur "la consultation d'atlas d'anatomie, de crânes et de clichés radiologiques sagittaux", est très différente de celle de FANT. Elle montre en particulier une forte constriction dans la zone des narines. Cette forte constriction conduit à une fréquence basse du premier formant nasal. La longueur est de 12.5 cm. De plus, un amortissement important a été donné au premier formant (200 Hz de largeur de bande pour FN1). Sans que cela soit absolument explicite dans le texte, il est vraisemblable que la simulation de [ŋ] a joué un rôle dans ce choix. Les simulations ne sont pas comparées à des réalisations naturelles, et les conclusions reprennent d'une manière un peu circulaire les choix qui ont dicté la sélection des paramètres initiaux :

- 1 - Réduction de l'intensité du premier formant et élévation de sa fréquence. Cette réduction est due au placement de FN1 sous F1. A1, lors du couplage, vient naturellement réduire l'amplitude de F'1. Le fort amortissement de FN1 l'empêche d'être visible et distinct de F'1.
 - 2 - Réduction de l'intensité de la voyelle, qui ne reçoit pas d'explication théorique.
 - 3 - Apparition d'un zéro entre 700 et 1800 Hz, conséquence automatique du couplage.
- Beaucoup de chercheurs, observant un pic en basse fréquence (250 - 400 Hz) sur le

spectre de voyelles nasales, ont invoqué les simulations de HOUSE & STEVENS pour l'assimiler à FN1.

8.3 - BJUGGREN & FANT (1964)

Ce très bref article, publié dans les Quaterly Progress Report de Stockholm, est notre seule source de connaissance précise sur la fonction d'aire du conduit nasal (figure 1.48 à 1.50). Il présente les données anatomiques utilisées mais non décrites dans FANT (1960). Pour le sujet masculin, on trouve les sections d'un moule de plastique réalisé par un anatomiste sur un cadavre et la fonction d'aire obtenue en additionnant les aires des deux fosses. La valeur de la circonférence est donnée pour permettre le calcul du "facteur de forme" qui atteint 3.5 par endroit. Pour les deux sujets féminins, seul les fonctions d'aire sont présentées, mais elles distinguent les fosses gauche et droite. Il est important de noter que l'aire aux narines pour les sujets féminins est encore plus importante (2.2 et 2.8 cm² pour la somme) que pour le sujet masculin (1.5 cm²). La longueur le long du "centre de gravité" est d'environ 11 cm pour le sujet masculin et de 9 et 10 cm pour les femmes. Aucune allusion n'est faite quant à une possible distortion due à l'insertion du plastique, ou à une modification des tissus post-mortem ou par le procédé de conservation. L'assymétrie des aires des deux fosses est soulignée, mais elle apparaît somme toute modérée. Des différences importantes sont visibles pour le rhinopharynx, qui ont déjà été soulignées à la figure 1.8. La première fréquence de résonance du conduit nasal masculin fermé à l'extrémité vélaire est de 500 Hz d'après FANT, ce qui paraît trop faible d'au moins 100 Hz d'après toutes les simulations effectuées : MAEDA (1982 a), MERLIER (1984), FENG (1986) et nous même plus loin. Le calcul semble avoir été vérifié à l'aide de la formule de Helmholtz. Cette valeur convient mieux à un conduit de 12.5 cm de long , comme dans FANT (1960). FANT souligne pourtant que cette valeur est sans doute inférieure, en raison de la "présence d'éléments de masse aux coudes et aux orifices internes , liés au paramètre de radiation de l'équation d'onde ... Les individus aux nez étroits sont susceptibles d'avoir des fréquences encore plus basses...". Il est clair que la fréquence relativement élevée de la première résonance est un problème qui a préoccupé l'auteur.

8.4 - FUJIMURA & LINDQVIST (1971)

Cette étude, d'un intérêt capital, et qui traite de bien d'autres problèmes que de celui de la nasalité, mérite une analyse attentive. Elle reprend des matériaux présentés quelques années

Fig. 1.48

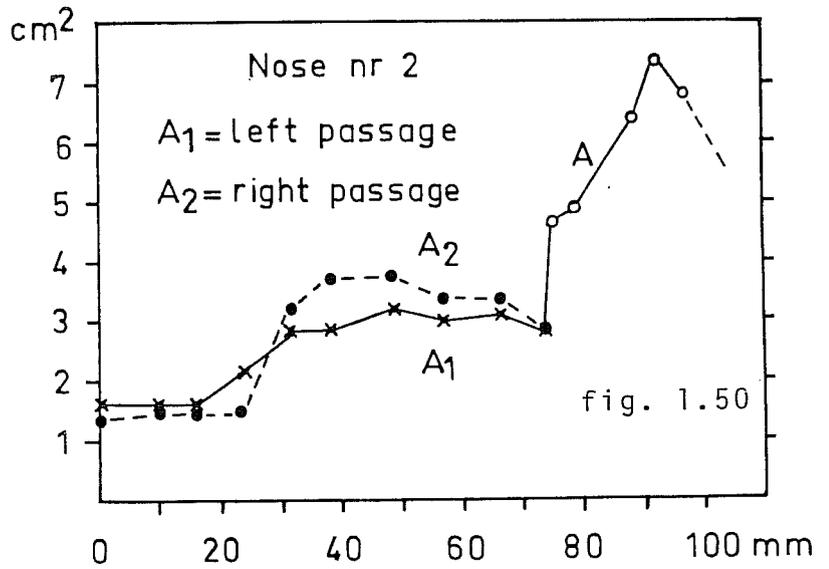
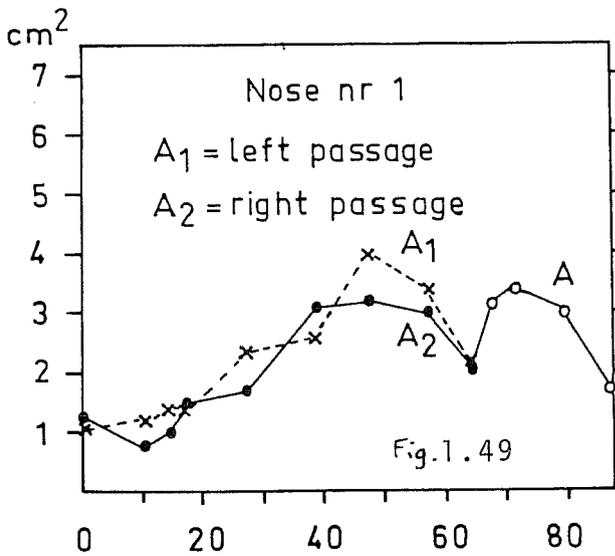
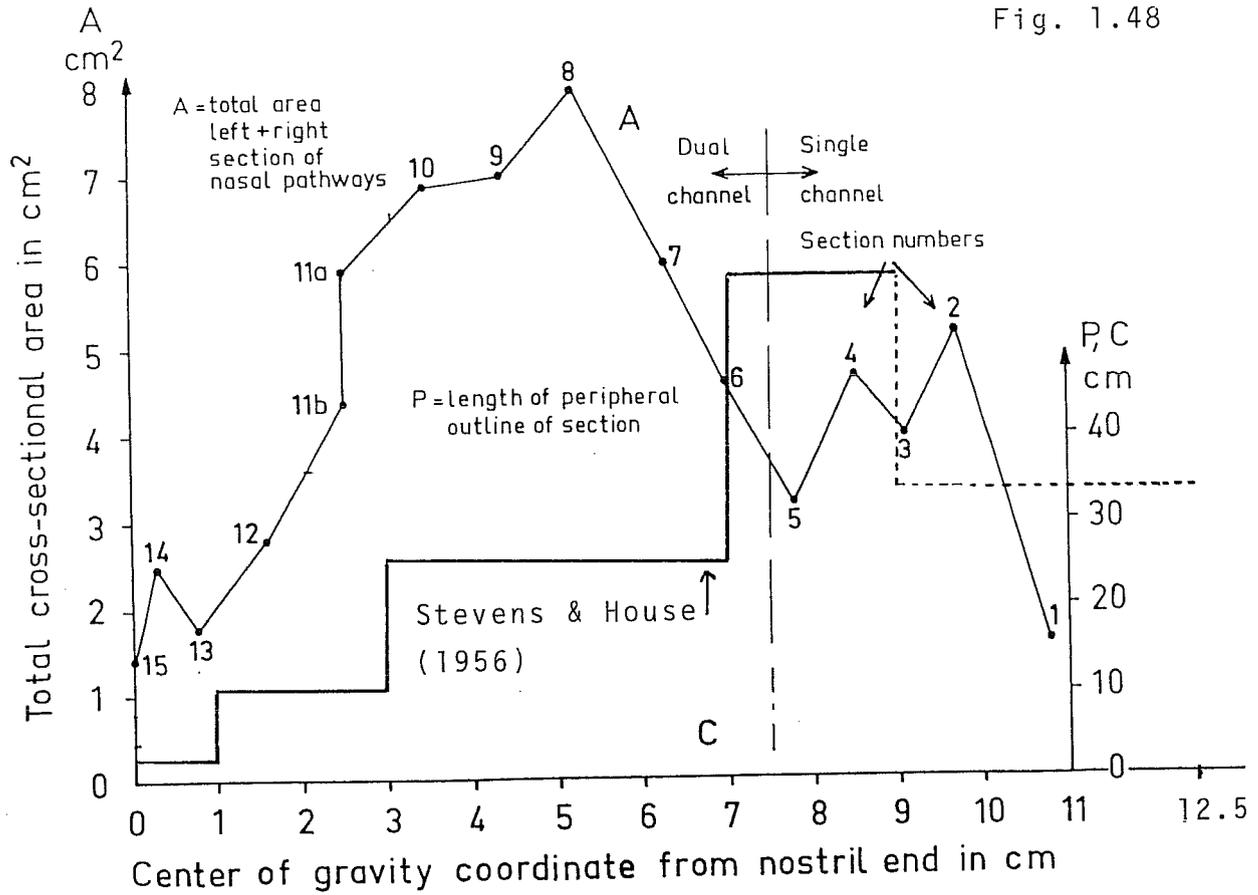


Figure 1.48 - Fonction d'aire nasale (in BJUGGREN & FANT 1964), et fonction d'aire de STEVENS & HOUSE (1956).

Figures 1.49 & 1.50 - Fonctions d'aire nasales de 2 sujets féminins (in BJUGGREN & FANT 1964).

auparavant (FUJIMURA & LINDQVIST 1964a, 1964b). Les auteurs présentent des courbes de réponse obtenues à l'aide d'un vibreur externe appliqué au niveau du larynx. La distortion spectrale inhérente est supprimée par le calcul d'une correction par une technique d'analyse par synthèse itérative. Le vibreur externe permet de court-circuiter les effets spectraux de la source et d'obtenir une bien meilleure résolution. L'étude contient en outre le résumé très complet de la théorie acoustique du couplage nasal mentionné plus haut. La figure 1.51 (figure 1.14 et table II de l'article) compare une version orale (a) et nasale (b) de la voyelle neutre [ə]. Les courbes (c) et (d) sont les résultats de l'analyse par synthèse avec et sans la correction de non-linéarité. Il est clair que la différence essentielle réside dans l'apparition d'une paire pôle / zéro (270 / 310 Hz) dans la version nasalisée que les auteurs identifient comme FN1. On notera également un pôle supplémentaire vers 1830 Hz, suivi d'un fort zéro, car il n'y a aucune raison théorique pour que F'3 soit décalé vers le bas après couplage. Il s'agit donc de FN2. L'analyse par synthèse a de légères difficultés à modéliser la zone comprise entre 700 et 1700 Hz. Selon les auteurs, c'est l'indice de la présence d'une deuxième paire pôle / zéro. L'amplitude de F'1 est également réduite. Nous voulons discuter l'hypothèse que la paire de basse fréquence est la trace d'une résonance liée à une cavité sinusale, et que les irrégularités entre F'1 et F'2 sont causées par le véritable FN1. La difficulté essentielle est que ni le pic (FN1), ni le zéro (A1) ne sont clairement visibles. Il faut admettre que F1N coïncide exactement avec F'1 (517 Hz) ou avec F'2 à 1230 Hz, ce qui est relativement élevé. Les irrégularités peuvent également être dues à d'autres paires pôle / zéro liées à des cavités sinusales différentes. En résumé, nous devons admettre soit que FN1 est beaucoup plus bas que ne le laisse supposer les simulations de FANT, impliquant une forte constriction nasale, soit, ce qui paraît de prime abord moins probable, qu'il est de fréquence élevée, impliquant une large ouverture nasale et un conduit nasal court. Notre figure 1.52 (figure 1.15 de l'original) présente les 3 voyelles [ā, ō, ~u]. Les configurations notées "og"(open glottis) ne nous concernent pas directement. On retrouve un pic et une vallée vers 370-400 Hz et 450-480 Hz respectivement, et F'1 est décalé vers les hautes fréquences. Aucun autre indice clair n'apparaît sous 2000 Hz, excepté la réduction d'amplitude de F'1, vraisemblablement sous l'effet du zéro très proche au-dessous. F'2 en revanche est bien visible vers 2 KHz, sous la forme d'un "F3 éclaté" ("split F3") pour [~u]. Une possibilité, si on ne veut pas admettre que le premier pic est FN1, est de considérer que FN1 se trouve par hasard pour les trois voyelles à la même fréquence (chaque fois différente!) que F'1 ou F'2, le zéro nasal (A1) tombant dans la vallée entre F'2 et FN2. Un indice (?) pourrait exister : F'2 est plus intense que F2 pour [~u] et [ō], ce qui suggère la présence d'énergie supplémentaire apportée par FN1, et la largeur de bande de F'2 de [ā] est large, comme dans le cas de deux pics de fréquence très proche. Une autre variante est

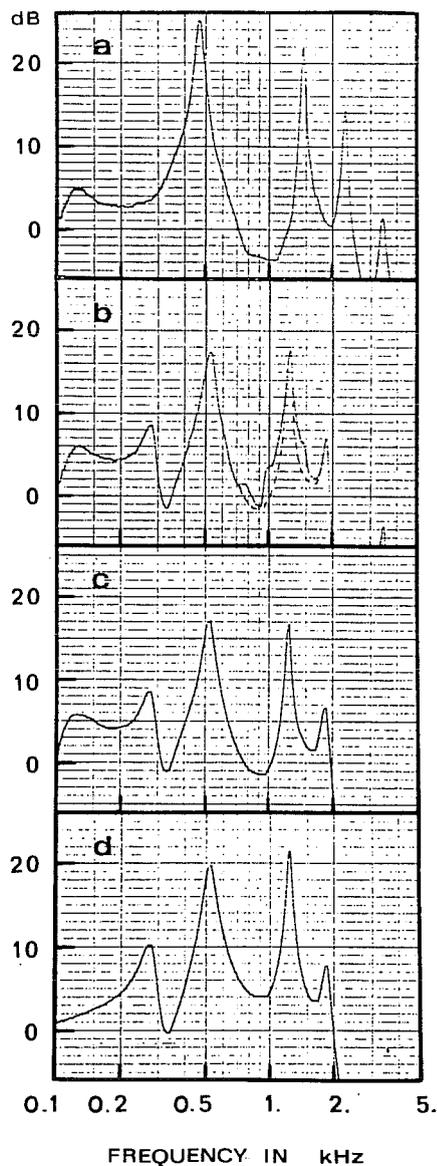


FIG. 14. Data of matching articulations of a neutral vowel with and without nasalization. (a) Recorded response curve for [ɜ̃]. (b) Same for [ɜ̃] (the broken line is an overlaid tracing of Curve c. (c) Synthesized curve for [ɜ̃]. (4) Same without the correction function (the correction function is given in Fig. 5, Curve 625-II).

TABLE III. Pole and zero frequencies for non-nasalized vowel [ɜ̃] and a nasalized vowel with approximately the same tongue articulation [ɜ̃̃].

	pole 1	pole 2	pole 3	extra pole	zero
[ɜ̃]	470 (33)	1450 (42)	2300 (62)
[ɜ̃̃]	517 (56)	1230 (55)	1830 (154)	270 (46)	312 (60)

Figure 1.51 - Fonctions de transfert d'une voyelle [ə] orale et nasale mesurées par un vibreur externe (in FUJIMURA & LINDQVIST 1971).

FUJIMURA AND LINDQVIST

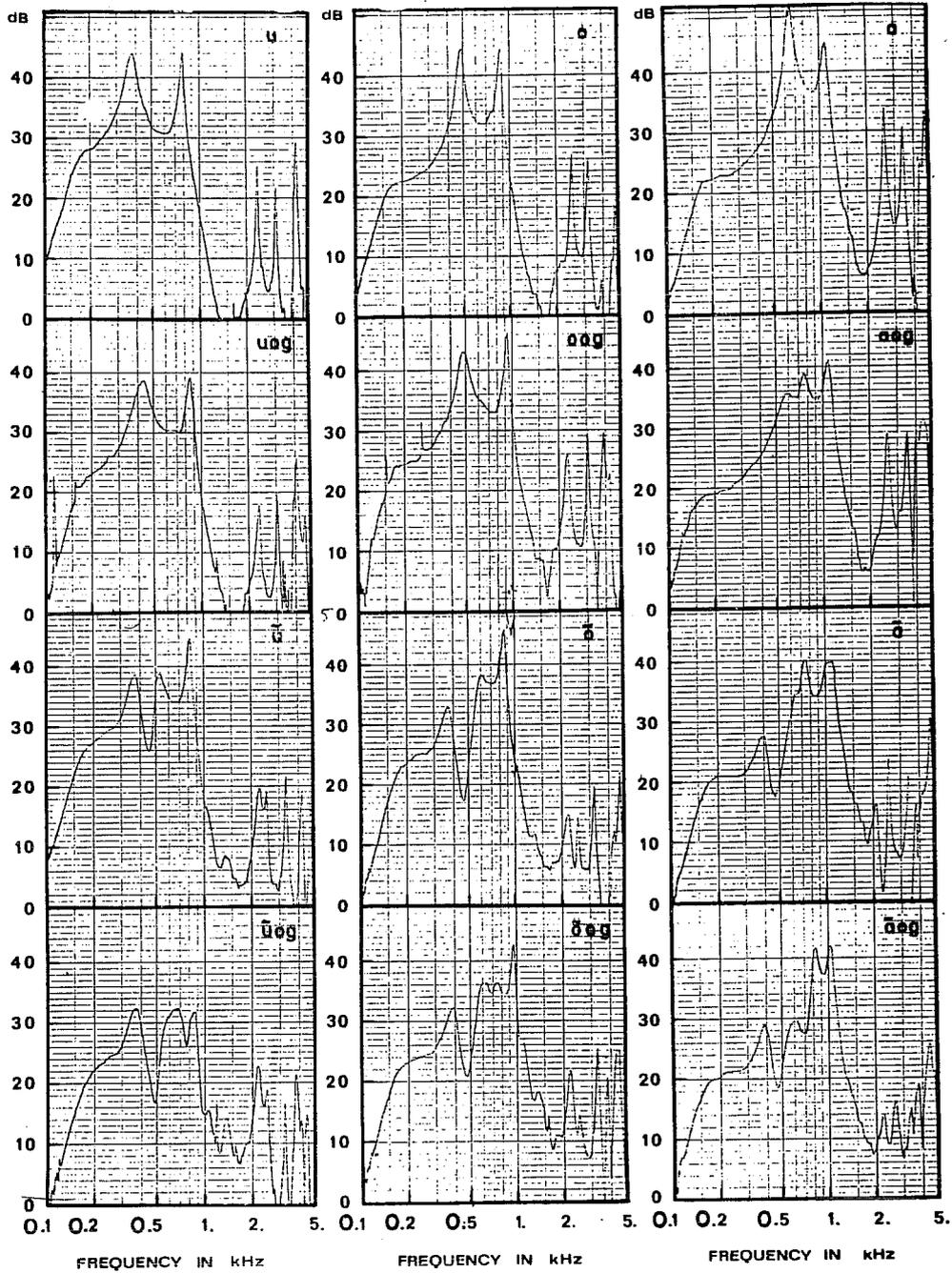


FIG. 15. Comparisons of nonnasalized, open glottis, and nasalized conditions, where "og" stands for open glottis.

. Figure 1.52 - Fonctions de transfert des voyelles |u|, |o| et |a| orales et nasales mesurées à l'aide d'un vibreur externe (in FUJIMURA & LINDQVIST 1971).

de voir la trace de FN1 dans le pic ou l'épaule du flanc gauche du deuxième pic de [ã]. Pour les autres voyelles, FN1 se trouverait confondu avec F'1 vers 550 - 600 Hz. Une dernière possibilité serait que le deuxième pic soit FN1, et que F'1 soit confondu avec F'2. Cette étude, qui a exercé une forte influence sur de nombreux travaux postérieurs, a toujours constitué la meilleure preuve de l'existence d'un premier formant nasal de fréquence relativement basse. Cette interprétation est plausible, mais il est peut-être possible d'offrir une interprétation différente, si elle est fondée sur d'autres arguments solides. On ne peut que regretter que les superbes tracés de FUJIMURA & LINDQVIST ne soient pas plus explicites.

8.5 - MRAYATI (1975; 1976)

Le chapitre de la thèse de MRAYATI consacré aux voyelles nasales reste le seul travail d'ensemble disponible pour le français. Après une revue succincte de quelques études antérieures, il fournit les schémas de réactance pour les conduits de FANT et STEVENS & HOUSE à différents degrés de couplage. L'examen de cette figure (reproduite ici figure 1.53) est très instructive. Rappelons que les fréquences où la réactance est infinie correspondent aux formants de nasalité (FN_i) et celles où elle est nulle aux zéros (A_i). Le conduit de STEVENS et HOUSE produit un FN1 bien plus bas que celui de FANT : 300 contre plus de 500 Hz. On remarque que l'influence du degré de couplage est faible sur la position fréquentielle de FN1. En revanche, pour les deux conduits, les FN2 sont identiques pour les valeurs significatives de couplage. L'allure presque verticale de la branche de la réactance vers FN2 aux faibles valeurs de couplage montre que le zéro est très proche et que la contribution du pôle sera négligeable dans ce cas. Pour FN1, le zéro associé, beaucoup plus proche de FN1 que de FN2, reste toujours à une certaine distance de FN1 pour tous les degrés de couplage.

Les simulations des voyelles nasales françaises reposent sur la forme du conduit de STEVENS & HOUSE. Bien qu'aucune raison ne soit donnée, il est évident que ce sont les résultats des mesures sur les voyelles naturelles qui ont dicté ce choix. Ces mesures, faites par L.P.C., donnent presque toutes un formant vers 350 - 450 Hz. Seul un conduit nasal possédant une forte constriction antérieure peut en rendre compte. Les conclusions de l'auteur en découlent tout naturellement. L'ordre des premiers pics et de la vallée spectrale est donc stable pour MRAYATI : FN1 - A1 - F'1, sauf peut-être pour [õ], où F1 est proche de la première fréquence caractéristique du conduit nasal. L'amplitude de F'1 est faible car A1 est à proximité. Nous avons déjà critiqué (LONCHAMP 1978) le spectre calculé de [ã] (figure 1.12). Les mesures L.P.C. pour 5 sujets sont ensuite présentées,

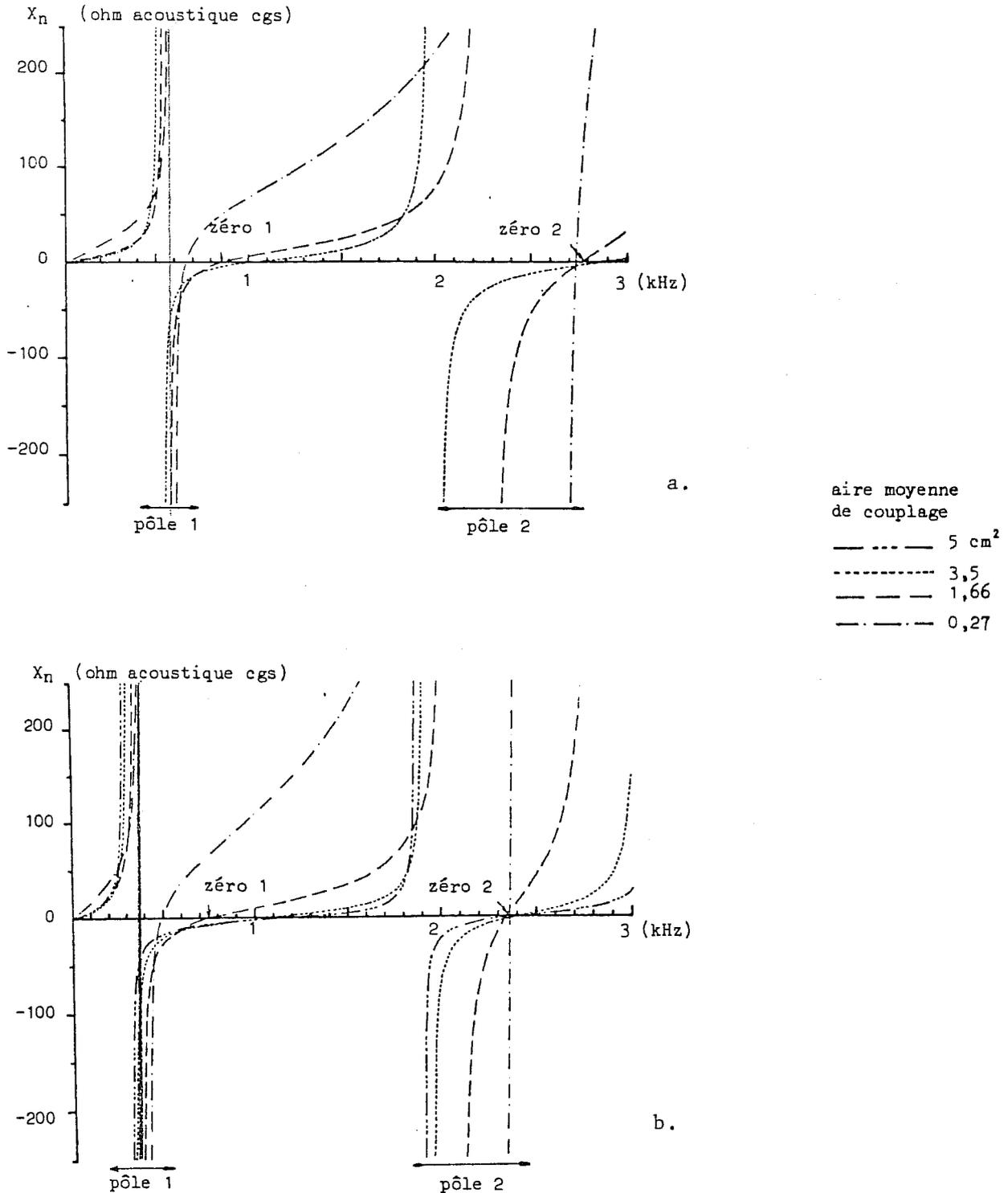


Figure 7. Réactance d'entrée du conduit nasal au point de couplage, d'après les dimensions données par :

a : FANT (1960),

b : STEVENS et HOUSE (1961).

· Figure 1.53 - in MRAYATI (1975): cf aussi figure 1.29 B.

ainsi qu'une comparaison entre les voyelles orales et nasales d'un sujet. Nous reviendrons plus loin sur ces données. Dès à présent, on peut faire remarquer que l'analyse L.P.C., fondée sur un modèle tout pôle d'ordre limité, n'est pas l'outil le plus approprié. Cette analyse peut, comme le montre la figure 1.54, fournir une estimation variable pour des spectres voisins. Ces deux [ɛ̃] possèdent un spectre F.F.T. très semblable. Pourtant, dans un cas, l'analyse L.P.C. fournit un pic vers 650 Hz, et dans l'autre deux pôles à 340 et 750 Hz de très fortes largeurs de bande. En fait, visuellement, les maxima se placent vers 540 et 720 Hz.

Le travail de MRAYATI a fourni le cadre de notre interprétation antérieure (LONCHAMP 1978) où nous n'attirions pas assez l'attention sur le fait que les analyses théoriques dépendaient du choix de la fonction d'aire nasale.

Dans une brève communication au Symposium 'Modèles articulatoires et Phonétique', GUERIN & MRAYATI (1977) montrent sur des simulations un effet négligeable du couplage nasal pour les voyelles [a] et [ɛ] : F1 s'abaisse de 10 et 3 Hz, F2 de 1 Hz. Le troisième formant de [a] est le plus sensible au couplage : il diminue de 110 Hz. Celui de [ɛ] ne chute que de 17 Hz. Aucun détail précis sur la forme du conduit nasal ou le degré de couplage n'est fourni. L'étude des fonctions de sensibilité confirme, selon les auteurs, cette conclusion. Ces résultats paraissent curieux. On sait que le couplage augmente la fréquence de F1. La seule explication est que la mesure de F1 concerne en fait FN1. La forme du conduit nasal et le degré de couplage placent sans doute FN1 à la même fréquence que le premier formant de la voyelle orale homologue. F'1 n'a sans doute pas été mesuré, et doit se trouver beaucoup plus haut. Pour s'approcher des fréquences mesurées sur les voyelles naturelles, GUERIN & MRAYATI confirment qu'il est nécessaire de modifier la forme de la cavité pharyngale.

8.6 - LINDQVIST & SUNDBERG (1976)

Dans un article capital, Lindqvist-Gauffin et Sundberg (1976) montrent que la fonction de transfert des sons nasalisés est trop complexe pour être due uniquement au résonateur nasal couplé au conduit oral. Ils font l'hypothèse que les cavités sinusales jouent un rôle non négligeable. Leurs données empiriques ont été obtenues en excitant le conduit vocal par une source externe au niveau du larynx (cf. FUJIMURA & LINDQVIST 1964, 1971) et le conduit nasal par une source externe couplée à un tube étroit inséré dans le conduit nasal jusqu'à 2 cm du voile. L'intérêt de cette technique la précision de la mesure de la fonction de transfert ne dépend plus des seules fréquences multiples de la fréquence fondamentale. Si on examine les courbes de réponse pour le conduit nasal seul (reproduites

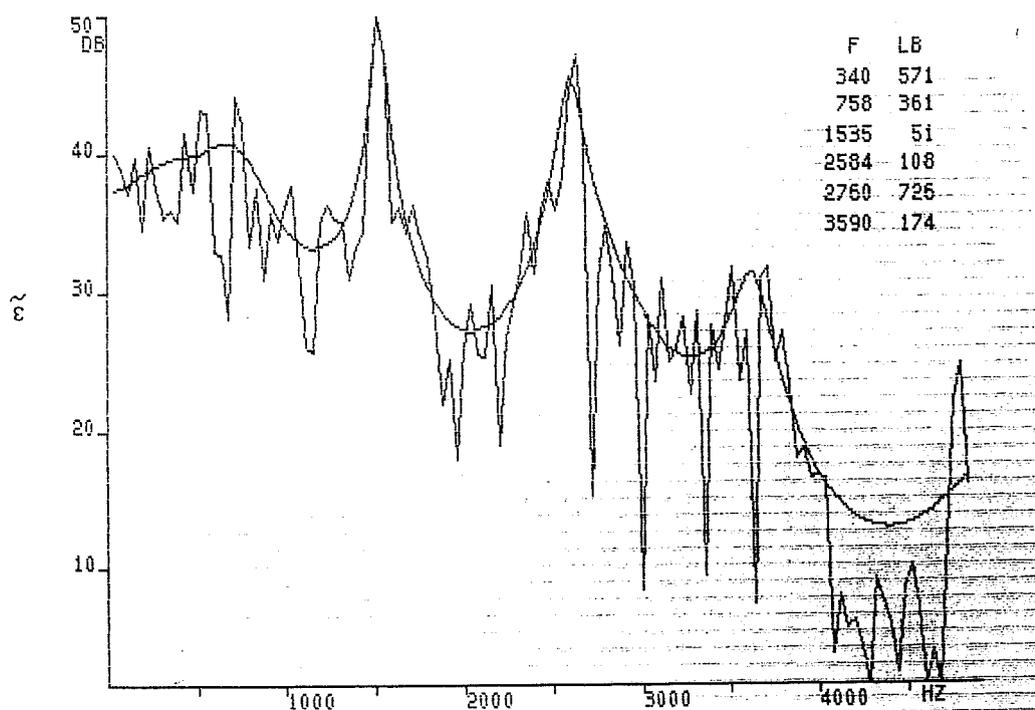
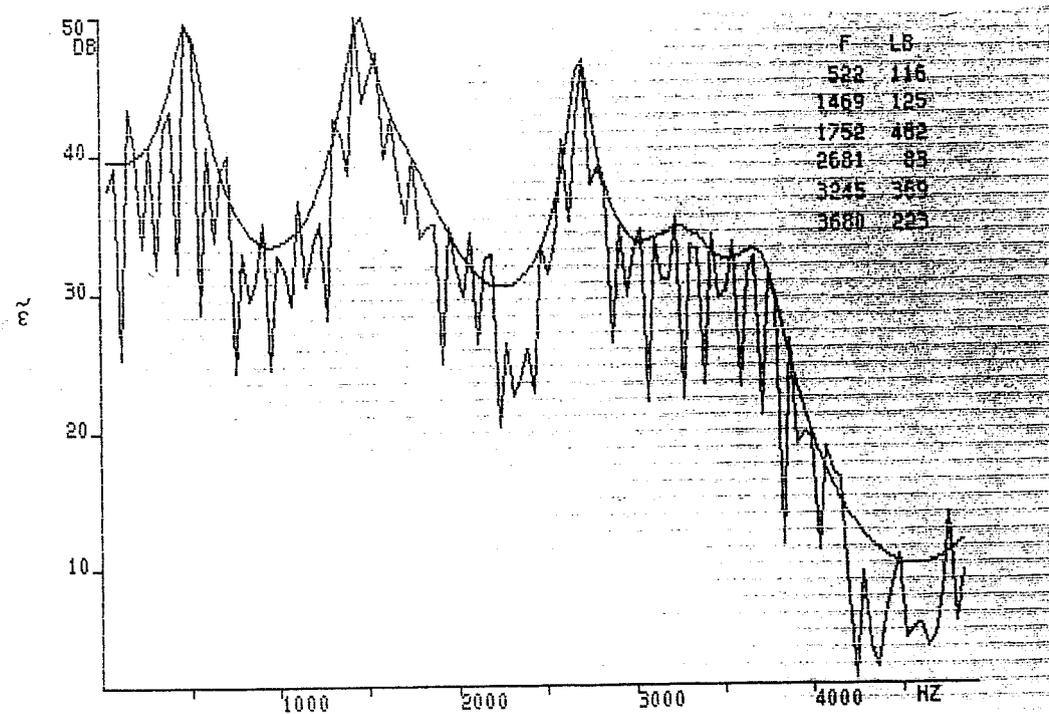


Figure 1.54 - Spectres L.P.C. et F.F.T. de deux voyelles $|\tilde{\epsilon}|$: notez la présence d'un ou de deux pics dans la zone de F'1.

ici à la figure 1.55), on note en basse fréquence pour le sujet J.L. (courbe a) un premier pic intense vers 450 Hz, suivi d'une bosse vers 620 Hz. Pour J.S. (courbe b), il n'y a pratiquement qu'un seul pic vers 570 Hz. Les deux courbes c et d (sujet J.S.) ont été obtenues après traitement de l'une des deux branches de la partie antérieure du conduit nasal à l'adrénaline provoquant un assèchement de la muqueuse nasale. On note l'apparition de nombreux petits pics entre 300 et 1200 Hz et la disparition du pic important vers 500 Hz. L'allure générale de la courbe présente un maximum vers 1 KHz. Ces données suggèrent que :

- Les petits pics sont dus au couplage de plusieurs cavités sinusales (maxillaires, frontales, ou autres) de tailles inégales. Il est bien sûr impossible au seul conduit nasal d'avoir une telle densité de pôles dans un gamme réduite de fréquence. La théorie indique qu'un conduit de 12 cm de long aura un pôle tous les 1400 Hz seulement. Le couplage augmente lorsque les orifices sont bien dégagés par l'assèchement de la muqueuse. Les nombreux zéros associés "aplatissent" le spectre dans la zone du premier formant nasal.

- Le premier formant nasal doit se situer vers 600 Hz. Pour J.L. la fréquence plus basse peut être due, soit à une taille plus grande du conduit ou à une constriction narinaire plus importante, soit à la présence d'une paire pôle-zéro liée à la résonance d'un sinus dont le zéro vient décaler FN1 vers le bas en affaiblissant la zone vers 500-600Hz. On trouve cette possibilité dans la simulation effectuée par les auteurs sur un modèle (figure 1.56). La courbe b, correspondant au conduit nasal de FANT muni de sinus maxillaires et frontaux, montre un pic F1N plus bas que sa position sur la courbe a, calculée sans sinus. Une assymétrie des deux branches nasales (courbe c) conduit également à doubler les pics correspondant aux résonances sinusales dans les cas extrêmes.

La figure 1.57 présente pour deux autres sujets les fonctions de transfert par excitation externe des 3 consonnes nasales [m], [n] et [ŋ]. Les zones en basse fréquence sont très similaires : un pic intense vers 270 Hz suivi d'un pic discret vers 500 Hz. Puis pour [ŋ], des maxima vers 1000 et 2000 Hz. Deux interprétations sont possibles. D'une part, on peut considérer le premier pic comme la première résonance des conduits oral et nasal couplés. Sa fréquence dépend de la longueur totale et de l'aire à l'extrémité ouverte (narines). Pour un conduit uniforme de 21 cm de long, cette fréquence est de 420 Hz environ, donc beaucoup plus élevée que celle présente ici. L'aire aux narines doit donc être faible pour faire baisser celle-ci. D'autre part, on peut envisager que le zéro d'une paire pôle / zéro liée à une résonance sinusale affaiblit les fréquences vers 400 Hz, c'est-à-dire le flanc droit du pic, décalant ainsi ce premier formant vers des fréquences plus basses. La fréquence observée n'est donc plus conforme à la véritable fréquence de la première résonance du conduit. Dans les deux cas, la petite bosse vers 500 Hz est le pôle sinusal, et le formant à 1000 Hz le deuxième formant du conduit. On peut objecter à la seconde

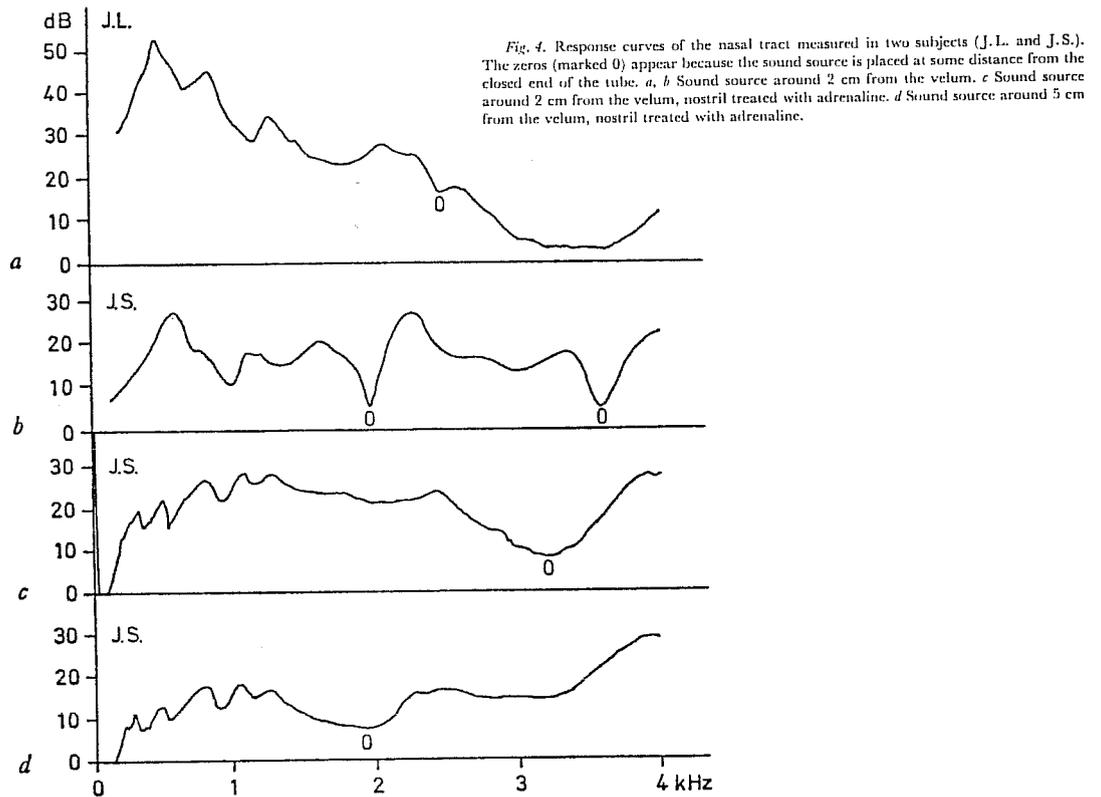


Figure 1.55 - Fonctions de transfert du conduit nasal de deux sujets (mesure par vibrateur externe; in LINDQVIST & SUNDBERG 1976).

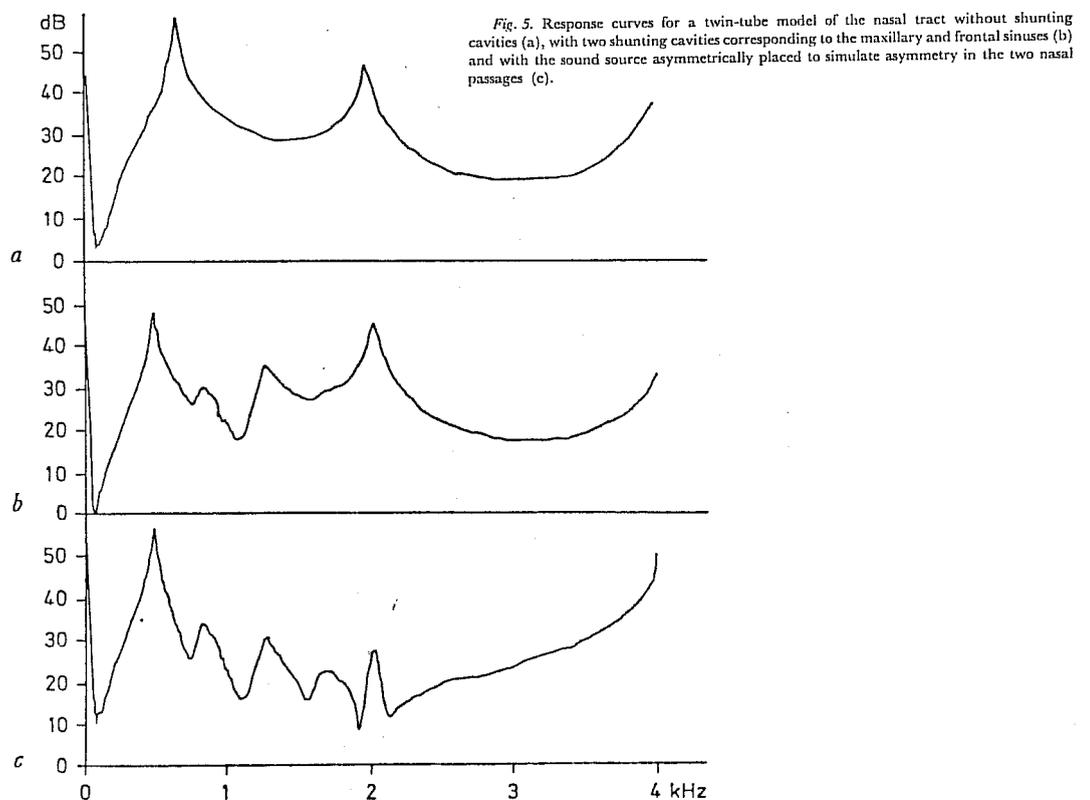


Figure 1.56 - Fonctions de transfert d'un modèle mécanique du conduit nasal (in LINDQVIST & SUNDBERG 1976).

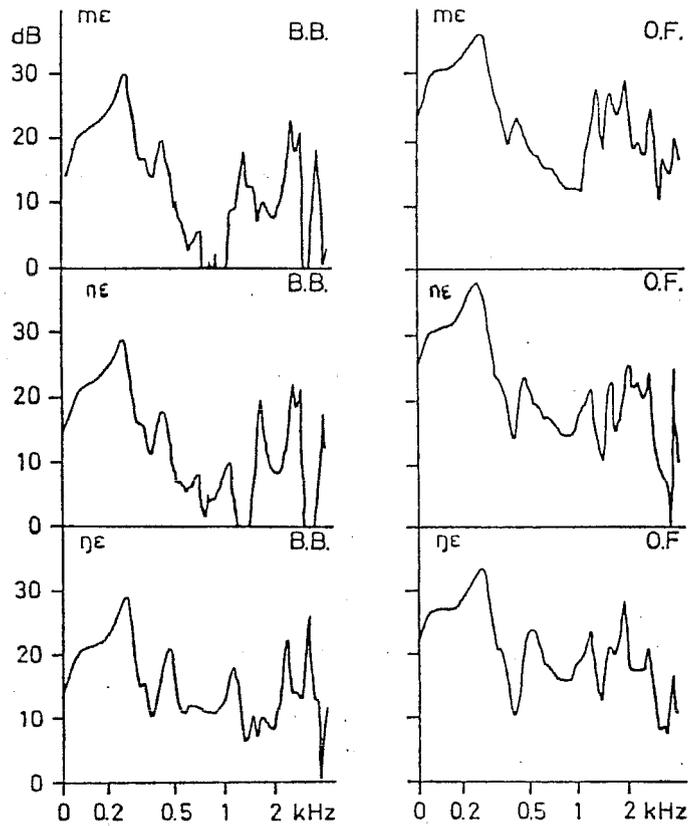


Figure 1.57 - Fonctions de transfert de 3 consonnes nasales par deux sujets (vibrateur externe) (in LINDQVIST & SUNDBERG 1976).

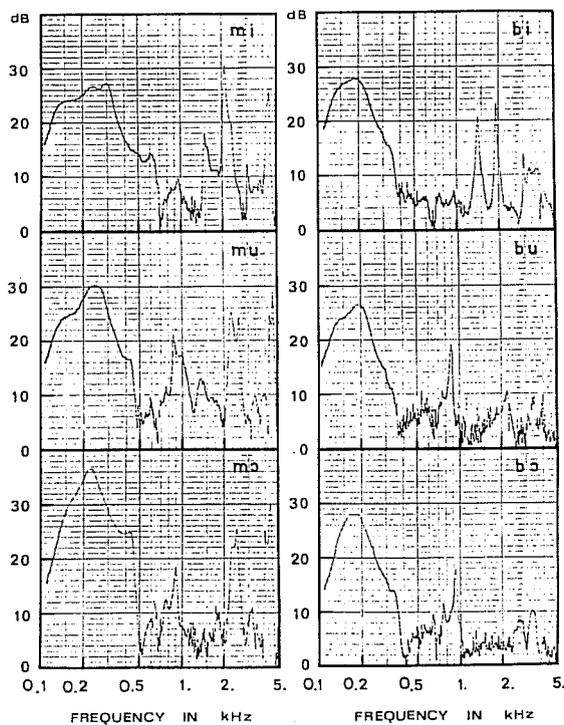


FIG. 16. Comparisons between nasals and stops with approximately the same articulation except for lowering of the velum.

Figure 1.58 - Fonctions de transfert de la consonne [m] (vibrateur externe) (in FUJIMURA & LINDQVIST 1971).

interprétation que les données de FUJIMURA & LINDQVIST (1971) (figure 1.58) présentent également un formant très bas, proche en fréquence de la barre de voisement d'un [b], alors qu'un déplacement dû à un zéro n'est pas évident. Il existe cependant une trace de pôle vers 400 - 500 Hz.

Parmi cette moisson d'informations, nous retiendrons qu'il existe une preuve difficilement réfutable d'un premier formant nasal vers 600 Hz pour le conduit nasal seul, ce qui implique comme on le verra un aire narinaire relativement importante. De plus on peut soupçonner l'état physiologique du sujet d'avoir une importance non négligeable sur la fonction d'aire nasale.

8.7 - LONCHAMP (1978)

Nous mesurons le manque de modestie s'attachant à la présentation d'un travail personnel dans cette revue (LONCHAMP 1978; version remaniée dans LONCHAMP 1979). Mais nous souhaitons préciser quelques points obscurs ou erronés de ce travail. Sur les spectrogrammes formant la base de ce travail, nous avons interprété la deuxième harmonique (vers $F_0 \times 2 = 250$ Hz) comme le premier formant nasal. Il est clair aujourd'hui qu'il s'agissait en fait du "formant" glottal que FANT devait décrire en détail à partir de 1979. Les voyelles avaient en effet été enregistrées à un niveau d'effort vocal relativement faible en studio, ce qui favorise l'émergence de cette caractéristique. Au contraire, les deux échantillons de la figure 6 (p. 16) provenaient d'un enregistrement direct devant le spectrographe à un effort vocal plus important : ils ne présentent plus ce "formant" glottal. Notre identification était renforcée par la prise en considération des travaux de DELATTRE (1954, 1965, 1968), HOUSE & STEVENS (1956), HATTORI & al. (1958) et FUJIMURA & LINDQVIST (1971) (où nous interprétons implicitement le pic de basse fréquence comme le premier formant nasal) qui assimilent tous un pic de très basse fréquence à FN1. Pour identifier les indices, nous nous appuyions fortement sur les analyses théoriques de MRAYATI (1975, 1976) qui, comme nous l'avons vu, utilise le conduit nasal de STEVENS & HOUSE. Dans ce cas, l'ordre des formants est fixé sans ambiguïté : F1N, F'1, F'2, FN2 et F'3 sauf peut-être pour [õ]. Nous montrerons plus loin que cet ordre n'est pas établi avec certitude pour tous les sujets. Nous reprenons l'interprétation des mesures au paragraphe 10.

8.8 - MAEDA (1982 a)

L'étude de MEADA sur le rôle des cavités sinusales dans la production des voyelles

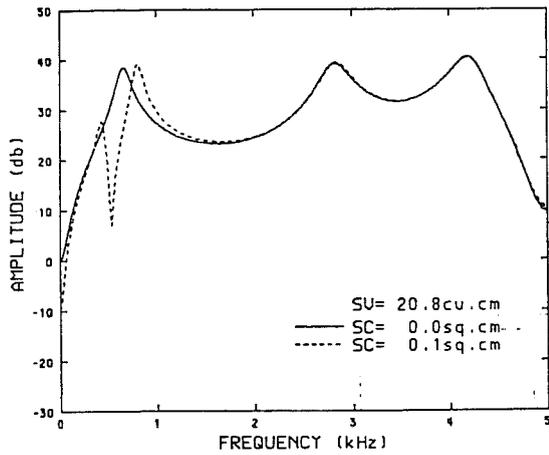


Fig. 3 The transfer function of the nasal tract with or without the sinus coupling. The innermost section is closed.

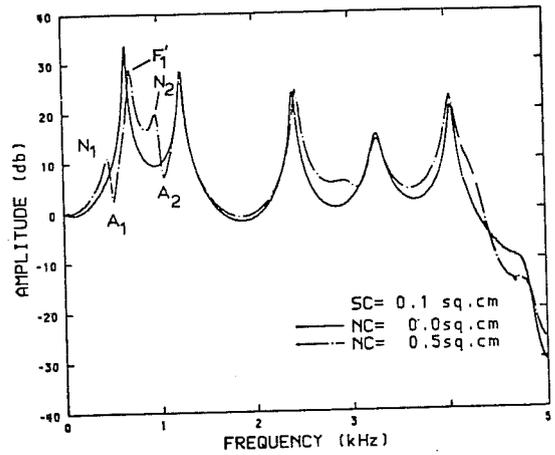


Fig. 4 The transfer function, under the closed glottis condition, of the vowel [Q] and that with nasal coupling, NC=0.5 sq cm. The sinus cavity is connected to the main nasal tract.

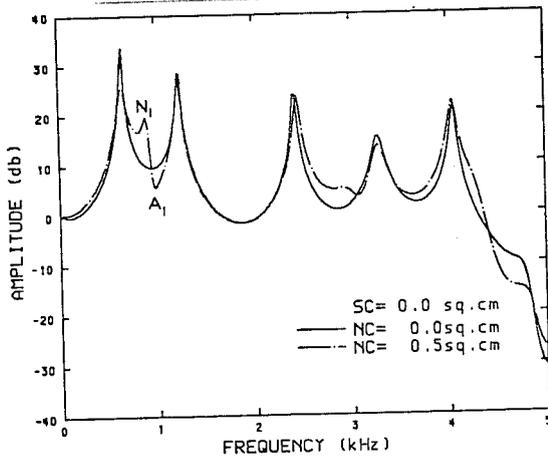


Fig. 6 The transfer function of the vowel [Q] and that with NC=0.5 sq cm. The sinus cavity is disconnected.

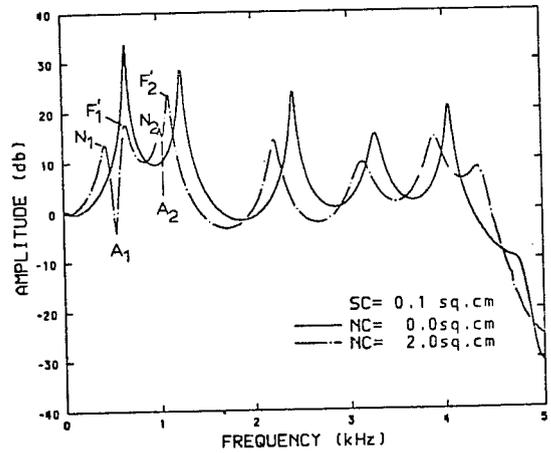


Fig. 5 The same as Fig. 4, except NC=2.0 sq.cm.

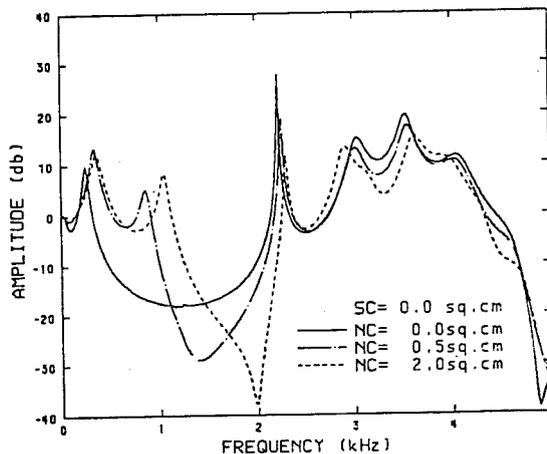


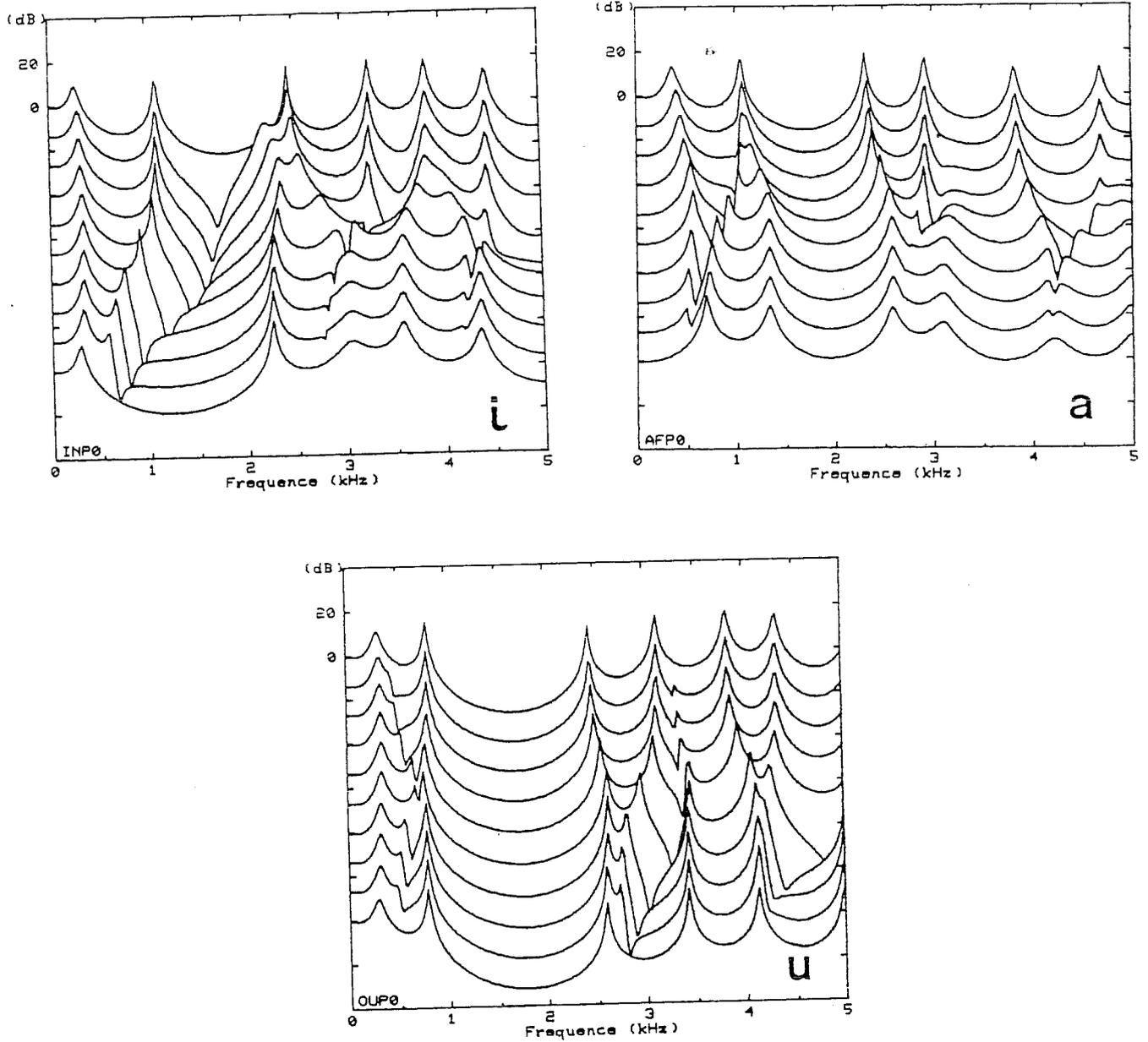
Fig. 7 The transfer function of [i], with NC = 0.5 and 2.0 sq cm. The sinus cavity is not connected in this example.

Figure 1.59 - Simulations de la fonction de transfert d'un conduit nasal avec et sans sinus (3) et des voyelles [ã] et [ĩ] (MAEDA 1982 a).

nasales découle d'un essai infructueux de synthèse des voyelles ouvertes à l'aide d'un modèle ne comportant qu'un tube nasal unique de forme voisine de celle de FANT. Remarquant que dans ce cas FN1 est placé plus haut que F'1 et s'appuyant sur les travaux de FUJIMURA & LINQVIST, l'auteur indique que l'abaissement de FN1 ne peut résulter d'une mauvaise estimation de la longueur : FN1 à 450 Hz exigerait un conduit nasal de 16 cm, de 40% plus long que la mesure de FANT. Sans envisager la possibilité d'une réduction de l'aire de la partie antérieure qui reste fixée à 2cm^2 , il obtient cet abaissement par le couplage sur le conduit nasal d'une cavité sinusale de 20.8 cm^3 munie d'un col de 0.1 cm^2 et de 0.5 cm de longueur. Les résultats des simulations sont reproduits ici à la figure 1.59 (figure 3 à 7 de l'original). Pour le conduit nasal seul, FN1 vers 670 Hz est "fendu" par le zéro associé à la cavité sinusale : deux pics apparaissent vers 450 et 820 Hz (figure 3). Le premier a pour origine la résonance du sinus et le deuxième est le premier formant nasal modifié en fréquence par l'influence du zéro et par le couplage. Pour [ã], à un degré faible de couplage (figure 4), un premier pic "sinusal" (noté N1) apparaît sous F'1, alors que FN1 (noté N2) est visible entre F'1 et F'2. Pour un couplage plus important (figure 5), l'amplitude de N1 augmente, mais reste cependant 20 dB sous F'1! N2 s'approche de F'2 avec lequel il se confond presque. MAEDA souligne la ressemblance avec les données de FUJIMURA & LINDQVIST. L'abaissement de F'2, contraire à la théorie, est attribué à la constriction du conduit oral par l'abaissement du voile. L'échec d'une nasalisation des voyelles ouvertes par un conduit simple est expliqué par le fait que A1 vient annihiler F'2, dont la place est prise par FN1 (figure 6 : notez que pour le couplage choisi pour l'illustration, FN1 est encore loin de F'2 !), ce qui enlève tout indice de nasalité, les remplaçant par une simple modification du timbre oral en fonction de la position exacte de FN1 jouant le rôle de F2. Pour une voyelle fermée comme [~i], un conduit nasal simple (figure 7) provoque l'apparition d'un pic supplémentaire dans la vallée entre F'1 plus bas et F'2 plus haut contribuant à la perception de la nasalité. Ce travail ne comporte aucune comparaison avec des spectres de voyelles naturelles.

8.9 - FENG, ABRY & GUERIN (1985) ; FENG (1986)

La contribution de FENG, ABRY & GUERIN (1985 a, 1985 b), détaillée dans la thèse de FENG (1986), rouvre le dossier des indices de nasalité de basse fréquence. Deux études importantes (LHERM 1984, MERLIER 1984) menées également à Grenoble ne sont pas mentionnées car les parties traitant de la simulation des voyelles nasales contiennent des résultats numériquement incorrects. Après avoir souligné la difficulté de trouver des indices invariants pour l'ensemble des voyelles, et prôné comme nous-même



• Figure 1.60 - Evolution spectrale de la consonne nasale vélaire [ŋ] aux voyelles orales [i], [a] et [u] (in FENG & al. 1975).

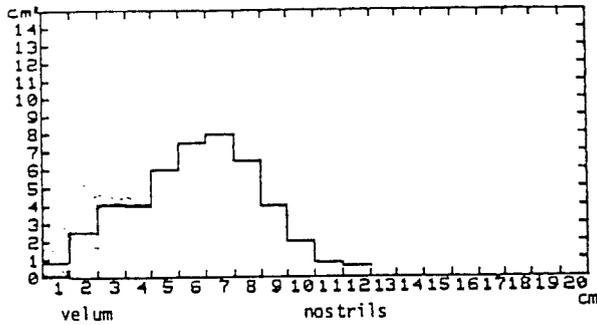
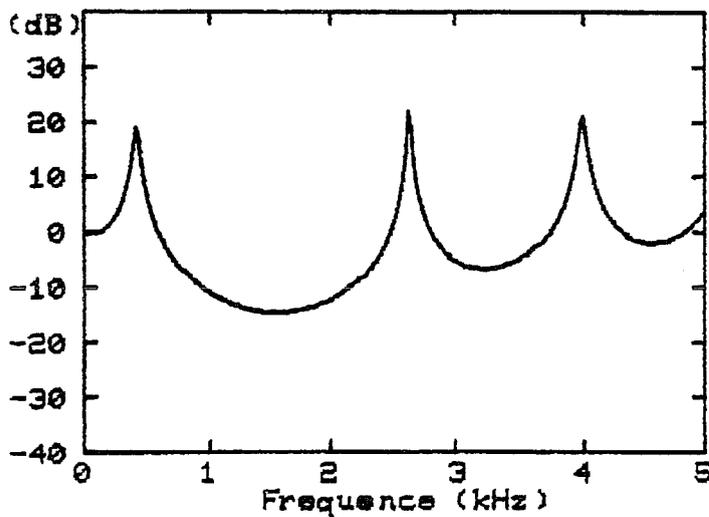


Figure 2-7. Fonction d'aire choisie pour le conduit nasal.
Noter que la sortie est 0,6 cm².



F1 = 445

Figure 2-8. Fonction de transfert du conduit nasal calculée
à partir de la fonction d'aire en Figure 2-7.

Figure 1.60 (Suite) - Fonction d'aire retenue pour les simulations de la page précédente et fonction de transfert. Notez la faible ouverture narinale (in FENG 1986).

une approche faisant un large appel à la simulation, FENG & al. prennent comme point de départ le spectre de la consonne [ŋ], cherchant les conditions nécessaires à la présence de deux pics vers 250 et 1000 Hz. Ils insistent sur la pertinence des travaux de MAEDA (1983), qui propose également comme indice de nasalité la présence de deux pics vers 250 et 1000 Hz dans une zone "vide" du triangle vocalique acoustique, à gauche de [u]. Mais il nous semble important de préciser que si les valeurs numériques sont les mêmes, les pics homologues de la consonne nasale vélaire et des voyelles nasales ne sont pas affiliés aux mêmes cavités : Pour MAEDA, le pic à 250 Hz est une résonance sinusale, et le pic de [ŋ] à la même fréquence est la résonance de l'ensemble du conduit nasal et pharyngal. FENG & al. notent d'ailleurs que le pic de 250 Hz est souvent invisible pour les voyelles ouvertes, et masqué par F'1 pour les voyelles fermées, alors que celui de 1000 Hz peut être masqué par F'2 pour les voyelles ouvertes.

L'étude de la consonne vélaire nasale est beaucoup plus simple que celle des voyelles, car le conduit vocal ne comporte pas de partie en dérivation. S'appuyant sur les nomogrammes de FANT (1960) et sur des simulations, ils concluent que seul un conduit oro - nasal possédant une constriction relativement importante dans la zone narinaire peut donner un formant aussi bas. L'étude des voyelles nasales se fait en étudiant le passage entre la voyelle orale et la consonne [ŋ]. La nasalité vocalique est donc étudiée comme un moment d'un geste articuloire qui conduit vers la consonne nasale vélaire, présente après voyelle nasale dans les français méridionaux. Les spectres des simulations des voyelles [i], [u] et [a] sans sinus et avec un conduit nasal à petit orifice sont reproduits à la figure 1.60. Pour [i], les formants oraux passent de 240, 2270 et 3240 Hz à 260, 2440 et 3150 Hz pour [ŋ]. Il est clair que le pic de 1kHz de la consonne est issu du formant nasal. Pour [u], l'évolution plus complexe n'est pas traitée dans ce travail ni dans FENG (1986).

Une première analyse de l'évolution des spectres pourrait être la suivante : Si le premier formant de la voyelle rejoint bien le premier formant de la consonne, le pic nasal entre F'1 et F'2 se dirige dans un premier temps vers F'2 avec lequel il finit par se confondre, puis il semble rebrousser chemin pour venir se confondre avec F'1! Mais alors le pic de 1000 Hz de la consonne n'est pas FN1, mais F2. Il faut donc que ce soit F'2 qui vienne rejoindre F'1. Pour [a], le formant nasal apparaît sous F1 et il vient former le premier pic de [ŋ], et non le deuxième comme précédemment. Le premier formant de la voyelle, quant à lui, repoussé par le formant nasal se dirige vers F'2 avec lequel il se confond pour donner le deuxième pic de la consonne vers 1000 Hz. Aucune rationalisation ne permet d'associer le formant nasal de la voyelle et le deuxième pic de [ŋ] dans ce cas. Les mêmes difficultés existent bien sûr si l'on part de la consonne nasale vers la voyelle comme il est fait très brièvement dans les travaux de nos collègues. L'origine de ces difficultés n'est pas difficile à trouver. Suivre un formant présuppose que son affiliation privilégiée avec une

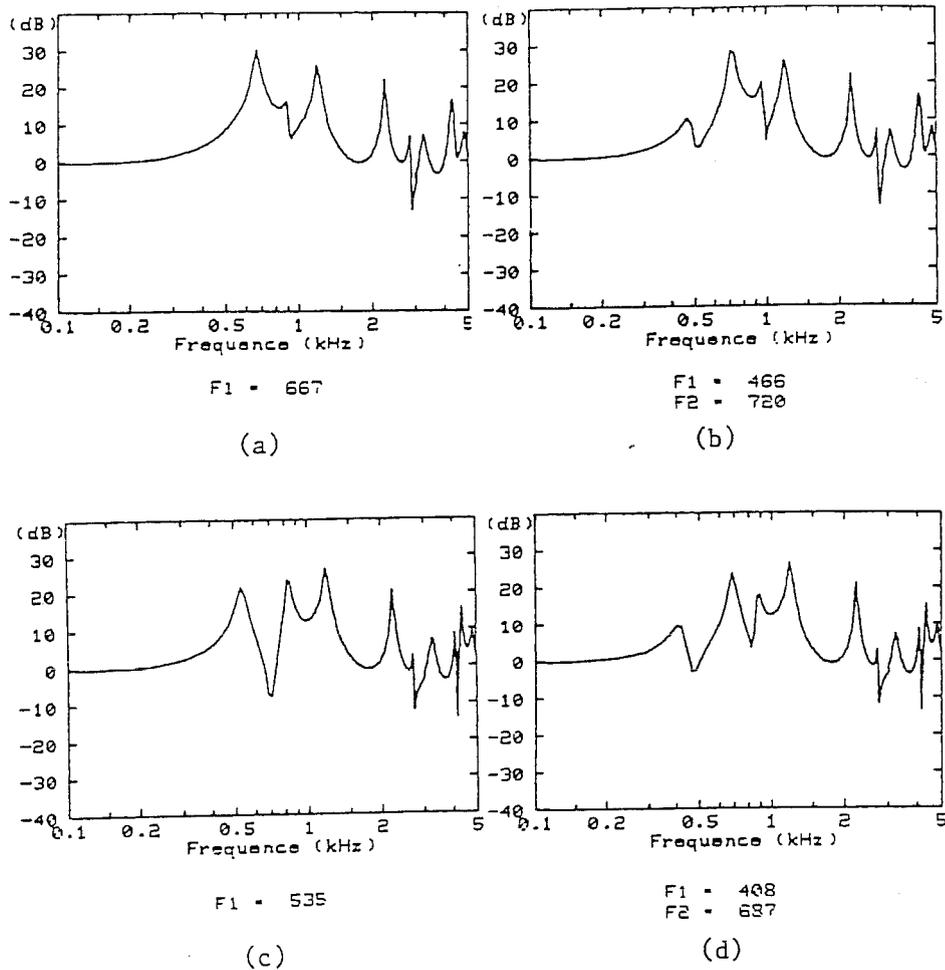


Fig. 7. Comparisons of the transfer functions for nasalized [a] (at a coupling stage of 0.4 cm^2), with different conditions :

- top : nose output area 2.0 cm^2 ,
- bottom : nose output area 0.6 cm^2 ,
- left : without sinus simulation,
- right : with a 20.8 cm^3 sinus maxillaries equivalent.

- 7-a. The 1st pole-zero pair is "wrongly" (MAEDA, 1982) between F1 and F2.
- 7-b. The sinus maxillaries add a more "correct" 1st pole-zero pair below F1 (MAEDA, 1982 and STEVENS & al., 1985).
- 7-c. The same positionning of the low pole-zero pair is obtained with a small nose output, without sinuses.
- 7-d. And is "optimized" (lowered) if one adds the sinus maxillaries.

Figure 1.61 - Comparaison de plusieurs fonctions de transfert de la voyelle [ã] (in FENG & al. 1975).

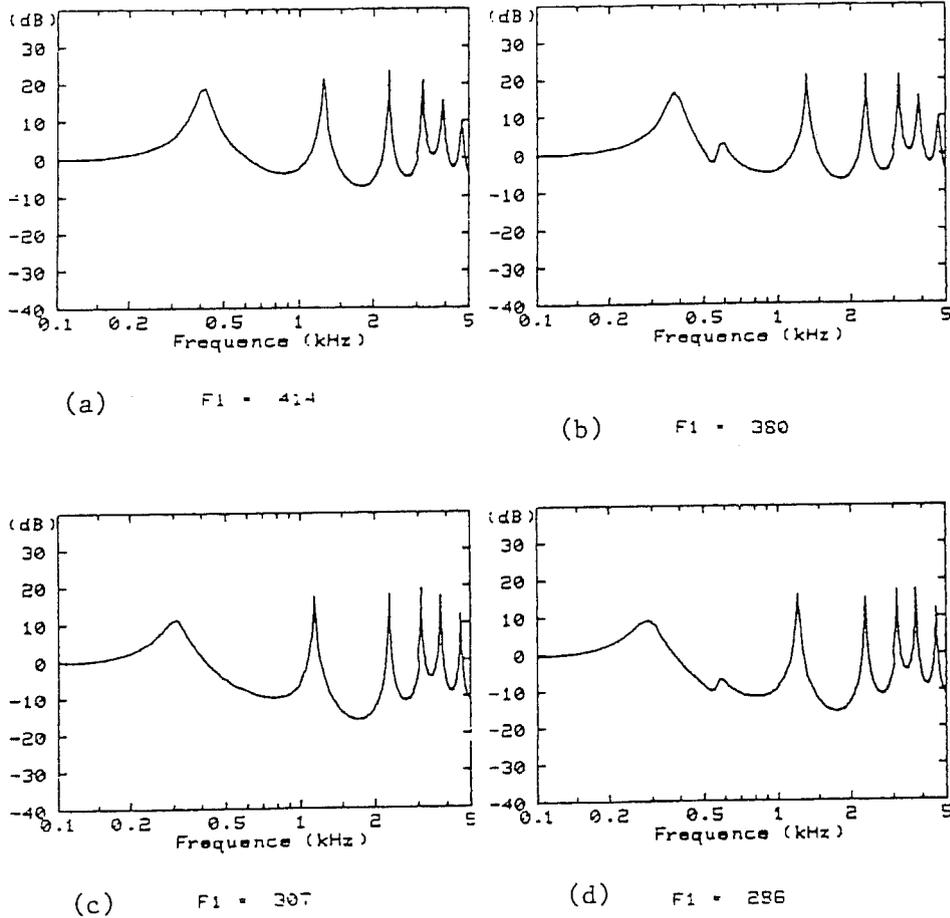


Fig. 10. Comparisons of transfer functions of the nasopharyngeal tract for [ε] (same output and sinus conditions as in Fig. 7).

Among the four solutions, the closest to sweep-tone measurement results (see Fig. 11) are the bottom ones - with a small nose output area.

Figure 1.62 - Comparaison de fonctions de transfert d'un [p] possédant la partie pharyngale de la voyelle [ε]. (in FENG & al. 1975). Les conditions sont décrites dans la légende de la figure 1.61.

caractéristique du conduit (cavité particulière ou longueur) reste constante. Ce n'est pas le cas ici où les formes de départ et d'arrivé sont fortement différentes. Une description plus correcte du passage de [ŋ] à [a] pourrait être la suivante : Le premier pic de la consonne dépend de la longueur totale du conduit nasal et pharyngal. Tant que le couplage du conduit nasal est important, il se produit une résonance de basse fréquence dont la fréquence varie faiblement entre 250 et 350 Hz. Lorsque le couplage diminue, cette résonance devient progressivement plus dépendante de la taille du seul conduit nasal. Le pic de 1 KHz de la consonne est lié au conduit nasal largement ouvert sur la pharynx. Lorsque le couplage diminue, l'importance de son affiliation avec le volume pharyngal et buccal s'amplifie, au détriment de l'affiliation avec le conduit nasal. Mais l'introduction d'une cavité antérieure de taille significative introduit un formant supplémentaire, ce qui produit un système à deux formants, qui vont donner F1 et F2. Pour un couplage important, les fréquences de F'1 et F'2 dépendent de trois cavités différentes : nasale, oral et pharyngale. Le degré de dépendance change rapidement en fonction du degré de couplage et de l'évolution de la forme de la cavité buccale. Le degré d'affiliation ne peut se mesurer qu'à l'aide des fonctions de sensibilité (CHARPENTIER 1986 par ex.). La continuité "visuelle" n'est pas le signe d'une affiliation constante.

La suite du travail de FENG & al. est une discussion du rôle des résonances sinusales. Rappelons que MAEDA (1982 a), à la suite de simulations avec un conduit nasal de large ouverture narinaire, suggérait que la résonance de basse fréquence (300 - 400 Hz), nécessaire à la perception de la nasalité pour [ã], était le fruit d'un couplage sinusal. L'utilisation d'une aire narinaire étroite rend le recours à cette solution inutile (figure 1.61c), puisque le formant nasal sera dans cette zone (ou juste au-dessus), alors qu'il se trouve au-dessus de F'1 (au delà de 800 Hz) chez MAEDA (figure 1.61 -a). Avec un sinus, la simulation dérivée de MAEDA (c) montre un petit pic à la bonne fréquence. Les pics suivants sont F'1, FN1 et F'2. La simulation à aire narinaire faible (d) le montre également, suivi de FN1, F'1 et F'2. L'ordre de FN1 et F'1 n'est pas le même dans les deux cas. FENG & al. tentent de montrer la supériorité de leur solution en suivant l'évolution des formants de la voyelle orale à la consonne [ŋ]. Mais comme nous avons essayé de le montrer plus haut, leur interprétation des résultats est discutable. Le fait que dans leur solution, l'augmentation du couplage abaisse ce qui va devenir le premier pic de la consonne, alors que dans la simulation dérivée de MAEDA, sa fréquence reste constante, ne nous semble pas un argument très fort. L'appel à l'évolution historique est sans objet dans ce contexte: les voyelles nasales françaises ont eu plutôt tendance à s'ouvrir (F1 croissant donc), et non à se fermer. La présence d'un appendice consonantique dans les français méridionaux (cf. LINTHORST 1973) n'est pas nécessairement liée à un abaissement graduel du premier pôle. De plus, cet abaissement se produit pour un fort

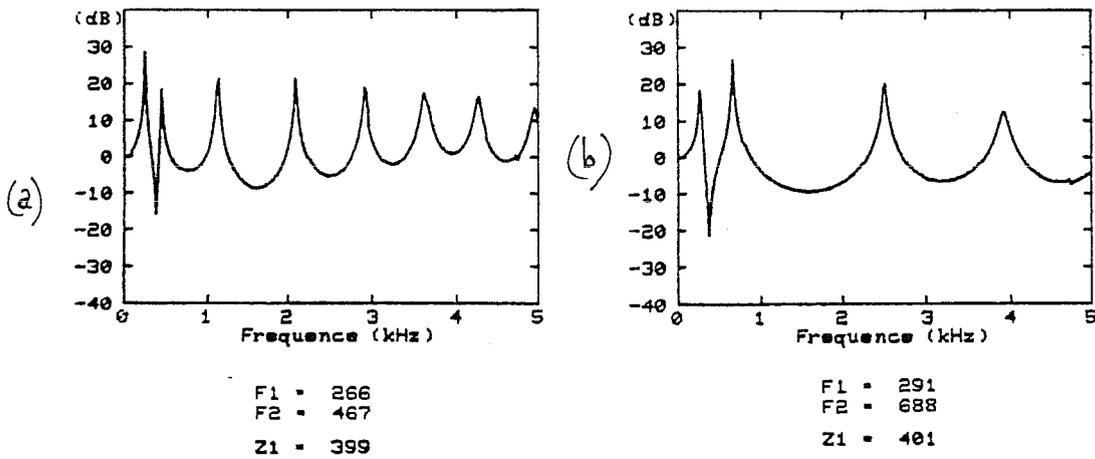
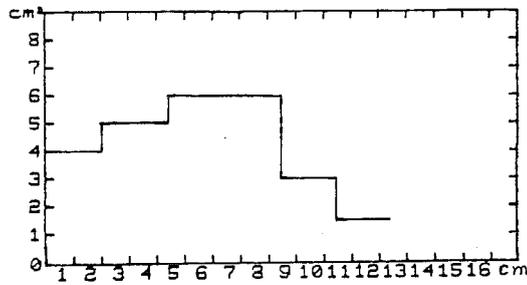


Figure 2-15. Résultats de simulation, selon les paramètres utilisés par FANT (1985), pour un conduit nasopharyngal (a) et pour le conduit nasal (b). Tous les deux sont avec l'introduction de sinus, accordé à 399 Hz. Noter que le premier pic en (a) est proche de celui en (b), dû à l'effet du zéro.

Figure 1.63 - Simulation de la fonction de transfert de la consonne $|\eta|$ (a) d'après les paramètres de FANT (1985) (in FENG 1986)

couplage quel que soit la forme du conduit nasal. Le test suivant, portant sur la simulation de [ŋ], est beaucoup plus probant. Sans un sinus, le conduit nasal à ouverture large produit un premier pic vers 500 Hz (figure 1.62 - a), très loin des mesures de LINQVIST & SUNDBERG (1976) qui le place sous 300 Hz. L'introduction d'un sinus améliore un peu la situation (b). Nous verrons plus loin qu'un choix différent de paramètres sinusaux permet d'obtenir de meilleurs résultats. Les courbes (c) et (d), résultant d'un conduit étroit avec et sans sinus, sont beaucoup plus satisfaisantes. Le recours à une résonance sinusale laisse ouvert le problème de la nasalisation chez l'enfant, car les sinus se développent lentement. En cas de coriza, ils sont vraisemblablement fermés par gonflement de la muqueuse. Mais poser le problème ainsi suppose que la nasalisation est entièrement déterminée par la présence ou l'absence de pics spécifiques. En anticipant, nous tenterons de montrer que la nasalité est également le produit d'une position spécifique des formants F'1 et F'2, sans que des pics ou vallées supplémentaires soient absolument nécessaires.

Nous souhaitons maintenant discuter plusieurs points de détail de la thèse de FENG (1986). La différence entre les fréquences de FN1 dans les simulations de MAEDA (1982a) et FANT (1985) n'est pas due à la taille différente de l'orifice narinaire (2 cm² contre 1.5 cm²), mais à la différence de longueur (11 cm contre 12.5 cm chez FANT) dont le rapport correspond exactement au rapport de fréquence (670 contre 560 Hz). Les propres simulations de FENG (cf. sa figure 2.5) le confirme d'ailleurs. Cette figure révèle qu'une baisse significative de la fréquence de FN1 ne se produit que si l'aire aux narines chute à 0.5 cm² ou moins. Dans les données de LINDQVIST & SUNDBERG (1976) utilisées pour choisir une forme de conduit nasal, FENG retient la valeur de 450 Hz comme fréquence de FN1 (sujet J.L. cf. figure 1.55). Nous avons vu que la valeur pour le sujet J.S., qui paraît moins perturbée par une paire pôle /zéro supplémentaire est de 570 Hz, ce qui correspond bien à la fréquence du conduit nasal de FANT si la longueur est fixée à 12.5 cm ! L'hypothèse d'une forte constriction aux narines (0.6 cm²), et une fréquence de FN1 égale à 445 Hz, ne s'impose donc pas de manière évidente. La forme exacte du reste du conduit n'a pas une grande importance. La fréquence du deuxième pic dépend de la longueur du conduit et d'une éventuelle constriction médiane. La simulation de la consonne [ŋ] montre que la coarticulation pharyngale de la voyelle associée n'a que peu d'influence sur les deux premiers formants.

La discussion du rôle potentiel des sinus mérite quelques commentaires. FENG (p. 69) tente de montrer qu'une résonance vers 400 Hz implique une taille de sinus incompatible avec les données anatomiques (45 cm²). Cette conclusion repose sur un choix arbitraire de la longueur et du diamètre du col. Nos résultats (§ 5.4) prouvent le contraire.

Il discute ensuite du problème crucial qu'est la possibilité d'associer le premier pic observé dans les voyelles nasales vers 450 Hz à l'effet d'une résonance sinusale qui scinderait

FN1, trop élevé, en deux parties (figure 1.60). Si le détail est trop complexe pour être rapporté ici, la conclusion est que cette tentative se solde par le dilemme suivant : On ne peut simuler avec un sinus de taille constante et un conduit nasal offrant un orifice aux narines large l'écart de fréquence important entre le pic vocalique (vers 450 Hz) et le premier formant de [ŋ], vers 250 Hz. Comme nous l'avons vu plus haut, l'effet de scission due à la paire pôle / zéro sinusale crée un pic de fréquence stable, 400 - 450 Hz pour un sinus accordé vers 500 Hz (MAEDA) ou 250 - 300 Hz pour un sinus accordé vers 400 Hz (FANT 1985). L'argument est certainement vrai pour la simulation de FENG, comme le montre la figure 1.63. Rappelons que nous montrerons plus loin par simulation que l'on peut obtenir par un choix convenable de tous les paramètres de couplage un pic de fréquence basse pour [ŋ] (vers 300 Hz) et un pic vocalique de fréquence plus élevée avec une fonction d'aire nasale à "grosses" narines. Mais il est vrai que le rôle des sinus n'est pas négligeable dans nos simulations. Nos résultats montreront également que la fréquence du premier pic de la consonne [ŋ] n'est pas toujours aussi basse que celle discutée ici.

La longueur de notre discussion montre l'importance des travaux de FENG & al. Le point essentiel est qu'il convient d'étudier la nasalité vocalique en veillant à ce que les hypothèses faites soient compatibles avec ce que nous savons de la nasalité consonantique. Une aire narinale faible est de ce point de vue une possibilité. Mais elle ne ressort pas de manière immédiate des documents anatomiques que nous avons discutés au paragraphe 3-5. La solution de MAEDA, qui fait jouer un rôle important aux résonances sinusales, ne peut être écartée définitivement.

8 . 10 - HAWKINS & STEVENS (1985)

Cette dernière étude n'est pas consacrée à une étude détaillée de l'acoustique de la nasalité, mais à celle de la perception, et sera de ce fait étudiée plus loin. Mais elle fournit des documents acoustiques d'un grand intérêt (figure 1.64). L'opposition voyelle nasale / voyelle non-nasale pour 5 timbres est illustrée par les réalisations d'un sujet masculin Gujarati. En basse fréquence, on note pour [~i] une légère élévation de F'1 et la présence d'un formant supplémentaire (FN1 ?) vers 1400 Hz, ainsi qu'une très légère bosse vers 900 Hz qui pourrait également être présente dans la voyelle non-nasale vers 800 Hz. Pour [~e], 3 pics remplacent le pic unique de l'orale. La nature du premier vers 200 Hz, n'est pas claire : formant glottal, résonance sinusale ? Sa présence constante pour les nasales, sauf [~i], et son absence pour les orales, invite à choisir la deuxième solution. Le second vers 500 Hz et le troisième à 1100 Hz sont F'1 et FN1, ou FN1 et F'1. Le problème de l'ordre fera l'objet d'une discussion détaillée plus loin et nous n'anticiperons pas sur cette

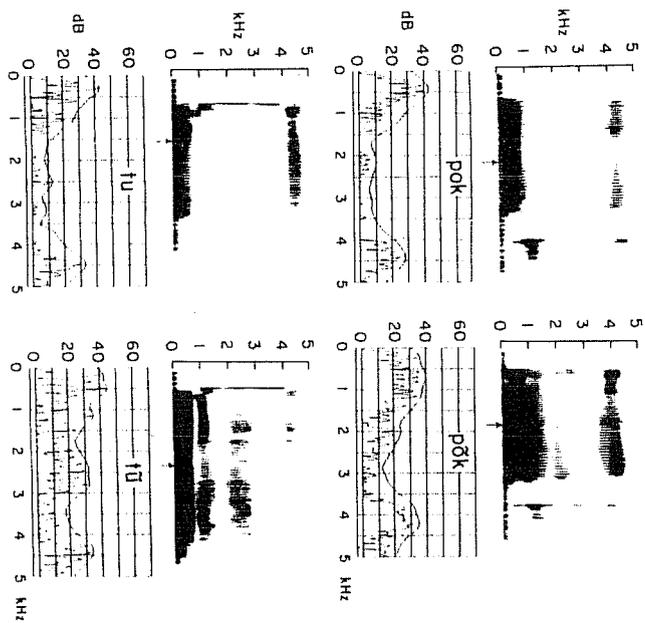
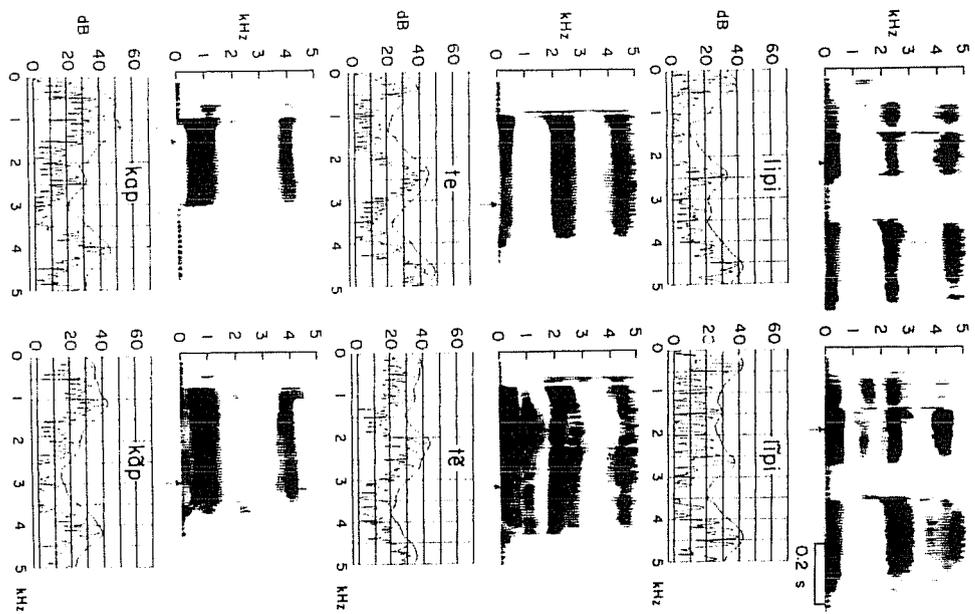


FIG. 1. Spectrograms of words illustrating the non-nasal-nasal distinction for vowels in Gujarati, together with spectra sampled within the vowels. The non-nasal vowels are in the left column, with the spectra immediately below the corresponding spectrograms, and the nasal vowels are in the right column. The time at which the spectrum is sampled is indicated by the arrow. Two spectra are shown in each panel, each representing a spectrum sampled over a 25.6-ms Hamming window. The lower curve is the discrete Fourier transform, and the upper smooth curve is a "pseudo-spectrum" for which each point is obtained by performing a weighted average of groups of points on the DFT over a bandwidth of 300 Hz. Speaker VS (male).

Figure 1.64 - Spectres F.F.T. et spectrogrammes de 5 couples de voyelles orale et nasale homologues : locuteur Gujarati (in HAWKINS & STEVENS 1985).

discussion. D'autres très petites irrégularités spectrales sont visibles pour les versions nasales et non-nasales. Pour [ã], on note 3 pics, à 200-250 Hz, 650 et 1050 Hz, alors que [a] ne possède que les 2 formants classiques, avec néanmoins une petite irrégularité vers 950 Hz. [o] n'a qu'un pic en basse fréquence, mais sa largeur de bande, (ainsi que la théorie !) nous incite à y voir F1 et F2. On retrouve pour [õ] 3 maxima vers 250, 650 et 1100 Hz. Mais le deuxième est particulièrement large et le même raisonnement devrait nous inciter à y distinguer deux formants. [~u] possède également 3 pics, vers 200, 500 et 1100 Hz, plus un vers 1500 Hz moins proéminent. A critères identiques, [u] possède 3 formants vers 300, 550 et 1000 Hz. Cette analyse parallèle doit nous rappeler que l'analyse des formants des voyelles orales repose souvent moins sur une analyse objective des maxima du spectre que sur l'identification de ceux prévus par la théorie classique. On mesure la difficulté pour les voyelles nasales quand les faits théoriques sont moins clairement établis!

8.11 - CONCLUSIONS

Il y a une certaine incohérence à vouloir résumer une synthèse ! Essayons néanmoins de dégager quelques faits qui seront les garde-fous des simulations et de l'interprétation. Ils sont volontairement en nombre très limité.

- 1 - La baisse de l'amplitude de F'1 peut être due au couplage, même en l'absence d'un zéro proche. On trouve une trace de FN1 vers 1000 Hz pour une voyelle [a] nasalisée (FANT)
- 2 - Il existe un paire pôle / zéro entre 270 et 400 Hz. Il n'est pas exclu, mais pas certain que FN1 ait une fréquence de 550 - 600 Hz, ou de 800 à 1000 Hz pour des voyelles postérieures (FUJIMURA & LINDQVIST)
- 3 - La fréquence de résonance du conduit nasal seul est vraisemblablement comprise entre 500 et 600 Hz. Des résonances sinusales sont essentielles pour expliquer la complexité du spectre en basse fréquence (LINDQVIST & SUNDBERG).
- 4 - La fonction d'aire de BJUGGREN & FANT ne permet pas de retrouver le spectre de [ŋ] par simulation (FENG, ABRY & GUERIN). Mais une modification radicale semble exclue d'après les clichés tomographiques.
- 5 - Toute solution doit rendre compte des résultats des analyses de voyelles naturelles (MRAYATI, LONCHAMP, HAWKINS & STEVENS).

9 - SIMULATIONS ET ANALYSES.

Cette partie, au coeur de ce travail, a pour objet de présenter les résultats d'un ensemble de simulations faisant varier certains paramètres articulatoires, et de dégager, par comparaison avec des spectres de voyelles (ou consonnes) naturelles, une série de conclusions sur la forme probable du conduit nasal, et, par conséquent, sur la nature et la position des indices de la nasalité vocalique. Ces informations seront mises à profit dans la partie suivante, qui analysera un corpus de voyelles naturelles. Pour toutes les simulations, les fonctions d'aire, codées O_i pour la partie orale, N_i pour la partie nasale, et S_i pour les paramètres des sinus, sont décrites aux annexes A, B et C respectivement. Rappelons (cf. § 7) que la source vocale est soit une pente de -6 dB / octave à partir de 250 Hz, ce qui provoque une légère discontinuité à cette fréquence, à ne pas confondre avec une résonance sinusale, soit le spectre du modèle de débit de FANT présenté au paragraphe 7.1. Le programme détermine par interpolation parabolique les fréquences et largeurs de bande de tous les pics, y compris des oscillations glottales, le cas échéant. Le programme d'analyse par transformée de Fourier sur une période détermine également la fréquence et une grossière estimation des largeurs de bande par interpolation parabolique sur les deux harmoniques adjacentes à un pic. Dans les discussions nous arrondissons les valeurs fréquentielles pour tenter de combattre l'illusion de précision que donne les simulations.

9.1 - SENSIBILITE DES FREQUENCES DE RESONANCE DU CONDUIT NASAL A LA MODIFICATION DE PARAMETRES ARTICULATOIRES.

Nous présentons tout d'abord (figure 1.65) la fonction de transfert du conduit nasal de BJUGGREN & FANT (1964) discrétisée de façon (sans doute excessivement) précise (N1). Elle ne comporte pas de sinus, et les deux fosses nasales sont symétriques. Les résonances se produisent à 670, 2200 et 4050 Hz. Pour FN1, ces valeurs sont bien sûr très voisines de celles de MERLIER (1984 : 691 Hz), LHERM (1982 : 680 Hz) ou MAEDA (1982 a : 670 Hz). Ces faibles variations sont dues autant à des différences dans la discrétisation (aires et longueur) qu'au choix des facteurs de forme et de perte. Les valeurs plus différentes des FN2 (MERLIER : 2676 Hz par ex) reflètent des choix différents dans la géométrie variable du passage rhinopharyngal. Les pertes par vibration des parois dépendent de paramètres dont les valeurs sont inconnus^x pour le conduit nasal. On leur donne par défaut les mêmes valeurs que pour le conduit oral. L'influence de ces pertes est mesurée en les supprimant et en comparant les résultats avec les précédents. Leur influence, faible, se traduit par une élévation de FN1 de 30 Hz seulement. L'introduction de deux sinus symétriques (S1) dont les dimensions sont fondées sur les résultats de AUST & DRETTNER (1975) provoque quelques^z modifications (figure 1.66). Une paire

u / ves

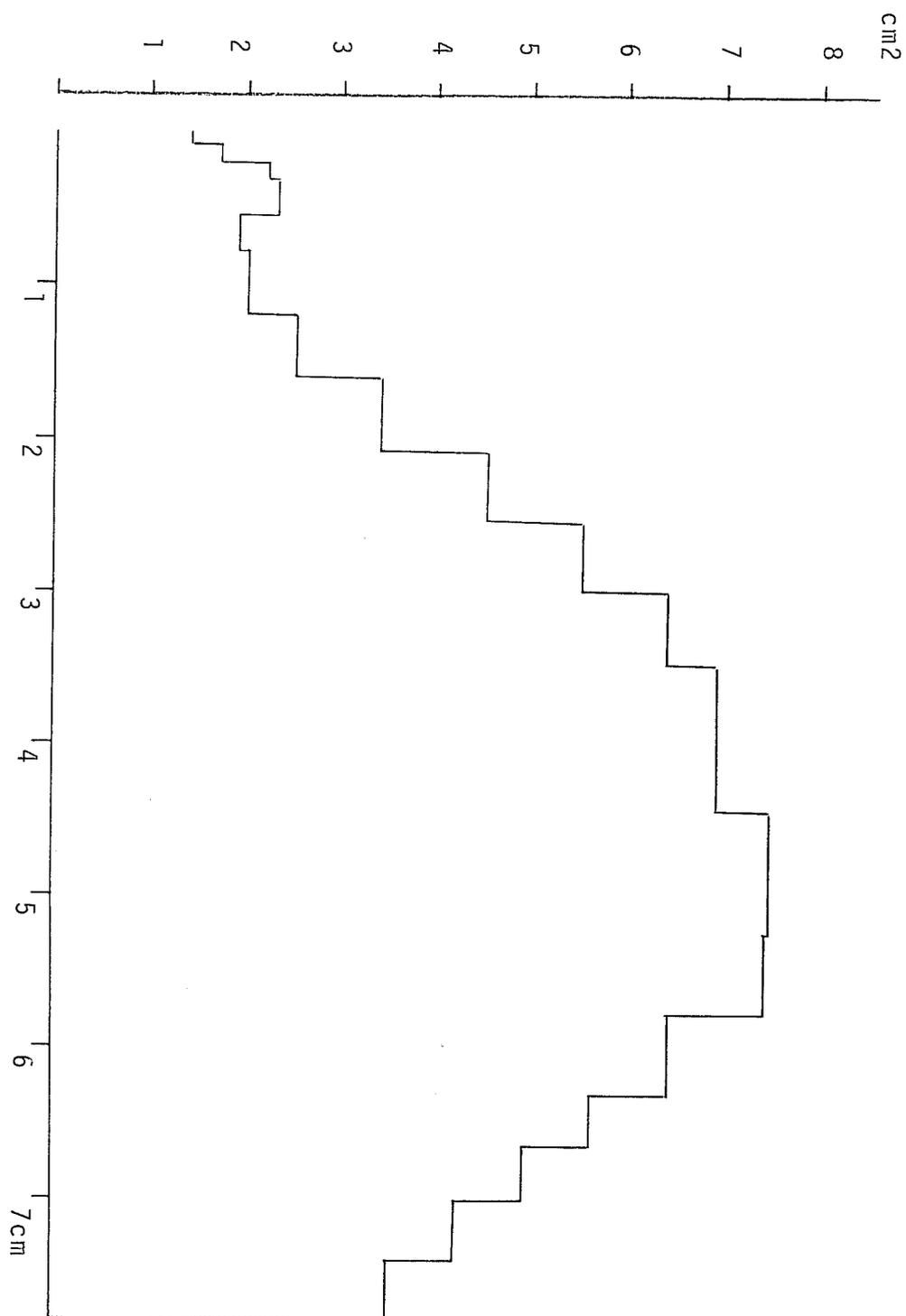


Figure 1.65 - Fonction d'aire de BJUGGREN & FANT (1964):
version discrétisée N1. (partie antérieure seulement).

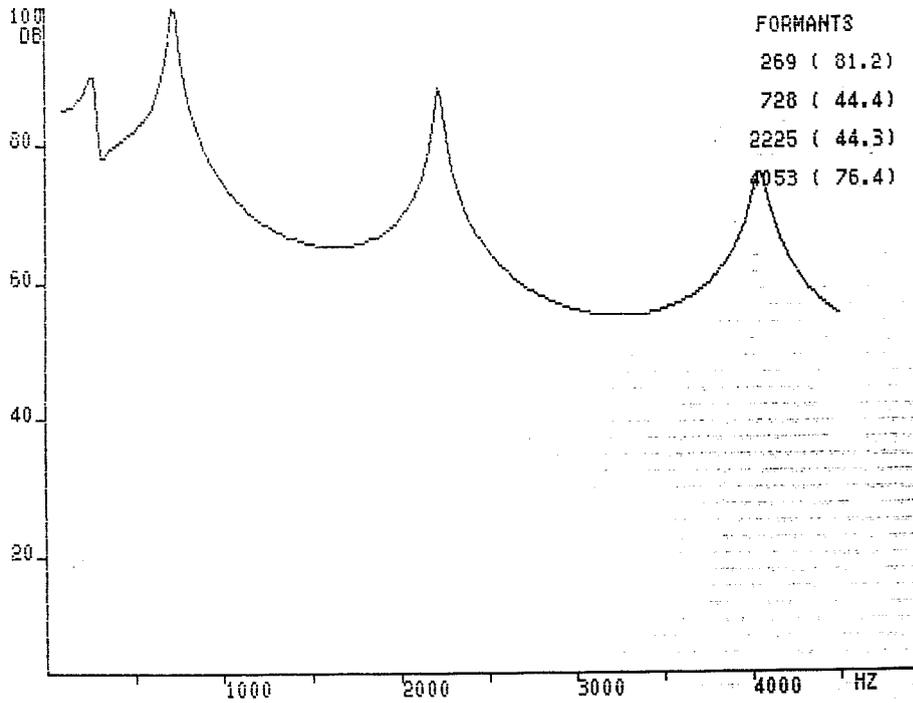


Figure 1.66 - Fonction de transfert du conduit nasal de BJUGGREN & FANT (1964) : deux sinus S1

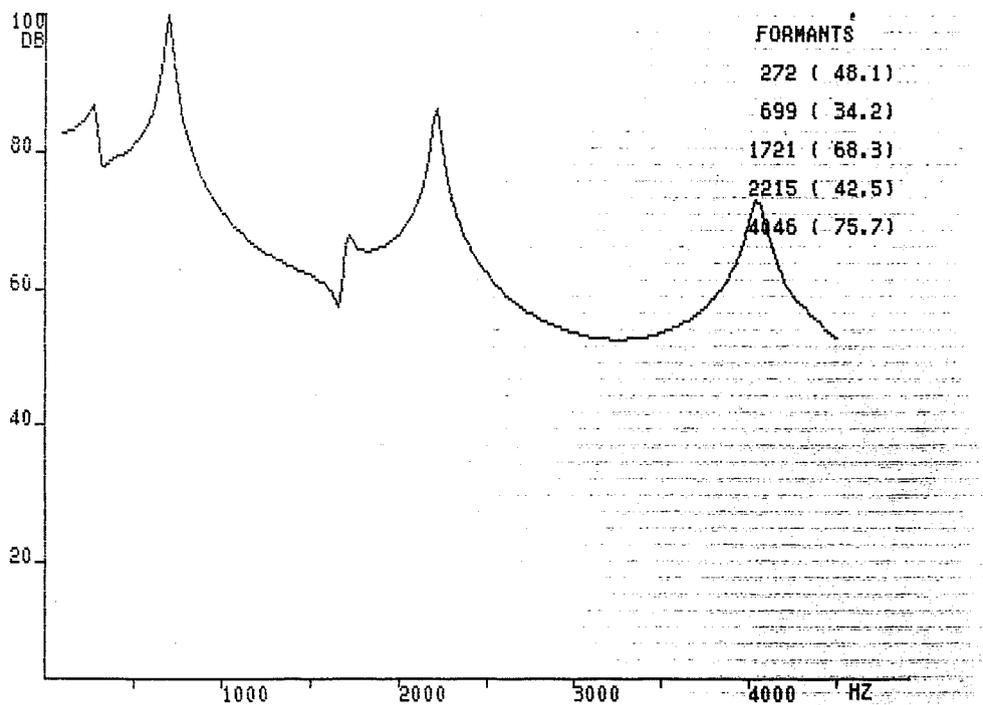


Figure 1.67 - Fonction de transfert du conduit de la figure 1.66 : suppression d'un sinus.

pôle / zéro discrète apparaît à 270 Hz pour le pic. La fréquence de FN1 est décalée de 60 Hz (10 %) vers les hautes fréquences par l'effet du couplage. Il faudrait en toute logique l'appeler F'N1 par analogie avec la notation adoptée pour les formants oraux décalés. Les décalages de FN2 et FN3 sont négligeables : 25 et 10 Hz. Il est intéressant de vérifier l'effet de la suppression d'un sinus. Les deux fosses nasales sont donc maintenant dissymétriques (figure 1.67). La fréquence de résonance du sinus unique reste bien sûr identique, mais la magnitude de la perturbation spectrale diminue. FN1 subit un décalage réduit de moitié (+ 30 Hz). Il apparaît enfin une paire pôle / zéro supplémentaire de faible amplitude vers 1700 Hz. Elle est due à la dissymétrie des deux fosses nasales. On démontre facilement que deux conduits identiques (sans pertes) au même point de couplage sur un troisième se comportent exactement comme un conduit unique ayant une aire double. L'impédance étant inversement proportionnelle à l'aire, doubler l'aire divise l'impédance par deux. Or l'impédance d'un branchement en parallèle de deux conduits identiques d'impédance Z_0 divise aussi l'impédance par deux :

$1/Z = 1/Z_0 + 1/Z_0 = 2/Z_0$ d'où $Z = Z_0/2$. Mais deux cavités différentes couplées sur le conduit rhinopharyngal ont deux résonances légèrement différentes dont l'une joue le rôle de formant nasal et l'autre de formant oral. Ceci se traduit par une paire pôle / zéro supplémentaire. Les deux FN1 ne sont pas visibles car de fréquence trop proche. En pratique, la mesure de la largeur de bande doit être supérieure, car elle est la somme des deux largeurs. Notre mesure est trop imprécise pour le vérifier dans ce cas (cf. § 9.2). On étudiera enfin (figure 1.68) le cas de deux sinus dissymétriques dont les volumes sont celui de S1 multiplié, ou divisé, par un facteur 1.25, le diamètre et la longueur du col étant pour le plus gros de 0.2 cm et 0.7 cm, et de 0.3 et 0.5 cm pour l'autre. Ce choix maximise la différence de fréquence selon la formule de Helmholtz, tout en restant bien à l'intérieur de la variabilité constatée anatomiquement : accroître le volume et la longueur du col, et réduire le diamètre du col diminuent la fréquence de résonance. On remarque un pic et surtout une vallée importante vers 400 et 450 Hz. Cette vallée est vraisemblablement creusée par le zéro du pôle de plus haute fréquence. Ce dernier comble le zéro associé au premier pic qu'on détecte à peine vers 185 Hz. Le diamètre plus important du petit sinus provoque un couplage plus important, une distance plus grande entre le pôle et le zéro et donc une perturbation spectrale plus importante. FN1 est repoussé à une fréquence légèrement plus élevée (+ 15 Hz) et il existe à nouveau une paire pôle / zéro supplémentaire vers 1750 Hz.

Les données anatomiques ont suggéré que l'aire des passages et méats de la zone des cornets était peut-être inférieure à celle mesurée sur moulage par FANT. Nous avons calculé l'effet spectral d'une réduction à 5 cm² de toutes les aires du conduit nasal (N1) supérieures à cette valeur. L'effet sur FN1 est très petit (+ 30 Hz), la fréquence de celui-ci

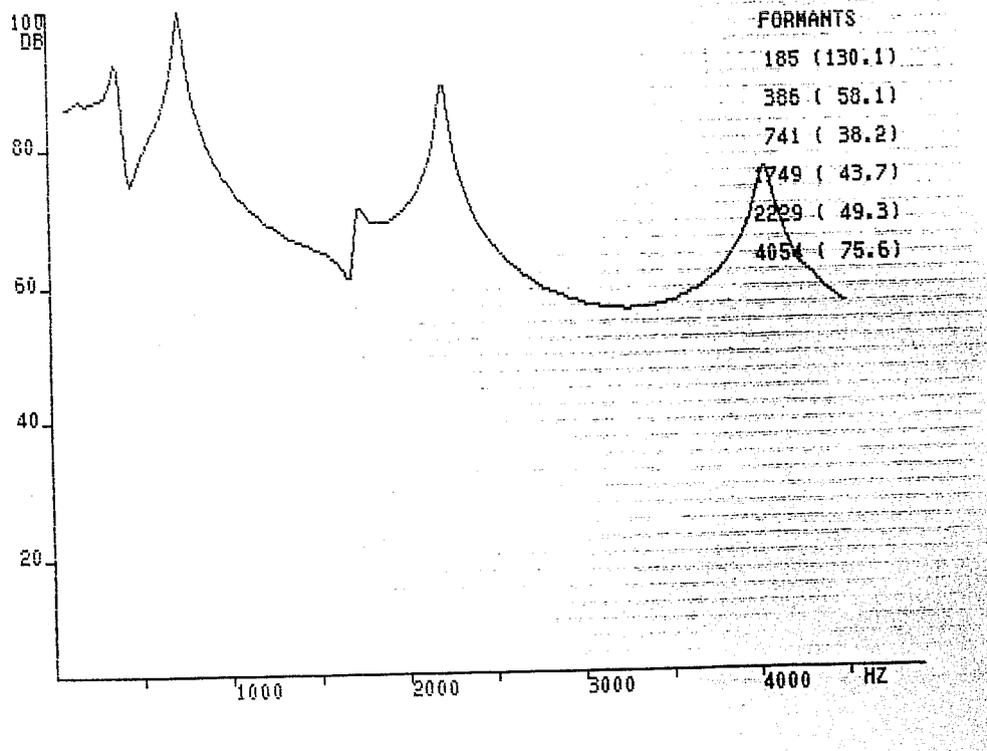


Figure 1.68 - Fonction de transfert du conduit nasal seul : deux sinus dissymétriques.

dépendant surtout de la longueur du conduit nasal. Il en est de même pour FN2 et FN3. En revanche les paramètres suivants influent significativement sur la fréquence de FN1. Une réduction du volume de la partie postérieure de 30 % élève FN1 de 5% (35 Hz). Un accroissement du volume produit l'effet inverse. On se souvient que l'un des sujets féminins de BJUGGREN & FANT (1964) montrait un volume postérieur très important. Il est donc probable que certains sujets possèdent une fréquence de FN1 plus basse que n'implique la taille de la seule zone narinaire. Une augmentation de la longueur totale du conduit le portant à 12.5 cm diminue la fréquence de FN1 de 90 Hz (-12%). Cette diminution est presque proportionnelle au rapport des longueur : 10.8 / 12.5 (-13.5%). Les formants supérieurs s'abaissent de 14 %.

La fonction d'aire décrite par STEVENS & HOUSE (1956) se caractérise à la fois, comme le montre le tableau suivant, par une longueur importante (12.5 cm) et une aire faible de la zone narinaire. Les sections sont décrites des narines vers le voile.

longueur section (cm)	1	2	4	2	1	2	0.5
aire (cm ²)	0.23	1.1	2.6	5.9	(3)	(2)	(1)

Les trois dernières aires, qui sont variables, ont été choisies arbitrairement. Les fréquences formantiques s'établissent à 356, 2098 et 3344 Hz. La fréquence de FN1 est presque deux fois plus basse que celle de la forme décrite par BJUGGREN & FANT. La faiblesse de l'aire de la première section en est responsable pour une très large part.

Pour l'essentiel, cette étude de la sensibilité des fréquences de résonance propres du conduit nasal confirme, ou révèle que :

- 1 - La fréquence de FN1 est déterminée par longueur totale du conduit, la taille de la zone proche des narines, et par le volume de la partie postérieure du conduit (rhinopharynx).
- 2 - Une dissymétrie importante des fosses nasales n'introduit pas de pôles supplémentaires. Ceux-ci sont vraisemblablement dus à la présence d'une ou plusieurs paires de sinus.

9.2 - DU ROLE DE L'ABAISSEMENT DU VOILE DANS LES VARIATIONS FORMANTIQUES ET DE LA DISSYMETRIE DES FOSSES NASALES.

L'abaissement du voile du palais provoque une diminution de l'aire de la zone vélaire du

conduit. En conséquence, les formants liés à la cavité orale seront modifiés même si la position linguale reste identique. BELL-BERTI (1983) a étudié sur un modèle articulatoire l'effet acoustique produit par une constriction vélaire. Les fonctions d'aire ne sont pas décrites. La modification du premier formant ne dépasse pas quelques Hz, sauf pour [u], où F1 chute de 50 Hz. Les fréquences de F2 de [i, a, ae, ʌ] sont relativement stables : +56 à -76 Hz. Seul le deuxième formant de [u] est abaissé de 110 Hz. Nous avons vérifié ces résultats dans le cas de [a], à partir de la fonction d'aire O3, en réduisant l'aire des sections 8 et 9 à 2 et 0.5 cm². Les fréquences des trois premiers formants chutent de 45, 82 et 203 Hz. Sauf pour F3, la baisse reste modérée. Cet effet s'oppose à la hausse provoquée par le couplage et tendra à réduire l'effet de la nasalité sur les fréquences formantiques.

LINDQVIST & SUNDBERG (1976) ont suggéré que la complexité spectrale de la fonction de transfert du conduit nasal provenait en partie d'une dissymétrie des fosses nasales. Les résultats de BJUGGREN & FANT n'indiquent que de faibles variations ($\pm 10\%$ au maximum). Nous avons testé cette hypothèse en simulant grossièrement un déplacement de la cloison nasale. Les aires des deux fosses ont été multipliées par 0.6 et 1.4, respectivement, ce qui représente une forte dissymétrie. Le seul effet visible (figure 1.69) est l'apparition, vers 1700 Hz d'une paire pôle - zéro supplémentaire dans la fonction de transfert nasale. On remarque aussi que la largeur de bande de FN1 reste constante. Ce résultat peu probant n'est pas incompréhensible. On sait que les fréquences formantiques sont fonction du rapport des aires, et non de leur valeur absolue. Une multiplication par un facteur constant ne modifie guère les fréquences de résonances des deux conduits. S'il est exact qu'une position excentrée de la cloison nasale ne bouleverse pas radicalement les rapports d'aire, il ne faut pas compter sur elle pour augmenter sensiblement le nombre des pôles. Cette augmentation est due essentiellement à la présence de cavités sinusales. Pour tenter d'obtenir un effet plus spectaculaire, nous avons simulé une occlusion de l'une des narines. Les résultats (figure 1.70) montrent à nouveau la présence d'une paire pôle - zéro supplémentaire vers 1700 Hz. Mais FN1 chute dans ce cas de 50 Hz. Il n'y a pas apparition de deux pics de basse fréquence, à la fréquence de résonance de chacune des cavités. Leur couplage au niveau du rhinopharynx est sans doute trop important. Ceci renforce notre conviction que les pics obtenus par LINDQVIST & SUNDBERG proviennent du couplage de cavités sinusales.

9.3 - SIMULATION ET ANALYSE DE LA CONSONNE NASALE VELAIRE.

Ce paragraphe est consacré à l'analyse de la consonne nasale vélaire [ŋ]. On a vu dans les sections précédentes que la simulation correcte du spectre de cette consonne était l'une des

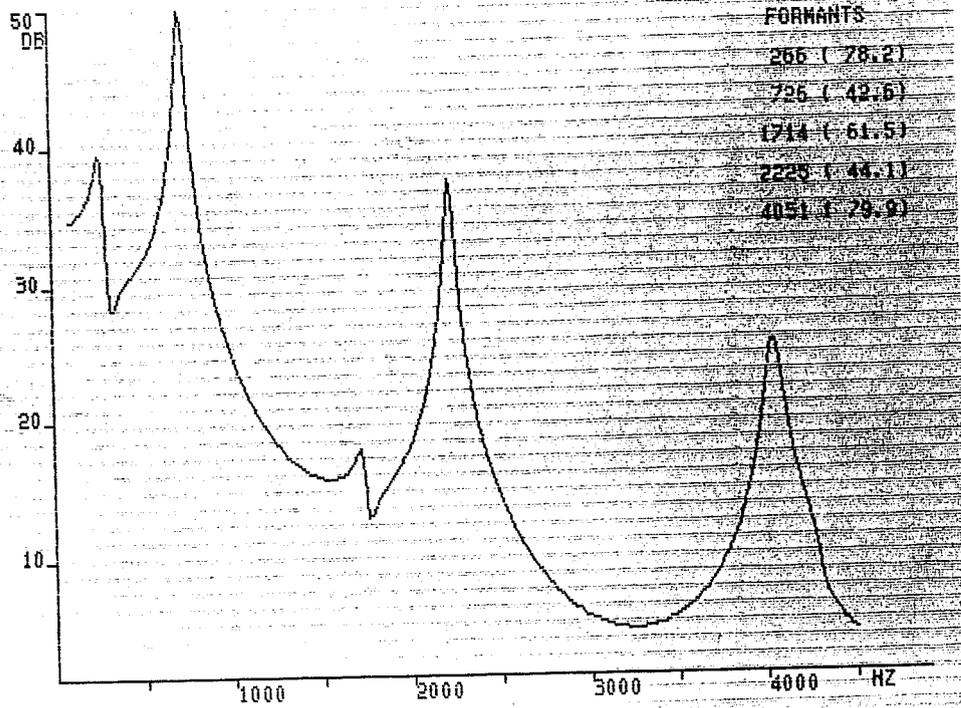


Figure 1.69 - Fonction de transfert du conduit nasal isolé : effet d'une dissymétrie de la cloison nasale.

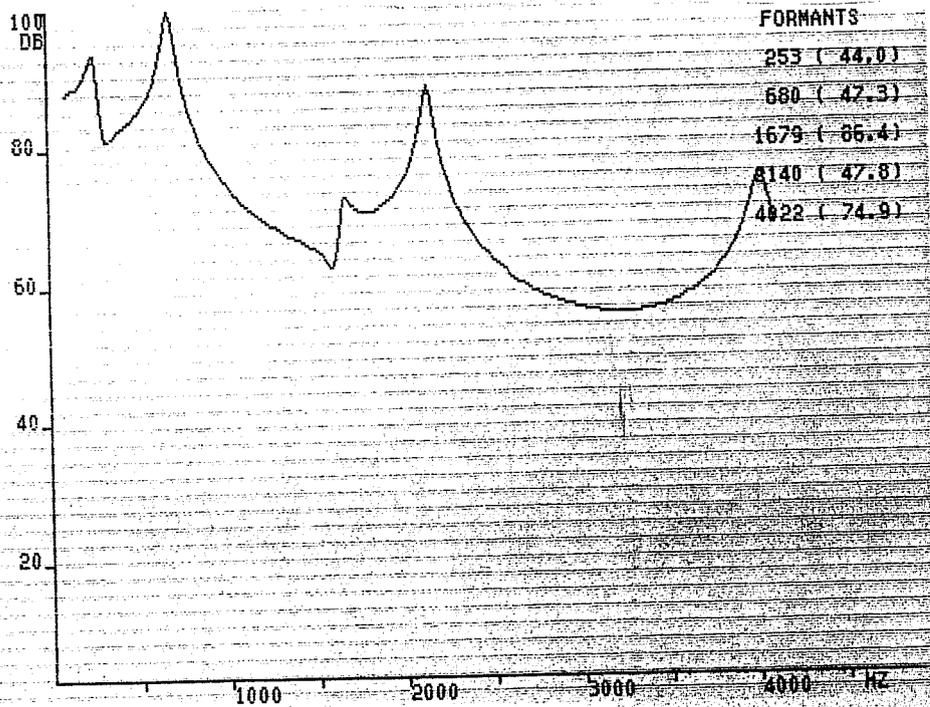


Figure 1.70 - Fonction de transfert du conduit nasal isolé : effet de l'occlusion d'une narine.

pierres de touche utilisées par FENG & al pour justifier leur choix d'une forme de conduit nasal à faible ouverture. Après avoir présenté l'articulation de cette consonne, nous décrirons ses caractéristiques spectrales à partir de quelques réalisations naturelles prononcées par l'auteur, en les comparant aux analyses publiées. Nous vérifierons qu'une simulation faisant appel au conduit nasal de BJUGGREN & FANT ne permet pas une simulation correcte. Nous proposerons enfin que quelques modifications plausibles de cette forme laissant intacte la forte ouverture nasale, ainsi qu'un choix adéquat de couplage sinusal permettent une simulation adéquate.

On peut vérifier sur la figure 1.71, d'après l'album de tracés radiologiques de KONECZA & ZAWADOWSKI (1951 ; figure 89), qu'un [ŋ], du moins lorsqu'il est prononcé de manière isolée, possède une très large ouverture vélaire et que la branche buccale est insignifiante, en raison de l'articulation très postérieure de la langue. Un simple branchement d'un conduit nasal sur un conduit pharyngal est donc justifié.

Les nasales vélaire n'ont pas fait l'objet de nombreuses études acoustiques. Au paragraphe 8-6, nous avons présenté et discuté les analyses par balayage fréquentiel externe de LINDQVIST & SUNDBERG (1976). Les spectres montrent un pic intense vers 270 Hz, suivi d'un pic plus discret à 500 Hz, et deux maxima vers 1000 et 2000 Hz. Pour FUJIMURA (1962), qui utilise une méthode d'analyse-par-synthèse, le premier pic se place selon les sujets entre 250 et 400 Hz. Pour le sujet K(enneth) S(tevens), notamment, les valeurs moyennes sont de 350, 1050, 1900 et 2750 Hz. La source glottale étant présente, il n'est pas interdit de penser que la fréquence fondamentale et / ou la deuxième harmonique, qui sont intenses, peuvent biaiser négativement les résultats des mesures dans certains cas, conduisant à la forte variabilité rapportée. Aucune mention n'est faite d'un petit pic vers 500 Hz. Quelques [ŋ (ð)] prononcés par l'auteur sont présentés à la figure 1.72. La variabilité de la partie basse est notable. Deux configurations sont possibles, qui coïncident curieusement avec deux séances d'enregistrement à des jours différents. Dans un cas (3 premiers spectres, n° 10 à 12), les trois premières harmoniques sont intenses, et séparées de la cinquième par un léger creux. Si l'on tient compte du "formant" glottal, le premier pic doit se trouver vers 300 Hz, et le deuxième vers 430 - 440 Hz, F_0 étant proche de 100 Hz. Le pic suivant se place vers 750 - 800 Hz, mais il semble à deux reprises scindé par un minimum spectral. L'un d'entre eux est peut-être une résonance sinusale, avec un zéro proche. Les maxima suivants se trouvent vers 2000 et 2500 Hz. Dans les autres cas (n° 24 à 26), les quatre premières harmoniques sont d'intensité décroissante, et il est très difficile de déterminer la fréquence de résonance. Il n'y a rien de notable vers 400 - 500 Hz, et même le pic vers 800 Hz n'est qu'à peine visible. Cette dichotomie suggère que la condition physiologique des fosses nasales du sujet joue un rôle non négligeable dans la forme du spectre. On ne peut guère faire l'hypothèse d'une variabilité

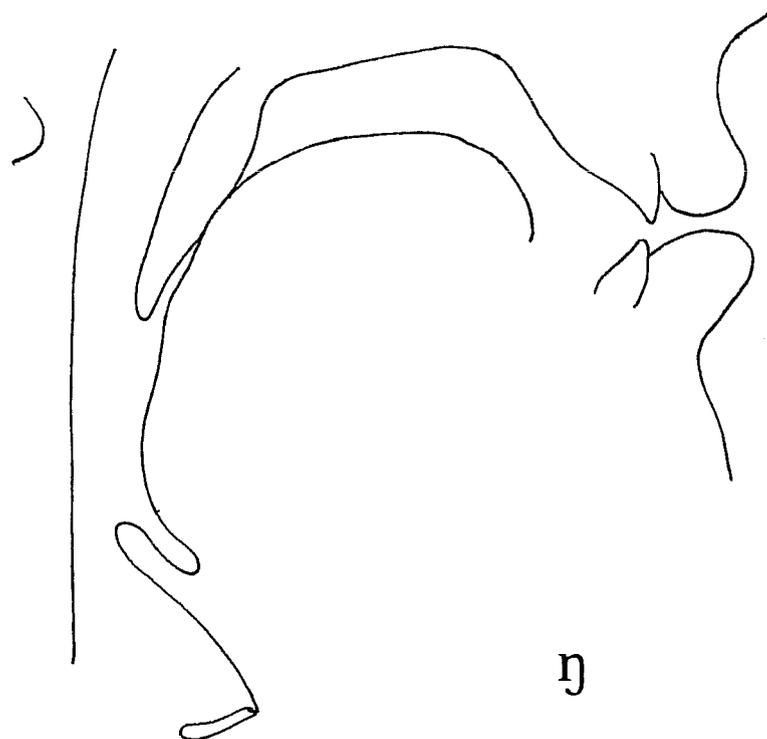


Figure 1.75 - Position articulatoire de la consonne [ŋ]
(in KONECZNA & ZAWADOWSKI 1951; fig. 89)

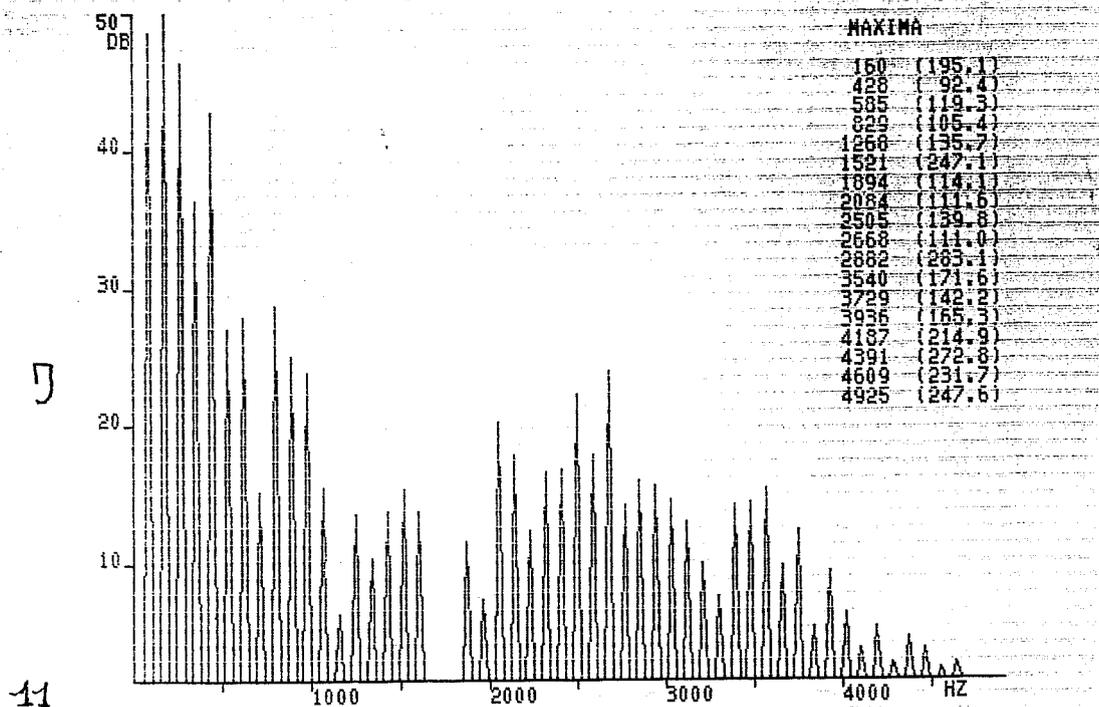
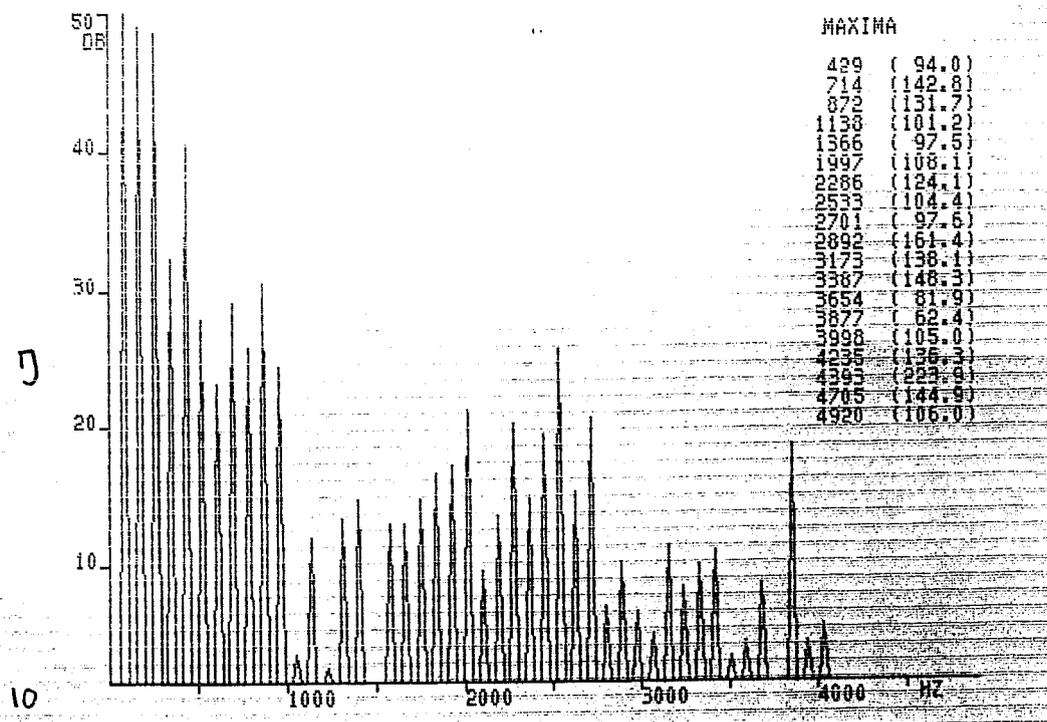


Figure 1.72 - Spectres par transformée de Fourier de consonnes [g] et [ŋ].

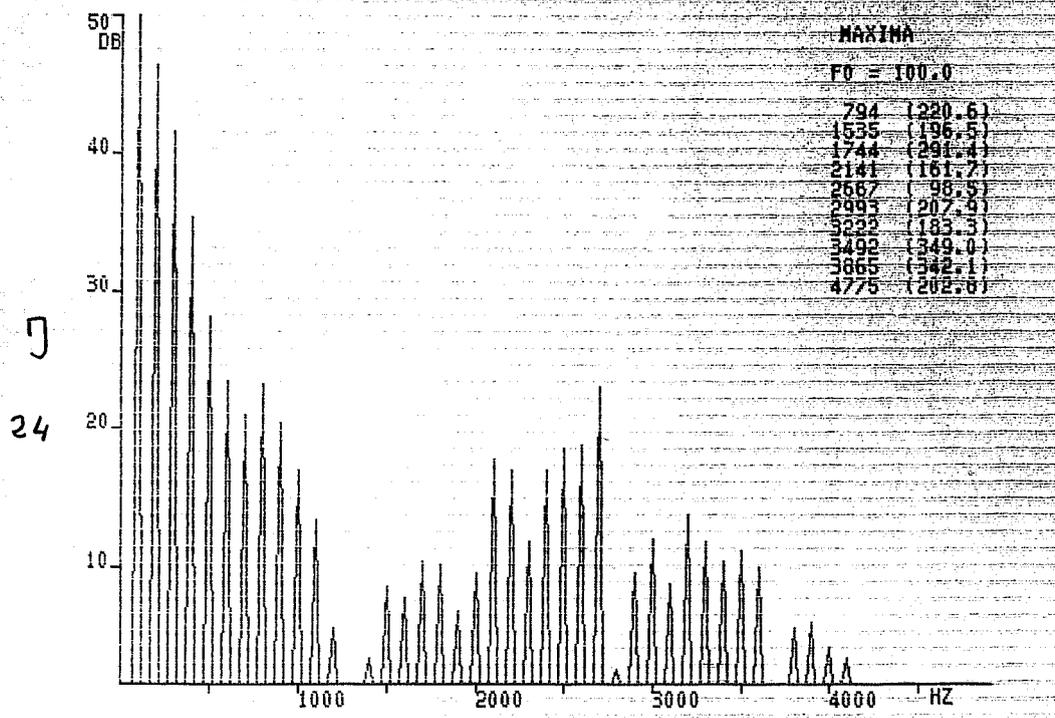
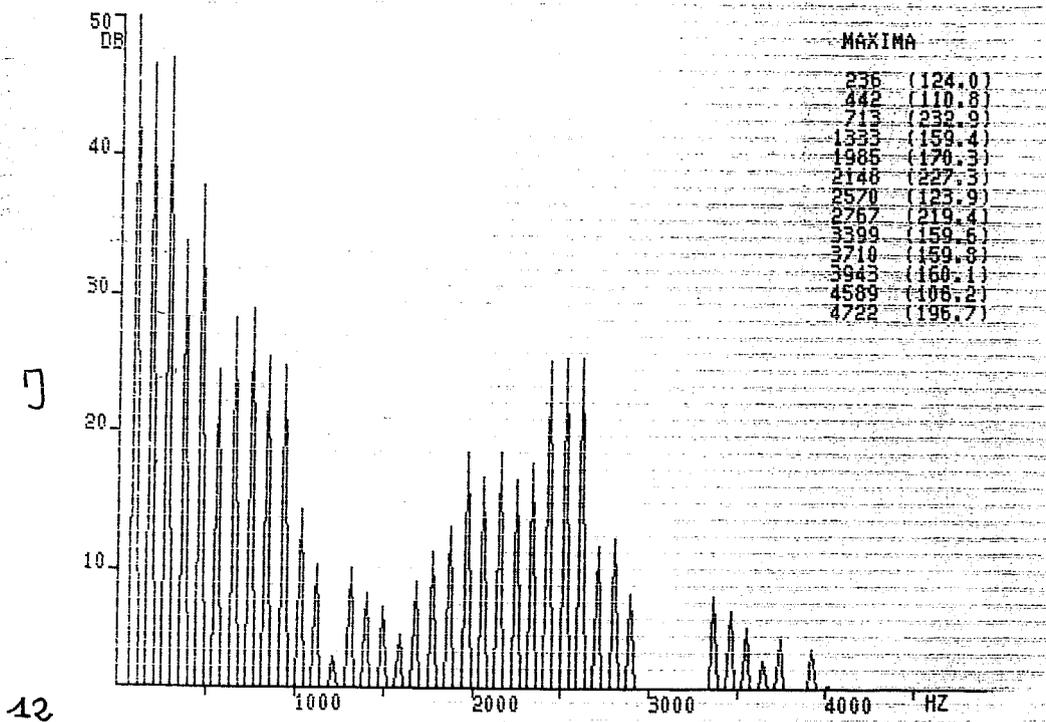


Fig. 1.72 (Suite).

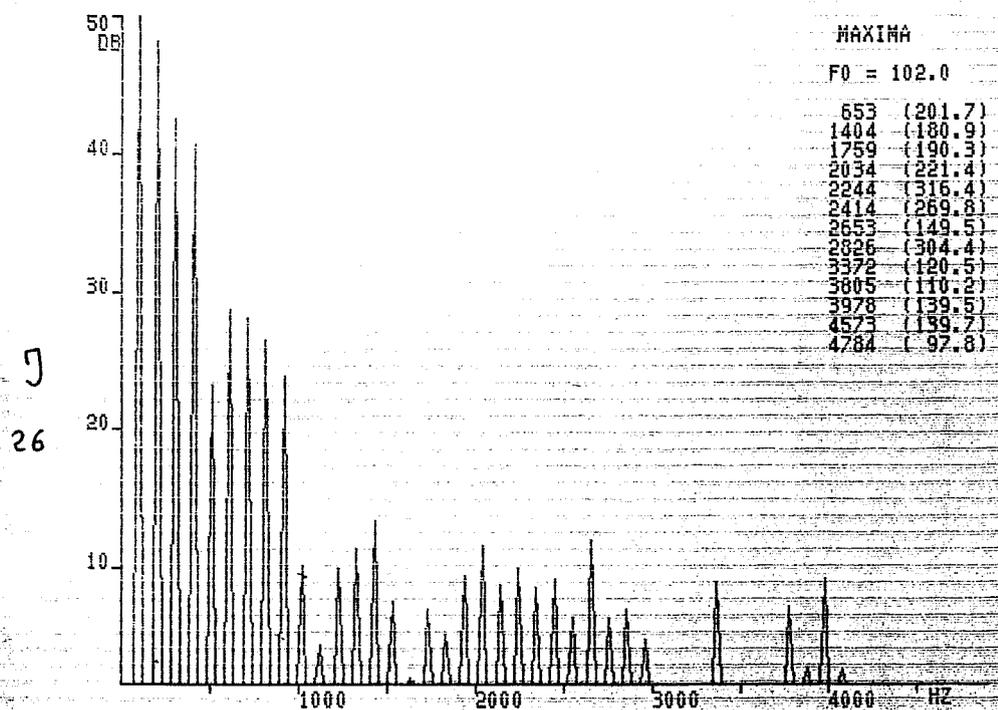
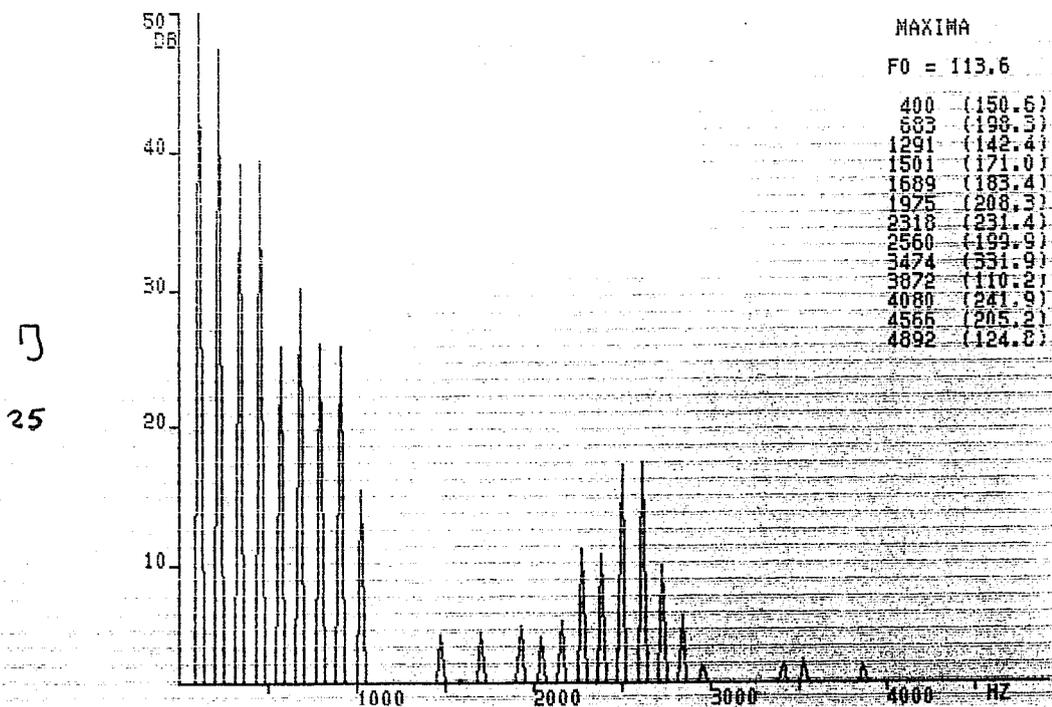


Figure 1.72 (Suite).

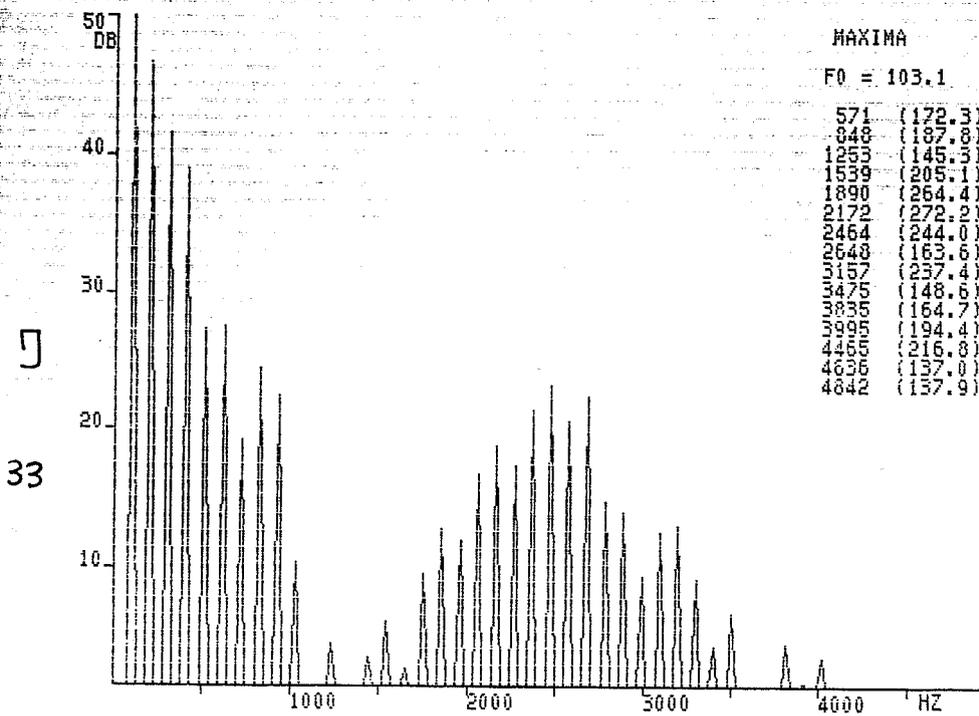
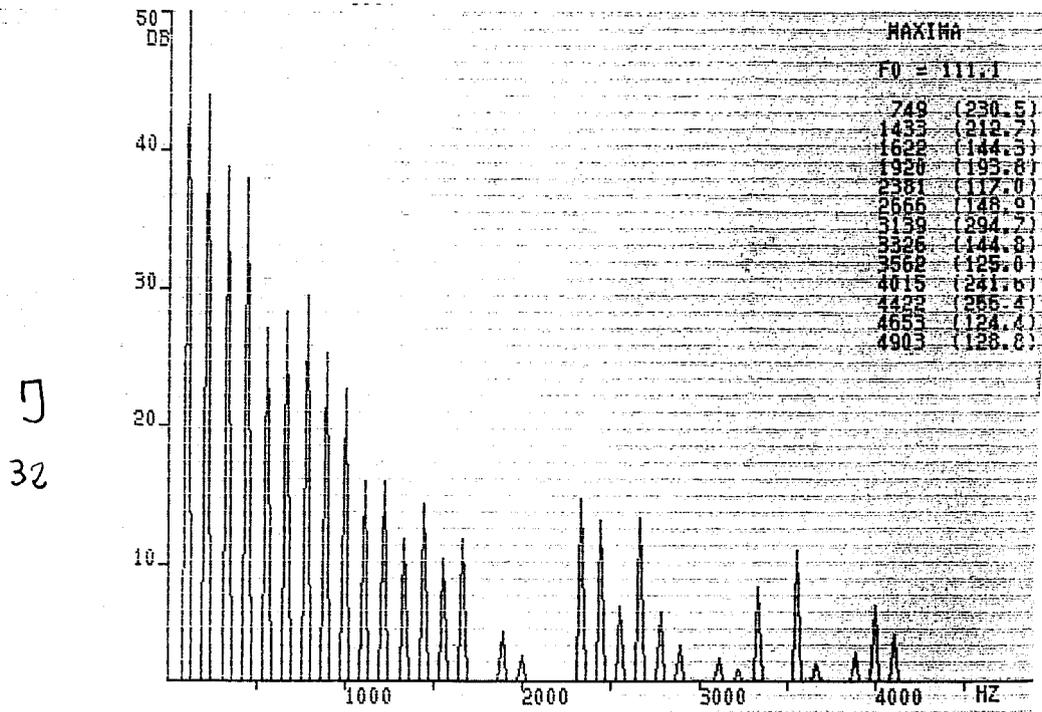


Figure 1.72 (Suite).

d'articulation, car les données pour une session montre une bonne stabilité. Les deux derniers spectres (n° 32, 33) ont été volontairement enregistrés pendant que le locuteur souffrait d'un léger coriza. Il montre la même allure que la deuxième série. Les quatre premières harmoniques sont intenses, ce qui place la première résonance près de 400 Hz. Le pic suivant est bas, entre 750 et 850 Hz. Les simulations présentées ci-dessous permettent de rendre compte de cette variabilité.

La structure de simulation retenue par FANT ne permet pas de réaliser l'occlusion de la zone buccale. Une paire pôle - zéro apparaît même si l'on réduit le passage buccal à une valeur très faible. Nous avons donc modifié le programme pour que la branche buccale soit en dérivation sur le conduit nasal, en non l'inverse comme précédemment

Les résultats de la simulation de [ŋ (ə)] (conduit oral O1, conduit nasal N1 de FANT avec un couplage de 2.4 cm², sinus S1) sont les suivants : la résonance des sinus apparaît à 260 Hz, suivie du premier formant à 534 Hz et du deuxième à 968 Hz, les formants supérieurs étant à 2350 et 3175 Hz. Si l'on compare ce spectre à ceux de LINDQVIST & SUNDBERG, ou à ceux des voyelles naturelles, l'écart est considérable. Le premier formant est beaucoup trop haut. La cavité sinusale employée ici possède un couplage trop faible, et une fréquence de résonance trop basse pour créer un pic significatif de basse fréquence. Il est impensable d'attribuer le pic de 270 Hz de LINDQVIST & SUNDBERG à une résonance sinusale en raison de son amplitude très importante et de la taille réduite du second vers 500 Hz, caractéristiques que l'on retrouve dans nos mesures sur des réalisations naturelles. Le conduit nasal de BJUGGREN & FANT (1964) ne permet pas une simulation correcte. Ceci est un élément majeur de la discussion de FENG & al.. Deux solutions se présentent. On peut changer la forme de l'extrémité du conduit nasal pour abaisser sa première fréquence de résonance. C'est la solution préconisée par FENG & al. (1985). Nous avons vu plus haut qu'elle se heurte à deux difficultés au moins. Les données anatomiques ne montrent pas d'erreurs flagrantes de la géométrie du limen nasi. Les estimations par balayage fréquentiel externe du premier pic de résonance (LINDQVIST & SUNDBERG) du conduit nasal seul ne sont pas compatibles avec une fréquence trop basse. Le sujet J.S montre une fréquence de 570 Hz, et si on accepte notre discussion des données de l'autre sujet, la fréquence n'est pas de 450 Hz, mais plus élevée. On est loin néanmoins des 700 Hz du conduit de BJUGGREN & FANT. La forme retenue par FENG & al. (figure 1.60) vise précisément à produire cette fréquence de 450 Hz. On peut citer un troisième problème. Pour rendre compte des données de FUJIMURA & LINDQVIST (1971), si on ne suppose pas un formant nasal très bas (vers 300 Hz), il faut que celui-ci soit dans deux gammes limitées pour pouvoir être confondu soit avec F'1, soit avec F'2.

La simulation de la consonne nasale vélaire de FANT (1960), numériquement satisfaisante

puisque la fréquence du premier formant est de 300 Hz, est fondée sur une hypothèse peu réaliste quant à la taille du pharynx. Nous proposons, après de nombreuses simulations exploratoires, que ces diverses contraintes peuvent être (approximativement) satisfaites par un conduit nasal

- 1 - un peu plus long que celui de BJUGGREN & FANT (12.5 cm, comme dans FANT 1960),
- 2 - possédant une partie postérieure de taille un peu plus importante, comme l'indique la fonction d'aire d'un de leurs sujets féminins, et largement couplée au pharynx,
- 3 - dont la taille des fosses au niveau des cornets est un peu inférieure aux valeurs classiques (cette modification ne déplace pas le formant dans la bonne direction !),
- 4 - et sur lequel sont branchés deux sinus possédant une fréquence de résonance qui est voisine de la fréquence du premier formant lorsque le conduit nasal est couplé.

Le conduit oral est supposé un peu plus long (18.5 cm) que dans la simulation précédente, pour tenir compte de la légère protrusion labiale qui accompagne le geste vélaire, et l'aire pharyngale un peu plus grande. En revanche on conserve une ouverture narinaire de 1.5 cm² pour le premier cm du conduit. Les fréquences de résonance propres de la partie nasale sont, sans sinus de 540, 1755 et 3470 Hz. Les résultats de cette simulation (conduit oral O2, conduit nasal N2, sinus S2) sont aux figures 1.74 et 1.75. Cette dernière, à échelle de fréquence logarithmique, doit faciliter la comparaison avec la figure 1.57, d'après LINDQVIST & SUNDBERG.

Le premier pic se place à 320 Hz. Il résulte de la scission par un zéro, lié à la résonance sinusale, du pic unique que forment la résonance sinusale et F1 approximativement à la même fréquence. Le calcul du pic sinusal par la formule de Helmholtz donne 345 Hz. Le zéro sinusal affaiblit les fréquences autour de 450 Hz, ne laissant subsister qu'une partie du flanc droit de FN1 qui apparaît sous la forme d'un pic moins intense (- 10 dB par rapport au premier) vers 475 Hz. Les formants suivants sont à des fréquences adéquates : 1180 et 2000 Hz environ. Le deuxième formant est néanmoins un peu trop haut. La présence de sinus est impérative pour obtenir une forme adéquate du pic de basse fréquence. S'ils sont supprimés, la fréquence du premier formant remonte à 385 Hz (figure 1.76), et le pic de 500 Hz disparaît.

FENG fait remarquer que l'utilisation d'un conduit à large ouverture narinaire nécessite l'introduction de sinus pour obtenir un pic de basse fréquence, vers 300 Hz. Nos simulations confirment cette conclusion. Mais nos mesures montrent également que le premier formant de la consonne vélaire est plus élevé lorsqu'un gonflement de la muqueuse

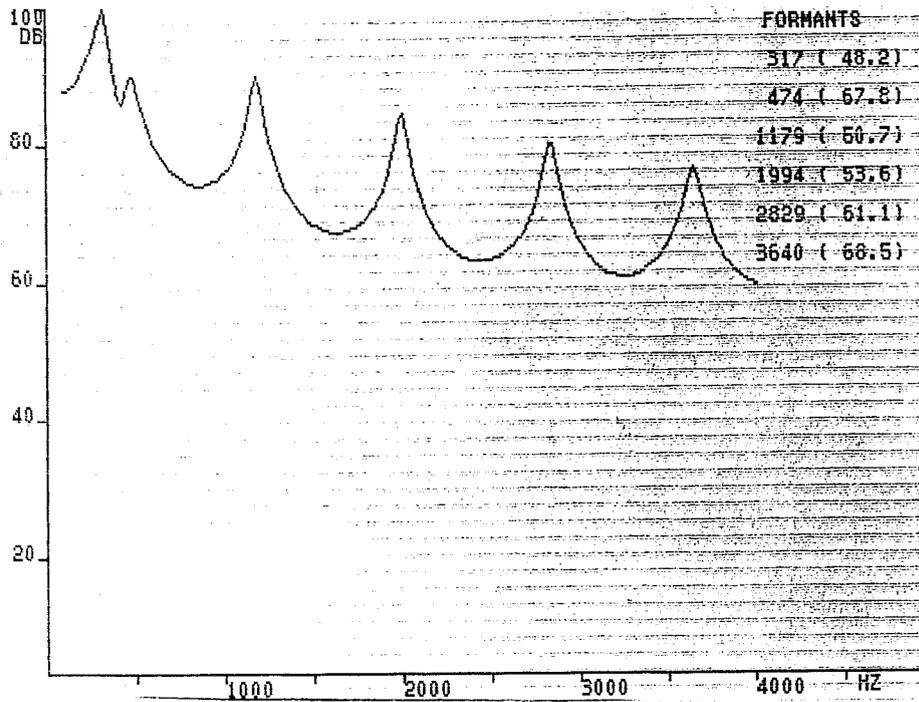


Figure 1.74 - Simulation du spectre d'une consonne [ŋ].

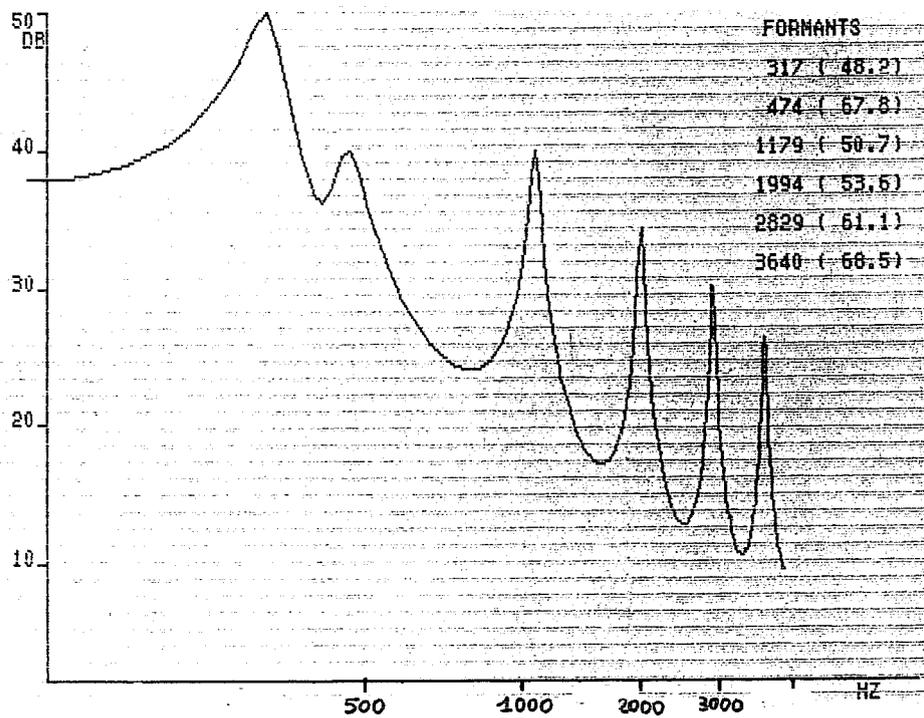


Figure 1.75 - Spectre de la figure 1.74 : échelle logarithmique.

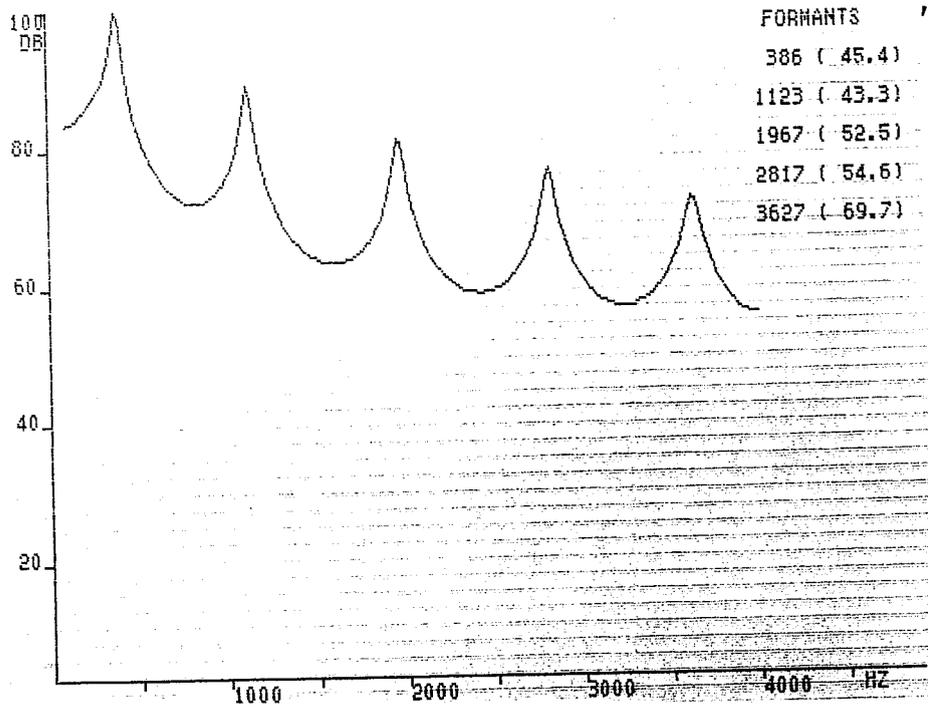


Figure 1.76A - Spectre d'une consonne [ŋ] : simulation sans sinus.

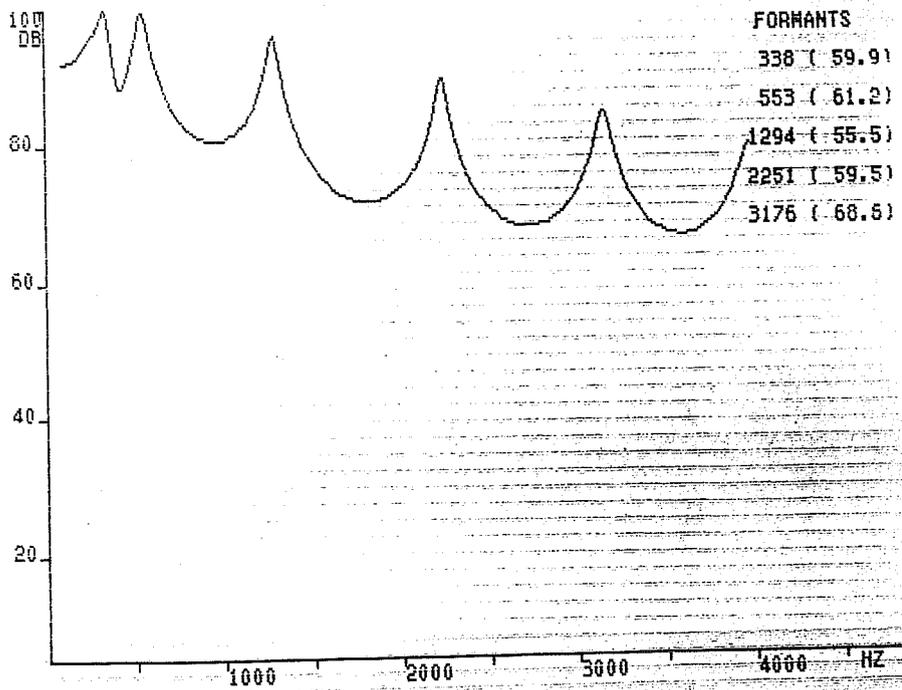


Figure 1.76 B - Spectre d'une consonne [ŋ] : fonction d'aire nasale de BJUGGREN & FANT, avec sinus.

nasale est présent. Nous attribuons cette élévation à la fermeture de l'ostium du sinus maxillaire. Cette augmentation de la fréquence est contraire au sens attendu de la variation formantique dans le cas d'une réduction de l'aire de la zone narinaire. Il ne paraît pas impossible de supposer que le gonflement des muqueuses accompagnant un léger coriza aura d'abord pour effet de réduire le diamètre de l'ostium maxillaire. La réduction de l'aire peut être trop faible pour changer les fréquences formantiques de façon significative. Les simulations de FENG montrent qu'une chute importante de la fréquence du premier formant ne se produit que pour une aire aux narines très réduite, d'environ 0.5 cm².

La présence des sinus branchés sur le conduit nasal de FANT ne conduit pas à résultats de simulation plus proche des mesures sur des consonnes naturelles (figure 1.76). On obtient deux pics d'égale amplitude, à 340 et 550 Hz. La fréquence du premier est identique à la valeur de la résonance sinusale calculée par l'équation de Helmholtz.

Ces résultats conduisent à des conclusions importantes :

- 1 - Une simulation adéquate du spectre de la consonne nasale vélaire peut être obtenue en utilisant un conduit nasal possédant une aire narinaire conforme aux mesures anatomiques disponibles. Mais la longueur du conduit nasal et le volume de la partie rhinopharyngale doivent être supérieurs aux mesures de BJUGGREN & FANT (1964).
- 2 - Le présence de sinus est nécessaire pour qu'un zéro vienne décaler la fréquence du premier pic de 380 à 320 Hz environ. Ce décalage est le produit du creusement du flanc droit du premier formant par le zéro sinusal. La présence d'un petit pic vers 500 Hz s'explique également par la présence de la paire pôle - zéro liée au sinus.
- 3 - Le gonflement de la muqueuse produit vraisemblablement une fermeture de l'ostium maxillaire. L'absence de résonance sinusale provoque un décalage du premier formant vers les fréquences hautes, attesté par les mesures des fréquences formantiques de réalisations naturelles.
- 4 - L'hypothèse d'une forte constriction narinaire pour rendre compte de la fréquence basse de F1 est inutile. Elle n'est compatible ni avec les données anatomiques, ni avec la fréquence de la première résonance propre du conduit de 570 Hz mesurée par LINDQVIST et SUNDBERG pour le sujet J.L., ni avec l'élévation de F1 constatée lorsque qu'un sujet présente un gonflement des muqueuses sous l'effet d'un léger coriza.

10 - IDENTIFICATION DES INDICES DE NASALITE DANS LES VOYELLES NATURELLES DU FRANCAIS.

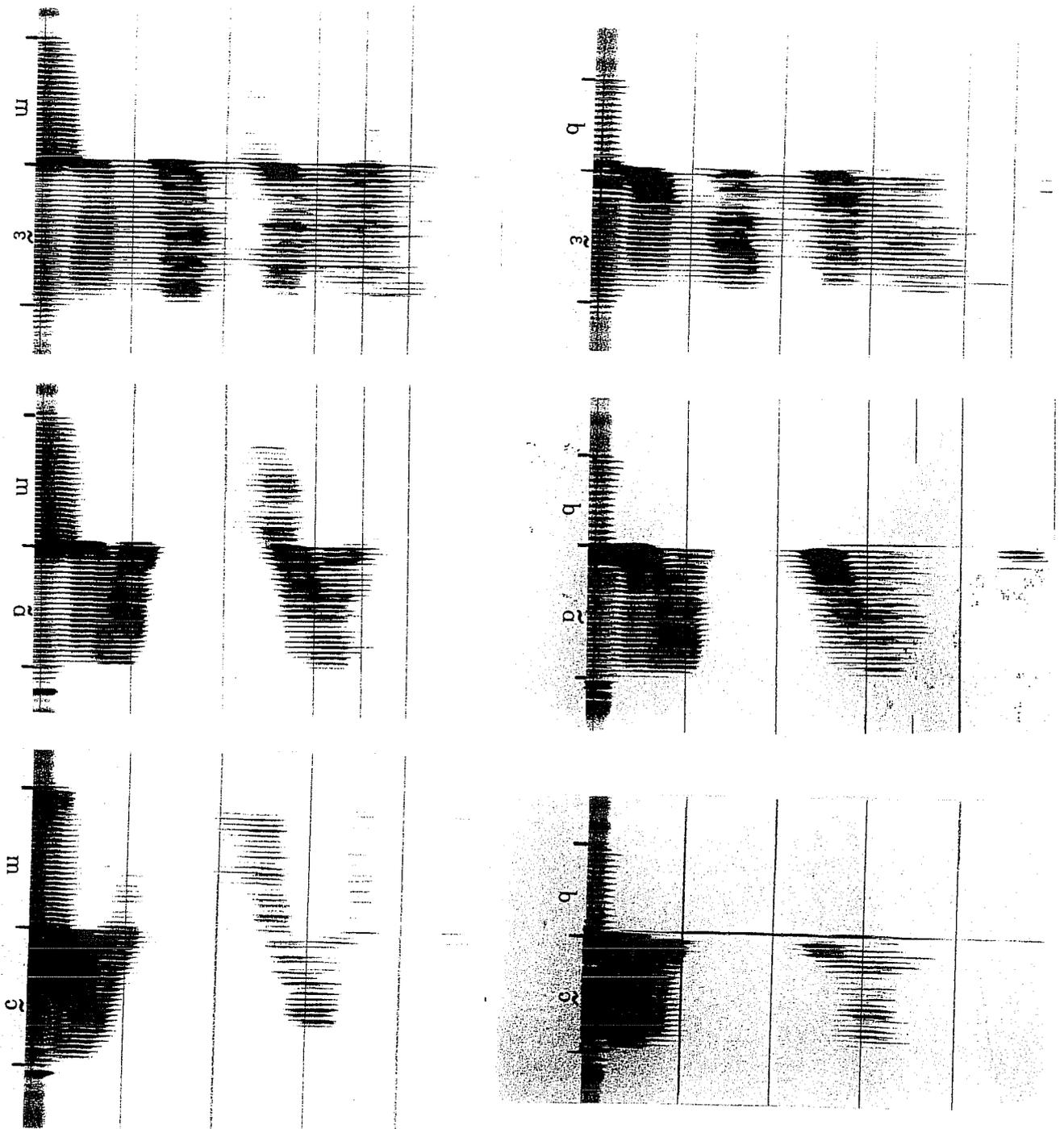
Dans cet avant-dernier paragraphe, nous tentons d'identifier, de manière raisonnée, les caractéristiques spectrales que présentent les spectres d'un corpus de voyelles nasales du français prononcé par l'auteur. Ce corpus ne présente pas certaines caractéristiques habituelles de ceux du domaine, et quelques explications s'imposent. Tout d'abord, il est de taille limitée, car nous souhaitons en faire une analyse exhaustive, au cas par cas, et non constituer des statistiques. A la lumière de notre expérience antérieure (LONCHAMP 1978), il est apparu qu'une dizaine de spectres par voyelle suffisait pour nourrir une discussion. Nous croyons que des statistiques doivent se fonder sur le magnifique corpus enregistré au C.N.E.T. sous l'égide du GRECO "Communication Parlée". Mais sa taille requiert en pratique des procédures (semi -) automatiques en vue de la segmentation et de l'analyse, qui n'existent pas encore. Il est clair que seules des analyses exploratoires comme celle-ci permettront de les dégager. L'utilisation d'un seul locuteur a été bien sûr dicté d'abord par des raisons pratiques. Mais nous pensions que les différences entre les locuteurs, telles que l'analyse des documents radiologiques et la nature variable du couplage nasal permettent de le prévoir, seraient de nature à rendre encore plus complexe l'examen des réalisations naturelles. Aujourd'hui, à posteriori, nous pensons avoir péché par excès de prudence. Pour atténuer l'impact négatif de ce choix, nous procéderons à la réanalyse des résultats présentés dans notre thèse de troisième cycle pour un autre locuteur, ainsi que de ceux de MRAYATI (1975).

Les indices étudiés sont les indices statiques, présents tout au long d'une voyelle nasale. La variabilité spectrale des voyelles nasales est souvent présentée comme un indice supplémentaire de nasalité, par opposition à la stabilité spectrale des voyelles orales. La réalité est plus nuancée. La figure 1.77 présente les spectrogrammes des 3 voyelles nasales du français précédées de [m] et [b]. La variabilité spectrale peut d'abord être due à une ouverture tardive du voile par rapport à la fin de la consonne précédente, la première partie de la voyelle restant orale. Si cela est vrai, la présence d'une consonne nasale en début de syllabe doit fortement réduire cette phase orale initiale, car le voile du palais sera en position basse dès le début de la voyelle. Or il apparaît que l'évolution spectrale ne diffère guère selon la nature orale ou nasale de la consonne précédente. La présence d'une phase orale initiale ne rend donc pas compte de la variabilité spectrale des spectrogrammes présentés. On peut être tenté de l'attribuer à une modification dans le degré d'abaissement du voile. Mais il faut noter que la structure acoustique de la voyelle [ɛ̃] reste très stable, si les transitions initiales montantes sont éliminées. La variabilité spectrale ne constitue donc pas un indice potentiel de nasalité pour l'ensemble des voyelles nasales du français. Elle est

présente pour les voyelles postérieures seulement. Il n'est donc pas injustifié de s'intéresser principalement aux indices statiques. Mais notre étude n'est certainement complète. Il conviendrait également de caractériser les durées intrinsèques qui distinguent clairement les voyelles orales et nasales.

10.1 - LE CORPUS

Le corpus a été constitué au fil des mois, par enregistrement direct sur l'ordinateur dans une pièce calme, le plus souvent tôt le matin pour profiter d'une voix reposée. Rétrospectivement, ce choix, de par la présence d'une large variabilité spectrale, nous apparaît comme un facteur décisif dans le succès de l'identification des indices. Comme on le verra, cette variabilité est absente de l'échelle de temps brève de l'enregistrement d'un corpus en une session. Avant chaque analyse acoustique, la réalisation phonétique a été scrutée auditivement pour s'assurer de sa conformité avec le timbre phonétique voulu. Une fréquence fondamentale grave, mais compatible avec l'étendue du registre du locuteur, a été recherchée afin d'augmenter la résolution spectrale. Celle-ci est indiquée sur chaque spectre, et varie de 90 à 105 Hz environ. Cette contrainte de hauteur impose au locuteur une voix relativement "faible", qui se traduit spectralement par une fréquence fondamentale très intense. Si une voix faible est le produit d'une onde de débit glottal à phase ouverte longue, le "formant" glottal doit être bas. La deuxième harmonique n'étant pratiquement jamais renforcée, on peut estimer sa fréquence à 150 Hz au plus. La durée de la phase d'ouverture est donc au minimum de 3.5 ms. Si la phase de fermeture est lente (K faible cf. § 7.1), la durée de la phase ouverte est d'au moins 70% de la durée de la période, valeur généralement retenue pour le coefficient d'ouverture des voyelles "faibles". Toutes les voyelles nasales ont été prononcées dans une syllabe [mV] pour éviter la descente du voile en cours de la réalisation vocalique. L'analyse a été réalisée au premier tiers de la durée de la voyelle, typiquement 70 à 100 ms après la fin de la consonne pour éviter les transitions consonantiques, mais aussi et surtout l'abaissement du voile qui précède la terminaison de la voyelles. Le calcul spectral n'a été effectué que si la voyelle présentait un plateau d'énergie dans les 70 à 100 premières ms, une excellente régularité temporelle des deux périodes adjacentes à celle analysée, et si l'écoute des trois périodes révélait un timbre bien défini. En revanche, aucune sélection n'a été effectuée sur la forme du spectre! De nombreuses analyses de deux périodes successives ont révélé une grande stabilité de la position et de la taille relative des pics, mais une variabilité de la forme et de la position exacte des vallées. Il faudrait à ce propos citer très longuement la lumineuse partie de la thèse de FENG (1986) consacrée aux modèles d'analyse pôle / zéro qui démontre de manière irréfutable l'impossibilité pratique d'estimer la position exacte d'un zéro dans les



• Figure 1.77 - Spectrogrammes des voyelles [ɛ], [ɑ] et [ɔ] précédées de la consonne [b] ou [m].

conditions réelles de production de la parole. Les quelques remarques sur la position probable des zéros se fondent plus sur l'observation d'effets induits comme l'affaiblissement d'un pic, ou un décalage fréquentiel, que sur l'observation directe d'une profonde vallée spectrale. Nous parlerons par la suite de spectres plats, pour caractériser la faible énergie de la zone de basse fréquence jusqu'à 1 KHz. L'utilisation de ces termes repose sur le critère (semi-)quantitatif suivant. La pente moyenne du spectre de - 6 dB / octave et la théorie acoustique de la production (FANT 1960 p. 55 - 59) font que l'amplitude relative de deux formants ayant les mêmes largeurs de bande est parfaitement déterminée par leur écart fréquentiel. Pour F1 et F2 à 500 et 1500 Hz, F2 est - 10 dB sous F1. Pour F1 = 500 et F2 = 800 Hz, la différence n'est que de - 4 dB. Si l'écart est différent, c'est qu'une caractéristique supplémentaire est présente : variation de la largeur de bande d'un formant ou présence d'un zéro (ou d'un pôle supplémentaire) modifiant par sa proximité l'un des formants. Un spectre plat désignera un spectre où l'écart est plus faible que la norme. Rappelons (cf. § 12.1) que la présence d'une paire pôle / zéro entre deux formants ne rend pas ipso facto le spectre plat comme nous l'avons écrit par erreur en 1978.

Le choix des spectres présentés ici est arbitraire, ainsi que leur numéro d'ordre. L'étude de chaque voyelle comprendra une description des formes spectrales jusqu'à 3000 - 3500 Hz, un tableau présentant les fréquences des pics, une analyse de résultats antérieurs, et notre interprétation. L'analyse cinéradiographique de nos productions (ZERLING (1984)) nous a convaincu de l'inutilité d'analyser la voyelle [õe] que nous ne distinguons plus de [ẽ].

10.2 - LA VOYELLE [ẽ̃]

10.2.1 - DESCRIPTION SPECTRALE

Les spectres sont reproduits à la figure 1.78. Ils ont été placés dans un ordre déterminé par la fréquence du premier pic commun présent dans tous les spectres dont nous identifierons finalement comme le premier formant nasal (FN1). L'utilisation des mots pic, maximum ou formant n'obéit dans cette partie qu'à des impératifs stylistiques.

Le spectre 1 montre, sous Fo très intense, un pic large vers 400 Hz, une harmonique plus intense vers 800 et une très légère irrégularité vers 1200 Hz, puis deux formants bien définis à 1500 et 2600 Hz. Le spectre est très plat puisque la différence n'est que de 3 dB entre les pics de 400 et 1500 Hz. Aucune vallée profonde n'existe sous 1500 Hz.

L'allure du spectre 2 est voisine. Le premier pic est très dissymétrique, ce qui suggère la présence d'un zéro juste à sa droite. L'harmonique de 800 Hz est très amplifiée. L'épaule gauche du pic de 1500 Hz est à la fréquence de la petite irrégularité spectrale du spectre 1.

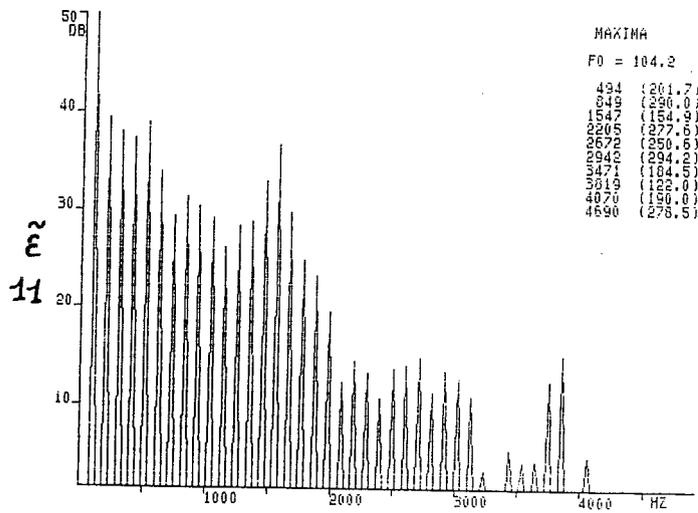
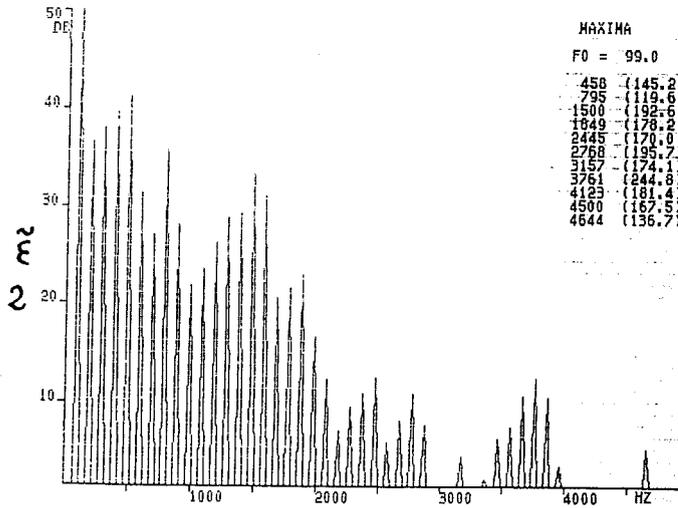
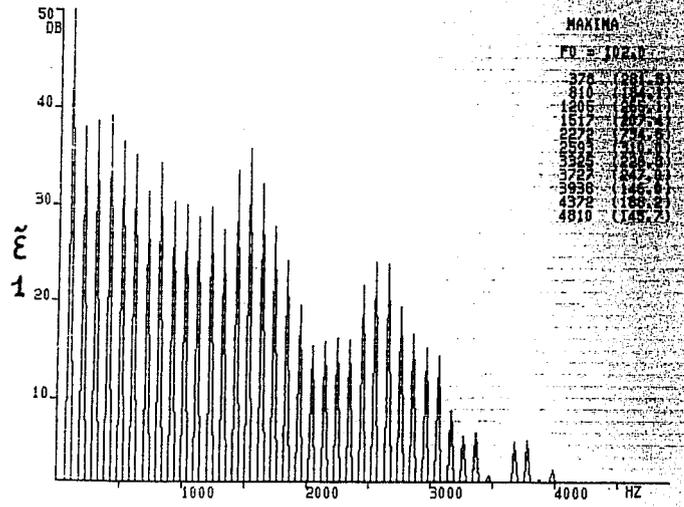


figure 1.78 - Spectres des voyelles [ɛ̃].

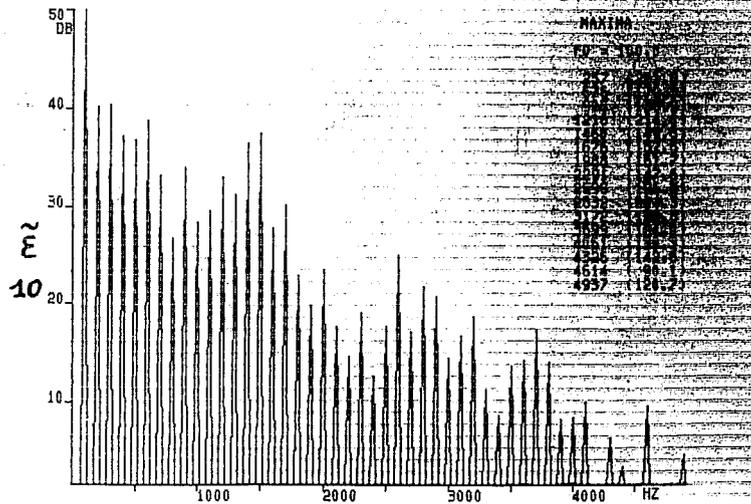
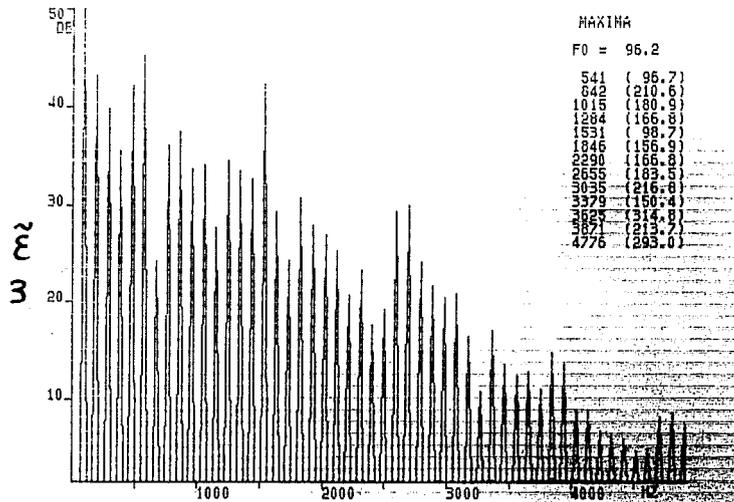
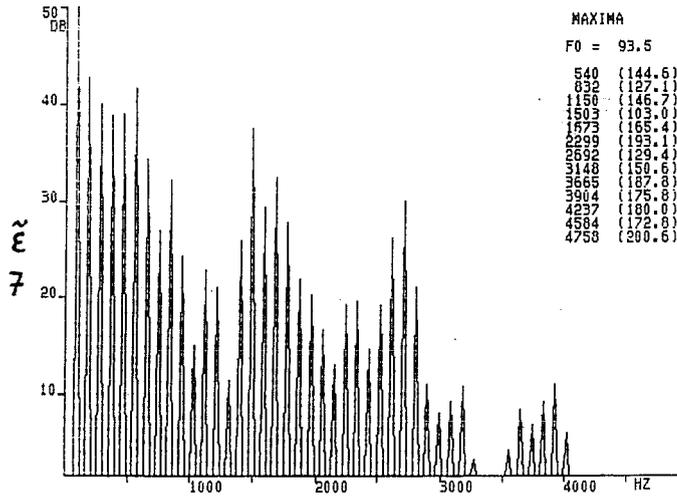


Figure 1.78 (suite).

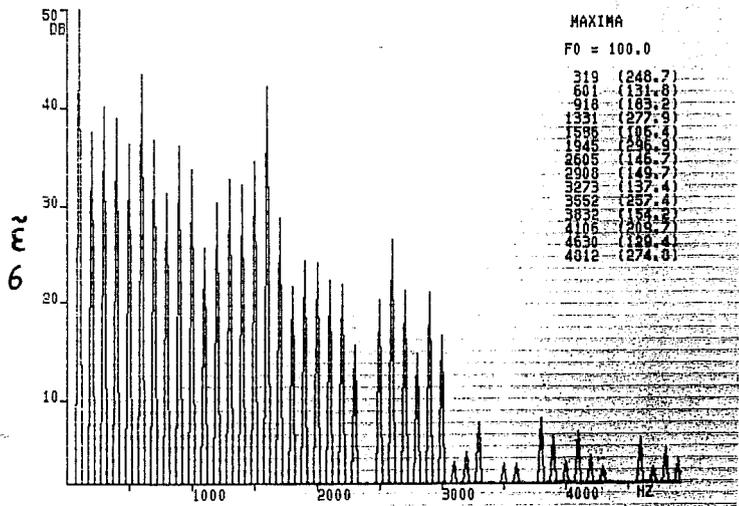
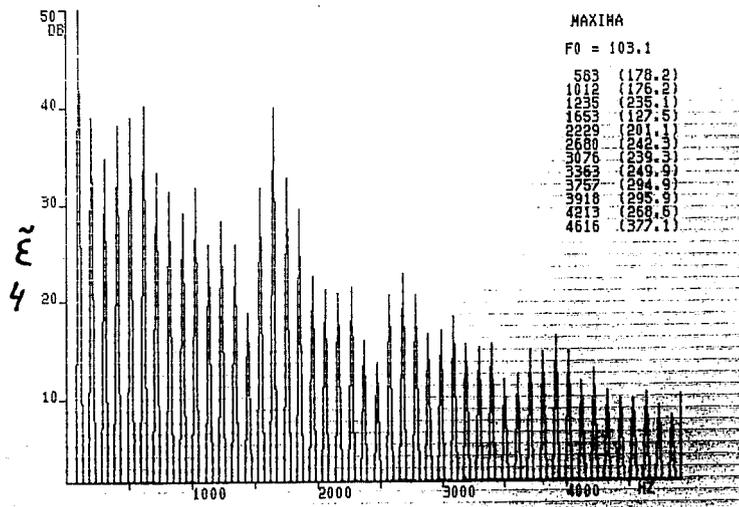
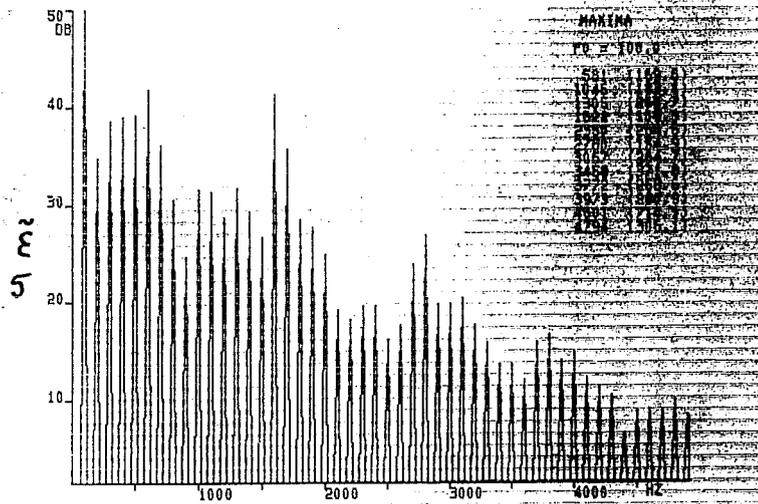


Figure 1.78 (suite).

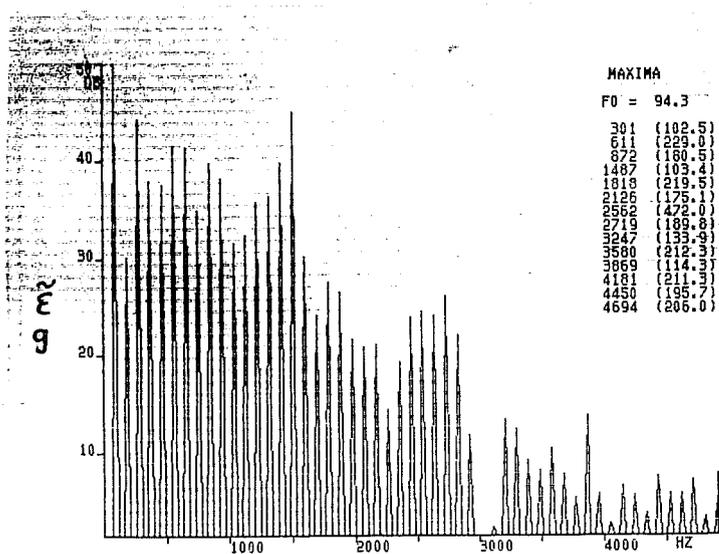
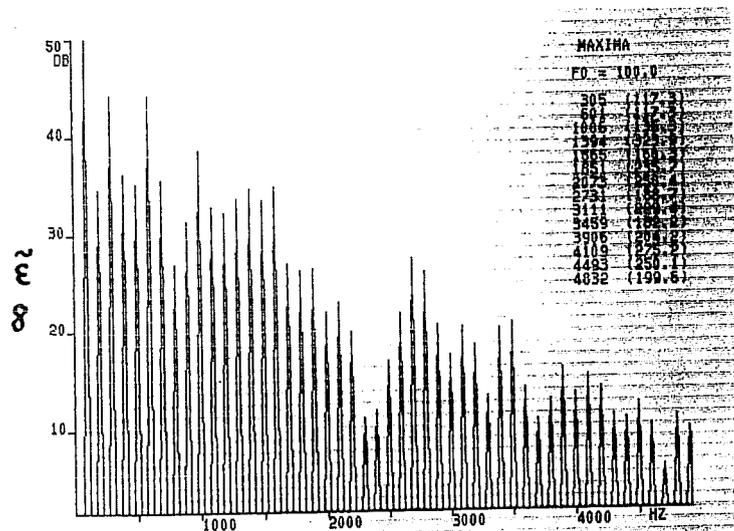


Figure 1.78 (Suite).

Au-dessus, on note plusieurs pics vers 1850, 2500 et 2800 Hz. Les deux derniers sont d'amplitude très faible. Les vallées sous 1.5 KHz sont plus creusées. Le spectre est beaucoup moins plat.

Le spectre 11 est pratiquement identique au spectre 1. Le premier pic est un peu plus haut (500 Hz), la bosse à 1200 Hz invisible. Les pics supérieurs (2200 et 2675 - 2950 Hz) sont faibles et discrets. L'absence de vallées et la platitude du spectre complètent la ressemblance.

Le spectre 7 se distingue des précédents par une énergie supérieure dans les fréquences élevées. On retrouve le pic de basse fréquence à 540 Hz, suivi d'un pic étroit à 840 Hz et deux harmoniques qui semblent émerger d'une profonde vallée (1150 Hz). Le pic suivant, intense, semble scindé en deux. Un maximum discret à 2300 Hz précède un formant intense à 2700 Hz. Cette voyelle, ainsi que la suivante, a été prononcée d'une voix volontairement plus forte que les précédentes. On sait (§ 7.1) qu'une augmentation du paramètre K lié à la raideur de la partie descendante de l'onde glottale, provoque l'élévation du niveau des fréquences élevées relativement à l'amplitude du formant glottal que l'on constate ici. Mais l'amplitude plus grande du pic de 2700 Hz ne peut s'expliquer par ce simple décalage.

Le spectre 3 se distingue des précédents par les creux qui séparent F_0 et le pic de 850 Hz du premier pic, ainsi que le remplissage de la zone fréquentielle entre 1000 et 1500 Hz, interrompu seulement par une harmonique plus basse. Le formant vers 1500 Hz, par sa dissymétrie, pourrait avoir éclaté sous l'effet d'un zéro en deux branches (1530 - 1850 Hz). Il ne reste plus qu'à mentionner les résonances à 2300 et 2650 Hz.

Pour la première fois, la mesure automatique des maxima du spectre 10 détecte une (minuscule) bosse en très basse fréquence (260 Hz). Puis on trouve la succession classique des formants à 580, 900, 1200, 1460 et 1675 Hz. On peut faire à nouveau l'hypothèse de la présence de zéros vers 1550 Hz, 2000 et 2600 Hz.

Le spectre 5 ne montre pas une bosse vers 260 Hz. Entre 200 et 500 Hz, un plateau de 3 harmoniques est précédé d'un faible creux. Le premier pic approche de 600 Hz, et les maxima suivants sont à plus de 1000 et 1300 Hz. La taille de l'harmonique à 1600 Hz donne un spectre presque plat. On note ensuite une petite bosse et un formant à 2350 et 2800 Hz. Tous les formants semblent à une fréquence légèrement plus haute que dans les exemples précédents.

Le spectre 4 ne diffère du précédent que par un détail. Le creux précédant le plateau est sur la troisième harmonique.

On voit nettement sur le spectre 6 une bosse vers 300 Hz. Le pic de 600 Hz est clairement visible. Les deux suivants (900 et 1300) le sont moins. Un zéro profond sépare le pic le plus élevé (2600 Hz) des deux précédents (1950 et 1600 Hz).

L'émission des deux voyelles suivante s'est faite d'une voix intense. Le spectre 8 montre trois harmoniques intenses à 300, 600, et 1000 Hz, alors que les pics à 1400 et 1550 Hz sont à peine visibles. Cette situation est exceptionnelle pour le deuxième. La configuration en haute fréquence est en revanche classique : 1850, 2075 et 2700 Hz.

Le dernier spectre (n° 9) montre à nouveau un maximum élevé à 300 Hz, puis deux larges pics vers 600 et 900 Hz. Le pic de 1500 Hz possède une épaule gauche qui cache sans doute le maximum que l'on trouve d'habitude à cette fréquence. Deux maxima proches (2550 et 2700 Hz) remplacent le pic unique le plus élevé, et il est précédé d'un pic vers 1800 Hz.

ns Le tableau suivant présente les fréquences de tout les maxima observés. Nous avons seulement regroupé dans la même case deux pics de fréquence proche (200 - 300 Hz) si nous avons signalé dans le texte que le formant semblait être scindé. Une étoile seule indique une épaule, et une valeur avec étoile une bosse insignifiante. On peut critiquer l'assignation de quelques valeurs à une case particulière. Ces hésitations concernent surtout les pics P6 et P7. La seule statistique possible en raison du nombre faible et variable de valeurs, et de la dispersion, est le calcul de la médiane.

spectre n°	Pic 1	P2	P3	P4	P5	P6	P7	P8
1		378	810	1205	1517		2272	2593
2		458	795	*	1500	1849		2445
11		494	849		1547		2205	2672
7		540	832	1150	1503		2299	2692
					1673			
3		541	842	1015	1531	1846	2290	2655
				1284				
10	257*	576	905	1216	1460	1988	2291	2598
					1676			2832
5	*	581	1046	1305	1622	*	2350	2780
4		583	1012	1235	1653		2229	2680
6	319	601	918	1331	1586	1945		2605
8	305	601	1006	1394	1565	1851	2073	2731
9	301	611	872	*	1487	1818	2126	2562
								2719
médiane	303	576	872	1235	1538	1850	2272	2663

TABLE I - FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [ɛ]

La comparaison avec les pôles L.P.C. données au paragraphe 4 (520 - 604 ; 1525 - 1505 ; (2600)- 2607) et qui proviennent du même locuteur montre une bonne correspondance avec les pics les plus intenses : P2, P5 et P8.

Il en est de même pour les mesures de LONCHAMP (1978, 1979). Ces mesures étaient entachées en basse fréquence d'un biais positif (+ 40 Hz vers 600 Hz, 20 Hz vers 1000 Hz) dû à une non-linéarité du marquage du spectrographe. Ce biais a été corrigé dans le tableau suivant. De plus, la pondération spectrale et le contrôle de gain automatique (cf. FANT 1960 p. 160) du spectrogramme diminuent l'intensité de la fréquence fondamentale, faisant ainsi apparaître Fo et la deuxième harmonique comme un formant de basse fréquence à 250 Hz. On pourrait peut-être également incriminer la courbe de réponse du

microphone et du magnétophone à 100 Hz. L'intensité de la deuxième harmonique est également accrue par le "formant" glottal placé plus haut que dans les données précédentes en raison d'un choix différent de F_0 (120 Hz). Nous avons supprimé ce pic dans le tableau. On pourra noter, en consultant LONCHAMP (1978, 1979) que leur lisibilité ne cède guère à celle des analyses numériques. La cohérence des deux ensembles de mesures révèle que le spectrographe était, et reste, un bon outil d'analyse, certes moins précis pour les estimations numériques que les spectres F.F.T. La cohérence des résultats indique que les indices spectraux sont stables à long terme pour un sujet. La numérotation des pics est celle du tableau précédent.

	P1	P2	P3	P4	P5	P6	P7	P8

spectres								
7		610	950	*	1550		2110	2650
10		600	930		1430		2040	2460
8		600	900		1500		2100	2520
2		610	890		1540		2110	2630
5		610	920		1510		2120	2620
5'		620	920		1540		2110	2630

médiane		610	920		1525		2110	2625
.....								
médiane table I		576	872		1538		2272	2663
LPC voix faible		520			1525			(2600)
LPC voix forte		604			1505			2607

TABLE II : FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [ɛ̃] - (sujet F.L., d'après Lonchamp 1978, corrigé)

L'examen des spectres montre une très étonnante stabilité des formes qui résulte sans doute de l'enregistrement du corpus en une seule session. On ne découvre aucune trace du pic 1, 4 ou 6. La cohérence des mesures, à 10 ans d'intervalle et par trois outils d'analyse distincts, nous autorise à prendre en compte les mesures du deuxième sujet qui sont rappelées ci-dessous. Le premier pic paraît nettement plus élevé chez ce sujet, alors que P5 (sans doute F'2) est plus bas de 80 Hz.

	P1	P2	P3	P4	P5	P6	P7	P8
spectres								
13		700	1000	1250*	1440	1800	2040	2580
12		770		1250	1450		1980	2720
8		770		1220*	1440	1640	1920	2650
							2150	2880
8'		720		*	1420		1900	2640*
15		720	920	1220	1490	*	1920	2640
2		720		*	1420		1890	2580
médiane		720	(960)	(1235)	1440	(1720)	1920	2640
sujet F.L (table II)		610	920		1525		2110	* 2625

TABLE III - FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [ɛ̃] (Sujet JPZ, d'après LONCHAMP 1978, corrigé)

10.2.2 INTERPRETATION

L'identification des indices spectraux repose sur les constatations suivantes. La première fréquence de résonance propre du conduit nasal n'est pas connue avec précision, mais sa fréquence est sans doute voisine de 550 Hz. Le premier formant de la voyelle orale de même articulation doit également se trouver, d'après les informations déduites de l'étude articulatoire, vers 550 Hz. La difficulté est donc que FN1 peut être juste au-dessus de F1, ce qui conduirait à interpréter le premier pic comme F'1, ou juste au-dessous, formant alors le premier pic. Dans tous les cas, ces deux formants seront légèrement décalés vers des fréquences plus élevées. Si l'on considère les mesures du sujet F.L., on constate que 2 voyelles ont un premier pic à des fréquences inférieures à 500 Hz (380, 460 Hz) et que la variabilité est forte.

Il est peu probable que les fréquences les plus basses, ainsi que la dispersion, soient dues à une variation de F'1, qui dépend de la forme de la cavité orale. Mais elles s'expliquent aisément si ce premier pic est le formant nasal, dont la fréquence dépend à la fois du couplage, et de l'état physiologique des fosses nasales du sujet.

Cette remarque forme "le noyau dur" de notre interprétation. La forme du conduit nasal de

FANT suppose une première fréquence de résonance nasale supérieure à 670 Hz, en raison du couplage. Si notre interprétation est correcte, notre conduit nasal doit être plus long et / ou de plus forte constriction narinaire, car sa fréquence de résonance est inférieure à 600 - 620 Hz, le conduit étant couplé. La fréquence la plus basse que nous avons mesurée (380 Hz) est proche de la fréquence de résonance du conduit nasal de FANT (sans sinus) à narines bouchées (340 Hz). C'est en grande partie la présence des pertes par vibration des parois qui interdit à FN1 de descendre à une fréquence nulle, même en cas d'occlusion complète.

Si le premier maximum (P2) est FN1, le deuxième (P3) est nécessairement le formant oral décalé (F'1) qui atteint presque 900 Hz. Le 4^{ème} pic (P5), juste au-dessus de 1500 Hz, correspond parfaitement à la fréquence attendue pour le deuxième formant oral (F'2) : sa fréquence plus basse que celle d'un [ε] est due au recul de la langue que nous avons noté dans l'examen des clichés radiologiques. La présence du premier zéro est révélée par la forme très variable du spectre entre 1000 et 1300 Hz. Selon sa fréquence exacte, qui dépend à la fois du degré de couplage et de l'aire variable du conduit nasal, fonction de l'état physiologique de la muqueuse nasale, il creusera une vallée plus ou moins profonde. Cette vallée peut mettre en évidence un pic supplémentaire (P4) absent de plusieurs spectres, dans le cas où une irrégularité de la source glottale fait apparaître une ou plusieurs dépressions dans la même zone. On peut également songer à la présence d'une ou plusieurs résonances liées aux sinus frontaux. Ce pic n'est jamais visible sur les spectrogrammes (cf. tableau II). La bande passante du filtre utilisé en spectrographie en bande étroite étant de 45 Hz, le temps d'intégration est d'environ 22 ms, soit deux périodes environ. Cette intégration tend à lisser légèrement le spectre, effaçant en particulier les irrégularités spectrales aléatoires dues à la source glottale. Si le pic P4 dépend de la forme particulière de la source pendant une période, on comprend pourquoi ce pic n'est pas visible sur les sections spectrographiques. Cette remarque ne clôt cependant pas le dossier car le sujet JPZ présente également un pic P4 dans quelques sections spectrographiques.

La seconde résonance du conduit nasal doit se trouver vers 2000 - 2200 Hz. On observe en effet un pic discret (P7) vers 2275 Hz, mais il en existe également un légèrement plus bas vers 1850 Hz (P6). Il paraît difficile de trancher en raison du faible niveau et de la variabilité. Les spectrogrammes, en revanche, ne montrent que le pic P7. C'est donc P7 que nous identifierons avec prudence à FN2. La fréquence de P6 est voisine de la fréquence du pôle que l'on obtient lorsque le conduit nasal n'est pas symétrique. Enfin, P8, stable et intense, est certainement F'3. Mais il existe souvent un pic plus faible quelques centaines de Hz plus haut, dont l'origine est vraisemblablement la troisième résonance nasale.

Reste à identifier l'origine du premier pic (P1). Il ne semble présent que lorsque FN1 est à fréquence élevée et quand le sujet utilise une voix intense. Il paraît impossible de trancher

entre deux hypothèses, qui d'ailleurs ne s'excluent pas. On peut songer au premier rebond du spectre qui suit le formant glottal. Il ne nous semble pas possible de l'identifier avec le formant glottal lui-même car sa fréquence moyenne de 300 Hz supposerait une phase montante de l'onde glottale inférieure à 20% de la durée de la période, ce qui est beaucoup trop faible. La deuxième hypothèse est que P1 est la trace d'une résonance sinusale. Sa présence dans les seuls cas où FN1 est élevé s'expliquerait par le fait que l'ouverture de l'ostium suppose une muqueuse nasale non tumescence, tout comme le large passage qu'implique un formant nasal élevé. Les analyses spectrographiques ne sont d'aucun secours en raison des distorsions en basse fréquence mentionnées plus haut.

les pics mesurés sur les réalisations du sujet JPZ peuvent s'interpréter de la même manière. On remarque que F'2 est à une fréquence plus basse, ce qui implique que le recul de la langue est plus marqué que pour le sujet précédent. Ce fait n'est pourtant pas très apparent sur les clichés radiologiques (cf. § 4). On peut hésiter à identifier le deuxième formant nasal : la résonance de 1920 Hz paraît de fréquence trop faible. Elle implique un fort volume de la cavité nasale, ce qui semble en contradiction avec un formant nasal placé plus haut que pour le premier sujet, qui suppose, quant à lui, une cavité plus courte ou plus ouverte. La fréquence de FN1 ne peut, en effet, être inférieure à 720 Hz après couplage. La fréquence du pic P7 (2640 Hz) paraît mieux convenir, le troisième formant étant alors rejeté vers 3300 Hz. Comme pour le pic P6 (1850 Hz) du sujet précédent, les pics discrets vers 1900 Hz n'ont pas une affiliation claire. Quant au pic P4, F'1 ou FN1 selon que FN1 correspond ou non à P2, il est très peu proéminent. Il tend à se confondre dans plusieurs cas avec F'2.

Il est important de constater que nous n'avons aucune assurance que le pic P2, de 100 Hz plus élevé que la fréquence du même formant pour le premier sujet, soit bien le formant nasal. Il pourrait s'agir de F'1, P3 étant alors le premier formant nasal. L'absence de variabilité, conséquence de l'enregistrement des voyelles au même moment, interdit toute conclusion. Il apparaît cependant que l'impossibilité de préciser l'affiliation du premier pic proéminent est sans conséquence pratique.

Nous allons montrer par simulation que la forme du spectre ne dépend pas de la position relative du premier formant nasal et du formant oral. En d'autres termes, la forme spectrale de la voyelle nasale reste identique, que le premier pic soit FN1 ou F'1.

Nous avons d'abord cherché par tâtonnements une fonction d'aire orale ayant des fréquences formantiques légèrement plus basses que la voyelle nasale. Celles du conduit retenu (O4) sont à 575, 1480 et 2560 Hz. Ces fréquences tiennent compte de la constriction vélaire. La figure 1.79 présente le spectre après couplage du conduit nasal de FANT (N1; sans sinus) avec un passage rhinopharyngal de 2.4 cm². Comme la fréquence propre de ce conduit nasal (670 Hz) est supérieure à F1, le premier pic est F'1, à 605 Hz,

et le deuxième est FN1, à 1100 Hz. Le couplage élève F'1 de 30 Hz, et FN1 de plus de 400 Hz. Ils sont séparés par un zéro discret à 1050 Hz, qui affaiblit FN1. On observe ensuite F'2 (1500 Hz), FN2 (2200 Hz), puis F'3 (3125 Hz), séparé de FN2 par un ample zéro à 2800 Hz, et enfin, vraisemblablement, FN3.

La correspondance avec les mesures des voyelles naturelles est bonne. Le premier et le troisième pic, à 605 et 1500 Hz, sont aux fréquences attendues : 576 et 1538 Hz. En revanche, FN1, à 1100 Hz, est plus haut que le deuxième pic mesuré (900 Hz). L'amplitude de FN2 est trop forte, et F'3 est décalé par un zéro trop profond qui indique sans doute une aire de couplage trop grande.

Nous avons ensuite modifié arbitrairement le conduit nasal pour forcer FN1 à descendre sous F1 (conduit N3). Rappelons que la théorie du couplage nasal (§ 5.3) indique que l'ordre des formants demeure, après couplage, celui des fréquences de résonance avant couplage. Ce second conduit est plus long (12.5 cm) que le précédent, et possède une constriction dans la zone des narines. Les formants de ce conduit nasal sont à 504, 1740 et 3455 Hz. Le spectre de la voyelle couplée à ce conduit est présenté à la figure 1.80. Le premier pic, qui est FN1 dans ce cas, s'élève par couplage jusqu'à 550 Hz (+ 45 Hz). F'1, dont l'amplitude est réduite par le premier zéro à 950 Hz, se décale jusqu'à 1020 Hz (+ 465 Hz). On note que l'effet du couplage est beaucoup plus important pour le formant placé le plus élevé, que ce soit FN1 ou F1. La théorie du couplage prévoit que la position du premier pic est toujours comprise entre sa fréquence propre et celle de la première résonance propre de l'autre cavité. Lorsque la première fréquence propre des deux conduits est voisine, la fréquence du premier pic spectral sera peu modifiée par un échange des affiliations. En revanche, si l'aire du conduit nasal diminue, sa fréquence propre s'abaissera, et le premier pic spectral avec lui. Il est peu vraisemblable que le degré d'aperture du conduit vocal soit aussi sujet à de fortes variations. Il est donc probable que le premier pic proéminent soit le formant nasal.

La fréquence de F'2 n'est que faiblement modifiée (1495 Hz). La similitude des formes spectrales de ces deux simulations est frappante. Rien ne permet de dire notamment que FN1 et F'1 ont échangé leur position. La stabilité spectrale résulte du fait que les pics proéminents sont à des fréquences très voisines quelle que soit l'affiliation du premier.

Les deux simulations suivantes font appel à la source glottale décrite par FANT (cf. § 7.1). Le conduit nasal utilisé est le conduit "long" (N3). La valeur du paramètre K, contrôlant la pente de la partie décroissante de l'onde glottale, est fixée à 0.7 pour la simulation présentée à la figure 1.81a et à 1.2 pour celle de la figure 1.81b. Pour la première, on note, sous le formant glottal intense de très basse fréquence, un premier rebond du spectre à 350 Hz, absent de la deuxième simulation. Il est instructif de comparer ces spectres à celui obtenu en ajoutant deux cavités sinusales (figure 1.97 : sinus S2). Le pic sinusal apparaît à

290 Hz, et sa fréquence, comme son amplitude, ne sont pas très différents des caractéristiques du premier rebond du spectre de la simulation de la figure 1.8 la. On peut bien sûr combiner les effets de la source et de la résonance sinusale (figure 1.82 : sinus S2 ; $K = 0.7$). On obtient une petite irrégularité visible à 310 Hz, dominée par le formant glottal et le formant suivant.

Les deux dernières simulations font appel au conduit nasal utilisé dans la simulation de la nasale vélaire. Il convient en effet de vérifier qu'un même conduit nasal permet de réaliser une simulation correcte des voyelles et des consonnes nasales. Pour un couplage de 2.4 cm, et le sinus S2, les trois premiers pics se placent à 630, 1120 et 1500 Hz. L'amplitude du second pic est faible, et sa fréquence toujours un peu élevée si on la compare aux mesures physiques. En revanche, la positions des deux pics majeurs est très voisine de celle des voyelles naturelles. Un profond zéro rejette F'3 à une fréquence élevée, ce qui pourrait indiquer un couplage trop important. Pour vérifier cette hypothèse, nous avons réduit le couplage en diminuant l'aire des deux dernières sections du conduit nasal N2 à 2.5 et 1.1 cm². Nous avons également modifié la fréquence de résonance du sinus (S3). Les résultats (figure 83) montrent que le couplage plus faible réduit bien l'amplitude du zéro à 2400 Hz. Le troisième formant s'abaisse, et sa fréquence (2750) s'approche des fréquences mesurées. L'emploi d'un sinus dont le col est plus court et le diamètre plus grand augmente la fréquence du pic associé (400 Hz). En conséquence, le premier formant est plus élevé (680 Hz) et s'approche des valeurs relevées pour le sujet J.P.Z. Il apparaît que les différences entre les réalisations des deux sujets ne dépendent pas forcément d'une articulation ou d'une forme de conduit nasal très différente. Une taille dissemblable des sinus pourrait fournir une explication satisfaisante. De même, la variabilité de la fréquence de FN1 pour le premier sujet pourrait refléter un changement dans le degré de couplage des sinus, plutôt qu'une variation de l'aire de la zone narinaire. L'analyse de la consonne nasale vélaire nous avait déjà conduit à proposer la même explication. L'amplitude du deuxième pic reste faible, mais sa fréquence (1060 Hz) se rapproche des valeurs mesurées. Globalement, les spectres de simulation sont très proches des spectres physiques, et il est probable qu'aucun indice essentiel de la nasalité n'a été omis.

Il convient de mentionner un dernier aspect de nos mesures et simulations, dont l'importance sera soulignée plus loin. La présence de trois formants entre 500 et 1500 Hz modifie l'amplitude relative des formants par rapport à ceux d'une voyelle orale qui posséderait deux formants à 500 et 1500 Hz.

Il est intéressant de comparer les résultats de nos mesures à ceux des études de BOGNAR & FUJISAKI (1986) et de MRAYATI (1975). Le tableau suivant permet la comparaison avec les résultats de BOGNAR & FUJISAKI. On remarquera que Z1 et Z2 notent le premier et deuxième zéro (antiformant).

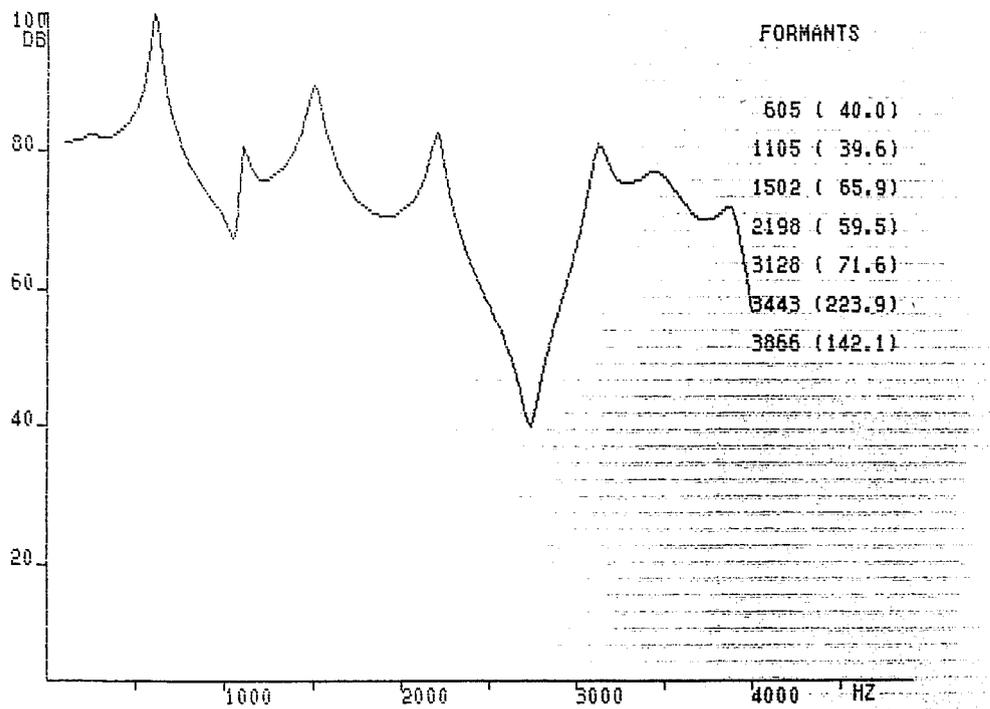


Figure 1.79 - Simulation du spectre de la voyelle [ẽ] : conduit de BJUGGREN & FANT.

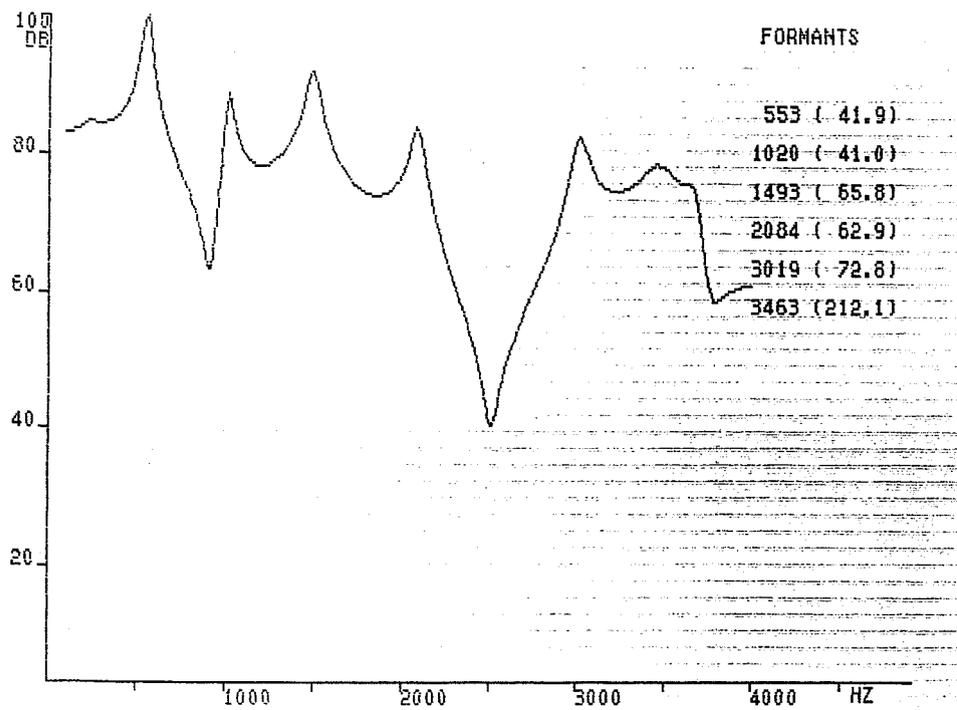


Figure 1.80 - Voyelle [ẽ] : conduit nasal 'long'.

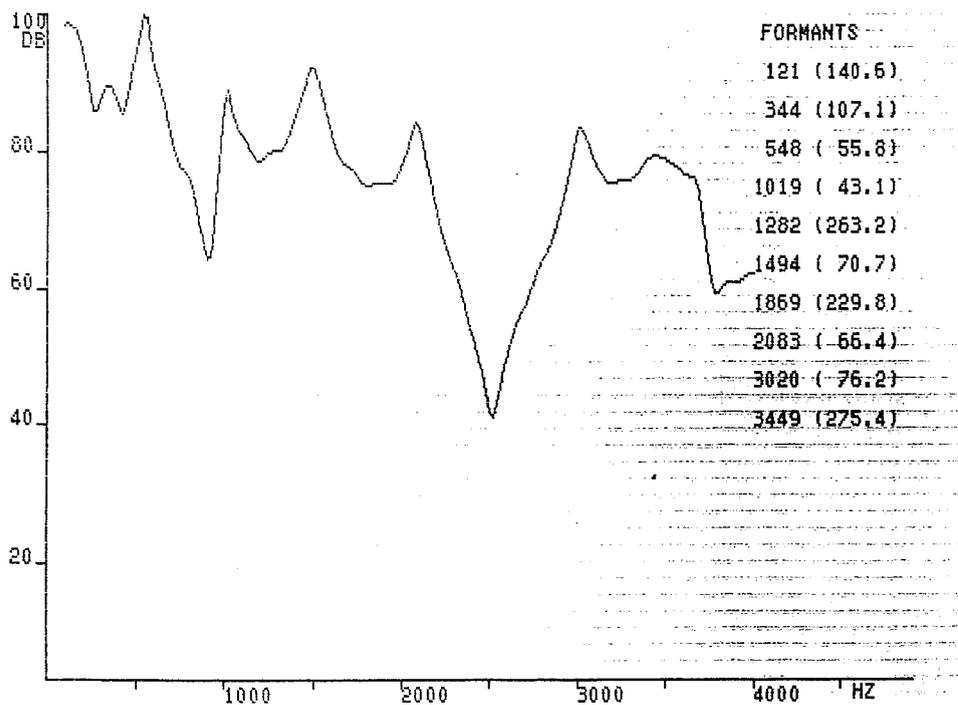


Figure 1.81 A - voyelle [ɛ̃] : Source glottale de FANT (1983)
 $F_g = 125$ Hz, $K = 0.7$.

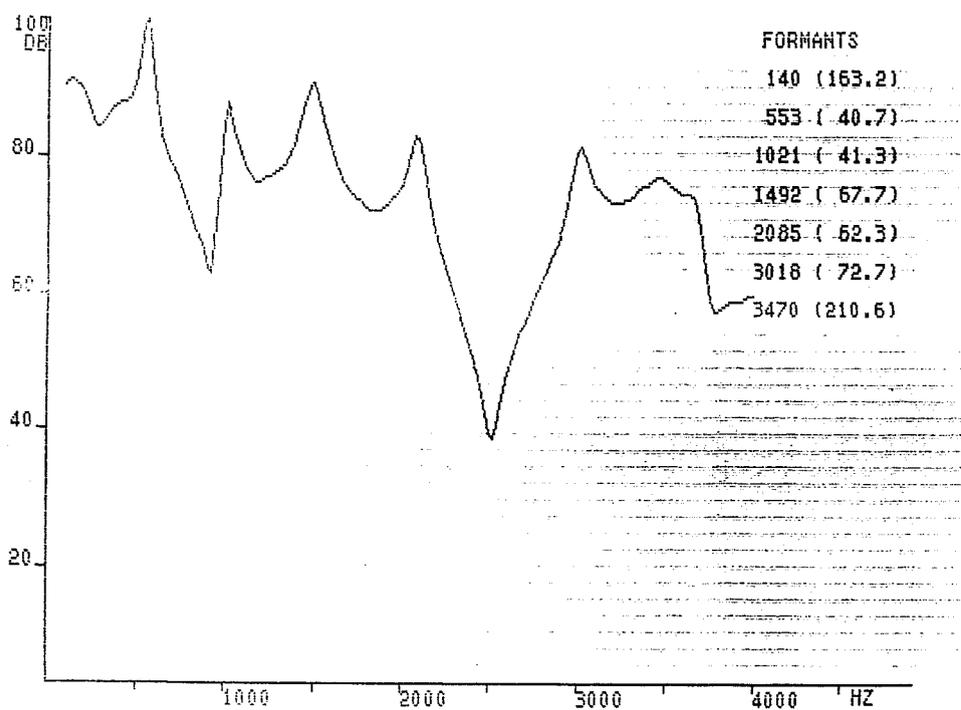


Figure 1.81 B - $K = 1.2$.

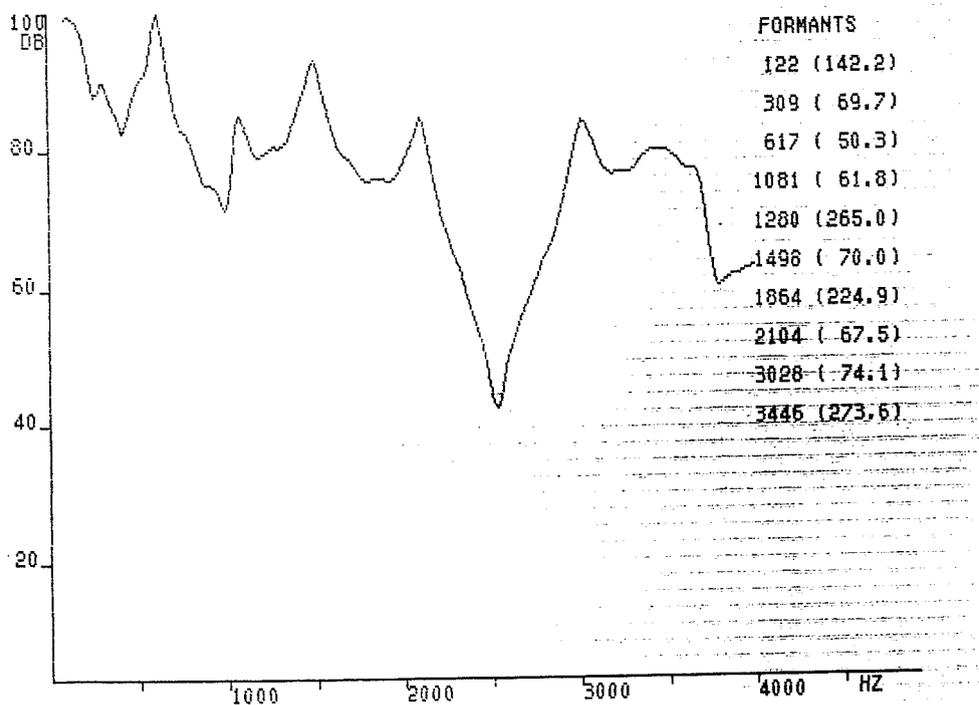


Figure 1.82 - Voyelle [ẽ] : Source vocale $F_g = 125$ Hz;
 $K = 0.7$; Sinus S2.

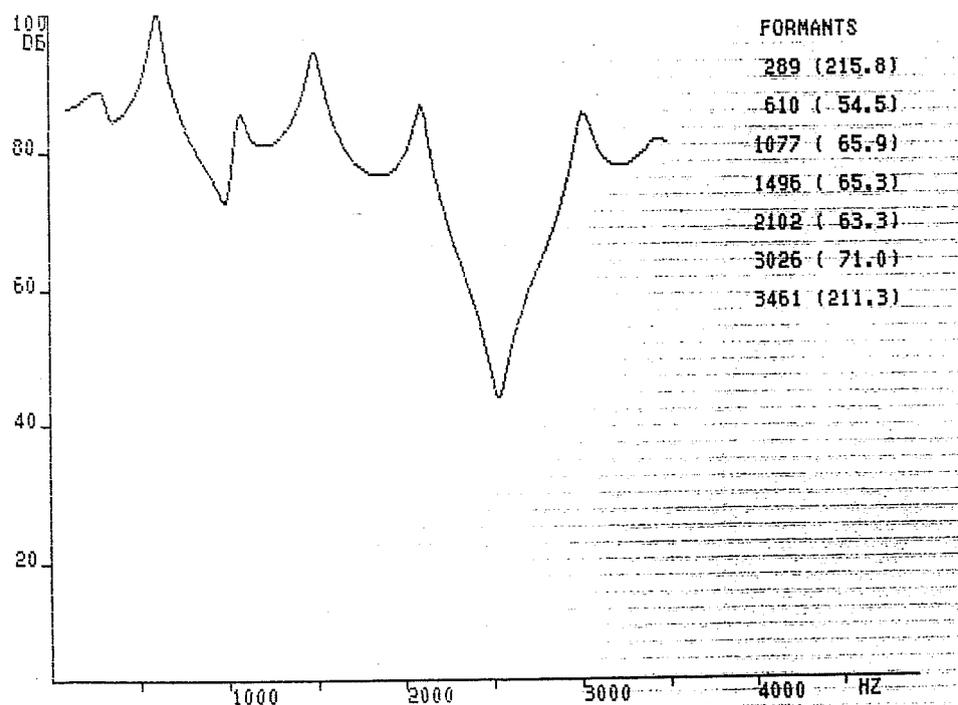


Figure 1.83 - Voyelle [ẽ] : Conduit nasal de la consonne nasale vélaire.

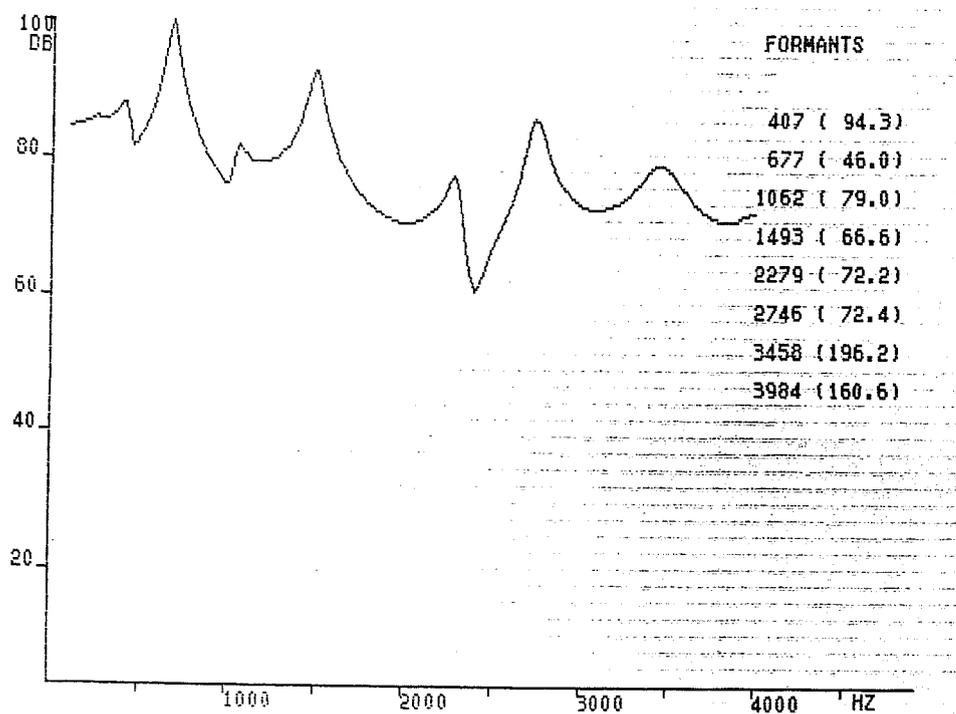


Figure 1.97 - Voyelle [ẽ] : même conduit qu'à la figure 1.83, mais couplage réduit (1.1 cm²)

ẽ	P1	Z1	F'1	F'2	P2	Z2	F'3
	270	350	690	1680	2180	2260	2760
table I	303		576	872	1538	2272	2663
table II			610	920	1525	2110	2625
table III			720	960*	1440	2640?	3300

La formant noté ici F'1 est vraisemblablement pour nous FN1(table I et II). Sa fréquence est proche des valeurs mesurées pour le sujet J.P.Z. (table III). Il est probable que P1 corresponde soit à un rebond du spectre, soit à une résonance sinusale. Le peu d'importance du pic proche de 1000 Hz est souligné par le fait que la technique d'analyse-par-synthèse ne le détecte pas. Le fréquence de F'2 est assez variable, ce qui pourrait indiquer une variabilité de la position de la langue sur l'axe antérieur - postérieur. l'accord est globalement très satisfaisant. On ne peut en dire autant des résultats de MRAYATI, obtenus par analyse L.P.C. Nous rappelons les moyennes de 5 locuteurs masculins dans trois types de contexte consonantique (occlusives sonores, occlusives sourdes, consonnes nasales), les résultats d'un sujet (R.C.), et ceux de nos mesures pour les deux sujets.

ẽ	FN1	F'1	F'2	F'3		
occ. son.	409	735	1663	3155		
occ. sd.	430	718	1649	3119	5 locuteurs masculins	
nasales	413	754	1697	3004		
occ. son.	350	786	1748	3666		
occ. sd.	362	882	1790	3612	sujet R.C.	
nasales	352	1022	1903	3252		
table I	303	576	872	1538	2272	2663
table II		610	920	1525	2110	2625
table III		720	960*	1440	2640?	3300?

Si les écarts-type des mesures des 5 sujets sont énormes dans de nombreux cas, c'est moins vrai des mesures du sujet R.C. L'écart-type de F'2 après nasale atteint pourtant 240 Hz pour ce sujet, et celui de F'1 233 Hz. Il est donc peu probable que les variations individuelles puissent expliquer totalement les différences. Très vraisemblablement, il faut incriminer l'analyse L.P.C., et particulièrement un choix d'un nombre de coefficients de

prédiction trop faible. Nous avons déjà illustré (figure 1.54) la variabilité des résultats de l'analyse L.P.C. pour des voyelles nasales dont le spectre F.F.T. est peu différent.

10.2.3 - CONCLUSIONS

Cette analyse de la voyelle [ẽ] nous a permis de dégager plusieurs conclusions intéressantes. Tout d'abord, la variabilité du premier formant, que l'on constate en comparant les enregistrements faits à différentes occasions, ne s'explique que si ce formant est le premier formant nasal. Sa fréquence dépend quelque peu du couplage, mais surtout de l'état physiologique de la muqueuse nasale. Un gonflement de celle-ci réduit l'aire de la zone narinare, abaissant la fréquence de FN1. Il est également probable que le degré de couplage des cavités sinusales est hautement sensible à un gonflement de la muqueuse. La présence ou l'absence d'une résonance sinusale retentit de façon significative sur la position de la première résonance. Ces facteurs expliquent sans doute l'importance des écarts mesurés pour la fréquence du premier pic proéminent.

Les simulations montrent également que la forme du spectre ne dépend pas de l'ordre des formants. Que la fréquence propre du premier formant nasal soit supérieure ou inférieure à la fréquence du premier formant du conduit nasal, le spectre de la voyelle [ẽ] montrera un pic proéminent entre 500 et 700 Hz, un pic discret vers 1000 Hz, puis F'2 entre 1450 et 1600 Hz. L'étendue des plages de variation semble supérieure à celles des voyelles orales, car, outre le premier formant, F'2 semble posséder une grande latitude de réalisation, dont l'origine est articulatoire. Le deuxième pic, vers 1000 Hz, peut à l'occasion se confondre avec F'2. En haute fréquence, outre une irrégularité spectrale vers 1800 - 2000 Hz, qui pourrait correspondre à une paire pôle - zéro liée à une dissymétrie des fosses nasales, on trouve dans l'ordre FN2, entre 2000 et 2200 Hz, F'3, décalé vers le haut par le deuxième zéro de couplage (Z2), au dessus de 2500 Hz, puis F'3 vers 3000 Hz. Quant au pic de très basse fréquence (300 Hz), il est probable qu'il représente soit le premier rebond spectral suivant le formant glottal, soit une résonance sinusale, soit les deux à la fois.

10.3 - LA VOYELLE [ã]

10.3.1. DESCRIPTION DU CORPUS

Les voyelles [ã] sont présentées à la figure 1.84. Le spectre 2 montre 3 pics sous 1 kHz. Le premier (473 Hz) est à l'extrémité d'une rampe montante qui suit la fondamentale. Il n'y a aucune trace de formant glottal. Le second et le troisième pic sont à 160 Hz de distance seulement (820 et 980 Hz). Le niveau chute rapidement de plus de 20 dB vers 1 kHz. Un

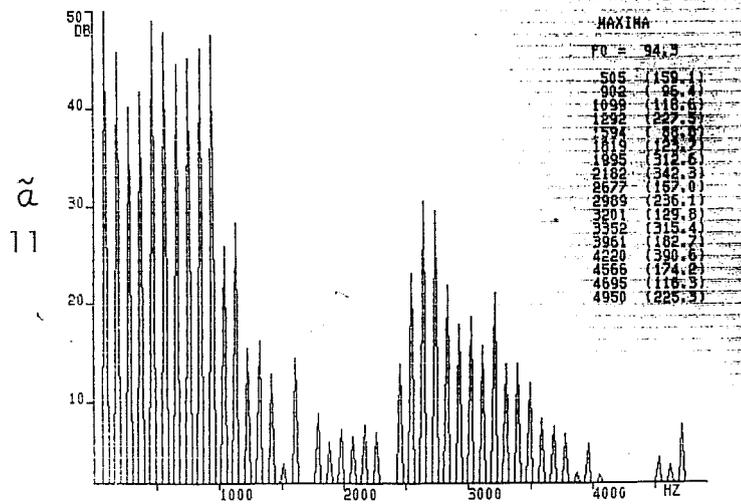
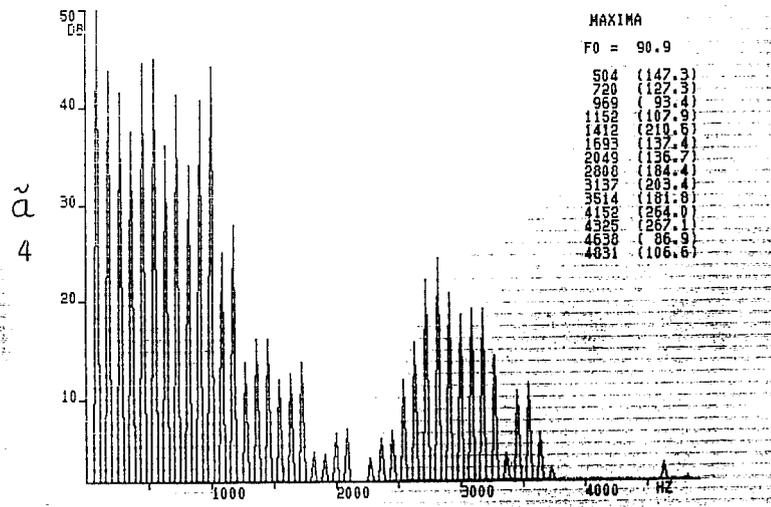
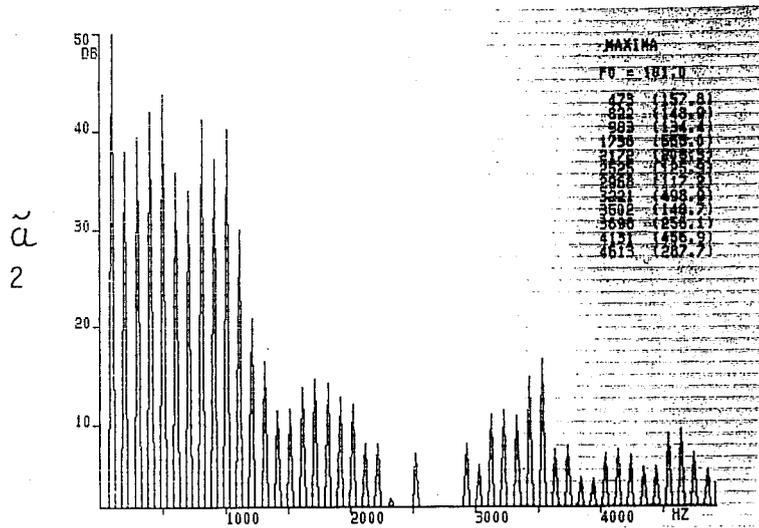


Figure 1.84 Spectres des voyelles [ã].

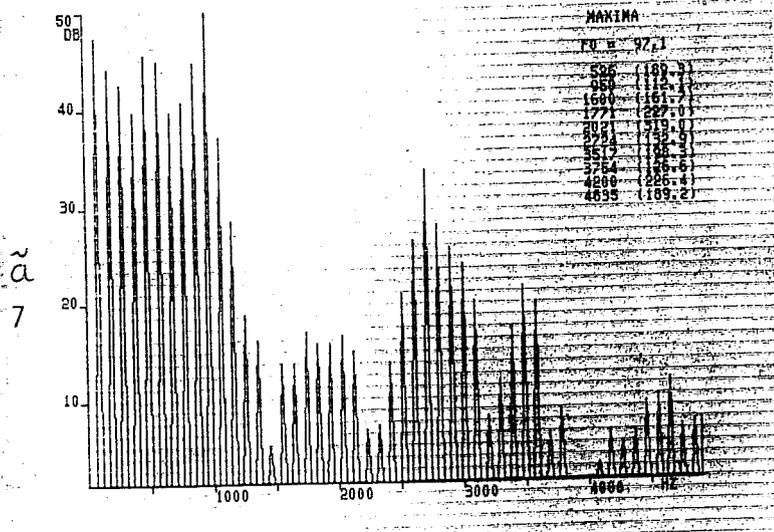
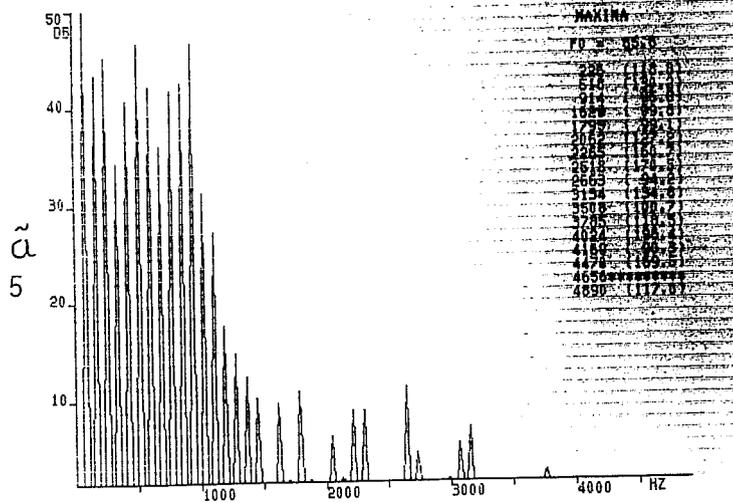
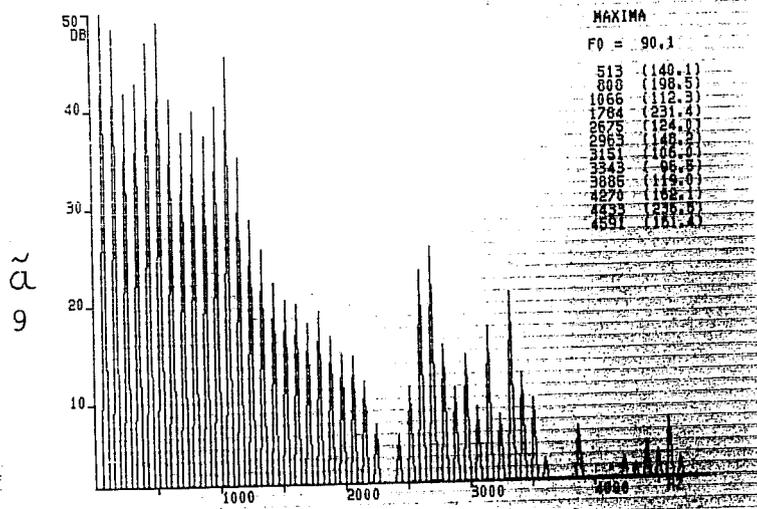


Figure 1.84 (Suite).

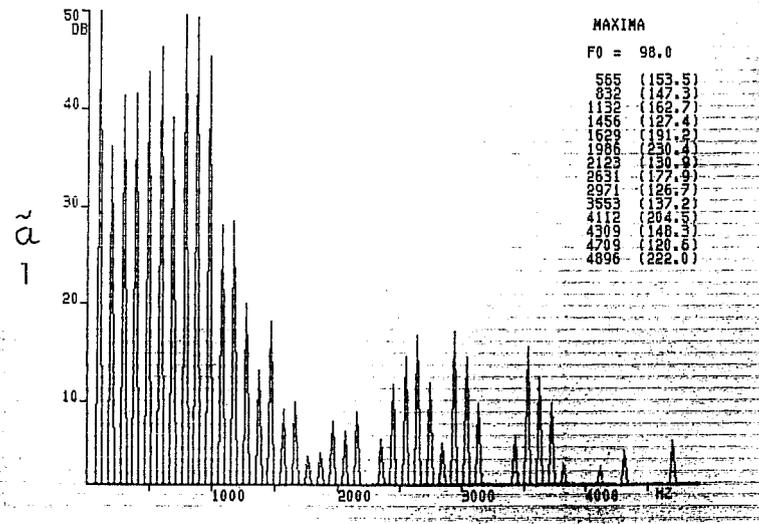
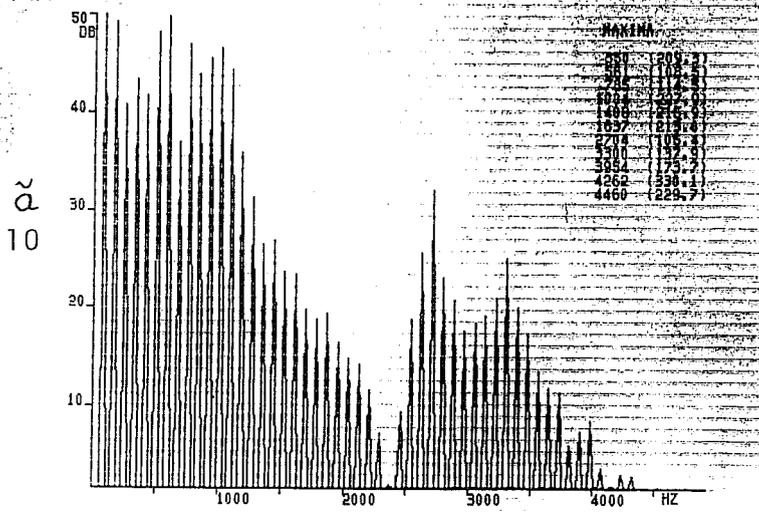
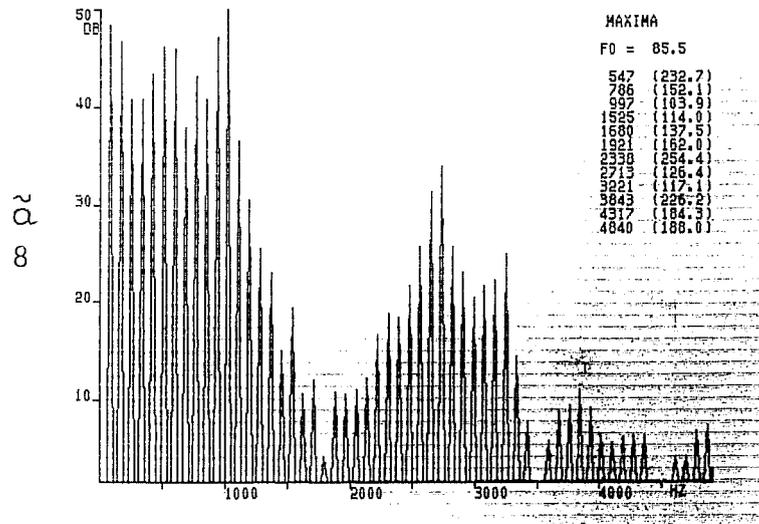


Figure 1.84 - Spectres des voyelles [ã].

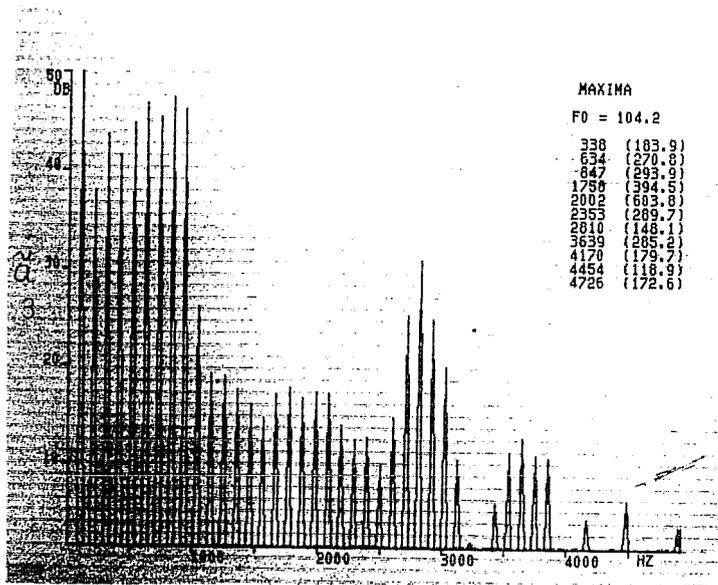
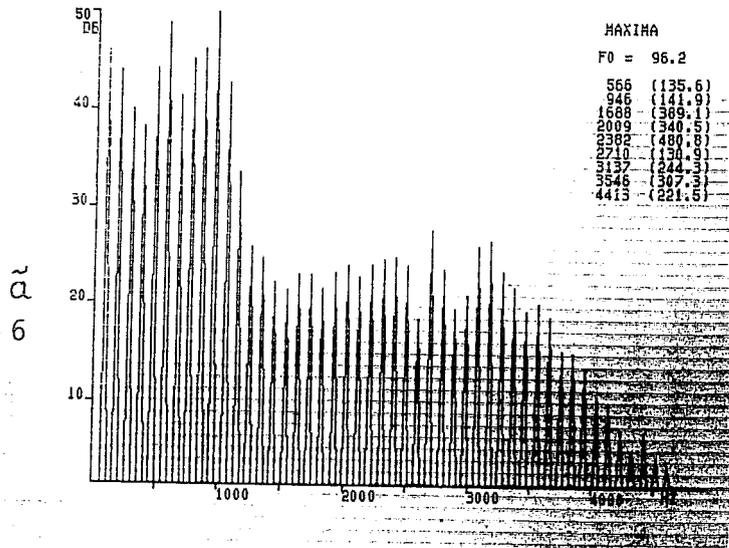


Figure 1.84 (Suite).

large dôme s'étale vers 1700 Hz. Le seul pic caractéristique de haute fréquence est à 3500 Hz. Le spectre suivant (n°4) possède également trois pics de basse fréquence. Le premier est légèrement plus haut, et la séparation du second et du troisième est plus importante (250 Hz). Les maxima suivants, en marches d'escalier, sont peu caractéristiques et précèdent une zone plus intense contenant deux maxima (2800 - 3150 Hz). Le deuxième pic disparaît pour la voyelle n° 11, remplacé par une rampe ascendante vers le troisième. Un pic intense de haute fréquence à 2675 Hz domine celui de 3200 Hz et plusieurs maxima discrets aux fréquences moyennes. La caractéristique du spectre 9 est la présence d'un zéro vers 2200 Hz. Ses flancs sont aussi nets que dans les simulations. A nouveau, le deuxième pic est presque invisible à 800 Hz, encadré de deux formants à 510 et 1050 Hz. Ce dernier est 100 Hz au dessus de la fréquence mesurée sur les spectres précédents. En haute fréquence, le formant de 2675 Hz domine celui de 3150 Hz. Pour la première fois, un pic discret se dégage en basse fréquence vers 250 Hz. Seuls deux maxima sont visibles plus haut. Au-dessus de 1000 Hz, il est impossible de préciser la structure formantique tant le niveau est faible. Le spectre n° 7 ne possède que deux formants sous 1000 Hz, à 525 et 950 Hz. Mais les formants supérieurs sont bien marqués, vers 2700 et 3500 Hz. Ils dominent un plateau aux flancs escarpés de 1500 à 2000 Hz. On retrouve 3 pics pour la voyelle suivante (n° 8), à 550, 780 et 1000 Hz. Le deuxième n'est qu'une irrégularité spectrale. Un zéro s'enfonce profondément à 1800 Hz, précédant un premier formant intense à 2700 Hz, et un deuxième plus dissymétrique à 3200 Hz, sous doute suivi d'un deuxième zéro. On compte 4 pics de basse fréquence pour le spectre 10. Si le premier est juste visible (350 Hz), le deuxième et le quatrième sont bien caractérisés. Le troisième est mis en évidence par le brusque affaiblissement de la 8^{ème} harmonique. La forme typique d'un ample zéro occupe le spectre entre 1300 et 2500 Hz. Les formants supérieurs, à 2700 et 3300 Hz, sont particulièrement faciles à identifier. Les 10 premières harmoniques de la voyelle n° 1, à l'exception de la fréquence fondamentale et de la 7^{ème}, forment une rampe montante. On hésite à appeler pics les maxima placés à 565 et de 832 Hz. En haute fréquence, 3 pics se partagent la zone comprise entre 2500 et 3500 Hz. Le spectre 6 est remarquable par sa platitude et son niveau élevé entre 1300 et 3200 Hz. Seules deux faibles irrégularités à 2700 et 3150 Hz correspondent aux pics majestueux de la voyelle 10. En basse fréquence, deux pics à 560 et 950 Hz émergent clairement. Le spectre 3 possède un formant de basse fréquence (338 Hz), sous un deuxième de fréquence particulièrement élevée (630 Hz) au même niveau que le troisième à 847 Hz. Un plateau occupe les fréquences moyennes, entre 1600 et 2200 Hz, alors qu'un pic intense (2800 Hz) est séparé du suivant à 3640 Hz par un zéro.

Le tableau IV rassemble les fréquences de tous les maxima visibles. Devant le nombre variable de pics de basse fréquence, nous avons été amené à regrouper dans une catégorie

P'3 les pics de fréquence proches de 900 Hz des voyelles ne possédant qu'un seul autre formant vers 500 Hz (P2). Lorsque 3 pics sont visibles, ils sont classés de P2 à P4. Les classes P5 à P7 rassemblent un assortiment représentatif de pics de faible amplitude. Dans quelques cas, l'attribution à l'une ou l'autre classe est arbitraire.

spectre n°	P1	P2	P3 (P'3)	P4	P5	P6	P7	P8
2		473	822	983		1736		2525*
4		504	720	969	1152 1412	1693	2049	2808
11		505	902		1292	1594	1819	2677
9		513	808	1066		1784		2675
5	226	518	914			1795		2663
7		526	950			1771	2021	2724
8		547	786	997	1525		2338	2713
10	350	561	785	1004		1837		2704
1		565	832		1456			2631
6		566	946					2710
3	338	634	847			1758		2810
médiane	(300)*	526	786*	997*	*	1766	*	2690
			908*					

TABLE IV - FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [ā]

Comme pour la voyelle précédente, les fréquences formantiques mesurées dans notre thèse de 3ème cycle sont proches de celles du tableau IV.

	P1	P2	P3	P4	P5	P6
spectres						
9		610	890	1860*	2130	3000
6		600	880	*	2160	3050
1		600	900	*	2100	3020
7		610	900	*	2160	2930
11		620	900	*	1920*	3000
12		600	870	1750*	2140	2820
						3050
médiane		605	895	(1800)*	2140	3000
médiane	(300)*	526	786 - 997	1766	*	2690
table IV			908			

TABLE V : FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [ā] - (sujet F.L., d'après Lonchamp 1978, corrigé)

A nouveau le pic P2 est légèrement plus haut dans les mesures spectrographiques. Le pic P3 est toujours unique, et sa fréquence est identique à celle du pic P'3 du tableau précédent. Si P5 est presque toujours visible, P4 n'est en fait qu'une petite irrégularité spectrale à la fréquence mal définie. Quant à P6, il est 300 Hz plus haut. Les mesures du second sujet sont rassemblées dans le tableau suivant.

	P1	P2	P3	P4	P5	P6	P7
spectres							
10		720	950	1050*	1920		3350
17		740	940	1160	1960	3000	3600*
2		750	900*	*	1800		3330
16		740	930		1880	2840	3340
					2400*		
14		720	940	1180	1920	2850	3360
11		750	910	1180	2000	3000	3450
médiane		740	935	1170	1920	2900	3360
médiane	(300)*	526	786 - 997	1766 *		2690	
table IV			908				
table V		605	895		2140	3000	

TABLE VI : FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [ã] - (sujet J.P.Z., d'après Lonchamp 1978, corrigé)

10.3.2. INTERPRETATION

Comme pour la voyelle précédente, la fréquence du premier pic proéminent (P2) n'est pas stable. Elle s'inscrit entre 475 et 635 Hz. La médiane (526 Hz) est un peu plus basse que la médiane du premier formant nasal de la voyelle [ẽ] (576 Hz). Le pic correspondant des spectrogrammes est à 605 Hz, valeur identique à la médiane du premier formant de la voyelle [ẽ]. Le comportement identique de ce formant pour les deux voyelles, ainsi que la proximité des fréquences mesurées, nous incitent à identifier P2 comme le premier formant nasal (FN1). F'1 est rejeté vers les fréquences hautes. Il se confond dans la moitié des réalisations avec F'2. Lorsque F'1 et F'2 sont confondus, le pic proéminent est à une fréquence moyenne de 900 Hz, ce qui implique une position très reculée de la langue et une forte constriction pharyngale, que confirment les clichés radiologiques. Lorsque les deux pics sont distincts, ils sont séparés de 200 à 250 Hz environ. Il est probable que la fréquence exacte du premier zéro, vers 900 Hz, joue un rôle dans la présence ou l'absence de F'1. Si la fréquence du zéro coïncide avec F'1, celui-ci ne sera pas visible. C'est sans

doute le cas du spectre n° 7. La fréquence de F'1 est très voisine de celle de la voyelle [ɛ̃]. La différence essentielle entre [ã] et [ɛ̃] est la fréquence de F'2.

Il faut remarquer que la proximité de F'1 et F'2 renforce l'amplitude de ces deux composantes, qu'elles soient confondues ou non. Le formant nasal est souvent moins intense. Dans une voyelle orale dont les deux premiers formants sont à 500 et 1000 Hz, le premier formant domine le second d'environ 6 dB. Dans le cas de la voyelle nasale [ã], qui possède deux pics proéminents à ces fréquences, le second "formant" est souvent plus intense que le premier. Plus précisément, le groupe F'1 - F'2 domine souvent FN1. Voilà, selon nous, l'explication de l'affaiblissement du premier formant, mentionné comme indice de nasalité par DELATTRE (1954) et de nombreux chercheurs. On voit qu'il s'agit en fait, non de l'affaiblissement de F'1, mais de l'amplitude plus faible de FN1 relativement au bloc F'1 - F'2. Cette démonstration reste d'ailleurs vraie si le premier pic est F'1. La prééminence du second pic sera assurée, dans ce cas, par la proximité de FN1 et F'2. Mais la présence du premier zéro de couplage peut, lorsque sa fréquence coïncide avec F'2, réduire l'amplitude de ce dernier. Le mécanisme décrit paraît plus efficace pour expliquer la réduction apparente de l'intensité du premier pic qu'un accroissement de la largeur de bande dû aux pertes acoustiques importantes dans le conduit nasal.

Comme dans le cas précédent, on ne peut affirmer que le premier formant du sujet J.P.Z. est également FN1. De l'égalité des fréquences de ce premier pic pour les deux voyelles, on peut seulement conclure que ce formant a la même affiliation.

Les pics suivants, entre 1200 et 2400 Hz, sont nombreux et instables. Il ne paraît pas toujours possible de faire la part des irrégularités d'origine glottale et des résonances authentiques. FN2 en particulier, que l'on attend vers 2200 - 2300 Hz, n'est bien visible que sur les sections spectrographiques. La fréquence du pic P6 paraît trop faible pour être FN2. Parmi les interprétations possibles, celle d'un pôle lié à une dissymétrie des fosses nasales mérite d'être à nouveau mentionnée. le pic P8 est, sans hésitation, F'3. Mais les sections spectrographiques montrent un F'3 plus élevé, vers 3000 Hz. Au delà, on note régulièrement un second maximum qui pourrait correspondre à FN3.

L'interprétation des résultats du sujet J.P.Z. (table VI) pose quelques problèmes. Comme pour la voyelle précédente, le pic P2 est à une fréquence plus élevée (740 Hz) que celle du premier sujet. Mais dans les deux cas, la fréquence est la même pour les deux voyelles nasales. Il peut s'agir ici de FN1 ou de F'1. Les deux pics suivants sont à des fréquences très proches. La séparation moyenne n'est que de 235 Hz. Le premier de ces deux formants est F'1 ou FN1, le deuxième F'2. La fréquence un peu plus haute de F'2 indique une articulation linguale un peu moins postérieure que celle premier sujet. On hésite à nouveau à proposer le pic P5 (1920 Hz) comme trace de FN2. Sa fréquence est relativement basse, ce qui semble contradictoire avec la présence d'un premier formant

nasal élevé. P6 est un autre candidat, mais sa fréquence paraît au contraire trop élevée (2900 Hz). Le pic homologue de la voyelle [ɛ̃] se plaçait à 2600 Hz. Une fréquence proche de 3000 Hz conviendrait mieux à celle du troisième formant. L'amplitude de P6 est néanmoins beaucoup plus faible que celle du pic P7, 450 Hz plus haut, qui paraît, de par sa prééminence, un meilleur candidat pour F'3. Globalement, ces résultats suggèrent que le conduit nasal du sujet J.P.Z. est plus court que celui du premier sujet. Mais la présence d'une résonance supplémentaire à 1900 Hz, dont l'origine est peut être à chercher dans une dissymétrie du conduit nasal, pourrait expliquer en partie le décalage constaté de FN2. Cette paire pôle - zéro jouerait ici le même rôle pour FN2 que le formant nasal par rapport à F'1 pour le premier sujet.

Nos simulations se fondent sur une forme de conduit oral (O5) dont les fréquences de résonances sont de 610, 1100 et 2480 Hz. Comme pour la voyelle précédente, ce conduit a été obtenu par des ajustements successifs sur une forme initiale empruntée à LHERM (1984). Pour le conduit nasal de FANT (N1), dont les aires des deux dernières sections nasales ont été fixées à 2.1 et 1.1 cm², et le sinus S1, le premier pic du spectre simulé est F'1. La fréquence propre du conduit nasal est en effet supérieure à celle du conduit oral (figure 1.85). Le couplage augmente sa fréquence de 85 Hz, et il approche de 700 Hz. Le pic suivant à 1030 Hz est vraisemblablement formé par FN1 et F'2, le zéro de couplage provoquant une très légère irrégularité spectrale sur son flanc droit. Il est difficile de décider si le formant suivant, à 2400 Hz, est FN2 ou F'3. La nasalisation devant augmenter légèrement la fréquence de F'3 (2480 Hz pour la voyelle orale), il est plus probable qu'il s'agit de FN2, dont la fréquence propre est de 2200 Hz. Quoi qu'il en soit, le couplage et le zéro décalent l'un d'entre eux qui se confond presque avec F'4. En très basse fréquence, la résonance sinusale forme une bosse peu distincte vers 260 Hz. L'enveloppe spectrale de la voyelle simulée à l'aide du conduit nasal "long" (N3 ; sinus S2, couplage 2.4 cm²) est presque identique (figure 1.86), alors que le premier formant est FN1 Sa fréquence est un peu plus basse que précédemment (630 Hz). F'1 est la petite irrégularité à peine visible à 900 Hz. F'2 se place à une fréquence un peu plus élevée (1075). Le troisième pic proéminent (2040 Hz) ne peut être que FN2, car sa fréquence est plus basse que la fréquence du troisième formant de la voyelle orale. Or le couplage ne peut qu'augmenter la fréquence d'un formant, et non la diminuer. Un zéro bien marqué sépare FN2 de F'3 (2750 Hz), très proche de FN3 vers 3050 Hz. Le sinus de taille plus réduite que dans la simulation précédente est responsable du pic de très basse fréquence à 285 Hz. Ces simulations confirment que l'allure générale du spectre des voyelles nasales ne dépend pas de l'ordre des formants de basse fréquence. Le premier pic peut être indifféremment FN1 ou F'1. La seule différence notable entre les spectres mesurés et les spectres simulés tient à l'amplitude un peu forte du premier pic relativement à F'2. Il est vraisemblable que

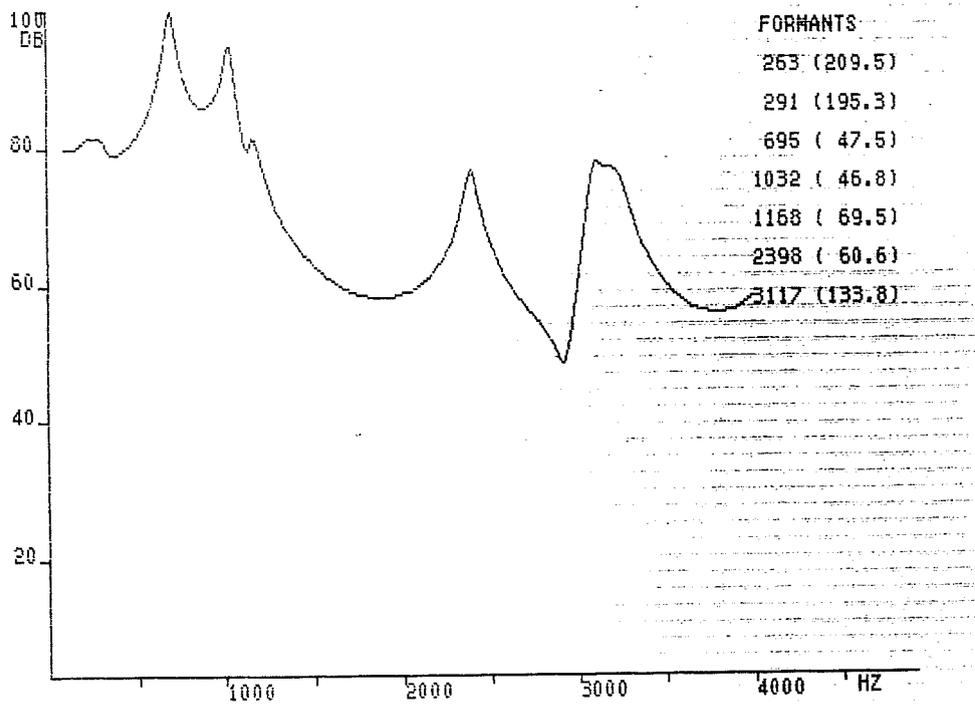


Figure 1.85 - Simulation du spectre de la voyelle [ã] : conduit nasal de FANT.

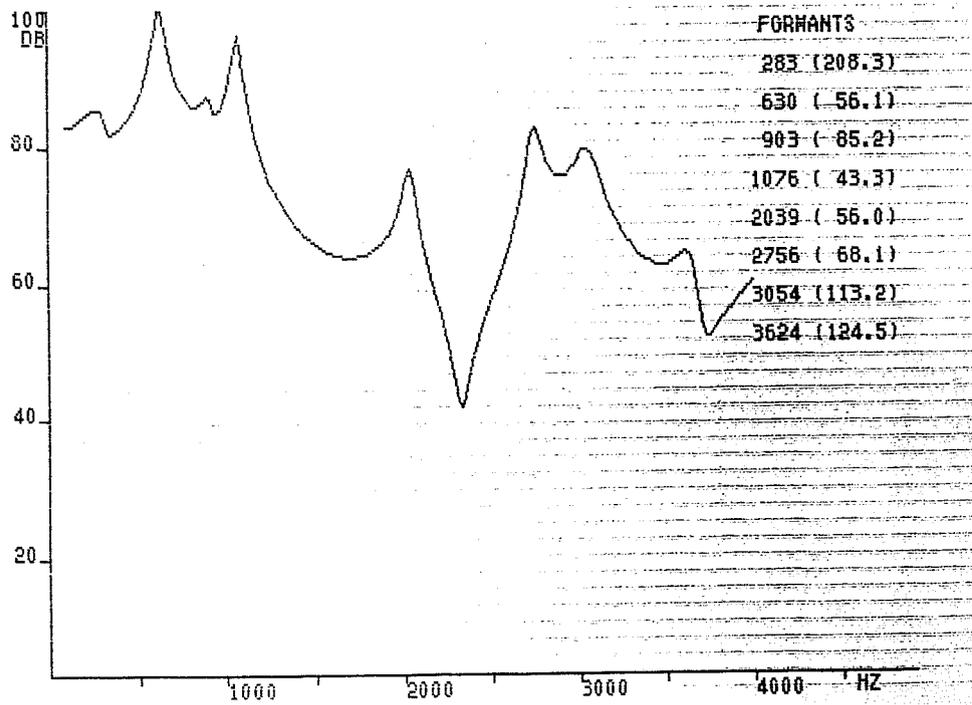


Figure 1.86 - Spectre de la voyelle [ã] : conduit nasal 'long'.

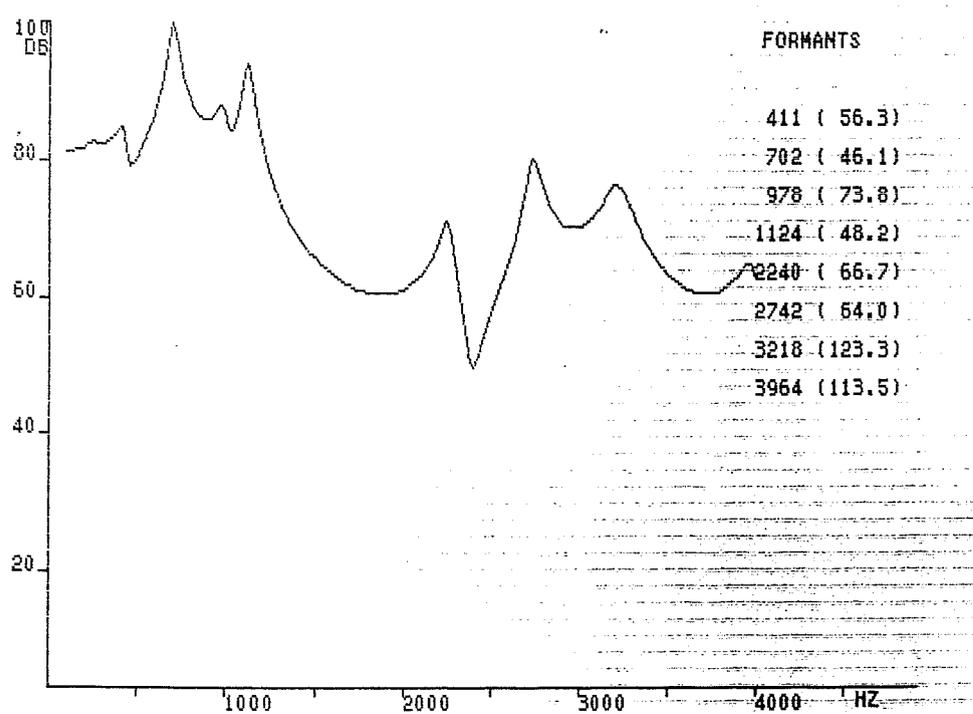


Figure 1.87 - Voyelle [ã] : conduit nasal de la consonne nasale vélaire; couplage faible.

notre choix de la valeur de couplage place, dans les deux cas, le premier zéro un peu trop près de F'2, réduisant ainsi son amplitude. La dernière simulation (figure 1.87) fait appel au conduit nasal utilisé pour la simulation de la nasale vélaire (N2 ; aires des deux dernières sections : 2.5, 1.1 cm²). L'emploi d'un sinus de taille réduite (S3) est sans doute responsable de l'élévation de FN1, qui dépasse de peu 700 Hz. F'1 est toujours aussi discret, vers 1000 Hz, et F'2 est à une fréquence un peu trop élevée (1125 Hz). Son amplitude est faible, et on peut à nouveau soupçonner le zéro de couplage d'en être responsable. La fréquence de FN1 est voisine des valeurs mesurées pour le sujet J.P.Z. Les paramètres des formants supérieurs, dans l'ordre FN2, F'3 et FN3, sont conformes en amplitude et en fréquence à ceux des réalisations naturelles. Toutes les simulations, quelle que soit la taille ou la forme exacte du conduit nasal, montrent un bon accord avec les réalisations naturelles.

Les mesures de BOGNAR & FUJISAKI sont en bon accord avec les nôtres. Le fréquence de F'2 est particulièrement basse, ce qui suggère une articulation très reculée. Le pic P1 ne peut être le formant nasal. Même si le conduit nasal était complètement obturé aux narines, sa fréquence de résonance couplée serait supérieure à 300 Hz. La fréquence de P1 convient au formant glottal. P2 est vraisemblablement la trace de FN2, qui n'est pas évident dans toutes nos données. F'3 est à mi-distance des mesures des tableaux IV et V, ce qui pourrait impliquer une large variabilité due à la position exacte du second zéro de couplage.

ã	P1	Z1	F'1	F'2		P2	Z2	F'3
	190	380	590	790		2060	2430	2840
table IV	300*		526	786-997 908	1766	*		2690
table V			605	895	(1800)	2140		3000
table VI			740	935	1920	2900?		3360?

Les résultats de MRAYATI sont rappelés dans le tableau suivant.

\tilde{a}	FN1	F'1	F'2	F'3	
occ. son.	381	619	1057	3274	
occ. sd.	407	597	1039	3252	5 locuteurs masculins
nasales	368	587	1055	3180	

occ. son.	362	801	1093	3492	
occ. sd.	363	789	1118	3468	sujet R.C.
nasales	354	820	1199	3386	

table IV	300*	526	786-997	1766	* 2690
			908		
table V		605	895	(1800)	2140 3000
table VI		740	935		1920? 2900? 3360?

Contrairement à ce que nous constatons pour la voyelle [ɛ̃], on note une relative concordance pour certains formants. Les moyennes de F'1 et de F'2 pour les 5 locuteurs sont très proches de nos estimations des fréquences de FN1 et F'2. En revanche, les valeurs élevées de F'1 pour le sujet R.C. implique à la fois une ouverture excessivement ouverte et un conduit nasal extrêmement court, de l'ordre de 8 cm. Il est plus vraisemblable que le pic LPC se place à une fréquence comprise entre FN1 et F'1 (600 - 900 Hz). On peut soupçonner à nouveau que ce résultat est dû à un nombre insuffisant de coefficients de prédiction qui force l'algorithme à représenter deux pics distincts par un large maximum spectral.

10.3.3. - CONCLUSIONS

La discussion de la structure acoustique de la voyelle [ã] nous a conduit à proposer une nouvelle explication de l'indice de nasalité décrit il y a plus de 30 ans par DELATTRE et connu sous le nom d' "affaiblissement de F1". Que le formant en cause soit FN1 ou F'1, son amplitude est différente de celle d'une voyelle orale car l'amplitude du formant suivant est amplifiée par la proximité, ou même la fusion, de deux résonances : F'2, et FN1 ou F'1.

Acoustiquement, la voyelle [ã] se caractérise par un premier formant entre 600 et 700 Hz, lorsque l'aire du conduit nasal n'est pas réduite par un gonflement des muqueuses dû par exemple à un rhume. Ce formant est suivi de deux pics proches, vers 800 et 1000 Hz, ou

même d'un pic unique entre 900 et 1000 Hz, souvent plus proéminent(s) que le premier formant. Aux fréquences élevées, FN2 est marqué par un pic discret vers 2200 Hz. F'3 est plus intense, en raison de la proximité de FN3. Comme pour la voyelle [é], il conviendrait d'étudier plus avant l'origine d'une résonance supplémentaire qui se place entre 1700 et 1900 Hz, et qui pourrait être responsable de la variabilité des formes spectrales de haute fréquence.

Nos simulations confirment que l'ordre des formants FN1 et F'1 n'a qu'une influence négligeable sur le spectre de la voyelle [ã], dont la compacité spectrale est remarquable.

10.4 - LA VOYELLE [õ]

10.4.1 - Description du corpus

Les spectres des 12 [õ] retenus arbitrairement sont présentés à la figure 1.88. Le spectre 27 montre une deuxième harmonique intense, indice possible de la présence du formant glottal, puis un pic intense à 435 Hz, suivi d'un plateau déprimé de 5 harmoniques s'étendant jusqu'à 0.9 kHz, d'où n'émerge qu'une harmonique plus intense à 800 Hz. La pente spectrale est ensuite excessivement raide, ce qui fait penser à la présence d'un zéro. Quelques petits pics d'amplitude très faible, 40 dB sous le plus intense, jalonnent la zone comprise entre 1 et 2.7 kHz. Les plus visibles sont à 1600 et 2000 Hz. Deux pics proches apparaissent de part et d'autre de 3000 Hz. Le spectre 20 ne se distingue du précédent que par un affaiblissement important de la 9^{ème} harmonique, qui dégage la 10^{ème} à 900 Hz. On peut penser à l'effet d'un zéro, mais sa largeur de bande doit être très faible pour que l'effet soit aussi localisé. La fréquence du deuxième pic est plus basse de 100 Hz, comme celle des pics de haute fréquence, à 2750 et 3000 Hz. Le spectre 9, à une fréquence fondamentale plus élevée, ne montre pas de maximum vers 700 - 800 Hz. L'écartement plus important des harmoniques ne permet pas de savoir si cette disparition est réelle, ou due seulement au plus faible échantillonnage spectral. A nouveau, un zéro semble dégager un pic de faible amplitude à 1050 Hz. Un formant unique est visible vers 3 KHz. Le spectre 16 est celui d'une voyelle prononcée d'une voix volontairement intense. La deuxième harmonique est beaucoup plus faible que la troisième, qui forme un pic supplémentaire vers 300 Hz. Il est vraisemblable, mais pas certain, que ce pic supplémentaire soit d'origine glottale : une phase montante plus courte de l'onde glottale conduit à un formant glottal de fréquence plus élevée. Mais il peut s'agir aussi d'une résonance sinusale. Le deuxième maximum, à 650 Hz, est à une fréquence plus basse que dans les exemples précédents. La pente spectrale au delà de 700 Hz est plus faible, avec un pic discret à 925 Hz, en pleine pente. Hormis un formant très faible à 1700 Hz, le seul pic

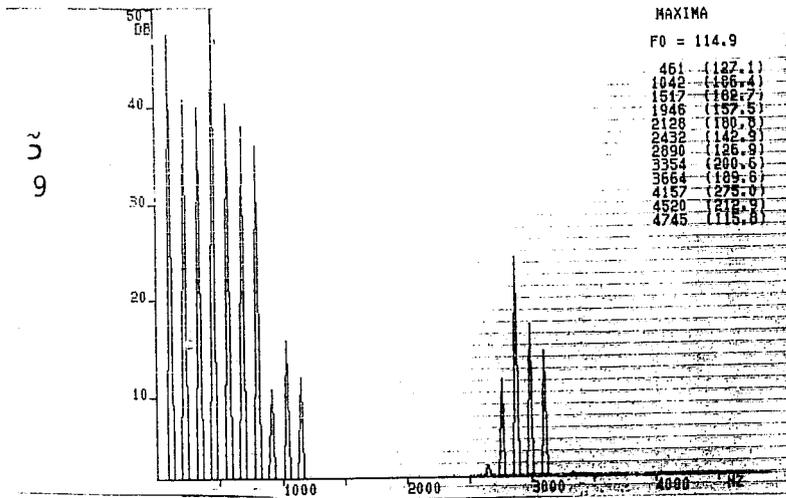
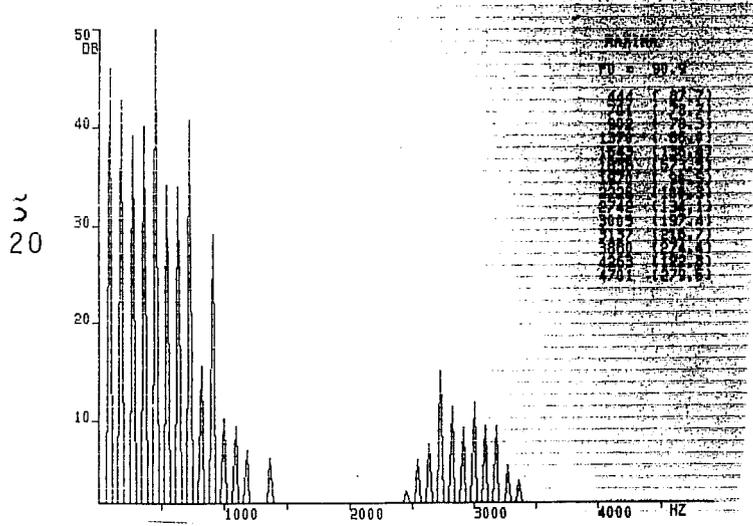
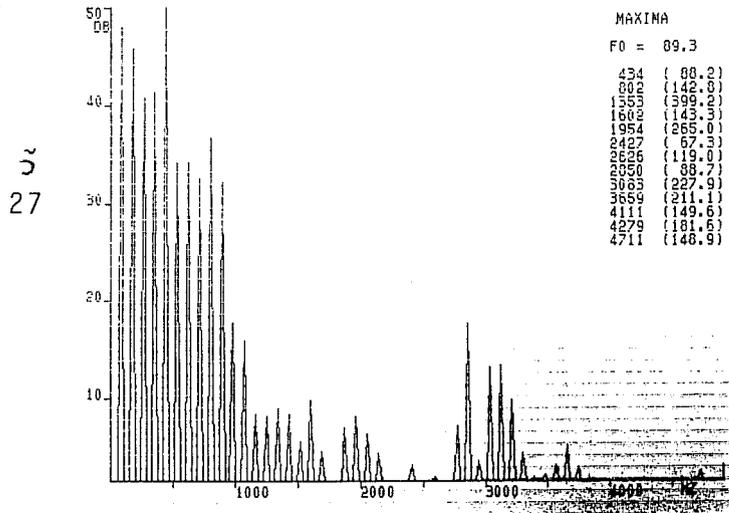


Figure 1.88 - Spectres des voyelles [ɔ̃].

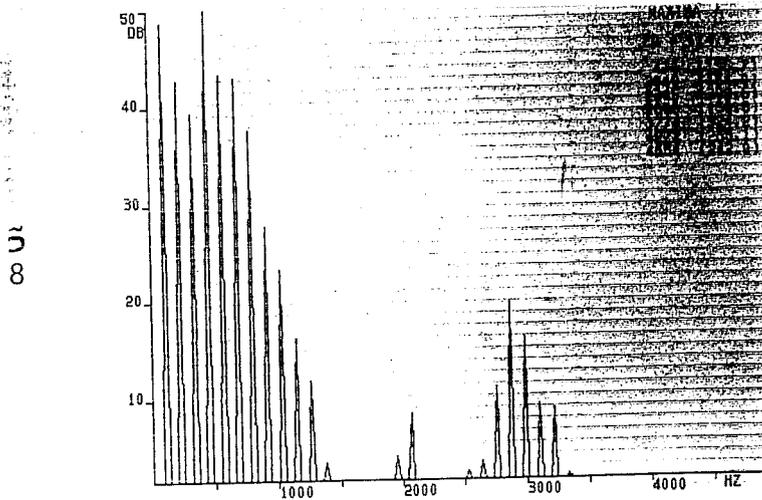
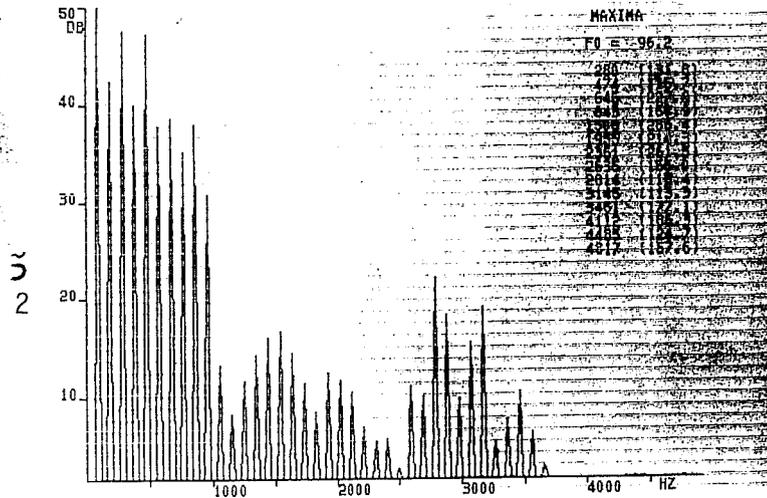
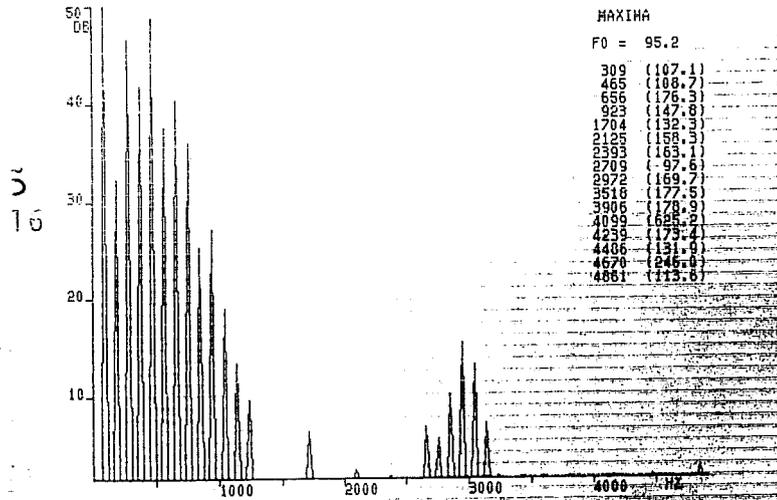


Figure 1.88 (Suite).

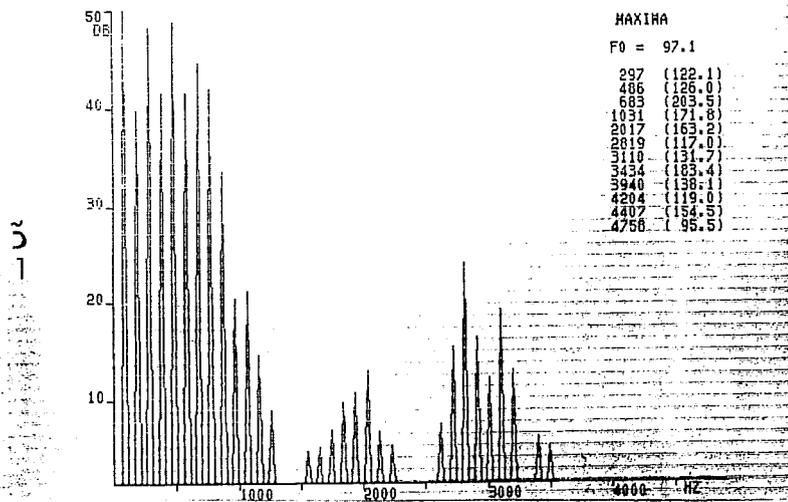
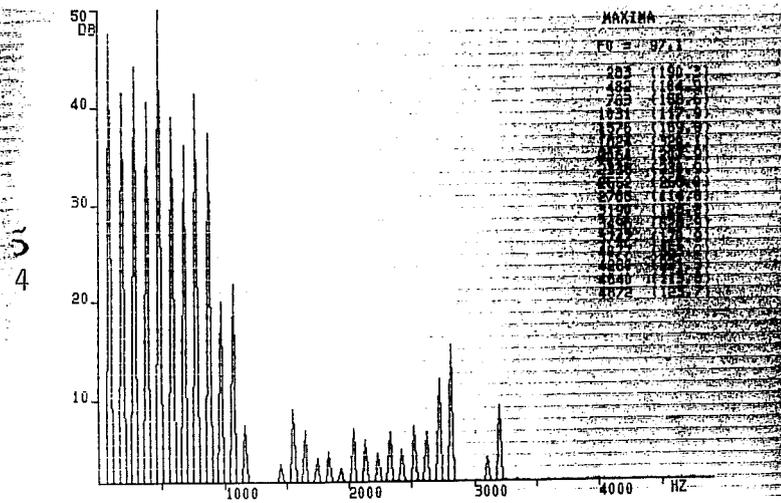
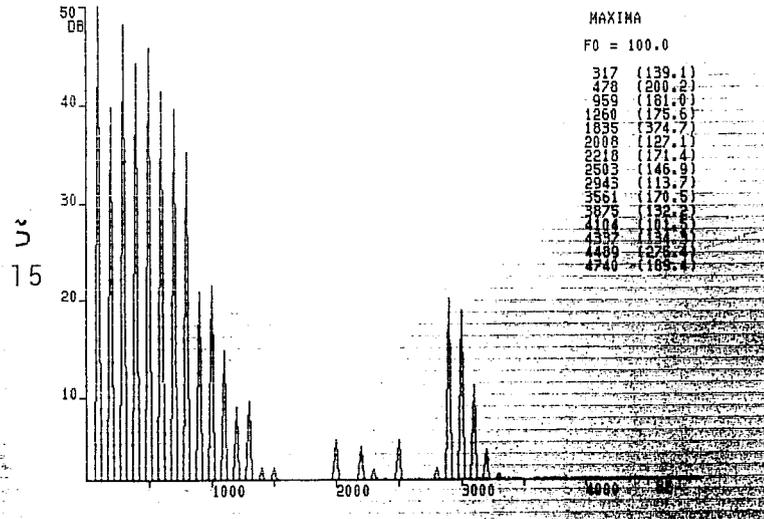
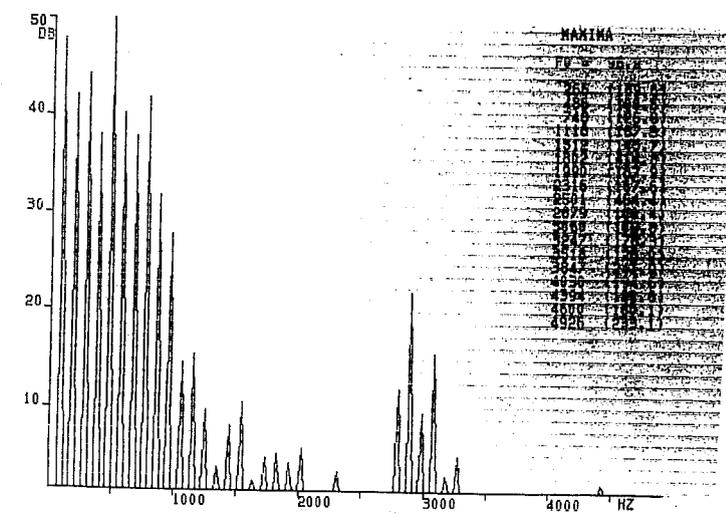
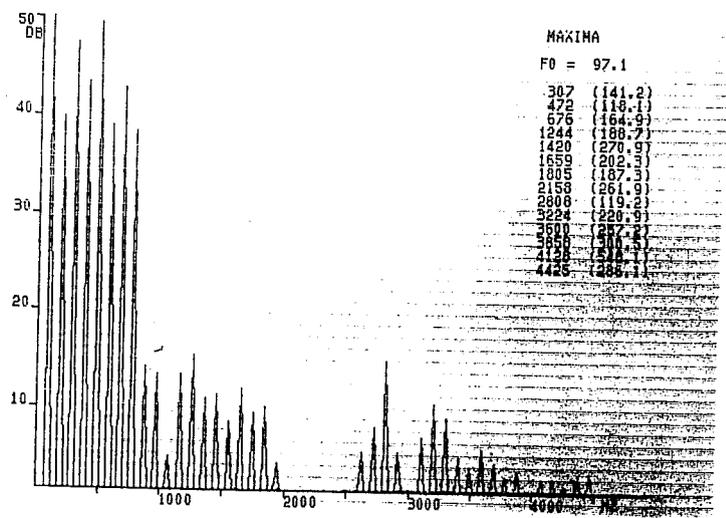


Figure 1.88 (Suite).

37



60



51

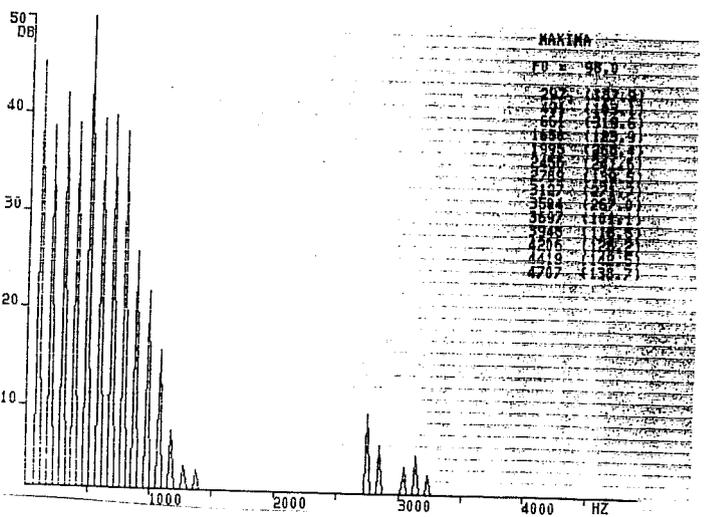


Figure 1.88 (Suite).

de haute fréquence est à 3000 Hz. Le spectre 6 possède également un pic de basse fréquence à 300 Hz. Le niveau spectral baisse de 35 dB entre l'harmonique 8 et 9. De petits maxima ponctuent la zone de haute fréquence à, 1250, 1650, 2800 et 3250 Hz. La caractéristique essentielle du spectre 2 est la valeur élevée (850 Hz) du dernier pic intense. Aux fréquences moyennes, le formant de 1500 Hz est plus intense que d'ordinaire. Un minimum spectral semble régulièrement présent entre 900 et 1100 Hz. Le pic de basse fréquence est absent de la voyelle 8, à une fréquence fondamentale relativement plus élevée. Comme à l'exemple 9, aucun formant n'est visible entre 600 et 900 Hz, et la pente spectrale est plus douce. Le zéro semble à une fréquence supérieure, car les composantes entre 1400 et 1900 Hz sont à plus de 50 dB sous le niveau de l'harmonique la plus intense. Le spectre suivant présente la même allure, en dehors de la présence d'un formant à 320 Hz. Le pic le plus élevé, à 3000 Hz, est intense dans les deux cas. Les trois spectres suivants (n° 4, 1 et 7) sont très semblables. Les 2 premiers pics sont à 300 et 480 Hz. La fréquence du troisième oscille entre 680 et 780 Hz. Au-dessus, on notera une épaule vers 1000 ou 1100 Hz, de petites irrégularités dont la plus constante est à 2000 Hz, et deux pics proches au-dessous et au-dessus de 3000 Hz. Le dernier spectre (5) ne présente pas de pic net vers 700 Hz.

Le tableau VII présente les fréquences de tout les maxima observés. La 6ème colonne rassemble sans doute des données disparates. Seuls les pics P2, P3 et P7 sont proéminents.

spectre n°	P1	P2	P3	P4	P5	P6	P7	P8
27		434	802	*	1602	1954	2850	
20		444	701	902			2742	3003
9		461	750*	1042	1517	1946	2890	
						2128		
						2432		
16	309	465	656	923	1704	2125	2972	
6	307	472	676	1244?	1659	1865	2808	3224
2	280	474	654	843	1508	1955	2814	3145
8		474	*	*		2049	2896	
15	317	478	700*	959		2008	2943	
						2218		
						2500		
4	283	482	783	1031	1576	2064	2785	3190
1	297	486	683	1031		2017	2819	
7	266	486	748	1118	1512	1990	2879	3060
5	297	491	661		1650	1995	2769	3127
médiane	297	474	700	1031	1590	2000*	2835	3136

TABLE VII - FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [õ]

Les mesures spectrographiques pour le même sujet sont présentées à la table VIII. P3 n'étant présent que sous la forme d'une "épaule", la détermination difficile de sa fréquence est signalée par une étoile. La concordance entre les médianes des tableaux VII et VIII est excellente.

	P1	P2	P3	P5	P6	P7	P8
spectres							
17		520	680*	1400	2040	3000	3360
16		500	680*	1440	2100*	2900*	3300*
13		480	650*	1680	2000*	2900	3180
2		480	650*		2100	3060	3350
17		490	670*	1440	2080	2910	3200
1		480	650*		2040	3000	3240
médiane		485	670*	1500*	2060	2950	3270
médiane table VII	297	474	700	1590	2000*	2835	3136

TABLE VIII : FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [õ] - (sujet F.L., d'après Lonchamp 1978, corrigé)

Les mesures formantiques effectuées sur les spectrogrammes des voyelles du sujet J.P.Z. sont rassemblées dans le tableau suivant. On peut d'ores et déjà noter qu'aucun pic ne correspond à P2.

	P1 - P2	P3	P4	P5	P6	P7	P8	P9
spectres								
1		680*	950	1190	1930	2770	3250	3520
2		680*	960	1180	1880	2580	3340	3600
4		680*	960	1190	2100	2360*	3100	3660
18		690	900*	1200*	2160	*	*	3650
19		710	1000	1320	1920	2320	3120	3600
20		660	1080		2040	2760	3240	3550
médiane		680*	960	1190	2015*	2500*	3240	3630
médiane 297 - 474	700	1031	1590	2000*	2835	3136		
table VII								

TABLE IX : FREQUENCES DES PICS SPECTRAUX DE
LA VOYELLE [õ] - (sujet J.P.Z., d'après Lonchamp 1978, corrigé)

10.4.2 - INTERPRETATION

Un examen du tableau VII montre immédiatement que la fréquence du premier pic proéminent (P2) est très stable. Si l'on élimine les deux valeurs extrêmes, la variation totale n'est que de 42 Hz. L'écart-type de la série entière est de 17.3 Hz. Cette stabilité contraste avec la variabilité du premier pic des voyelles précédentes. Il est évident que la fréquence est beaucoup plus basse. Elle est inférieure à la fréquence de résonance propre du conduit nasal que nous avons utilisé dans la simulation de la consonne nasale, et même à celle du conduit N3, arbitrairement étroit dans la zone narinaire. Il est donc plus que probable que ce pic est le premier formant oral décalé (F'1). Il paraît impossible de plaider que le pic P1 est F'1. Sa position est excessivement basse si l'on tient compte des effets de couplage qui ont tendance à élever les fréquences de résonance. Il est absent de plusieurs spectres dont la fréquence fondamentale est supérieure à 110 Hz, mais également du spectre n° 20, dont la fréquence fondamentale n'est que de 91 Hz. A nouveau, on hésite entre une origine glottale, ou l'affiliation aux sinus maxillaires. La réponse est peut-être qu'il reflète l'une et l'autre cause.

Le pic suivant (P3) est plus variable. Sa fréquence moyenne de 700 Hz est très basse, inférieure à celle mesurée pour le deuxième formant de [u] . Il peut avoir, théoriquement,

deux origines: FN1 ou F'2. Il paraît difficile de trancher, d'autant que ces deux pics peuvent être confondus. Mais un autre élément doit être intégré dans la discussion : l'extrême raideur de la pente spectrale qui suit ce pic suggère la présence d'un zéro vers 800 Hz. La pente peut atteindre 30 dB pour une différence fréquentielle de 100 Hz. Si l'ordre classique pôle - zéro est bien conservé ici, FN1 doit précéder le zéro, et sa fréquence doit être inférieure à 800 Hz. Mais l'analyse théorique présentée au paragraphe 5.3 montre que si deux formants sont proches, ce qui est le cas ici de F1N et F'2, le zéro peut précéder le pôle. Dans ce cas, FN1 devrait avoir une fréquence supérieure à 800 Hz, comme celle de P4. Mais la présence de P4 au dessus du second pic spectral ne simplifie pas l'analyse. Ce pic n'est pas visible sur tout les spectres. Il peut d'ailleurs s'agir d'un artefact lié à la présence du zéro creusant une dépression dans le flanc droit du second pic ! Le nombre d'inconnues est trop grand pour qu'une solution unique soit proposée. Les plus plausibles sont :

- a - FN1 et F'2 confondus en un pic unique, P4 n'étant qu'un artefact ;
- b - FN1 comme pic unique, avec F'2 décalé par couplage et ne subsistant qu'à l'état de trace (P4)

Il paraît cependant peu probable que la fréquence propre de F'2 soit inférieure à la fréquence propre de FN1 (550 - 600 Hz). Comme un formant ne peut en "doubler" un autre sous l'effet du couplage nasal, l'ordre F'2 - FN1 est peu plausible.

Le seul pic proéminent de haute fréquence est P7, vers 2850 Hz. Il s'agit sans doute de F'3. Entre 1000 et 2500 Hz, un grand nombre de pics de très faible amplitude sont visibles. FN2 doit figurer parmi eux, mais on ne peut que noter une plus grande densité d'estimations vers 2000 Hz. Comme dans les voyelles précédentes, F'3 semble suivi d'un pic proche, qui est sans doute FN3.

Il est impossible de retrouver un pic correspondant à P2 sur les spectres du sujet J.P.Z. Quatre harmoniques, entre 375 et 750 Hz, forment un plateau en pente douce vers les hautes fréquences. La 6ème harmonique est un peu plus intense, et semble définir un pic vers 680 Hz. Le pic noté P5 est quelquefois intense, et séparé du plateau précédent par un zéro. On peut supposer que F'1 et FN1 sont proches, et que le premier zéro sépare FN1 de F'2, qu'il annihile à l'occasion. Mais si FN1 est élevé pour ce sujet, on peut aussi envisager que F'2 soit plus bas que FN1. Notons qu'une forte constriction vélaire et pharyngale peuvent expliquer une fréquence très basse de F'2.

Les simulations sont fondées sur le conduit O6. Les fréquences de résonance sont un peu hautes, en raison sans doute d'une aire aux lèvres trop forte, et d'une constriction vélo-pharyngale trop faible. Les formants de ce conduit sont à 475, 830 et 2410 Hz. La figure 1.89 illustre les résultats de la simulation effectuée avec le conduit nasal de FANT (N1 ; sinus S1 ; aire des deux dernières sections nasales : 2.1, 1.1 cm²). Le premier pic est

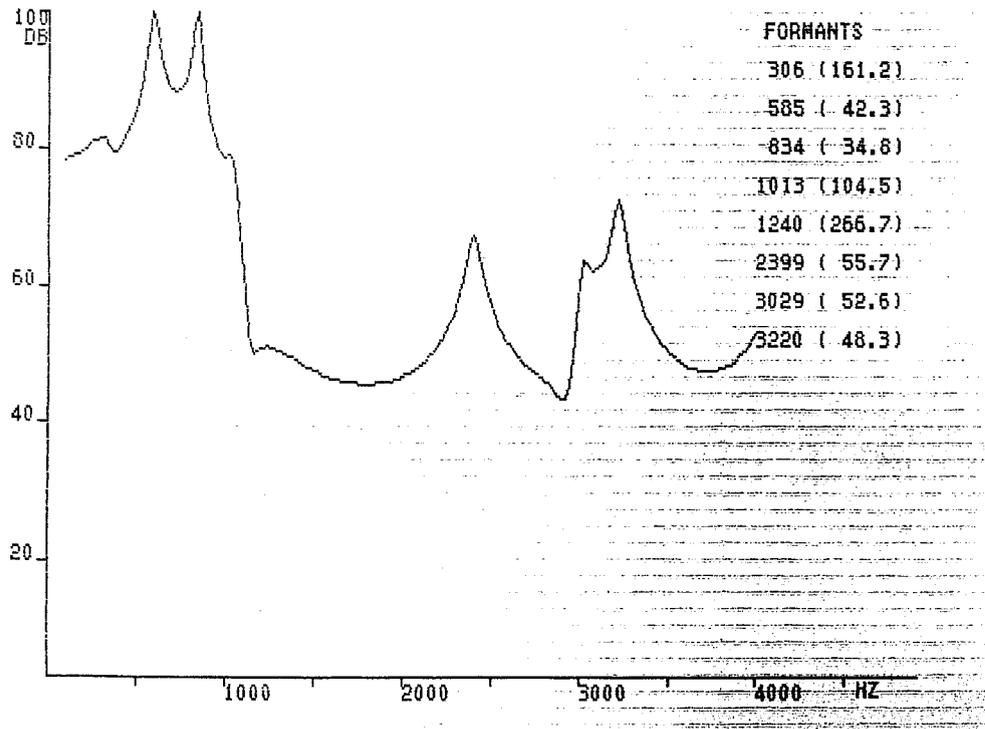


Figure 1.89 - Simulation du spectre de la voyelle [ɔ̃] : conduit nasal de FANT.

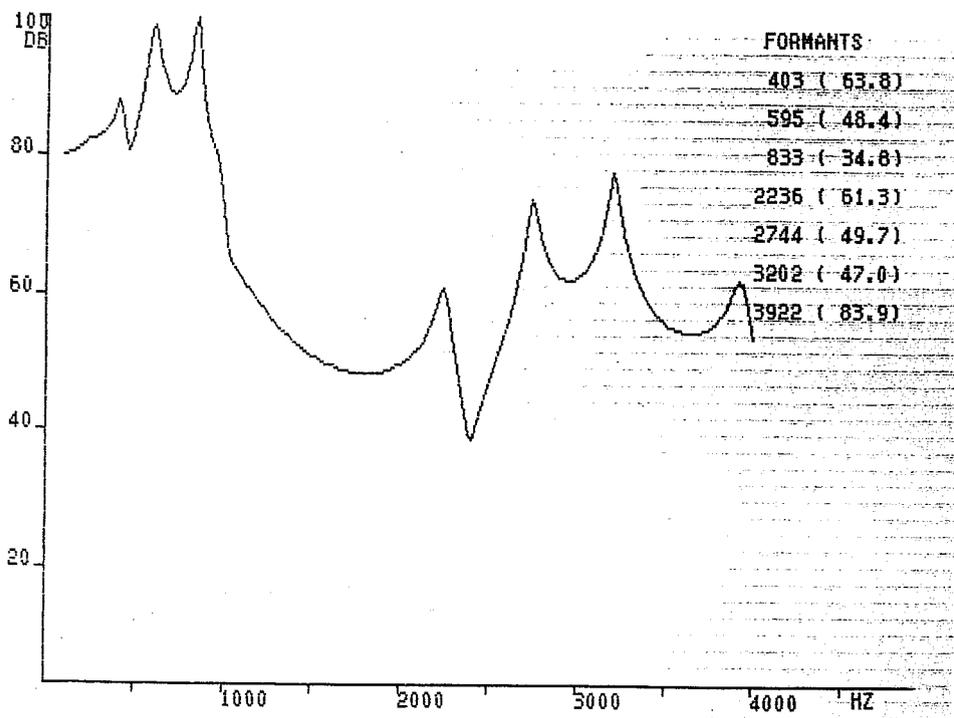


Figure 1.90 - Voyelle [ɔ̃] : conduit nasal 'long'.

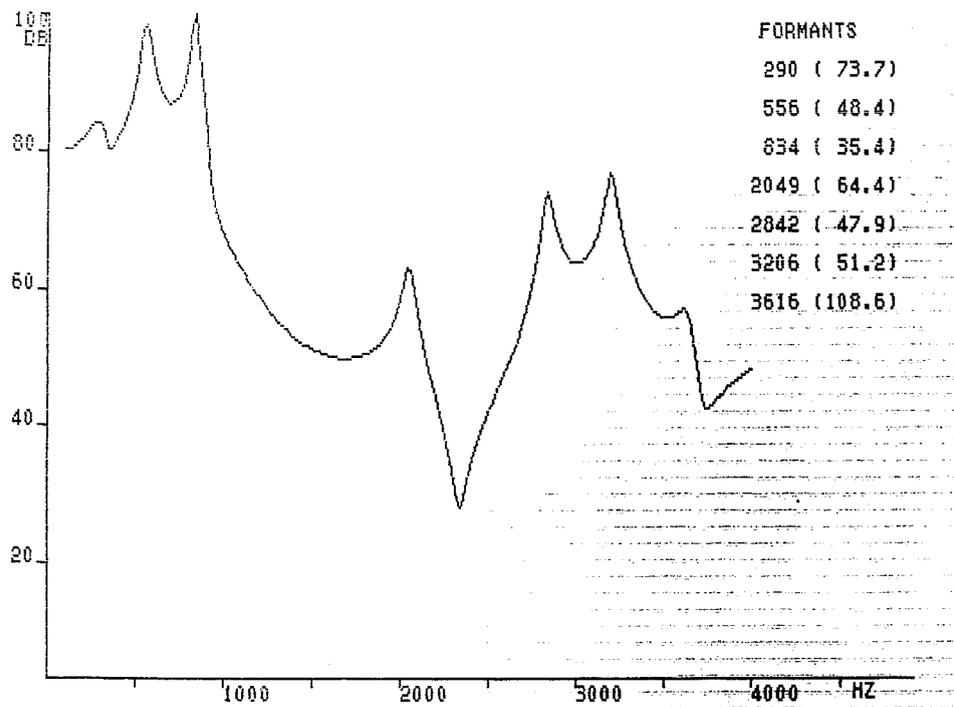


Figure 1.91 - Voyelle [õ] : conduit nasal de la consonne nasale vélaire (couplage faible).

585 Hz, plus haut que dans les mesures sur les voyelles, mais nous avons déjà noté que le conduit utilisé possède une fréquence propre trop élevée. Le pic suivant (835 Hz) est FN1, car sa fréquence propre est inférieure à la fréquence de F2 sans couplage. F'2 n'est visible que par la petite irrégularité spectrale sur le flanc extrêmement raide de FN1, dont est responsable le zéro de couplage. Cette simulation, bien que les fréquences soient différentes des fréquences réelles, possède bien des attributs des spectres réels, et montre la plausibilité d'un F'2 de faible intensité sur le flanc droit de F1N. F'3, qui atteint 3030 Hz, se trouve rejeté contre F'4 (3220 Hz) par le zéro associé à FN2 (2400 Hz). La simulation suivante (figure 1.90) fait appel au conduit nasal "long" (N3 ; Sinus S2, couplage de 2.4 cm²). On notera particulièrement l'absence de toute trace de F'2. FN1 et F'2 sont parfaitement confondus en un pic unique à 835 Hz. Ceci montre que l'autre solution envisagée pour rendre compte du spectre des voyelles naturelles est également plausible : FN1 et F'2 peuvent fusionner. La dernière simulation est fondée sur le conduit nasal N2 (Sinus S3, couplage 2.5 et 1.1 cm²). La figure 1.91 montre une épaule très discrète vers 1000 Hz, qui doit être la trace de F'2. Le couplage plus faible assure à F'3 (2750 Hz) une fréquence proche de celle mesurée. Une dernière remarque s'impose. La fréquence du second pic est pratiquement identique, à 1 Hz près, quel que soit le conduit nasal utilisé. Cette fréquence est la deuxième fréquence propre du conduit oral sans couplage. Le formant nasal, dont la première fréquence propre est inférieure, ne peut dépasser cette limite, comme la théorie l'enseigne.

Les simulations sont particulièrement utiles dans le cas difficile que constitue la voyelle [õ]. Elles montrent en particulier que le second pic spectral est FN1 seul, ou FN1 confondu avec F'2, et non F'2 seul comme on aurait pu le penser au vu des seuls spectres.

Les résultats de BOGNAR & FUJISAKI sont très voisins des nôtres, comme le montre le tableau ci-dessous, et cet accord montre l'intérêt des deux méthodes d'analyse employées.

	P1	Z1	F'1	F'2	P2	Z2	F'3
õ	170	390	450	690	2250	2360	2900
table VII	297		474	700	2000		2835
table VIII			485	670*	2060		2950
table IX				680*	2015		3240

Comme pour les voyelles précédentes, le pic P1 ne peut être le formant nasal. Les résultats de MRAYATI sont rappelés dans le tableau ci-dessous.

\tilde{o}	FN1	F'1	F'2	F'3			
occ. son.	353		808	3296			
occ. sd.	373		712	3412	5 locuteurs masculins		
nasales	350		688	3452			

occ. son.	431		764	3065			
occ. sd.	442		711	3049	sujet R.C.		
nasales	419		733	3040			

table VII	297	474	700	(1590)	2000	2835	3136
table VIII		485	670*	(1500*)	2060	2950	3270
table IX			680*	(1190)	2015	2500*	3240

Pour cette voyelle, il est possible d'attribuer à F'1 le formant noté FN1 dans le cas du sujet R.C. Mais les mesures moyennes sur les 5 sujets sont toujours biaisées négativement. La fréquence du deuxième formant est également voisine de notre estimation, aussi bien pour les 5 sujets que pour R.C.

10.4.3 - CONCLUSIONS.

Les caractéristiques essentielles de la voyelle [ō] résident dans les faits suivants :

- 1 - Le premier formant visible sur les spectres est très certainement le premier formant oral (F'1). Sa fréquence est en effet inférieure à l'estimation que nous avons faite de la première fréquence de résonance du conduit nasal du sujet, qui ne peut être inférieure à 550 Hz. Or le couplage ne peut avoir pour effet de faire baisser la fréquence d'un formant.
- 2 - Le second pic visible, à une fréquence très basse (700 Hz environ), est vraisemblablement le premier formant nasal (FN1), avec lequel le second formant (F'2) pourrait venir se confondre pour former un pic unique. Il n'est exclu que F'2 soit, dans d'autres cas, un pic discret sur le flanc droit, à la pente excessivement raide, du second pic proéminent. Cette pente résulte de la présence du premier zéro, vers 800 Hz. Il paraît peu probable que la fréquence de résonance de F2 puisse être si basse qu'elle se place en dessous de la fréquence de FN1 avant couplage. Il

faudrait pour cela qu'elle atteigne moins de 550 Hz.

- 3 - Il paraît exclu que le pic peu intense, visible vers 300 Hz, soit FN1 ou F'1. Comme dans les voyelles précédentes, il peut s'agir d'une résonance sinusale et / ou d'un rebond du spectre glottal.

La voyelle [õ] est donc une voyelle compacte, dont la fréquence basse du deuxième formant s'explique par le fait qu'il est en réalité, au moins dans certains cas, le premier formant nasal.

11 - LES INDICES ACOUSTIQUES DE LA NASALITE VOCALIQUE : UN ESSAI DE SYNTHÈSE

Dans l'introduction, nous définissons le but de ce travail de la manière suivante : caractériser les indices acoustiques de la nasalité vocalique, c'est-à-dire identifier l'origine et la nature des indices présents dans les spectres de réalisations naturelles. Pratiquement, cette identification s'est opérée par référence aux modifications prévisibles qu'apporte la nasalisation sur le spectre vocalique. Elle a été confirmée par la comparaison de spectres de réalisations naturelles avec des spectres déduits de nos connaissances anatomiques et articulatoires, par l'intermédiaire d'un modèle de simulation.

La théorie acoustique de la nasalisation est bien connue depuis les travaux de FANT et de FUJIMURA. Notre travail n'apporte, bien sûr, aucune contribution nouvelle à l'analyse théorique.

On rappellera, presque pour mémoire, les trois caractéristiques essentielles des effets du couplage d'un conduit nasal sur le conduit oral :

- 1 - Apparition de plusieurs paires pôle - zéro, notamment dans la zone du premier et troisième formant des voyelles orales.
- 2 - Décalage de tous les formants, mais à des degrés divers, vers les fréquences hautes
- 3 - L'ordre d'apparition des formants est lié à leur fréquence de résonance pour les conduits oral et nasal non couplés.

L'examen comparatif des nombreux travaux consacrés à l'articulation et à l'acoustique des voyelles nasales, ainsi que la consultation de sources d'informations sur l'anatomie des fosses nasales, nous a permis de glaner quelques faits ou conclusions, dont certaines nous

paraissent méconnues :

- 1 - S'il existe bien une paire pôle / zéro entre 270 et 400 Hz, il n'est pas exclu que FN1 ait une fréquence de 550 - 600 Hz pour les voyelles postérieures (FUJIMURA & LINDQVIST)
- 2 - La fréquence de résonance du conduit nasal seul est vraisemblablement comprise entre 500 et 600 Hz. Des résonances sinusales sont essentielles pour expliquer la complexité du spectre en basse fréquence (LINDQVIST & SUNDBERG), dont la résonance de 300 - 400 Hz mentionnée ci-dessus.
- 3 - L'utilisation de la fonction d'aire de BJUGGREN & FANT ne permet pas de retrouver le spectre de [ŋ] par simulation (FENG, ABRY & GUERIN). Mais une modification radicale de l'aire de la zone narinaire semble exclue d'après ce que montrent les clichés tomographiques.

Nos propres simulations, qui ont d'abord porté sur la consonne nasale vélaire, révèlent que :

- 1 - Une simulation adéquate du spectre de la consonne nasale vélaire peut être obtenue en utilisant un conduit nasal possédant une aire narinaire conforme aux mesures anatomiques disponibles. Mais la longueur du conduit nasal et le volume de la partie rhinopharyngale doivent être supérieurs aux mesures de BJUGGREN & FANT (1964).
- 2 - Le présence de sinus est nécessaire pour qu'un zéro vienne décaler la fréquence du premier pic de 380 à 320 Hz environ.
- 3 - Le gonflement de la muqueuse peut produire une fermeture de l'ostium maxillaire. L'absence de résonance sinusale provoque un décalage du premier formant vers les fréquences hautes, attesté par les mesures des fréquences formantiques de réalisations naturelles.
- 4 - L'hypothèse d'une forte constriction narinaire pour rendre compte de la fréquence basse de F1 est inutile. Elle n'est compatible ni avec les données anatomiques, ni avec la fréquence de la première résonance propre du conduit de 570 Hz mesurée par LINDQVIST et SUNDBERG pour le sujet J.L., ni avec l'élévation de F1 constatée lorsque qu'un sujet présente un gonflement des muqueuses sous l'effet d'un léger coriza.

L'interprétation de l'origine et de la nature des structures acoustiques présentes dans quelques dizaines de voyelles nasales, confirmée par le bon accord entre les formes spectrales des simulations et des voyelles naturelles, montrent que les indices spécifiques de nasalité vocalique sont discrets. Une composante spectrale nous semble pouvoir être qualifiée de spécifique si elle n'apparaît que dans le spectre des voyelles nasales, si elle est présente dans le spectre de toutes les voyelles nasales, et, enfin, si ses caractéristiques sont relativement invariantes.

Si la voyelle [ɛ̃] montre effectivement trois formants entre 500 et 1500 Hz, alors qu'une voyelle orale ne peut en posséder que deux dans le même intervalle, ce n'est pas toujours le cas de la voyelle [ã]. Quant à [õ], elle ne présente souvent, sous 1000 Hz, qu'un seul formant et une "épaule" sur le flanc droit de ce dernier. La présence d'un formant supplémentaire dans la région de F1 n'est donc pas un indice spécifique. Le second formant nasal, vers 2000 Hz, est d'amplitude faible et ne constitue pas une caractéristique physique marquante. Les résonances sinusales sont très difficiles à identifier. La principale, liée sans doute au couplage des sinus maxillaires, possède une fréquence très basse (300 Hz), dans une zone où existent déjà des irrégularités spectrales d'origine glottale. Les zéros ne sont qu'exceptionnellement marqués par des minima spectraux caractéristiques.

Il ne paraît donc pas possible de proposer un mécanisme ou un algorithme de détection de la nasalité, en vue de la reconnaissance automatique de la parole par exemple, qui s'appuierait sur la détection de structures acoustiques spécifiques, toujours présentes et relativement invariantes.

Pour les voyelles françaises étudiées, la détection de la nasalité ne nous semble possible que si l'on tient compte de la fréquence et de l'amplitude relative des formants. Ces deux paramètres dépendent à la fois de la position articulaire linguale, qui distingue les voyelles nasales de la voyelle orale transcrite avec le même symbole phonétique, et des effets du couplage, qui se traduisent ainsi d'une manière indirecte. Pour la voyelle [ã], notamment, nous avons offert une nouvelle explication de l'affaiblissement de F1, signalé comme indice de nasalité il y a très longtemps par DELATTRE. Pour les deux autres voyelles nasales, l'amplitude des formants principaux est également modifiée, soit en raison de la présence d'un formant supplémentaire, soit par l'effet du premier zéro de couplage.

Il apparaît possible d'affirmer que le premier formant des voyelles [ɛ̃] et [ã] est le premier formant nasal (FN1) pour le sujet étudié. On rend compte ainsi de la variabilité fréquentielle du premier pic, que l'on peut attribuer à une variation de l'aire du conduit nasal et du degré de couplage des sinus, fonction de l'état physiologique de la muqueuse

nasale. En revanche, le premier formant de la voyelle [õ] est F'1.

Nous avons pu montrer qu'une inversion de l'ordre des formants F'1 et FN1 ne provoquait pas de modification notable de la forme du spectre des voyelles [ẽ] et [ã]. Ces voyelles sont donc relativement insensibles à une variation modérée de la fréquence de la première résonance nasale. Comme la fréquence du deuxième formant est principalement contrôlée par l'articulation linguale, intrinsèquement stable, les voyelles nasales possèdent donc des fréquences formantiques spécifiques. Elles sont à la fois différentes de celles des voyelles orales, car la position linguale est différente, et relativement stables, sauf peut-être en cas d'occlusion complète du conduit nasal.

Cette brève synthèse est bien sûr incomplète car elle ne prend en compte que les caractéristiques acoustiques de la nasalité vocalique. Le dernier chapitre offre, à la lumière de résultats très récents sur la perception de la nasalité, quelques réflexions sur les indices perceptifs de la nasalité vocalique, qui intègrent les faits saillants du domaine acoustique.

12 - PERCEPTION DES INDICES DE NASALITE VOCALIQUE

Les paragraphes suivants ont pour principale ambition de faire le point sur une série de travaux très récents qui s'efforcent de préciser la nature des indices perceptifs de la nasalité vocalique. Nous croyons être en mesure de faire en conclusion une synthèse originale, après une analyse critique de ces études. L'ordre de présentation des travaux n'est pas essentiellement chronologique, mais plutôt thématique, cherchant à identifier les problèmes communs aux diverses perspectives.

12.1 - HAWKINS & STEVENS (1985)

Cette brève présentation ne donne qu'un pâle reflet de la richesse, et de la complexité, des arguments évoqués. Cette étude illustre parfaitement les difficultés inhérentes à une étude des indices acoustiques de la nasalité vocalique. La première est le choix de l'indice, ou des indices, à manipuler. La liste des candidats potentiels établie par HAWKINS & STEVENS comprend :

- une paire pôle - zéro supplémentaire vers F1,
- le décalage des formants vers les fréquences hautes,
- l'augmentation des largeurs de bande,
- une ou plusieurs paires pôle - zéro liées aux résonance sinusales.

Une revue rapide des travaux antérieurs révèle que :

- la distinction voyelle nasale / non nasale est possible même pour des sujets dont la langue maternelle ne possède pas de voyelles nasales phonologiques (cf. BUTCHER 1976, WRIGHT 1986 par exemple)
- un timbre nasal peut s'obtenir en ajoutant un pic vers 250 Hz (DELATTRE 1954) et / ou une paire pôle - zéro vers F1 (TAKEUCHI & al. 1975). Une amplitude plus grande des formants supérieurs accroît, dans certains cas, les jugements de nasalité, qui restent néanmoins assez faibles.
- Dans des simulations employant un modèle articulatoire, on a pu mettre en évidence qu'une nasalisation adéquate des voyelles ouvertes nécessitait un plus grand passage velopharyngal que les voyelles fermées (MAEDA 1982 a).

Mais il faut tenir compte du fait que, notre connaissance de la géométrie des fosses nasales et du conduit oral étant ce qu'elle est, les indices acoustiques de nasalité issus d'un modèle articulatoire sont pas forcément aussi adéquats pour ces deux catégories.

Ouvrons une parenthèse pour préciser les causes acoustiques des modifications de l'amplitude des formants supérieurs qui accompagnent la nasalité. FANT (1956) a démontré que l'amplitude des formants est complètement définie par la distance entre les fréquences formantiques. D'une l'analyse théorique de l'acoustique de la production, il déduit la règle suivante : doubler la fréquence d'un formant augmente de + 6 dB le niveau des formants supérieurs. Dans le cas des voyelles nasales, la présence d'un formant nasal au-dessus de F1 réduit la distance entre F2 et FN1, et élève donc l'amplitude des formants supérieurs. Même dans le cas où FN1 est placé au-dessous de F1, le couplage provoque une élévation de F'1, qui se rapproche des formants supérieurs. Le problème se complique si l'on reconnaît que la présence de pôles supplémentaires liés aux cavités sinusales doit également provoquer le même effet.

Mais cette élévation de l'amplitude des formants supérieurs est contrecarrée par le rôle des zéros qui, comme nous l'avons rappelé au paragraphe 8.1, provoque un abaissement du niveau des fréquences supérieures, dont la magnitude dépend de l'écart fréquentiel entre le pôle et le zéro liés. Il est clair qu'il est impossible de proposer une règle générale portant sur la forme globale du spectre. Une connaissance précise du nombre, et de la fréquence, des pôles et des zéros est nécessaire pour prévoir l'amplitude des formants supérieurs. Seules des mesures de voyelles naturelles permettraient de déterminer, si la variabilité individuelle n'est pas trop forte, l'effet global de la nasalisation sur le spectre. En l'absence de statistiques comparatives, notons que les simulations de MAEDA (1982 b) montrent, entre autres choses, un "aplatissement" du spectre dans la région de F1, ou de F1 - F2, selon les voyelles.

HAWKINS & STEVENS testent tout d'abord l'hypothèse d'un indice de nasalité commun à toutes les voyelles : une paire pôle - zéro dans la région de F1, avec un

ajustement éventuel de la fréquence de F'1 imitant la différence constatée chez un locuteur Gujarati entre des voyelles orales et nasales de même timbre. La paire pôle - zéro était au-dessus de F'1 pour les voyelles [~i], [~e], [õ] et [~u], et en dessous pour [ã]. La fréquence du pôle et la valeur de la séparation entre le pôle et le zéro étaient les variables expérimentales. L'absence de corrélation entre les jugements des 5 auditeurs indiens et des 3 auditeurs gujarati reflète, selon les auteurs, la prise en compte, dans le jugement des auditeurs indiens, d'un autre critère que celui de nasalité. Ceci est discutable. On pourrait supposer que les locuteurs n'acceptent de considérer une voyelle comme nasale qu'à la condition qu'elle possède des fréquences formantiques particulières. La position de ceux-ci serait un indice de nasalité, mais différent de celui envisagé par HAWKINS & STEVENS. L'intersection des voyelles synthétiques jugées comme très nasales par les auditeurs américain et gujarati se produit lorsque la fréquence du zéro est au milieu de l'intervalle fréquentiel entre FN1 et F'1, séparée d'un moins 100 Hz de la fréquence de FN1. Les voyelles les plus nasales ont les caractéristiques suivantes :

	F'1	Z1	FN1
~i :	275	600	900 Hz
~e :	350	500	700
õ :	400	500	650
~u :	250	475	650
ã :	850	600	450

Ce choix pose un problème crucial. Le formant nasal est largement inférieur à la fréquence de F'1 pour [ã], alors qu'il est beaucoup plus élevé pour les autres voyelles. Une fréquence de 450 Hz après couplage suppose un conduit nasal à sortie très étroite, qui n'est pas anatomiquement très plausible. Il n'est pas certain également que la fréquence de FN1 d'un tel conduit nasal soit cohérente avec les fréquences de FN1 retenues pour les autres voyelles. On mesure ici la première difficulté d'une étude perceptive : comment effectuer le choix des paramètres de synthèse, ou anatomiques, dans le cas de l'utilisation préalable d'un modèle articulatoire. Il paraît raisonnable de demander que les paramètres choisis soient compatibles avec les structures acoustiques qu'un locuteur unique peut émettre. Ceci n'est pas certain pour les paramètres retenues ici.

Ces voyelles servent à une expérience d'identification et de discrimination pour des séries allant d'une voyelle orale à la voyelle nasale décrite. La discussion révèle que la différence entre [o] et [õ] est plutôt perçue comme une différence de timbre : voyelle ouverte / non ouverte. On peut faire l'hypothèse que le formant nasal à 650 Hz joue le rôle de F'1 dans cet exemple, F'1 plus bas (400 Hz) étant pris comme F1N. La frontière phonétique orale /

nasale est franchie lorsque la séparation entre FN1 et Z1 est voisine de 100 Hz. Mais on peut craindre que cette conclusion ne découle en fait des conditions expérimentales. Pour cette tâche à choix forcé, imaginons que le sujet se fixe comme critère de répondre "nasal" chaque fois que le spectre présente une différence de timbre perceptible, qu'elle soit identifiable ou non comme une trace de nasalisation. Cette différence se produira logiquement lorsque la perturbation spectrale sera suffisante, c'est-à-dire pour une même valeur de séparation de la paire pôle - zéro.

Il est tout à fait vrai, comme le notent HAWKINS & STEVENS, que les résultats sont interprétables de la manière suivante : l'introduction d'une paire pôle - zéro supplémentaire réduit le degré de proéminence de F1. Leur conclusion, plus précisément, est qu'il existe un indice fondamental de nasalité, commun à toutes les voyelles, auquel le système auditif répond de manière spécifique : le degré de proéminence du pic spectral dans la zone de F1. Un degré réduit de proéminence de F1 diminuerait le nombre de neurones auditifs dont la fréquence de décharge est synchronisée sur F1.

Mais, outre le fait que la manière de mesurer la proéminence d'un pic spectral n'est pas décrite avec précision, il est évident que toutes les paires réduisant la proéminence de F1 ne provoquent pas l'apparition d'un timbre nasal. Identifier les indices de nasalité suppose que l'on détermine les raisons pour lesquelles certaines paires pôle - zéro sont plus "efficaces" que d'autres. En outre, l'examen des spectres naturels n'est guère encourageant pour cette conclusion. Pour [i], le formant nasal n'est qu'une bosse sur le flanc de F1 qui demeure très proéminent (MAEDA 1982 a), à l'inverse de ce qu'on trouve pour les voyelles [ɛ]. Beaucoup de [ā] et de [ō] ne possèdent que deux formants dans la région couvrant F1 et F2, tous les deux proéminents. Dans la discussion est évoquée longuement l'hypothèse, également reprise dans la conclusion, qu'un indice commun de la nasalité est accompagné d'indices auxiliaires. Le premier est une modification de la forme globale du spectre, que nous avons discutée plus haut. Citant LONCHAMP (1978) à l'appui de leur résultat concernant [ō], HAWKINS & STEVENS suggèrent enfin que le degré d'aperture influence le degré de nasalité perçue. Plus précisément, la nasalité produirait une variation du centre de gravité de la proéminence spectrale de basse fréquence, conduisant à un changement du degré d'aperture perçue. Si nous interprétons correctement ces deux formulations, le degré d'aperture influence le degré de nasalisation car la nasalisation s'accompagne d'ordinaire d'un changement d'aperture. Ces formulations ne reflètent pas exactement notre position, qui fera l'objet du paragraphe final.

12.2 - CHENG & GUERIN (1987 b)

Cette étude possède beaucoup de points communs avec celle de HAWKINS & STEVENS. La conséquence acoustique principale de la nasalisation est définie comme l'apparition d'un triplet pôle - zéro - pôle à la place du seul F1. Si les auteurs reconnaissent que les positions des formants et du zéro sont liées, car elles résultent d'un geste articulatoire particulier, cette dépendance est mise entre parenthèses dans l'étude perceptive. Le point de départ du travail est la sélection, par écoute informelle, de prototypes de synthèse possédant un timbre nasal de bonne qualité pour chacune des 3 voyelles nasales du français, ainsi que pour [-i]. Les paramètres retenus pour les voyelles du français sont les suivants:

	FN1	Z1	F'1	F'2
- [ɛ̃]	325	800	875	1600
- [ã]	375	650	800	1000
- [õ]	400	800	900	1300

On remarque immédiatement que FN1 est placé très bas, à une fréquence voisine de ce que nous avons identifié comme une résonance sinusale, alors que Z1 et F'1 sont relativement proches, à une fréquence élevée. F'1 ne sera donc pas très proéminent, sauf si F'2 est bas. Il serait facile de faire la liste des différences, et des correspondances, entre nos mesures acoustiques, ou les fréquences de notre synthèse sans paire pôle - zéro (figure 1.93) et celles-ci. Mais on ne peut tirer de conclusions de la seule constatation de divergences.

L'étude perceptive proprement dite est fondée sur l'analyse de la variation dans le degré de nasalité perçue lorsque les paramètres de synthèse sont modifiés.

Voici notre interprétation des résultats essentiels, reproduit ici figures 1.94 et 1.95. Pour chaque paramètre, la description s'applique à la voyelle dont les autres paramètres sont aux fréquences présentées dans le tableau ci-dessus. Pour [õ] et [ã], la fréquence de F1N peut varier entre 200 et 500 Hz sans diminuer le degré de nasalité. Pour [ɛ̃], il ne doit pas dépasser 300 - 350 Hz. On peut se demander si la conclusion de CHENG & GUERIN est correcte. Ils indiquent que la présence d'un formant vers 300 Hz est nécessaire pour l'obtention d'un timbre nasal. Ne faudrait-il pas plutôt conclure que, tant qu'un premier pic demeure en basse fréquence, il n'empêche pas la perception d'un indice de nasalité porté par les autres caractéristiques de la voyelle ? Si sa fréquence devient élevée, il vient bousculer les indices de nasalité que constituent la position et / ou l'amplitude des formants F'1 et F'2. Pour [õ] et [ã], la proximité de F'1 et F'2 les rend proéminents, et plus résistants à une variation de FN1. Le pic isolé de F'1 de la voyelle [ɛ̃], moins élevé, est plus sensible à la proximité de FN1. La position de F'2 est cruciale pour [ã] et [õ], tout

	F 2										
	1800	1700	1600	1500	1400	1300	1200	1100	1000	900	800
400								ɨ̃ : 10 ɛ̃ : 5			ɨ̃ : 5
500		ɨ̃ : 5	ɨ̃ : 10	ɨ̃ : 3	ɛ̃ : 5	ɨ̃ : 6	œ : 5	ɨ̃ : 15 u		ɨ̃ : 5 ɨ̃ : 5 ɨ̃ : 5	ɨ̃ : 6
600	ɨ̃ : 6	ɨ̃ : 5 ɨ̃ : 5	ɨ̃ : 10 ɨ̃ : 10		ɨ̃ : 5		ɨ̃ : 3	ɨ̃ : 6	ɨ̃ : 10		ɨ̃ : 12
700	ɨ̃ : 5	ɨ̃ : 6 ɨ̃ : 6 ɨ̃ : 6	ɨ̃ : 12 ɨ̃ : 4	ɨ̃ : 12 ɨ̃ : 12	ɨ̃ : 10 ɨ̃ : 5	ɨ̃ : 10 ɨ̃ : 10	ɨ̃ : 10 ɨ̃ : 5	ɨ̃ : 3 ɨ̃ : 3	ɨ̃ : 15 ɨ̃ : 5 ɨ̃ : 5	ɨ̃ : 40 ɨ̃ : 5 ɨ̃	ɨ̃ : 8 ɨ̃ : 4 ɨ̃ : 8
800	ɨ̃ : 10 ɨ̃ : 10 ɨ̃ : 10	ɨ̃ : 20 ɨ̃ : 25 ɨ̃ : 5 ɨ̃	ɨ̃ : 6 ɨ̃ : 12 ɨ̃ : 12	ɨ̃ : 35 ɨ̃ : 6 ɨ̃ : 6 ɨ̃	ɨ̃ : 10 ɨ̃ : 20	ɨ̃ : 4 ɨ̃ : 8	ɨ̃ : 6	ɨ̃ : 6	ɨ̃ : 6 ɨ̃ : 6	ɨ̃ : 35 ɨ̃ : 3 ɨ̃ : 3 ɨ̃	
875		ɨ̃ : 6 ɨ̃ : 12 ɨ̃ : 18	ɨ̃ : 5 ɨ̃ : 10	ɨ̃ : 10		ɨ̃ : 3 ɨ̃ : 3		ɨ̃ : 10	ɨ̃ : 6		

NB: Les résultats sont en %

Table I - X a

Figure 1.93 - Identification de voyelles synthétiques sans indice de nasalité (in LONCHAMP 1978).

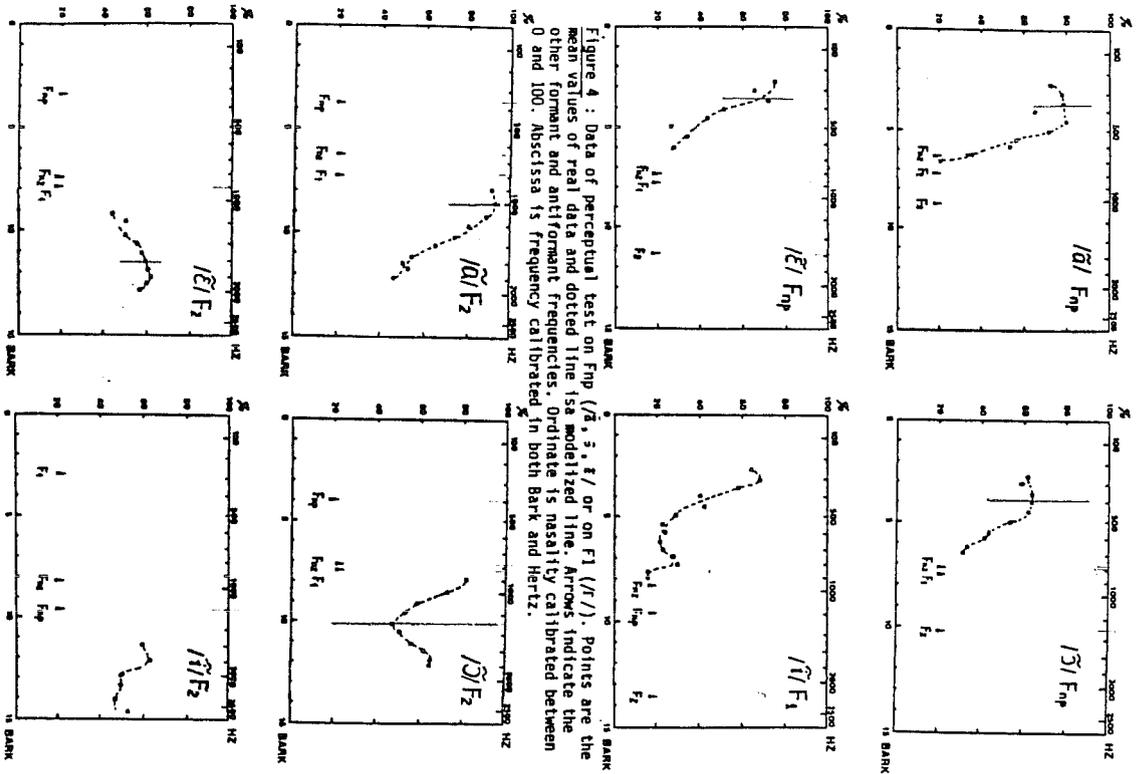


Figure 4 : Data of perceptual tests on Fnp (\tilde{e} , \tilde{i} , \tilde{e} or on F1 (\tilde{i})). Points are the mean values of real data and dotted line 15a modelized line. Arrows indicate the other formant and antiformant frequencies. Ordinate is nasality calibrated between 0 and 100. Abscissa is frequency calibrated in both Bark and Hertz.

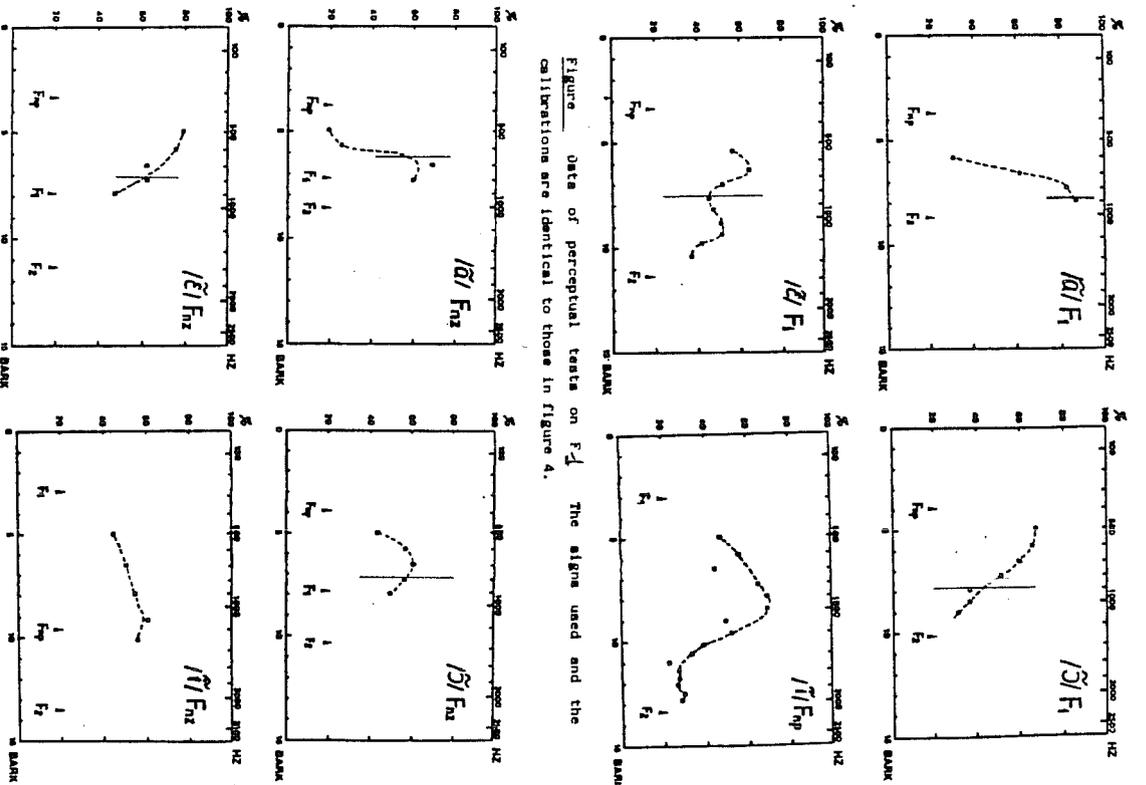


Figure 5: Data of perceptual tests on F2. The signs used and the calibrations are identical to those in Figure 4.

Figures 1.94 et 1.95 - Variations du degré de nasalité en fonction de la fréquence de F₁, Z₁, F'₂ et F'₂ pour 4 voyelles nasales (in CHENG & GUERIN 1987 b).

comme celle de F'1, ce qui concorde avec l'interprétation précédente. La position du deuxième formant du prototype [õ] ne semble pas optimale quant au degré de nasalité, mais il n'est pas certain qu'un ajustement conserverait le timbre constant. On peut faire la même remarque pour F'1. Abaisser si peu que ce soit la fréquence de Z1 détruit toute nasalité. Ceci implique que le niveau exact de F'1, très proche du zéro dans les trois voyelles, est capital. C'est moins vrai pour [õ], ce qui pourrait signifier que F'2 joue un rôle plus important. Pour [ẽ], la nasalité est d'autant plus forte que l'amplitude de FN1 est réduite par le zéro, ce qui doit poser problème à ceux qui suggèrent qu'un FN1 de fréquence très basse est un indice crucial de la nasalité vocalique.

La conclusion de CHENG & GUERIN est que la nasalité résulte d'un équilibre entre deux "masses" acoustiques, l'une englobant le formant glottal et FN1, ici à une fréquence très basse, et l'autre étant constitué de F'1 et F'2 pour les voyelles [ã] et [õ] dans lesquelles ils sont proches, ou de F'1 seul lorsque F'2 est plus lointain.

La difficulté de cette interprétation est que FN1 ne nous semble pas, dans cette expérience, à une position réaliste. Sa position n'est adéquate que pour une résonance sinusale. Son rôle paraît d'ailleurs limité dans le cas de la voyelle [ẽ], où, plus la fréquence de Z1 s'approche de lui et sa prééminence diminue, plus la voyelle est nasale. Notre interprétation du rôle de FN1, dans cette expérience, est que sa présence et sa position est indifférente, tant qu'il ne perturbe pas la fréquence et / ou la position de F'1 et F'2, ou qu'il n'est pas trop proéminent. Si l'on prend au sérieux l'hypothèse d'un équilibre entre deux masses issues d'une intégration large-bande, l'inclusion du formant glottal dans la "masse" inférieure a des conséquences sérieuses. Une variation du niveau du formant glottal pourrait mettre en péril la perception de la nasalité en modifiant l'équilibre. L'expérience devrait être faite, et nous prévoyons l'absence complète d'effet.

Nous sommes en revanche d'accord sur le rôle essentiel de la seconde masse, non en terme d'équilibre avec une masse placée plus bas, mais en terme de position et / ou de forme précise du maximum spectral. Nous croyons que les résultats discutés ici peuvent s'interpréter ainsi : ne touchez pas à F'1 et F'2. Cette conclusion sera discutée plus loin, au paragraphe 12.6.

12.3 - MAEDA (1984)

Dans un premier rapport, MAEDA (1982 b, 1983) concluait que l'indice majeur de la nasalité était un aplatissement du spectre dans la zone où se placent F1 et F2. Quelques mois plus tard, il proposait dans une communication intitulée de manière évocatrice "Une paire de pics spectraux comme corrélats acoustiques de la nasalisation des voyelles" des conclusions assez voisines de celles que tireront plus tard CHENG & GUERIN. Les

voyelles orales et nasales présentent toutes un pic en très basse fréquence (200 - 350 Hz): F1 et le formant glottal pour les fermées, le formant glottal seul pour les ouvertes. Les voyelles nasales ont de plus un pic plus élevé, à plus de 1000 Hz, dont la fréquence est spécifique : F'2, c'est-à-dire F2 décalé sous l'effet du couplage pour les ouvertes et les postérieures, ou le formant nasal pour les antérieures fermées. Cette présentation néglige la présence d'un formant nasal pour les voyelles ouvertes. D'après MAEDA, la fréquence de F'2 est bien un indice de nasalité pour la voyelle [u] : une fréquence plus élevée que la norme la fait percevoir comme voyelle nasale. Quand l'un des deux pics est éliminé par filtrage, la nasalisation s'efface. Enfin, le filtrage des fréquences élevées d'un [i] le fait percevoir comme un [~u].

MAEDA remarque que si les fréquences de ces deux pics sont reportées dans le plan F1 / F2, ils occupent une zone entre [u] et [y], souvent vide de voyelles orales. Le [u], présent en Japonais par exemple, possède un premier formant plus élevé. En utilisant un modèle articulatoire, l'auteur montre que la réalisation d'une voyelle orale dans cette zone exige une grande précision articulatoire. La théorie quantale de STEVENS prévoit donc que cette zone sera évitée. Nous avons montré (LONCHAMP 1978) qu'une simple délabialisation de [u] ne suffit pas pour l'atteindre. La fréquence de F1 augmente alors dans une large proportion, comme le montre l'exemple de la voyelle délabialisée [u] du Japonais. Mais il faut remarquer que le [u] français entre consonnes dentales, comme dans "doute" par exemple (LONCHAMP, sous presse), voit, sans modification de F1, son deuxième formant s'élever par coarticulation. Cette réalisation se place dans cette zone "nasale", vers 300 et 1200 - 1300 Hz, sans qu'il y ait nasalisation.

L'hypothèse de MAEDA est néanmoins séduisante. La réserve ne nous émettons touche à la nécessité d'inclure le pic de basse fréquence. Si un tel pic est présent pour toutes les voyelles, il ne peut être un indice de nasalité. Pour les voyelles ouvertes, une écoute informelle de voyelles nasales filtrées en très basse fréquence n'a pas montré une diminution nette du degré de nasalité. Le timbre des nasales est conservé au téléphone. En revanche, le deuxième pic, que se soit FN1 pour les voyelles antérieures fermées, F'2 pour les postérieures ou le bloc F'1 - F'2 pour les ouvertes paraît déterminant.

12.4 - BOGNAR & FUJISAKI (1986)

Ce travail est de grande importance car c'est le seul, à notre connaissance, qui tente de séparer les effets d'une perturbation du spectre par adjonction de paires pôle - zéro, de ceux qui proviennent d'un placement spécifique des fréquences formantiques. Il est dommage cependant que la première paire pôle - zéro ait été placée à une fréquence très

basse, d'après les résultats d'analyses de voyelles naturelles présentés au paragraphe 10. Le timbre des stimuli synthétiques variait de [ε] à [ɛ̃]. L'intervalle de variation des paramètres de synthèse était limité à celui mesuré pour les voyelles orale et nasale : F1 croissant de 610 à 690 Hz ; F2 décroissant de 1810 à 1680 Hz ; F3 croissant de 2550 à 2760 Hz. La séparation des fréquences des paires pôle - zéro était variée de façon indépendante, la première limitée à 80 Hz (270 et 350 Hz respectivement pour le pôle et le zéro). La deuxième paire pôle - zéro vers 2200 Hz était peu marquée, car séparée également de 80 Hz au maximum. Un seul sujet français et un sujet japonais ont pris part au test de perception. Les réponses du sujet japonais indiquent clairement que seul l'espacement des pôles et des zéros influe sur le degré de nasalité perçue. Le déplacement formantique est sans effet. Les réponses du sujet français varient selon qu'il utilise volontairement un critère phonétique ou phonologique : présence / absence de nasalité ou timbre phonologique [ε] / [ɛ̃]. Les réponses phonétiques pour la séparation pôle - zéro maximale mais sans modification des formants de la voyelle [ε] montrent un taux de nasalité de 90%. Le seul déplacement formantique induit 50 % de réponses nasales. Mais les réponses phonologiques [ɛ̃] n'atteignent 90 % que pour la séparation maximale des pôles et zéros et le déplacement formantique le plus grand.

Ces résultats confirment qu'en français, mais non en japonais, la position formantique est un indice de nasalité. Vraisemblablement, la présence en français de voyelles phonologiques nasales implique l'existence de prototypes perceptifs dont les maxima spectraux sont à des fréquences spécifiques. Dans sa conception, cette expérience à choix forcé suscite la même interrogation que celle de HAWKINS & STEVENS : peut-on être sûr que les réponses à une séparation pôle - zéro sont fondées en totalité sur la perception de la nasalité, et non en partie sur une modification globale du timbre ? Il ne faut pas non plus perdre de vue que cette étude ne porte que sur l'une des voyelles nasales du français.

12.5 - WRIGHT (1986), KAWASAKI (1986)

Le chapitre que WRIGHT (1986) consacre à l'espace perceptif des voyelles nasales couvre une variété de sujets qui rend difficile une présentation complète. Il rappelle tout d'abord que les systèmes phonologiques dont le nombre de voyelles nasales est égal au nombre de voyelles orales sont aussi fréquents que les systèmes où les nasales sont moins nombreuses. Ce sont les voyelles médianes qui manquent le plus souvent, et le français est atypique de ce point de vue (RUHLEN 1973). Synchroniquement, dans les alternances morphophonologiques, les voyelles fermées s'ouvrent, et les ouvertes se ferment (BEDDOR (1982)). Diachroniquement, on connaît bien l'exemple du français et de plusieurs dialectes chinois qui montrent l'ouverture des voyelles fermées sous l'effet de la

nasalisation, mais les exemples attestés sont trop peu nombreux pour que cette ouverture soit posée comme une règle générale.

L'expérience décrite par WRIGHT est une analyse perceptive multidimensionnelle, par la méthode INDSCAL, d'un ensemble de voyelles orales et nasales. Les sujets sont américains. A l'aide d'un système de palatographie optique à capteurs infra-rouge, il a pu s'assurer que les voyelles orales et nasales de même timbre avaient été produites avec une forme identique de la partie antérieure de la langue. Sur les deux premières dimensions, les voyelles se placent conformément au triangle acoustique ou articulatoire. La troisième dimension sépare les voyelles orales et nasales. Il tente ensuite d'interpréter les petites différences dans la position des voyelles orale et nasale de même timbre sur le plan des deux premières dimensions. Nos propres expériences de l'analyse multidimensionnelle des voyelles nous rendent sceptique sur la validité de ces écarts (LONCHAMP 1981). On sait que les jugements subjectifs de distance sont soumis à un effet de plafonnement : les distances longues sont sous-estimées relativement aux distances courtes. Pour un couple de voyelles orale et nasale de même timbre, les distances de chaque élément de la paire aux autres voyelles seront très semblables, impliquant que les deux voyelles de la paire sont très proches dans l'espace. Mais la mesure directe de la distance des voyelles d'une même paire ne montrera pas une distance négligeable. Cette contradiction provoque une distorsion des positions dans l'espace perceptif. Les hypothèses d'orthogonalité des dimensions et de calcul des distances par une équation de forme euclidienne ne sont pas adéquates. Empiriquement, on constate une grande sensibilité à de petites différences dans les estimations de distance, qui ne sont probablement pas significatives. Cette distorsion due à l'effet plafond est trop faible pour remettre en cause l'interprétation des dimensions, mais elle interdit à nos yeux une analyse trop fine des positions relatives dans l'espace multidimensionnel (centralisation de [~i], ouverture de [õ] par exemple). Il ne faut pas s'étonner des contradictions apparentes entre les analyses acoustique et perceptive.

WRIGHT conclût une très longue discussion, faisant appel notamment au concept de spectre auditif global, en indiquant que les timbres des voyelles nasales sont moins contrastés entre eux que les timbres des orales correspondantes. Cette conclusion paraît intuitivement correcte, mais le détail de la démonstration nous semble discutable.

KAWASAKI (1986) teste expérimentalement, sur des voyelles naturelles, l'hypothèse que la nasalisation d'une voyelle est plus perceptible lorsque la consonne nasale adjacente, responsable phonétiquement de la nasalisation, est atténuée ou supprimée. Les résultats corroborent cette hypothèse, qui peut expliquer le paradoxe linguistique suivant : l'opposition phonologique oral / nasal est souvent neutralisée lorsque le contexte comprend une consonne nasale. Cette expérience révèle aussi que les auditeurs compensent perceptivement la présence d'une nasalisation automatique de la voyelle sous l'effet d'une

consonne nasale en la percevant comme très faiblement nasale, sinon comme orale. Le degré de nasalisation perçue n'est donc pas seulement fonction des informations spectrales disponibles.

12.6 - LES INDICES PERCEPTIF DE LA NASALITE VOCALIQUE : QUELQUES REFLEXIONS.

Ce paragraphe propose quelques réflexions sur les indices perceptifs de la nasalité vocalique. L'abaissement du voile du palais provoque une "perturbation" du spectre de la voyelle. On sait que les caractéristiques essentielles de cette perturbation sont en basse fréquence :

- l'apparition d'un formant supplémentaire (FN1) vers 500 - 600 Hz ;
- La présence d'un zéro 100 à 200 Hz plus haut, modifiant éventuellement l'amplitude ou la position d'un formant se trouvant dans sa zone d'influence ;
- une élévation de la fréquence de F'1, que celui-ci soit placé plus haut ou plus bas que FN1, conséquence automatique du couplage ;
- une élévation de même nature de F'2, s'il est de fréquence basse ;
- une légère modification de F'1 et F'2 due à la diminution de l'airé de la partie vélaire du conduit, conséquence de l'abaissement du voile ;
- le présence d'un ou plusieurs pics discrets liés aux résonances sinusales.

Le résultat auditif des simulations de MAEDA (1982 a), ainsi que les résultats détaillés de HAWKINS & STEVENS, indiquent que cette perturbation spectrale, perçue comme une nasalisation, est relativement audible pour les voyelles fermées, mais moins pour les voyelles ouvertes. Pour celles-ci, cette constatation est vraie que FN1 soit placé au-dessus de F'1, comme dans les simulations de MAEDA (1982 a), ou au-dessous comme chez HAWKINS & STEVENS. D'après ces derniers, il n'existe, dans le cas d'une voyelle dont le premier formant est à 770 Hz par exemple, aucune combinaison de pôle de fréquence comprise entre 300 et 500 Hz et de zéro placé 100 à 200 Hz plus haut donnant une franche impression de nasalité. Nos propres résultats (LONCHAMP 1978 ; figure 1.93) montrent en revanche que sans paire pôle - zéro, certaines combinaisons de F1 et F2 conduisent à la perception d'une nasalité pour des voyelles ouvertes ayant un premier formant à 700 ou 800 Hz.

On remarque que ces positions dans le plan F1 / F2 (800 - 1700 et 800 - 1500 pour [ɛ] ;

700 - 900 et 800 - 900 pour [ã]) ne sont pas occupées physiquement par des voyelles orales. Ces fréquences formantiques semblent d'ailleurs difficiles à produire dans les limites qu'impose l'anatomie du conduit vocal. Les prototypes de CHENG & GUERIN, choisis pour leur haut degré de nasalité, possèdent également un premier formant élevé, entre 800 et 900 Hz. Nous avons cru pouvoir résumer notre interprétation de leur expérience par la formule : "ne touchez pas à F'1 et F'2". Si l'on considère que le FN1 de basse fréquence retenu par CHENG & GUERIN n'est pas un indice puissant de nasalité, on peut se demander si l'absence d'indices de nasalité ne conduit pas à une compensation se traduisant par l'exagération de la position de la fréquence de F'1, qui serait un autre indice de nasalisation. Mais il ne faut pas perdre de vue que dans l'expérience de CHENG & GUERIN, la proximité du zéro réduit l'amplitude de F'1. Son rôle premier, pour les voyelles [ã] et [õ], pourrait être d'assurer la proéminence de F'2. Sans cette proéminence de F'2, nos résultats (figure 1.93) ne montrent au mieux que 50 % de réponses nasales.

Examinons le cas de la nasalisation d'une voyelle ouverte de timbre [a] dont les deux premiers formants sont à 700 et 1300 Hz. Un premier formant nasal viendra se placer vers 600 - 650 Hz, même si la résonance propre du conduit nasal est plus basse, car il est automatiquement soumis à une élévation due au couplage. Le premier formant oral, dont l'amplitude peut être réduite en raison de la proximité du premier zéro, se décale vers les fréquences hautes pour atteindre 1000 Hz. Nous faisons l'hypothèse que son importance perceptive est faible, tant en ce qui concerne le degré de nasalisation perçue que le timbre vocalique proprement dit. Le deuxième formant reste à une fréquence voisine de celle qu'il occupe pour la voyelle orale. Le timbre est celui d'une voyelle faiblement nasale possédant des formants à 600 et 1350 Hz. Il est relativement semblable à celui de la voyelle orale d'origine, dont les formants étaient à 700 et 1400 Hz. Lorsque FN1 se place à une fréquence nettement plus basse, comme dans le cas où l'aire du conduit nasal est réduite par gonflement des muqueuses, le premier formant est néanmoins décalé vers les hautes fréquences par l'effet de couplage. Son timbre sera donc également quelque peu différent du timbre de la voyelle orale correspondante.

Il est possible d'augmenter la différence de timbre en modifiant la fréquence de F'2 par un déplacement de la langue sur l'axe antérieur - postérieur. Reculer la langue, ou plus précisément, réduire la fonction d'aire au niveau pharyngal, abaisse la fréquence du deuxième formant. Il peut descendre ainsi jusqu'à 1000 Hz environ et venir se confondre avec F'1. Comme nous l'avons décrit plus haut, cette proximité assure une forte proéminence au(x) pic(s) proches(s) de 1000 Hz. Le premier pic, à 600 - 650 Hz, paraît en conséquence affaibli. Or on sait que "l'affaiblissement de F1", qui touche en fait le plus souvent le pic de 600 Hz, c'est-à-dire FN1 d'après notre interprétation, est un indice de nasalité. La voyelle nasale ainsi articulée s'éloigne de [a] et se rapproche de [ɔ̃]. Elle s'en

distingue néanmoins par un premier formant plus élevé et un timbre nasal. Modifier le degré d'aperture, en revanche, n'a pas d'effet si FN1 reste le premier pic du spectre : FN1 ne dépend pratiquement que de la forme du conduit nasal. Une telle modification, dans la mesure où elle n'affecte pas F'2, ne conduira qu'à une variation de la fréquence et de l'amplitude du pic intermédiaire (F'1) dont nous faisons l'hypothèse qu'il est perceptivement de faible importance.

Cette voyelle nasale se caractérisera donc à la fois par un timbre nasal et un timbre vocalique spécifique, lié à une position particulière des formants. Cette interprétation fait jouer un rôle important à la distance perceptive. L'augmentation de celle-ci "exige", en quelque sorte, le recul de la langue pour assurer une différenciation suffisante de [a] et de [ã].

On peut appliquer le même scénario au cas de [ɛ̃]. Un simple abaissement du voile pour une voyelle de timbre [ɛ], dont les formants sont à 550 et 1700 Hz, ne modifie pas la fréquence du premier pic, qui peut être soit FN1, soit F'1. La fréquence de F'2 reste voisine de celle de F2. Le timbre de la voyelle orale et de la voyelle nasale ne se distingue que par les indices propres de nasalité. Un recul de la langue modifiera la fréquence de F'2, assurant ainsi une plus grande distance perceptive entre les voyelles orale et nasale.

Pour les voyelles antérieures fermées, un formant nasal apparaît sur le flanc droit du premier formant. Le couplage augmente également la fréquence de F'1. Les tests perceptifs semblent montrer que ces indices sont suffisants pour assurer un degré important de nasalité.

Pour les voyelles postérieures, comme [ɔ̃], F'1 est décalé par couplage vers les hautes fréquences. Il se place vers 450 Hz, donc sous FN1 qui demeure à 600 Hz. Le zéro associé à FN1 réduit l'amplitude de F'2. Le second pic spectral a donc une fréquence très basse. Cette combinaison d'un F'1 relativement élevé et d'un second pic placé très bas assure à [ɔ̃] un timbre caractéristique.

Cette analyse des indices perceptifs de la nasalité, qui n'est à l'heure actuelle qu'une hypothèse, peut, et doit être testée expérimentalement. Pour les voyelles ouvertes, la vérification doit porter principalement sur le faible rôle perceptif du pic intermédiaire du groupe de formants de basse fréquence. Ce pic intermédiaire, vers 900 - 1000 Hz, est soit le premier formant oral décalé, soit le formant nasal. Il est encadré de deux formants dont les fréquences sont de 600 et 1500 Hz pour [ɛ̃], par exemple. Notre interprétation suggère qu'une modification de son amplitude, tant qu'elle reste inférieure ou égale à celle des deux autres pics, et / ou une modification de sa fréquence, n'apporte ni une forte augmentation du degré de nasalité perçue, ni un changement radical du timbre. Ces manipulations ne doivent cependant pas modifier l'amplitude importante de F'2. Cette prééminence, liée à la proximité de F'1, pourrait être l'indice de nasalité que l'on connaît plutôt sous le nom

d'affaiblissement de F1. Techniquement, l'utilisation d'une synthèse par addition d'harmoniques permettrait un contrôle indépendant des paramètres de fréquence et d'amplitude.

Ce test doit impérativement comporter une évaluation de degré de nasalité par rapport à une voyelle orale et une voyelle nasale. Il pourrait prendre la forme d'une estimation du degré de nasalité sur une échelle subjective, la présentation du stimulus-test se faisant en position centrale d'un triplet comportant une voyelle orale et une voyelle nasale. Ce test pourrait avantageusement être étendu à tous les indices potentiels de la nasalité. Mais il convient de veiller scrupuleusement à ce que les modifications présentées restent dans des limites compatibles avec les mesures acoustiques ou les prédictions d'un modèle articulatoire réaliste, particulièrement en ce qui concerne la forme du conduit nasal.

Notre interprétation du rôle des indices de nasalité vocalique pêche peut-être par l'importance qu'elle accorde à la forme spectrale des voyelles nasales françaises. Pour les voyelles ouvertes notamment, elle implique que dans les langues possédant des voyelles nasales phonologiques, le geste articulatoire lingual sera modifié afin d'accroître l'écart fréquentiel entre le deuxième formant des voyelles orales et nasales de même timbre. Cette prédiction peut être testée en comparant les fréquences du deuxième formant des voyelles orales et nasales, ou en examinant directement les positions articulatoires sur des clichés radiologiques. Si cette prédiction se révèle fautive, il conviendra de décider si notre analyse est totalement incorrecte, ou si le français possède une "marge de sécurité" importante, assurant des distances perceptives plus grandes que les valeurs qu'imposent la seule nécessité de distinguer le timbre des voyelles orales et nasales.

12.7 - CONCLUSIONS

Cette dernière partie de notre étude a donc proposé un cadre théorique rendant compte de plusieurs résultats expérimentaux concernant la perception des indices de nasalité vocalique. Schématiquement, la perception du caractère nasal d'une voyelle nous semble dépendre à la fois de caractéristiques propres à la nasalité, mais également de la perception d'un timbre particulier lié à une position spécifique des formants. Dans le cas des voyelles ouvertes, une modification de la position articulatoire de la langue sur l'axe antérieur - postérieur, qui contrôle la fréquence du deuxième formant, nous semble un élément décisif pour assurer une distance perceptive suffisante entre la voyelle orale d'origine et la voyelle nasale.

Cette proposition conduit à des prédictions pouvant être réfutées expérimentalement. On connaît encore mal l'importance relative des indices de nasalité, comme l'affaiblissement du premier pic spectral, ou la présence, dans certains cas, de trois maxima spectraux dans

la zone couvrant les deux premiers formants des voyelles orales. Nous suggérons enfin une forme de test auditif permettant d'établir l'importance relative de ces indices spécifiques, par une comparaison simultanée de stimuli contrôlés avec des voyelles orales et nasales.

Nous serions pleinement satisfait si ces quelques propositions pouvaient constituer un point de départ pour d'autres travaux.

ANNEXE A : FONCTIONS D'AIRE ORALES

N° section Conduit oral (cm²) / longueur (cm)

	O1[η]	O2 [η]	O3 [a]	O4 [ɛ]
lèvres				
1			5.0 / 1.09	6.0 / 1
2			5.7 / 1.09	5.5 / 1
3			8.1 / 1.09	5.0 / 1
4			8.1 / 1.09	4.5 / 1
5			7.7 / 1.09	4.5 / 1
6			7.3 / 1.09	4.0 / 1
7			5.2 / 1.09	3.0 / 1
8			3.5 / 1.09	2.0 / 1
9			2.5 / 1.09	1.5 / 1
	-----	-----		-----
10	4.0 / 1	5.5 / 1.09	2.4 / 1.09	3.0 / 1
11	5.0 / 1	6.0 / 1.09	2.3 / 1.09	3.0 / 1
12	5.0 / 1	6.0 / 1.09	2.2 / 1.09	3.0 / 1
13	5.0 / 1	6.0 / 1.09	1.8 / 1.09	3.5 / 1
14	4.0 / 1	6.0 / 1.09	1.0 / 1.09	4.0 / 1
15	3.0 / 1	4.0 / 1.09	3.0 / 1.09	4.0 / 1
16	1.5 / 1	3.0 / 1.09	2.0 / 1.09	2.0 / 1
17	2.3 / 1	3.3 / 1.09	2.5 / 1.09	2.5 / 1
larynx				
longueur	17 cm	18.5 cm	18.5 cm	17 cm

La ligne pointillée indique l'emplacement du branchement nasal.

N° section Conduit oral (cm²) / longueur (cm)

	O5 [ā]	O6 [ō]
lèvres	5.0 / 1.06	1.5 / 1.1
1	6.0 / 1.06	4.0 / 1.1
2	8.0 / 1.06	10.0 / 1.1
3	8.0 / 1.06	10.0 / 1.1
4	8.0 / 1.06	10.0 / 1.1
5	7.0 / 1.06	7.5 / 1.1
6	5.5 / 1.06	4.5 / 1.1
8	2.5 / 1.06	2.0 / 1.1
9	1.0 / 1.06	0.8 / 1.1
	-----	-----
10	1.5 / 1.06	1.2 / 1.1
11	1.5 / 1.06	1.5 / 1.1
12	1.5 / 1.06	1.8 / 1.1
13	1.7 / 1.06	2.5 / 1.1
14	2.0 / 1.06	3.0 / 1.1
15	2.8 / 1.06	3.5 / 1.1
16	2.0 / 1.06	2.3 / 1.1
17	2.5 / 1.06	2.5 / 1.1

larynx

longueur 18 cm

19 cm

La ligne pointillée indique l'emplacement du branchement nasal.

ANNEXE B : FONCTIONS D'AIRE NASALES

N° section Conduit nasal (cm²) / longueur (mm)

Référence	N1	N2	N3
1 narines	1.4 / 0.7	1.5 / 0.7	0.3 / 0.7
2	1.7 / 1	1.5 / 1.0	0.3 / 1.0
3	2.2 / 1	1.5 / 1.0	2.2 / 1.0
4	2.3 / 2.4	1.5 / 2.4	2.3 / 2.4
5	1.9 / 2.4	1.5 / 2.4	1.9 / 2.4
6	2.0 / 4	1.5 / 4.1	2.0 / 4.1
7	2.5 / 4	2.0 / 4.1	2.5 / 4.1
8	3.4 / 4.7	3.4 / 4.8	3.4 / 4.8
9	4.5 / 4.7	4.5 / 4.8	4.5 / 4.8
10	5.5 / 4.7	5.5 / 4.8	5.5 / 4.8
11	6.4 / 4.7	6.0 / 4.8	6.4 / 4.8
12	6.9 / 9.5	6.0 / 9.8	6.9 / 9.8
13	7.5 / 8.2	6.0 / 8.4	7.5 / 8.4
14	7.4 / 5.5	6.0 / 5.7	7.4 / 5.5
15	6.4 / 5.5	6.0 / 5.7	6.4 / 5.7
16	5.6 / 3.5	5.6 / 3.6	5.6 / 3.6
17	4.9 / 3.5	4.9 / 3.6	4.9 / 3.6
18	4.2 / 4	4.2 / 4.1	4.2 / 4.1
19	3.5 / 4	4.0 / 4.1	3.5 / 4.1
	-----	-----	-----
longueur	78 mm	80 mm	80 mm

20	3.5 / 3.3	4.5 / 4.95	3.5 / 4.95
21	4.3 / 3.3	5.3 / 4.95	4.3 / 4.95
22	4.5 / 3	5.5 / 4.5	4.5 / 4.5
23	4.1 / 3	5.1 / 4.5	4.1 / 4.5
24	4.2 / 3.3	5.2 / 4.95	4.2 / 4.95
25	4.8 / 3.3	5.8 / 4.95	4.8 / 4.95
26	4.2 / 5.4	5.2 / 8.1	4.2 / 8.1
27 voile	2.4 / 5.4	5.2 / 8.1	2.4 / 8.1
longueur	30 mm	45 mm	45 mm
longueur totale	108 mm	125 mm	125 mm

la ligne pointillée sépare les fosses nasales proprement dites (des narines à l'extrémité postérieure du septum nasal) du conduit rhinopharyngal. LES AIRES INDIQUEES SONT LA SOMME DES AIRES DES DEUX FOSSES NASALES.

ANNEXE C : SINUS

SINUS 1 - volume : 24.8 cm³

longueur du col : 0.6 cm

diamètre du col : 2.4 mm

SINUS 2 - volume : 20.8 cm³

longueur du col : 0.4 cm

diamètre du col : 2.5 mm

SINUS 3 - volume : 20.8 cm³

longueur du col : 0.32 cm

diamètre du col : 3 mm

BIBLIOGRAPHIE DE LA PREMIERE PARTIE

SIGLES des REVUES ou CONFERENCES

- Bull. Inst. Phon. Grenoble : Bulletin de l'Institut de Phonétique de Grenoble - France
 Bull. L.C.P. : Bulletin du Laboratoire de la Communication Parlée - Grenoble - France
 ICASSP - IEEE : (Proceedings of the) International Conference on Acoustics, Speech and
 Signal Processing - The Institute of Electrical and Electronics Engineers. - USA
 JASA : Journal of the Acoustical Society of America - USA
 JASJap : Journal of the Acoustical Society of Japan - Japon
 J.E.P. : X èmes Journées d'Etude sur la Parole du Groupe "Communication Parlée" du
 GALF (Groupement des Acousticiens de Langue Française, maintenant SFA,
 Société Française d'Acoustique) - France
 JSHR : Journal of Speech and Hearing Research - USA
 JSHDis : Journal of Speech and Hearing Disorders- USA
 J. of Ph. : Journal of Phonetics - London - G.B.
 MIT - QPR- RLE : Massachusetts Institute of Technology - Quaterly Progress Reports -
 Research Laboratory in Electronics - U.S.A.
 MIT - SCG - WP : Massachusetts Institute of Technology - Speech Communication Group
 - Working papers - U.S.A.
 STL-QPSR-RIT : Speech Transmission Lab. - Quaterly Progress Report - Royal Institute
 of Technology - Stockholm - Suède.
 Tr. Inst. Phon. Strasbourg : Travaux de l'Institut de Phonétique de Strasbourg - France
 Verbum - Tr. Inst. Phon. Nancy II : Travaux de l'institut de Phonétique de Nancy II -
 France.

- ANANTHAPADMANABHA T.V (1984) "Acoustic analysis of voice source dynamics",
STL-QPSR-RIT, 2-3/84, 1 - 24.
 ANANTHAPADMANABHA T.V. & FANT G. (1982) "Calculation of true glottal flow
 and its components", STL-QPSR-RIT, 1/82, 1 - 30.

- ATAL B.S., CHANG J.J, MATHEWS M.V. & TUKEY J.M. (1978) " Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique", JASA, 63(5), 1535 - 1555.
- AUST R. & DRETTNER B. (1975) "The patency of the maxillary ostium in relation to body posture", Acta Otolaryngologica, 80, 444 - 446.
- BADIN P. & FANT G. (1984) "Notes on vocal tract computation", STL-QPSR-RIT, 2-3 /1984 53 - 107.
- BEDDOR P.S. (1982) Phonological and phonetic effects of nasalization on vowel height, Ph.D., U. of Minnesota.
- BEDDOR P.S. & STRANGE W. (1982) "Cross-Language study of perception of the oral - nasal distinction", JASA, 71, 1551-1561
- BELL-BERTI F. (1983) "Velar position, port size and vowel spectra", Actes du 11ème Congrès International d'Acoustique, Paris, vol. 4, 19 - 21.
- BENQUEREL A.P & LAFARGUE A. (1981) "Perception of vowel nasalization in French", J. of Ph., 9(3), 309-322.
- BERGERON R.T., OSBORN A.G. & SOM P.M. (1984) Head and neck imaging - Excluding the brain , C.V. Mosby & Co. , St. Louis - Toronto.
- BOGNAR E. (1980) Préliminaires anatomiques à l'étude phonétique des mouvements mandibulaires, Tr. Inst. Phon. Grenoble.
- BOGNAR E. & FUJISAKI H. (1986) "Analysis, synthesis and perception of the french nasal vowels", Proc. ICASSP-IEEE, Tokyo, 1601-1604.
- BJUGGREN B. & FANT G. (1964) "The nasal cavity structure", STL-QPSR-RIT, 4, 5-7.
- BOTHOREL A., SIMON P., WIOLAND F. & ZERLING J.P. (1986) Cinéradiographie des consonnes et voyelles du français, Tr. Inst. Phon. Strasbourg.
- BUTCHER A. (1976) "The influence of the native language on the perception of vowel quality", Arbeitsberichte Inst. Phonet., Universität Kiel, 6, 1 - 137.
- CARRE R. (1981) "Vocal source - vocal tract coupling : effects on the vowel spectrum", Proc. FASE, Venise, vol. 1, 115 - 118.
- CARRE R. & GUERIN B. (1985) "Vowel spectrum variability due to the source" in GUERIN B. & CARRE R. (eds.), Symposium franco-suédois sur la parole, GALF (SFA), Vol. 1, 1 - 21.
- CARRE R. & TRIBOLET J. (1974) "Acoustic characteristics of vowel nasalization", JASA, 55, 520(A).
- CHARPENTIER F. (1986) "Fonctions de sensibilité d'un modèle dissipatif", Bull. Inst. Phon. Grenoble, 15, 1 - 33.
- CHENG Y.M. & GUERIN B. (1987a) "Study of the source-filter interactive concept and

- its application to male and female speech synthesis", Bull. L.C.P., 1A, 29 - 66.
- CHENG Y. & GUERIN B. (1987b) "Nasal vowel study : formant structure, perceptual evaluation and neural representation in a model of the peripheral auditory system", Bull. L.C.P., Grenoble, 1, 91-132.
- CLEMENTE C.D. (1975) A regional atlas of the human body, Urban & so., 2ème ed.
- CRANEN B. & BOVES L; (1987) "On subglottal formant analysis", JASA, 81(3), 734 - 746.
- DELATTRE P. (1954) "Les attributs acoustiques de la nasalité vocalique et consonantique", Studia linguistica, 8, 103-109.
- DELATTRE P. (1965) "La nasalité vocalique en français et en anglais", French Review, 39, 92-109.
- DELATTRE P. (1968) "Divergences entre nasalité vocalique et consonantique en français", Word, 24, 64-72.
- FANT G. (1956) "On the predictability of formant levels and spectrum envelopes from formant frequencies", For Roman Jakobson, Mouton, The Hague, 109 - 120.
- FANT G. (1960) Acoustic theory of speech production, Mouton, The Hague, 2 ème ed., 1970.
- FANT G. (1979a) "Glottal source and excitation analysis", STL-QPSR-RIT, 1/79, 85 - 107.
- FANT G. (1979b) "Vocal source analysis - A progress report", STL-QPSR-RIT, 3-4/79, 31 - 53.
- FANT G. (1980) "Voice source dynamics", STL-QPSR-RIT, 2-3/80, 17 - 37.
- FANT G. (1981) "The source filter concept in voice production", STL-QPSR-RIT, 1/81, 21 - 37.
- FANT G. (1983) "Preliminaries to analysis of the human voice source", MIT-SCG-WP, 3, 17 - 40 (aussi in STL-QPSR-RIT, 4/1982).
- FANT G. (1985) "The vocal tract in your pocket calculator" in FROMKIN V. (ed.), Phonetic Linguistics, Ac. Press. N.Y., 55-77.(aussi dans STL-QPSR-RIT 2/1985 1-19)
- FANT G. & ANANTHAPADMANABHA T.V. (1982) "Truncation and superposition", STL-QPSR-RIT, 2-3/82, 1 - 17.
- FANT G., ISHIZAKA K., LINDQVIST J. & SUNDBERG J. (1972) "Subglottal formants", STL-QPSR-RIT, 1, 1-12.
- FANT G., LILJENCANTS J. & LIN Q.G. (1985) " A four-parameter model of glottal flow", STL-QPSR-RIT, 4/85, 1 -13.
- FANT G., LIN Q.G. & GOBL C. (1985) "Notes on glottal flow interaction", STL-QPSR-RIT, 2-3/85, 21 - 45.

- FENG G. (1986) Apport de la modélisation au traitement du signal de parole : le cas des voyelles nasales et la simulation des pôles et des zéros, Thèse de doctorat, U. de Grenoble. (aussi dans Bull. Inst. Phon. Grenoble, 16, 1987, 1 - 102)
- FENG G., ABRY C. & GUERIN B. (1985) "How to cope with nasal vowels : some acoustic boundary poles", Actes du Symposium Franco-Suédois sur la parole, in GUERIN B. & CARRE R. (eds), Grenoble, vol. 2, 123 - 142.
- FLANAGAN J.L. (1958) "Some properties of the glottal sound source", JSHR, 1, 99 - 116.
- FUJIMURA O. (1960) "Spectra of nasalized vowels", MIT-OPR-RLE, 58, 214-218.
- FUJIMURA O. (1962) "Analysis of nasal consonants", JASA, 34, 1865 - 1875.
- FUJIMURA O. & LINDQVIST J. (1964a) "The sinewave response of the vocal tract", STL-QPSR-RIT, 1, 1-7.
- FUJIMURA O. & LINDQVIST J. (1964b) "Experiments on vocal tract transfer", STL-QPSR-RIT, 3, 1-7.
- FUJIMURA O. & LINDQVIST J. (1971) "Sweep-tone measurements of vocal tract characteristics", JASA, 49, 541-558.
- FUJISAKI H. & LJUNGQVIST M. (1986) "Proposal and evaluation of models for the glottal source waveform", ICASSP-IEEE, 1605 - 1608.
- GOERTZEL G. (1960) "Fourier analysis" in RALSTON A. & WILF H.S., Mathematical methods for digital computers, Vol. I, Wiley, N.Y., 258 - 262.
- GUERIN B. (1978) Contribution aux recherches sur la production de la parole - étude du fonctionnement de la source vocale, Doctorat d'Etat, INP - U. Scientifique et Médicale, Grenoble
- GUERIN B. & MRAYATI M. (1977) "Nasal vowels study : sensitivity functions", in CARRE R., DESCOUT R. & WAJSKOP M. (eds.) Actes du symposium 'Modèles articulatoires et Phonétique', GALF, 137 - 146.
- HAMADA M. (1983) "Nasal vowel identification using L.P.C. - based formant analysis", MIT-SCG-WP, 3, 41-54.
- HATTORI S., YAMAMOTO K. & FUJIMURA O. (1958) "Nasalization of vowels in relation to nasals", JASA, 30, 267-274.
- HAWKINS S. & STEVENS K.N. (1985) "Acoustic and perceptual correlates of the non-nasal - nasal distinction for vowels", JASA, 77, 1560-1575.
- HOUSE A.S. & STEVENS K.N. (1956) "Analog studies of the nasalization of vowels", JSHDis, 21, 218-232.
- JOOS M. (1948) "Acoustic Phonetics", Language, 24, 1 - 136.
- KAWASAKI H. (1986) "Phonetic explanation for phonological universals : the case of distinctive vowel nasalization", in OHALA J.J. & JAEGER J.J., Experimental

- phonology, Academic Press, N.Y., 81 - 103.
- KINSLER L.E. & FREY A.R. (1960) "Fundamentals of acoustics", 2ème ed., Wiley, N.Y.
- KONECZNA H. & ZAWADOWSKI W. (1951) Przekroje Rentgenographiczne Glosek Polskich, Varsovie
- LATCHAW R.E. (1985) Computed tomography of the head, neck and spine, Year Bk Med. Pub., Chicago.
- LHERM B. (1984) Réalisation d'un outil de simulation harmonique et de synthèse temporelle du type analogue du conduit vocal - Etude des voyelles nasales, Thèse D.I., I.N.P. de Grenoble.
- LINDQVIST-GAUFFIN J. & SUNDBERG J. (1976) "Acoustic properties of the nasal tract", Phonetica, 33, 161-168. (Version préliminaire in STL-QPSR-RIT 1/72, 13-17)
- LINTHORST P. (1973) Les voyelles nasales du français : Etudes phonétique et phonologique, Groningue.
- LONCHAMP F. (1978) Recherches sur les indices perceptifs des voyelles orales et nasales, Thèse de 3ème cycle, Université de Nancy II.
- LONCHAMP F. (1979) "Analyse acoustique des voyelles nasales françaises", Verbum - Travaux Institut Phonétique Nancy II, 2(1), 9 - 54.
- LONCHAMP F. (1981) "Analyse multidimensionnelle de l'espace perceptif des sons vocaliques", 11èmes J.E.P. - GALE, Montréal, 406 - 418.
- MAEDA S. (1982 a) "The role of the sinus cavities in the production of nasal vowels", Proc. ICASSP - IEEE, Paris, 911-914.
- MAEDA S. (1982 b) "Acoustic correlates of vowel nasalization : a simulation study", JASA, 72, S102.
- MAEDA S. (1983) "Acoustic cues of vowel nasalization : a simulation study", Rapport d'Activité, CNET.
- MAEDA S. (1984) "Une paire de pics comme corrélat acoustique de la nasalisation des voyelles", 13èmes J.E.P. du G.A.L.F., Bruxelles, 223-224.
- MERLIER R. (1984) Rayonnement externe des conduits oral et nasal, Thèse D.I., INP de Grenoble
- MRAYATI M. (1975) "Etude des voyelles nasales françaises", Bull. Inst. Phon. Grenoble, 4, 1-26.
- MRAYATI M. (1976) Contribution aux études sur la production de la parole, Thèse de doctorat (Sciences physiques), Grenoble
- PALACIOS E., FINE M. & HAUGHTON V.M. (1980) Multiplanar anatomy of the head and neck for computed tomography, Chichester, N.Y.

- PIALOUX P., VALTAT M., FREYSS T. & LEGENT P. (1975) Précis d'orthophonie, Masson.
- RUHLEN M. (1973) "Nasal vowels", Working papers in language universals : Stanford University, 12, 1 - 36.
- SMITH S. (1951) "Vocalization and added resonances", Folia phoniatica, 3, 165-169.
- SOM P.M. (1984) "The paranasal sinuses", in BERGERON & al., 1-142.
- STRENGER F. (1969) Les voyelles nasales françaises, Lund.
- TAKEUCHI S., KASUYA H & KIDO K. (1975) "On the acoustic correlate of nasality", J.A.S.Jap., 31(5), 298-309 (en japonais).
- TAKEUCHI S., KASUYA H & KIDO K. (1977) "A study on the effects of nasal and paranasal cavities on the spectra of nasal sounds", JASJap, 33(4), 163-172 (en japonais).
- WRIGHT J.T. (1986) "The behavior of nasalized vowels in the perceptual vowel space", in OHALA J.J. & JAEGER J.J., Experimental phonology, Academic Press, N.Y., 45 - 67.
- ZERLING J.P. (1984) "Phénomènes de nasalité et de nasalisation vocaliques : étude cinéradiographique pour deux locuteurs", Tr. Inst. Phon. de Strasbourg, 16, 241-266.



DEUXIEME PARTIE

LA MODIFICATION DU TIMBRE PAR LA FREQUENCE FONDAMENTALE

-

Contribution à l'étude de la
représentation interne du premier
formant

0 - INTRODUCTION : LA REPRESENTATION INTERNE DES FORMANTS

Cette deuxième partie est consacrée à une étude expérimentale de la modification du timbre vocalique qui accompagne un changement de fréquence fondamentale. En un mot, le phénomène à décrire et à interpréter est le suivant : lorsque la fréquence fondamentale d'une voyelle ouverte augmente, le timbre passe à celui d'une voyelle plus fermée, même si les fréquences formantiques restent les mêmes.

Cette observation est l'une de celles qui peuvent jeter un doute sur la validité de la notion de formant, non comme entité physique définie ici comme fréquence d'un pic de résonance observée sur un spectre, mais comme paramètre perceptif du timbre.

Le but de cette étude est donc d'apporter une contribution à l'étude de la représentation perceptive des maxima du spectre, plus précisément du premier d'entre eux (F1). Nous utiliserons le terme formant auditif pour cette représentation interne. Dans ce travail nous ne nous prononcerons pas sur la forme que peut prendre cette représentation interne. Le terme formant auditif doit donc être entendu comme une désignation commode pour l'un des indices perceptifs participant à la formation d'une sensation définie de timbre vocalique. Sans autre précision, il s'agit de l'indice qui corrèle avec la position du premier pic spectral des voyelles non nasales. Il est sans doute inutile de souligner l'intérêt qu'il y a à cerner avec plus de précision un concept aussi central du domaine de la perception humaine des sons linguistiques.

0.1 - FORMANTS ET FREQUENCE FONDAMENTALE

Physiquement, la position fréquentielle des pôles d'un système, qui est numériquement très proche de la fréquence formantique lorsque, comme dans la parole, la largeur de bande est faible, ne dépend pas des fréquences d'excitation : fréquence fondamentale, notée F_0 , et harmoniques, multiples entiers de celle-ci. La fonction de transfert, sur laquelle on observe des maxima aux fréquences formantiques, dépend, on le sait, de la forme du conduit vocal, et non des paramètres laryngés, en dehors de très faibles modifications dues au couplage supra-glottal. Une modification de la fréquence fondamentale ne conduit donc pas d'elle-même à un changement des fréquences de résonance observées, dans la limite évidente où les harmoniques sont en nombre suffisant pour définir les maxima du spectre : une fréquence fondamentale de 1000 Hz, comme dans les notes les plus aiguës d'une soprano, ne permet pas d'observer les pics d'un [u] par exemple, vers 350 et 1000 Hz. Ce cas limite n'est pas celui que nous envisageons ici, où la plage de variation de F_0 est comprise entre 100 et 300 Hz. Les harmoniques sont donc suffisamment nombreuses pour "définir" sans ambiguïté les formants au plan physique.

0.2 - LE FORMANT COMME INDICE PERCEPTIF : DISCUSSION DES CRITIQUES "CLASSIQUES".

Il existe d'autres arguments pour mettre en doute ou même nier le rôle perceptif des formants, comme le fait d'une manière un peu provocatrice BLADON (1982b). Sans vouloir entrer dans une discussion très détaillée, comme dans SCHWARTZ (1987), on peut répondre ce qui suit à la démonstration de BLADON, qui critique d'abord le caractère réductionniste d'un intérêt exclusif porté aux fréquences formantiques. Cette attitude conduirait à négliger leur amortissement, indice important des voyelles et consonnes nasales, les creux d'énergie des consonnes [l] et des fricatives, ou l'amplitude de F_0 pour les voyelles murmurées. On peut répondre que le succès de la synthèse formantique, même pour les voyelles nasales comme il a été vu dans la première partie, l'absence complète de sensibilité à l'amplitude des zéros spectraux, la sensibilité réduite à l'amplitude des formants (KAKUSHO & al. 1971, KLATT 1982) et à la pente globale du spectre, ainsi que la prise en compte du "formant" glottal, indiquent que le système perceptif est sensible au premier chef aux zones de forte intensité acoustique que sont les formants. Il conviendrait certainement de séparer plus nettement dans nos discussions le concept physique de pôle, c'est-à-dire de fréquence de résonance physique d'un système, de celui de formant, zone d'énergie intense d'extension plus ou moins large et visible dans un spectre. La confusion n'est à la rigueur permise que dans les voyelles "pures" à F_0 bas, où chaque pôle manifeste son existence sous la forme d'un formant. Il est vrai que l'amplitude réduite des formants est un indice du caractère consonantique sonant, mais les positions fréquentielles de ceux-ci, au moins autant que leur amplitude, déterminent leur identification précise. Quant à la pente globale du spectre de la barre des occlusives, aucun défenseur de la notion de formant n'a songé à l'analyser dogmatiquement en ces termes! Mais, dans ce cas également, la position des maxima d'énergie joue un rôle aussi important que la pente globale dans le processus d'identification.

En ce qui concerne l'argument d'"indéterminabilité", qui souligne que les formants ne sont pas faciles à mesurer à l'aide de nos instruments et qu'ils présentent des discontinuités et des fusions, ces difficultés de suivi ou de mesure n'empêchent pas les spectres d'avoir des zones de forte énergie, qui contrastent avec les zones de faible énergie! Il est indéniable que le problème de la "mesure" d'un formant par le système auditif se pose bien, et c'est justement l'ambition de ce travail que de l'étudier à partir d'une difficulté potentielle.

La dernière série de critiques porte sur l'adéquation perceptive de la notion de formant. La distance perceptive entre les voyelles ne corrèle pas avec les distances de Minkowski (distances de type "euclidien", mais avec un exposant quelconque) dérivées de leur position formantique. BLADON, citant LONCHAMP (1981) avec approbation, confirme à partir de

ses propres expériences (BLADON 1981), "l'effet plafond" de sous-estimation des distances longues. Comme SCHWARTZ, l'argument de la distance nous paraît faible. L'utilisation du modèle de BLADON & LINDBLOM (1981) conduit effectivement à une bonne prédiction des distances subjectives entre des voyelles très différentes. Après transformation fréquentielle en échelle Bark, on applique au spectre acoustique une convolution avec un filtre "auditif" de large bande, puis une transformation de l'intensité en phones / sones. L'échelle Bark sera présentée au § 2.10. Mais comme le note SCHWARTZ avec justesse, les mesures de poids phonétique de KLATT (1982) indiquent que les formants permettent une bien meilleure estimation des distances phonétiques. Il ne précise pas pourtant qu'il s'agit, dans ce cas, de mesures locales, c'est-à-dire s'appliquant à des stimuli de timbres (très) voisins. Un modèle à large bande d'intégration échoue dans ce cas, ce que nous confirmerons plus loin pour nos données. Nous sommes, comme SCHWARTZ, contre l'hypothèse à priori de l'unicité des représentations internes, mais il n'en découle pas que nous acceptons l'existence d'un mécanisme d'intégration spectrale large bande qui produirait un pic unique, appelé F'2, dans la bande couvrant les formants F2 à F4. Pour ne pas alourdir cette introduction, nous renvoyons en annexe une discussion d'une expérience de SCHWARTZ (1987) tentant d'offrir une preuve de l'existence de F'2. Il ne nous paraît pas que le concept de F'2 s'impose de manière irréfutable. Dans le cas où la représentation interne prendrait plusieurs formes, il faut préciser que nous nous intéresserons ici à la représentation la plus directement issue du codage des informations auditives par l'oreille interne.

0.3 - OBJECTIF DE L'ETUDE : LA REPRESENTATION PERCEPTIVE DES MAXIMA SPECTRAUX

Notre étude a donc pour objectif de préciser la représentation perceptive des maxima spectraux. Une meilleure compréhension de l'interaction entre les modifications de la fréquence fondamentale et les variations de timbre nous semble un élément décisif pour concevoir un modèle adéquat. Mais nous ne négligerons pas d'intégrer à notre discussion tout ce qu'apporte l'étude directe des formants auditifs par les méthodes de masquage, domaine dans lequel nous avons réalisé une expérience préliminaire qui sera décrite.

Le paragraphe suivant précise l'organisation de cette partie de notre travail et donne en quelques phrases un aperçu du résultat principal.

1 - ORGANISATION

Précisons l'organisation très simple des matériaux de ce chapitre. Dans un premier temps,

nous présenterons en détail l'ensemble des travaux comprenant des expériences ou des commentaires sur la modification du timbre vocalique liée à une variation de la fréquence fondamentale, en insistant sur la qualité des preuves expérimentales. Nous décrirons et commenterons également les études qui traitent de la modélisation des pics auditifs. Ces derniers comprennent notamment une expérience préliminaire de mesure du spectre auditif par masquage proactif que nous avons conduite en 1985. Puis nous présenterons les résultats d'une expérience d'identification phonétique de voyelles synthétiques à plusieurs fréquences fondamentales. Les résultats mettent en évidence une propriété jusqu'alors non reconnue du déplacement de la frontière phonétique sous l'effet de modifications de la fréquence fondamentale. Ce déplacement n'est pas une fonction monotonement croissante de F_0 , mais montre un point de rebroussement. Ce fait a été confirmé très récemment par les travaux de CARRE (1988). Cette propriété permet d'éliminer plusieurs modèles d'extraction des formants auditifs. Nous proposerons enfin quelques réflexions sur les contraintes que doit satisfaire un modèle conforme aux données perceptives. Il est vraisemblable notamment que la position du formant auditif est déterminée par la position de l'harmonique la plus proéminente et des harmoniques voisines si celles-ci sont moins de 10 dB sous la plus intense. Mais nous ne serons pas en mesure de décrire avec précision un tel modèle car la localisation des frontières phonétiques semble se déplacer légèrement en fonction de la fréquence fondamentale. La description d'un modèle détaillé implique d'autres études pour préciser l'ampleur et la cause de ces déplacements de frontière.

Une dernière précision nous semble nécessaire. Le problème de la modification du timbre vocalique sous l'effet de la fréquence fondamentale est relié par beaucoup de chercheurs à celui des différences formantiques observées entre les productions masculines, féminines et enfantines. Nous ne voulons pas poser à priori un lien téléologique entre ces domaines. La conservation d'une qualité vocale "masculine", "féminine" ou "enfantine", liée de toute évidence aux différences formantiques et à la forme de la source glottale, plus qu'à la hauteur de la fréquence fondamentale, interdit d'ailleurs toute "normalisation" complète des informations perceptives.

2 - L'INTERACTION DU TIMBRE VOCALIQUE ET DE LA FREQUENCE FONDAMENTALE : HYPOTHESES ET DONNEES EXPERIMENTALES.

Cette partie présente, par ordre chronologique, les hypothèses et confirmations expérimentales de la variation du timbre vocalique sous l'effet de la fréquence fondamentale. Elle analyse aussi les études moins nombreuses sur la représentation interne des maxima spectraux. Précisons qu'il ne s'agit pas d'une simple revue des travaux, où l'on rappelle simplement les conclusions des auteurs, mais, dans certains cas, d'une

réinterprétation des résultats à la lumière d'autres hypothèses, ou en s'appuyant sur d'autres résultats expérimentaux. Ces analyses ont dicté la forme et le contenu de l'expérience qui sera décrite plus loin.

2.1 - R.L. MILLER (1953)

Dans l'un des premiers articles d'après-guerre consacrés à la synthèse vocalique, MILLER décrit un synthétiseur harmonique à 100 composantes fréquentielles mis au point quelques années auparavant par H. FLETCHER des laboratoires de la BELL TELEPHONE. Le son synthétique était automatiquement enregistré dans un intervalle de silence d'une bande magnétique, entre un [s] et un [t] ! Le gros du travail porte sur la division perceptive de l'espace F1 / F2, le rôle de F3 et de l'amplitude des formants. De nombreuses remarques, d'un intérêt toujours actuel, seront perdues de vue lors du règne ultérieur du "pattern playback" des HASKINS Laboratories. L'utilisation systématique de deux fréquences fondamentales (140 et 288 Hz) révèle immédiatement une modification du timbre à fréquences formantiques égales. Les frontières phonétiques se déplacent dans la même direction que Fo. MILLER remarque que les voyelles ouvertes sont plus touchées que les voyelles fermées et le lien avec les productions masculine, féminine et enfantine est mentionné brièvement. Dès ce moment, il apparaît clairement que le timbre d'une voyelle ne dépend pas exclusivement de la fréquence des paramètres formantiques de la synthèse, et les caractéristiques essentielles de l'interaction timbre / Fo sont mises en évidence.

2.2 - FANT (1957) ; CARLSON, FANT & GRANSTRÖM (1970); FANT, CARLSON & GRANSTRÖM (1974)

Dans un rapport rédigé en suédois que nous citons de seconde main d'après TENENHOLTZ et FLOREN (1978 / 1979), FANT (1957) observe qu'une augmentation d'une octave de Fo (150 - 300 Hz) provoque une modification du timbre qui peut être compensée par une élévation de F1 et F2 de 50 à 100 Hz. La voyelle [ɑ] chantée par une soprano a approximativement les mêmes formants que le [a] d'une basse. L'identité phonétique ne peut être maintenue qu'en augmentant F1 proportionnellement à Fo.

Une conclusion similaire est dégagée par CARLSON, FANT & al. (1970) d'un test d'identification d'un ensemble de voyelles synthétiques à 100 - 120 et 200 - 240 Hz. La relation de compensation entre Fo d'une part, et F1 et F2 d'autre part, exige en moyenne un déplacement du paramètre F1 + F2 de 75 mels pour compenser l'effet du changement de la fréquence fondamentale d'une octave. Cette valeur de 75 mels correspond approximativement au quart de la distance moyenne entre deux phonèmes.

FANT & al. (1974) traitent de l'ambiguïté potentielle entre un [e] masculin et un [ø] féminin

qui peuvent avoir les mêmes fréquences formantiques. Leurs résultats montrent un déplacement de la frontière entre les voyelles arrondies et non arrondies quand la fréquence fondamentale passe de 110 à 220 Hz. Ce déplacement de frontière, ainsi que la valeur légèrement plus élevée des formants supérieurs (F3, F4), déplaçant vers le bas le centre de gravité du groupe F2 - F4, conservent un timbre de voyelle arrondie à la réalisation féminine. Une variation de Fo change donc le timbre d'une voyelle à fréquences formantiques constantes.

2.3 - FUJISAKI & KAWASHIMA (1968)

15 ans après les travaux de pionnier de MILLER, cette contribution au problème de la normalisation perceptive pose d'emblée que les différences anatomiques exigent en toute logique que les fréquences formantiques soient considérées, au plan perceptif, comme relatives et non absolues. Les formants supérieurs à F2 sont présentés comme plus stables. Pour F3, dont la fréquence dépend principalement de la taille du conduit pharyngolaryngé, les auteurs affirment l'existence d'une corrélation entre ce dernier et la fréquence fondamentale, fondée sur un rapport constant entre la taille des cordes vocales et du pharynx. A partir d'analyses spectrographiques sur 5 voyelles prononcées par 43 locuteurs de sexe et d'âge très différents, cette liaison statistique peut s'exprimer sous la forme $F3 = 6.2 (F_0 + 270 \text{ Hz})$ pour les quatre voyelles [u, e, o, a]. Des relations plus précises sont proposées pour les sous-ensembles des voyelles ouvertes et fermées. Mais l'examen de la figure 2 de l'article semble montrer que pour [e], Fo corrèle autant, et sans doute mieux, avec F1 et F2. Nous noterons encore, pour notre part, l'influence sur F3 de la labialisation des voyelles antérieures [y] et [ø], qui n'apparaît pas en Japonais et qui vient certainement perturber pour le français cette belle corrélation. Les expériences de perception montrent globalement que "la normalisation perceptive n'est pas complète si Fo et les formants supérieurs ne varient pas concurremment". Pour le problème qui nous concerne, on peut déduire de l'examen des figures (reproduites ici figure 2.1) les faits suivants.

La frontière phonétique entre [e] et [u] (cette dernière étant la voyelle délabialisée et centralisée du Japonais à F1 et F2 élevés [u]) se déplace lorsque Fo varie. Elle passe par exemple par les points de coordonnées $F1 = 570 \text{ Hz}$, $F2 = 1710 \text{ Hz}$ pour $F_0 = 130 \text{ Hz}$, mais par les points $F1 = 675 \text{ Hz}$, $F2 = 2025 \text{ Hz}$ pour $F_0 = 370 \text{ Hz}$. Pour F1, le déplacement de la frontière est une fonction linéaire de Fo. De 130 à 370 Hz, la frontière se déplace environ du tiers de l'accroissement de Fo. Le déplacement de la frontière phonétique dans la plan F1 / F2 est un glissement identique à celui que l'on observe entre les triangles des voyelles masculines et féminines. Pour notre étude, nous aurions préféré connaître les résultats d'une modification de la frontière à F2 constant, et non, comme ici, pour des stimuli dont F2 était toujours égal à 3 x F1. La frontière phonétique entre [a] et [o]

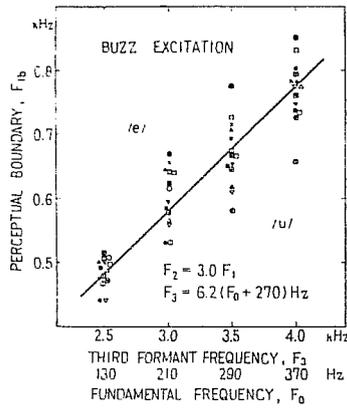


Fig. 5. Effect of simultaneous changes of F_0 , F_3 , F_4 , and K_{r4} on perceptual boundary F_{1b} in identification of /u/ and /e/.

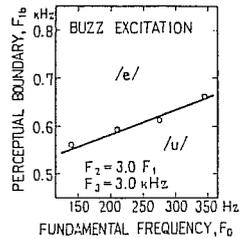
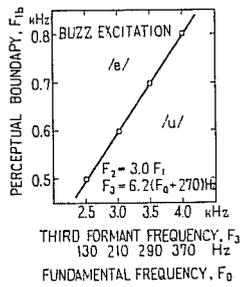


Fig. 6. Shift of perceptual boundary F_{1b} as function of fundamental frequency and higher formant frequencies in identification of /u/ and /e/.

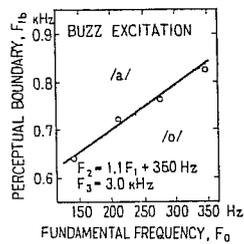
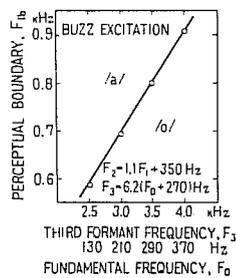


Fig. 7. Shift of perceptual boundary F_{1b} as function of fundamental frequency and higher formant frequencies in identification of /o/ and /a/.

Fig 2.1 - Variation de la frontière perceptive entre [e]-[u] ou [a]-[o] en fonction de la fréquence fondamentale seule ou de F_0 et F_3 . (FUJISAKI & KAWASHIMA 1968).

montre le même comportement. Pour $F_0 = 130$ Hz et 370 Hz, la limite passe de $F_1 = 630$, $F_2 = 1050$ à $F_1 = 820$, $F_2 = 1250$ Hz. L'accroissement de F_1 est à nouveau une fonction linéaire de F_0 , dont le coefficient de pente, plus fort que le précédent, est d'environ 0.8. Ces expériences confirment celles de MILLER, sans qu'il soit possible d'en déduire précisément l'influence de F_0 sur le timbre de voyelles physiquement identiques.

2.4 - SLAWSON (1968)

La série d'expériences décrites par SLAWSON reste une pièce maîtresse du dossier. Le point de départ est clair. Le timbre des voyelles, tout comme les timbres musicaux, ne changent que légèrement quand la fréquence fondamentale balaie une gamme très étendue. SLAWSON cherche à identifier les caractéristiques qui fondent cette quasi invariance. Il rappelle que STUMPF, dès 1926, constatait qu'un petit changement des fréquences formantiques suffisait à contrebalancer l'effet d'une variation de la fréquence fondamentale, alors qu'un changement proportionnel de F_0 et des formants altère profondément le timbre (CHIBA & KAJIYAMA 1941). Cette question est ensuite rapprochée de la différence entre les fréquences formantiques masculines, féminines et enfantines (cf. par ex. PETERSON & BARNEY 1952).

Huit expériences sont ensuite décrites, qui demandent toutes, sur une échelle numérique définie par le sujet, une estimation subjective de la dissimilarité entre un timbre de référence et un timbre-test. Les voyelles de référence, synthétisées d'après les mesures de PETERSON & BARNEY, étaient l'une des voyelles suivantes : [i, ae, a,), o] ou [u] à une fréquence fondamentale de 135 Hz. La voyelle-test était une version modifiée dont les deux premières fréquences formantiques étaient décalées d'un facteur variant de 0.95 à 1.3 par pas de 0.05. Les six séries principales incorporent une ou plusieurs des conditions suivantes : F_0 à 135 ou 270 Hz ; formants F_3 à F_5 fixes ou modifiés ; légère inflexion de F_0 (valeur terminale : -30 Hz) et attaque graduelle ou non. Pour la moitié des sujets, les sons à inflexion de F_0 et attaque douce étaient présentés comme des imitations de voyelles naturelles. Pour l'autre moitié, les sons à F_0 constant et attaque brutale étaient décrits comme des timbres musicaux. Une enquête postérieure a montré qu'aucun des sujets de ce deuxième groupe n'avait senti qu'il pouvait s'agir d'imitations de voyelles humaines, ce qui révèle le manque de naturel de ces stimuli synthétiques, aggravé sans doute par la durée choisie (700 ms). Deux expériences complémentaires seront brièvement décrites plus bas.

Les résultats principaux sont les suivants. Pour la version vocalique, la différence de timbre entre les voyelles ouvertes [ae, a,)] à 135 et 270 Hz est minimale lorsque le décalage formantique est de + 10 à +15 % (figure 2.2). Mais les scores montrent clairement que les timbres ne redeviennent pas exactement identiques. Pour les voyelles fermées, aucune

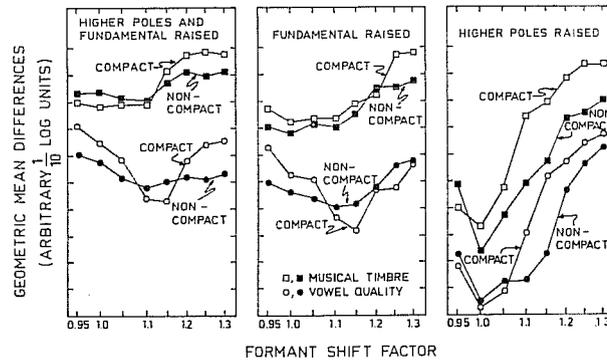


FIG. 8. Comparison of vowel quality and musical timbre judgments with respect to the compact-noncompact distinction (compact: empty points; noncompact: filled points). The vowel quality and musical timbre functions are displaced 0.2 log units from each other for clarity of presentation. The geometric means are calculated from the normalized estimates.

Figure 2.2 - Comparaison de l'écart subjectif de timbre en fonction de la fréquence formantique pour une élévation de F_0 de 135 à 270 Hz (in SLAWSON 1968).

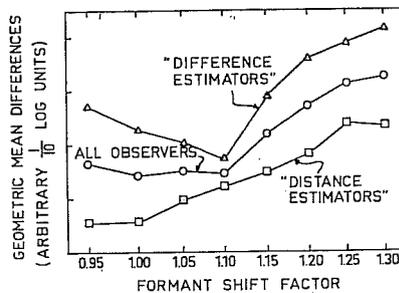


FIG. 7. Comparison of the musical timbre judgments of two sets of observers. The "difference estimators" were the 21 observers, in Expts IIa and IIb, whose minimum estimates were at formant shift factors of 1.05, 1.10, or 1.15. The "distance estimators" were the remaining 19 observers in those experiments. For clarity of presentation, the geometric means of the "difference estimators" are increased by 0.05 log units in relation to the over-all geometric means; those of the "distance estimators" are decreased by 0.05 log units. Otherwise, the coordinates are the same as in Fig. 4.

Figure 2.3 - Différentiation de deux types de réponses pour des stimuli à attaque brutale (même condition qu'à la figure 2.2) (in SLAWSON 1968).

modification formantique ne minimise clairement la différence de timbre. SLAWSON ne propose pas d'explication précise de la différence observée entre les voyelles ouvertes et fermées. On peut cependant remarquer que lorsque F1 est bas ([i] = 247 Hz; [u] = 270 Hz) et Fo élevé (270 Hz), un décalage de - 12 à + 80 Hz (facteur 0.95 à 1.30) ne conduit qu'à une variation de l'amplitude de F1, qui reste toujours défini par la fréquence de la seule fréquence fondamentale, les autres harmoniques étant de niveau très bas. Dans la plupart des cas, la fréquence physique du premier formant de [i] était d'ailleurs inférieure à la fréquence fondamentale. La variation des formants supérieurs est certes forte, mais on sait que le timbre des voyelles fermées dépend principalement de F1 : elles sont presque sur une même verticale dans le triangle du schéma F1 / F2, ce qui indique des F2 voisins. Pour les premiers formants des voyelles ouvertes ([ae] = 764 Hz, [a] = 733 Hz), la variation couvre une plage allant de -25 à + 230 Hz. En conséquence, l'harmonique la plus proche de F1 n'est pas la même aux modifications extrêmes. Les variations de F2 sont de la même ampleur, et participent aussi à la modification du timbre. Dans cette expérience, la manipulation des fréquences formantiques ne permet donc vraisemblablement une variation du timbre que pour les voyelles ouvertes. Enfin, la modification des formants supérieurs (F3 - F5) ne joue qu'un rôle mineur.

Les résultats pour les timbres musicaux à attaque brutale révèlent deux stratégies (figure 2.3) : certains sujets (dont un musicien professionnel) ont un comportement identique à celui constaté pour les timbres vocaliques, avec un minimum vers le facteur d'anamorphose de 1.10. SLAWSON les appelle "estimateurs de différence", en supposant qu'ils se sont attachés à mesurer les différences de timbre. Les autres sujets indiquent une différence minimale lorsque les fréquences formantiques coïncident (facteur 0.95 ou 1.0). Ces "estimateurs de distance" auraient comparé la distance entre les fréquences formantiques des spectres, sans tenir compte des différences de timbre. Il revient à R. CARRE (1988, cf. aussi CARRE & QUACH-TUAN 1986, CARRE & LANCIA 1975) d'avoir fourni, pour des expériences similaires et sans référence aux travaux de SLAWSON, une confirmation expérimentale que nous décrirons en détail plus bas. En un mot, un deuxième mécanisme de codage spectral, vraisemblablement fondé sur le taux de décharge moyen des fibres auditives, est utilisé lorsque dans le cas d'une attaque brutale. Ce mécanisme fournit une bonne estimation de la fréquence physique de F1. Cette explication n'attribue qu'un rôle mineur au "balayage" fréquentiel produit par une évolution lente de la fréquence fondamentale.

Deux expériences complémentaires sur des stimuli à attaque rapide présentés comme des voyelles ou des sons musicaux, pour des facteurs de décalage variant de 0.9 à 1.5 par pas de 0.1 et un Fo de 205 Hz seulement, révèlent un jugement de type "musical" à minimum de dissimilarité en cas d'identité des fréquences formantiques. La conclusion la plus claire est que le changement d'instruction concernant la nature des stimuli n'est pas responsable

des différences constatées.

La discussion de SLAWSON est centrée sur la différence entre les variations formantiques naturelles liées au sexe et à l'âge, qui dépasse largement 20% pour les voyelles ouvertes même si l'on ne tient compte que des locuteurs masculins et féminins, et les valeurs psychophysiques plus faibles constatées ici. "Ces résultats ne sont pas en accord avec ce que prévoit une interprétation fondée sur l'apprentissage" (ref. p. 100). SLAWSON note que c'est particulièrement vrai pour les instruments de musique dont les fréquences de résonance sont pratiquement fixes (violon, basson), ou changent proportionnellement à la fréquence fondamentale (flûte). La conclusion générale, présentée sous la forme d'une hypothèse, est que la fréquence d'excitation biaise légèrement la perception des fréquences formantiques.

On retrouve cependant, dans beaucoup de travaux ultérieurs, l'hypothèse que l'influence sur le timbre d'une élévation de la fréquence fondamentale est exactement contrebalancée par l'augmentation des fréquences formantiques liée au sexe ou à l'âge des locuteurs. Nos propres expériences, comme celles de SLAWSON, devraient permettre d'enterrer définitivement cette idée tenace.

2.5 - MUSHNIKOV & CHISTOVICH (1972, 1973)

Ces deux contributions, très souvent citées, n'ont pas vu leur conclusion confirmée par les travaux ultérieurs. KARNICKAYA & al. (1975) en donne une présentation plus pédagogique. Mais l'intérêt de la méthode employée impose une rapide présentation. La différence de timbre entre [i] et [e], qui dépend dans une large mesure de la fréquence du premier formant, requiert le "calcul" du formant auditif. L'idée de base de l'expérience est d'utiliser une harmonique relativement élevée (vers 500 Hz) et une ou plusieurs harmoniques de fréquence(s) plus basse(s) dans la zone de F1. Les sujets devaient ajuster le niveau de l'harmonique élevée afin de conserver à la voyelle synthétique un timbre ambigu, intermédiaire entre [i] et [e]. Les données expérimentales ne sont pas en accord avec l'hypothèse que l'identification phonétique est fondée sur la comparaison de la sonie dans de larges bandes de fréquence contiguës. Précisons que nous employons dans ce travail le terme sonie au sens large d'intensité (subjective) perçue, ce que traduit bien le terme anglais correspondant "loudness", opposé à intensité ou énergie, mesures physiques. Elles sont en revanche compatibles avec l'hypothèse que l'indice critique est la position fréquentielle du maximum de sonie. Le niveau de l'harmonique élevée ne dépend pas du nombre d'harmoniques de basse fréquence, mais seulement de la sonie de la plus élevée d'entre elles. Augmenter le nombre d'harmoniques accroît la sonie de la bande de large fréquence, car les harmoniques sont séparées de plus d'une bande critique, mais ne change en aucune

manière la sonie de l'harmonique la plus élevée. Notons que les expériences ultérieures mettront en évidence que le critère perceptif n'est pas la sonie de l'harmonique la plus proéminente, mais la somme pondérée des deux harmoniques les plus proéminentes lorsque la différence de niveau ne dépasse pas 10 à 15 dB. Cette conclusion est bien sûr compatible avec les données de MUSHNIKOV & CHISTOVICH qui ont manipulé l'amplitude des autres harmoniques, laissant au même niveau les deux plus intenses, c'est-à-dire celle de 500 Hz et la plus élevée de la bande de fréquence inférieure.

MUSHNIKOV & CHISTOVICH (1972) contient une autre expérience intéressante. Dans des voyelles synthétiques à deux harmoniques dans la zone de F1, ils manipulent la séparation fréquentielle entre ces deux harmoniques. La courbe de la sonie des deux harmoniques en fonction de leur écart fréquentiel présente la pente classique de + 6 dB / octave sous 600 Hz, due à la fonction de transfert de l'oreille externe et moyenne. Remarquons qu'il existe dans la littérature deux manières contradictoires de noter cette pente, + 6 ou -6 dB / octave, ce qui pourrait entraîner une confusion. Des hautes vers les basses fréquences, la pente est de - 6 dB / octave, alors qu'elle est évidemment de + 6 dB / octave des basses vers les hautes fréquences ! Une manière plus claire de s'exprimer est de dire que la sonie d'une composante diminue au fur et à mesure que sa fréquence s'abaisse sous 600 Hz.

Or les résultats de MUSHNIKOV & CHISTOVICH indiquent que la position de la frontière phonétique ne dépend pas du rapport de sonie entre les composantes. Il faut diminuer la sonie de celle de basse fréquence, d'environ 6 dB par octave, pour conserver la frontière phonétique au même endroit. MUSHNIKOV & CHISTOVICH appellent cette relation "equal significance curve", que nous traduirons par "courbe d'égale contribution (phonétique)". Ceci semble indiquer qu'une harmonique basse domine une harmonique haute avec une pente de - 6 dB / octave des basses vers les hautes fréquences. Il faut bien comprendre que cette relation est obtenue dans des conditions très différentes de celles que l'on observe pour les voyelles naturelles. La voyelle synthétique des auteurs soviétiques possède deux harmoniques proéminentes largement séparées, donc deux formants F1 ! Lorsque les harmoniques sont proches, le formant auditif est proche de la moyenne de leurs fréquences. Peut-être même, si l'on tient compte des résultats de ASSMANN & NEAREY (1987) décrits plus loin, est-il plus proche de l'harmonique la plus élevée. Lorsque la séparation augmente, le timbre change dans la mesure où le système perceptif détecte deux maxima dans la zone de F1. Réduire radicalement l'amplitude de l'harmonique de basse fréquence supprime l'un des deux F1. Le formant auditif se trouve donc vers la fréquence de l'harmonique élevée, comme dans le cas où les deux harmoniques sont proches, et le timbre originel est conservé.

2.6 - AINSWORTH (1974)

Si le premier travail de AINSWORTH sur l'effet de F_0 est trop global pour être utilisé (AINS WORTH 1971), deux contributions ultérieures méritent d'être présentées. AINSWORTH (1974) décrit la variation de la frontière phonétique en fonction de F_0 entre [u - U], et [U - ae], pour des stimuli synthétiques à F_2 covariant avec F_1 . La conclusion générale est que "... les résultats suggèrent que les sujets diffèrent grandement dans leur utilisation de F_0 pour normaliser les réalisations masculines, féminines et enfantines". En d'autres termes, certains sujets ne montrent aucune variation de la frontière pour F_0 variant de 110 à 280 Hz (6 Hz pour 100 Hz de variation de F_0 dans la paire [U - ae] pour l'un des sujets), alors que d'autres manifestent une forte dépendance (42 Hz pour 100 Hz pour deux sujets). Pour la paire [u - U] , un auditeur présente une variation de frontière de 192 Hz pour 100 Hz lorsque F_0 passe de 55 à 140 Hz , un autre de 102 Hz, alors que les deux "derniers" se contentent de 12 Hz ! Ces différences individuelles ne sont pas compréhensibles, même si l'on retient l'hypothèse d'AINS WORTH d'un calage de l'estimation de F_1 sur la fréquence de F_0 . La variation de la frontière est, pour l'un des sujets, supérieure à la variation de F_0 . Il faut peut-être incriminer la qualité médiocre de la synthèse, conduisant à un placement presque aléatoire de la frontière. On notera, sur les données individuelles d'un sujet, une variation négative de la frontière lorsque F_0 passe de 110 à 140 Hz, phénomène que nous retrouverons plus loin.

2.7 - CARLSON, FANT & GRANSTRÖM (1975)

Dans le paragraphe intitulé "Interaction de F_0 et F_1 ", cet article fondamental traite de l'estimation perceptive de la fréquence du premier formant à partir de la fréquence et de l'amplitude des harmoniques. Les auteurs soulignent tout d'abord que la sélectivité fréquentielle de l'oreille est telle que les harmoniques de basse fréquence doivent être présentes dans le spectre auditif. Les mesures les plus récentes, qui montrent que la sélectivité est encore plus forte que les premières estimations ne l'indiquaient, renforcent encore cette conclusion. Concrètement, une première approximation d'un modèle d'oreille doit comporter un banc de filtres composé de filtres relativement étroits permettant, comme dans le cas du filtrage spectrographique étroit, une représentation des harmoniques de basse fréquence. Deux hypothèses sont ensuite présentées. La première postule que l'auditeur est capable de reconstruire le pic de l'enveloppe spectrale à partir du niveau des harmoniques adjacentes, sans qu'une harmonique soit nécessairement présente à la fréquence formantique. La deuxième, proposée et testée par MUSHNIKOV & CHISTOVICH (1972), suppose que l'auditeur prend en compte la fréquence harmonique la plus "proéminente", en ignorant les autres, et qu'une harmonique de basse fréquence contribue

nettement plus à la position du formant auditif que son niveau physique faible ne le laisserait supposer. Cette formulation est d'ailleurs légèrement incorrecte dans la mesure où CARLSON & al. joignent deux conclusions distinctes, logiquement indépendantes, des auteurs soviétiques.

Les différences entre ces différentes hypothèses sont brillamment illustrées à la figure 2.4. Dans un spectre de sonie, le pic formantique se dégage bien, même s'il existe en basse fréquence un formant glottal intense. En revanche, un formant pondéré par la courbe d'"égale contribution phonétique" n'est présent que sous la forme d'une épaule sur une pente spectrale très raide. La partie expérimentale présente les résultats d'identification de 5 séries de voyelles synthétiques de timbre variant entre [i] et [e]. Chaque série, comprenant 5 voyelles dont les premiers formants s'étagent entre 250 et 350 Hz par pas de 50 Hz, correspond à une fréquence fondamentale différente, de 100 à 160 Hz par pas de 15 Hz. F_0 était plat, ou suivait un contour d'une déviation maximale de 4%. L'expérience a également été faite sur des stimuli comportant une légère oscillation périodique de F_0 ("jitter"). En théorie, le balayage fréquentiel qui en résulte devrait fournir à l'auditeur une information plus précise sur la position du pic formantique. CARLSON & al. notent cependant un accroissement de la variance des estimations. On peut comprendre cette détérioration si les auditeurs utilisent l'amplitude des harmoniques et si la mesure de l'amplitude nécessite le recueil d'informations auditives sur plus d'une période. Une variation à court terme de la fréquence fondamentale modifie à chaque période l'amplitude des harmoniques, ce qui peut accroître la variabilité de leur mesure.

Le résultat essentiel est que la frontière phonétique entre [i] et [e] montre une évolution parallèle à celle de F_0 . Elle passe régulièrement de 300 Hz pour $F_0 = 100$ Hz à plus de 350 Hz pour $F_0 = 160$ Hz, sauf entre 100 et 115 Hz, où elle paraît stable. La figure 2.5 présente également les prédictions fournies par l'hypothèse de l'utilisation de l'harmonique la plus intense d'un spectre pondéré par la courbe d'"égale contribution phonétique", et celles qui découlent du spectre de sonie. Si l'on "corrige" ces dernières en tenant compte de l'harmonique adjacente la plus intense d'une manière proportionnelle à leur sonie, on obtient une bonne prédiction de l'évolution de la frontière phonétique. L'estimation est moins bonne si l'on tient compte de la seule harmonique la plus proéminente (test d'une partie de l'hypothèse de MUSHNIKOV & CHISTOVICH), ou des trois harmoniques les plus intenses. L'usage de la double hypothèse de MUSHNIKOV & CHISTOVICH (harmonique la plus proéminente pondérée par la courbe d'égale contribution phonétique) fournit en revanche des prédictions très médiocres. Ces résultats suggèrent que "l'auditeur fait appel à un mécanisme d'interpolation pour estimer la fréquence formantique, et non au choix de la plus intense ou de la plus "contributive" de toutes" (ref. p. 67).

A posteriori, on peut regretter que l'expérience n'ait pas été poursuivie à des fréquences fondamentales plus élevées, car l'hypothèse d'une relation monotonement croissante entre

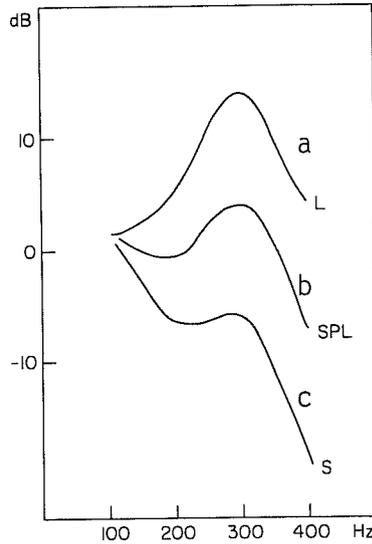


Figure 6 Envelope of first formant after different pre-emphasis.
 L: +6 dB/oct approx. equal loudness,
 SPL: unfiltered,
 S: -6 dB/oct approx. equal significance (see text).

- Figure 2.4 : Représentation schématique du formant auditif :
 a: spectre de sonie ; b : spectre physique ; c: 'égale contribution phonétique' (in CARLSON & al. 1975)

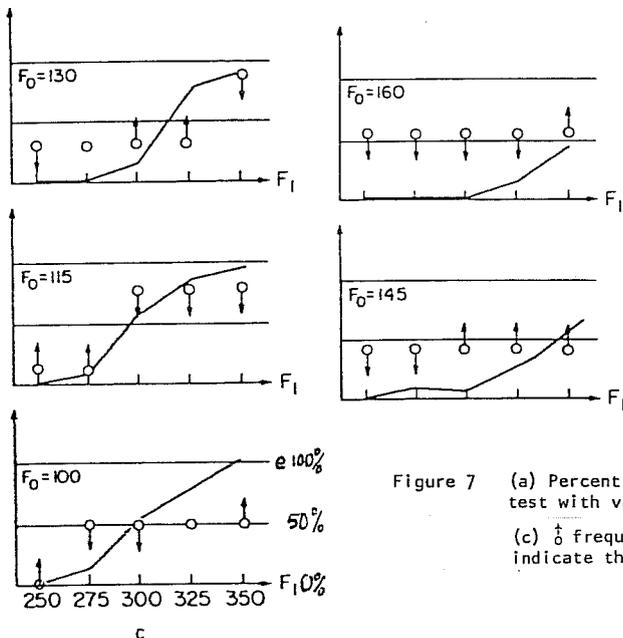


Figure 7 (a) Percent of [e] responses obtained from an identification test with varying F_0 and F_1 (— monotone pitch,
 (c) δ frequency value of the loudest harmonic. Arrow indicate the direction to the second loudest harmonic

- Figure 2.5 - Identification d'une série de stimuli synthétiques entre [i] et [e] en fonction de F_0 (in CARLSON & al. 1975).

la frontière phonétique et Fo aurait (vraisemblablement) été rejetée, comme chez CARRE (1988).

Cette étude, comme la plupart de celles venant du laboratoire de G. FANT, est un modèle du genre. Elle montre de façon claire que la fréquence du formant auditif est, selon toute vraisemblance, une somme pondérée de la contribution des harmoniques proches. Des données récentes de DARWIN & GARDNER (1985, 1986), présentées plus loin, indiquent cependant que des harmoniques plus lointaines pourraient également apporter dans certain cas une contribution à la position du formant auditif.

2.8 - SCOTT (1976)

Les conclusions de SCOTT (1976, cf. aussi SCOTT 1980), bien que publiées dans la plus grande revue internationale du domaine, ne sont pratiquement jamais mentionnées. Doit-on y voir là le poids d'une "idéologie" scientifique dominante, qui marginalise les résultats qui ne s'intègrent pas dans le paradigme dominant, au sens Kuhnien du terme ? Nous avons longtemps cru qu'elles fournissaient une explication satisfaisante de l'interaction timbre / Fo, et elles ont pesé dans notre décision d'étudier ce problème. Une analyse plus serrée nous a pourtant montré que les preuves n'étaient pas décisives.

L'idée centrale de SCOTT est que la variation temporelle à l'échelle de la période du signal de parole rend compte de la modification de frontière phonétique observée pour des stimuli ayant des fréquences fondamentales différentes. Pour une série de stimuli synthétiques entre [i] et [I] (exp. 1 ; figure 2.6), la frontière phonétique sépare les stimuli ayant deux passages par zéro positifs par période de ceux en possédant trois. La forme peu "classique" de l'onde temporelle suggère qu'une source impulsionnelle a été utilisée. Un contrôle sur un mannequin audiométrique a montré que cette caractéristique était conservée jusqu'au tympan, malgré les distorsions de phase appliquées par le canal auditif. Aucune mention n'est faite d'un effet observable dans la distribution des décharges nerveuses. Cette différence est utilisée pour prédire le sens de la variation de la frontière lorsque la durée de la période est modifiée (exp. 2). Pour 10 séries synthétisées à des fréquences fondamentales de 90 à 135 Hz par pas de 5 Hz, SCOTT montre que les séries de 90 à 115 Hz, pour lesquelles le nombre de déflexions négatives de la forme d'onde temporelle varie entre 3 et 4 (pour F1 variant de 270 à 390 Hz), ont une frontière placée plus bas que les séries de 120 à 135 Hz, dont le nombre de déflexions varie de 2 à 3.

Il existe, selon nous, une autre manière d'expliquer ces résultats. Le changement de timbre vers une qualité plus fermée se produit lorsque disparaît le dernier passage par zéro positif. Lorsque F2 est très éloigné de F1, comme ici, cette disparition se produit quand la fréquence de l'harmonique la plus proéminente en basse fréquence dépasse la fréquence de

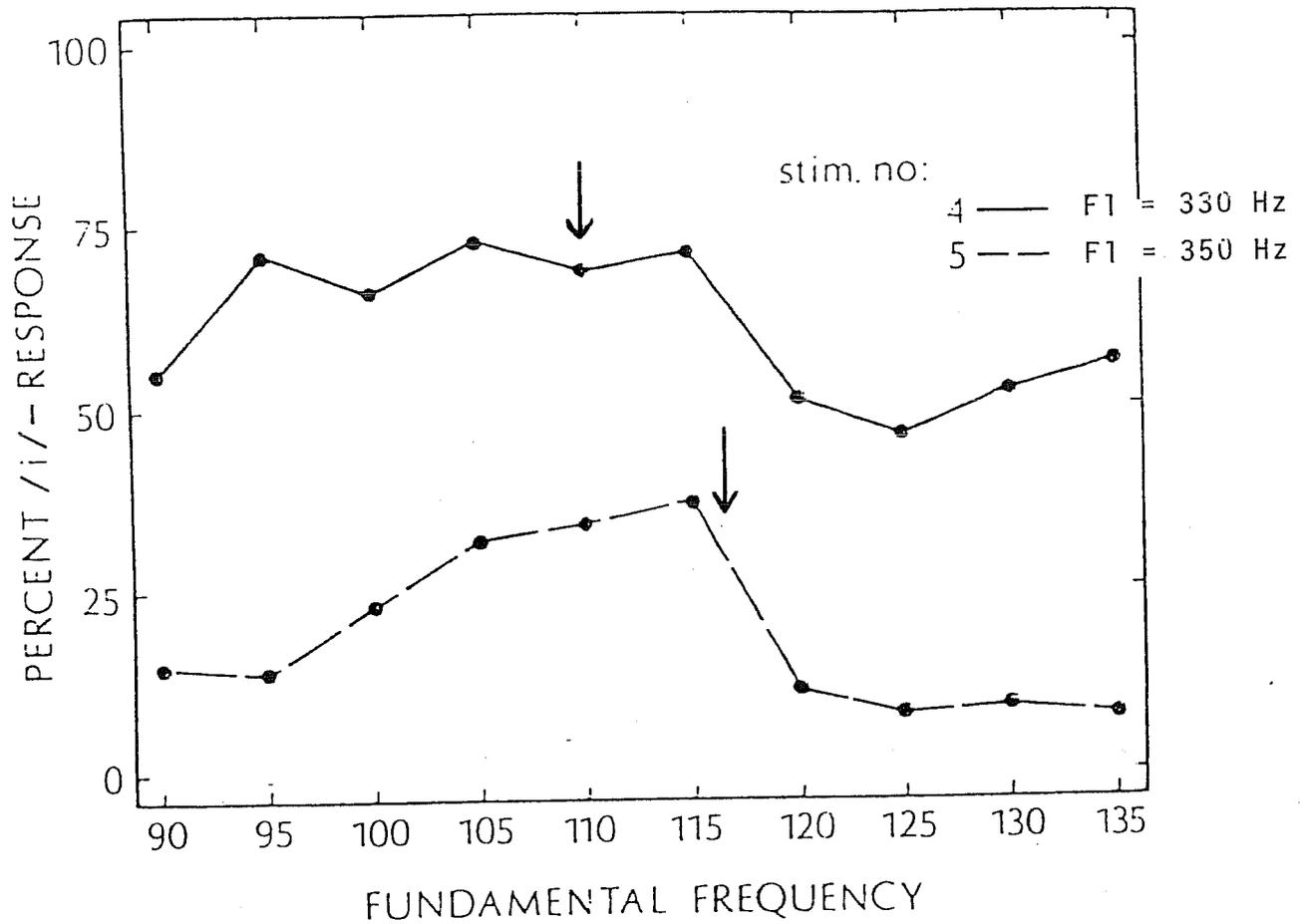


Figure 2.6 - Variation du pourcentage de réponses |i| pour des stimuli synthétiques ($F_1 = 330$ ou 350 Hz) en fonction de F_0 : la position de la flèche indique le dernier stimulus ayant 3 passages par zéro positifs pendant une période. (d'après SCOTT 1976).

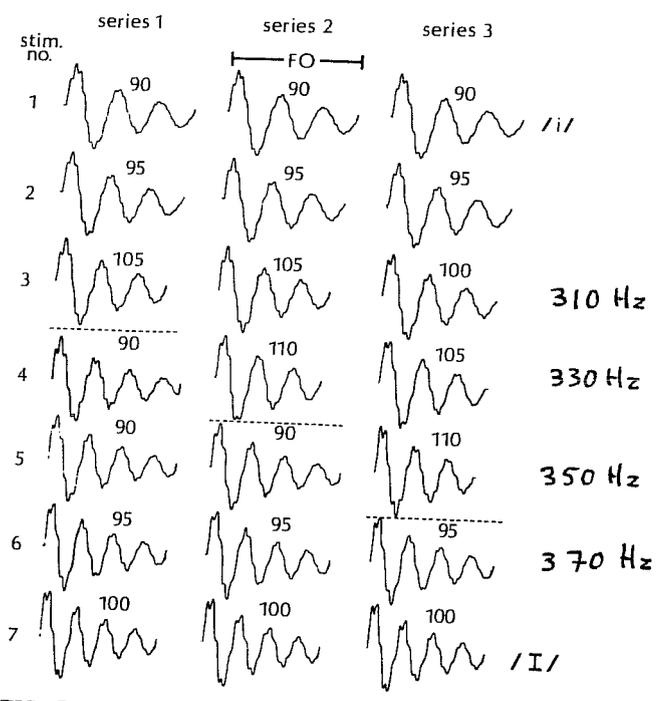


FIG. 7. Single fundamental frequency periods comprising series 1-3. Stimulus numbers correspond to the formant values shown in Fig. 2(a). Stimuli were chosen for the three series to give a waveform change from three cycles of F1 per fundamental period to four cycles at different points along the seven-step continuum.

Figure 2.7 - Variation du nombre d'oscillations temporelles pendant une période en fonction de F1 et Fo (in SCOTT 1976).

F1. L'oscillation amortie de l'harmonique-formant possède alors une pseudo-période un peu trop longue pour être un multiple entier de la fréquence fondamentale. La troisième expérience (figure 2.7) est très parlante. En manipulant la fréquence fondamentale à l'intérieur de chacune des séries, SCOTT provoque un changement du nombre de déflexions négatives entre des stimuli différents pour chacune des séries : il se produit entre les stimuli 3 et 4, ou 4 et 5 par exemple, pour la première et la deuxième série (cf. sur la figure 2.7 la position de la ligne pointillée). Pour une même fréquence du premier formant, les stimuli auront un nombre différent de déflexions négatives. La différence majeure de timbre se produit entre les stimuli ayant un nombre différent de déflexions négatives, et non à F1 constant. Les expériences suivantes, portant sur le rôle de la forme d'onde dans la création du timbre, ne nous concernent pas directement. Dans la discussion, SCOTT réfute une des hypothèses très brièvement mentionnée par CARLSON & al. La "fusion" de la fréquence fondamentale au timbre, qui tirerait celui-ci vers des fréquences plus basses, implique que plus la fréquence fondamentale est élevée, plus elle est proche du premier formant et plus elle contribue, de manière paradoxale, à faire "baisser" le timbre en augmentant le degré de "fusion". Cette explication n'est pas compatible avec une modification non-monotone de la frontière. La deuxième hypothèse est attribuée à CHISTOVICH (1971) et rend compte correctement, d'après SCOTT, des résultats de l'expérience 3 : la fréquence formantique ne peut changer que lorsqu'un changement quelconque amène l'harmonique adjacente à devenir la plus proéminente. Lorsqu'on élève F_0 , les harmoniques sont plus distantes, et il faut une plus grande modification de F1 pour que l'harmonique suivante devienne proéminente. Notons que, sous cette forme, cette hypothèse est incomplète. Si l'harmonique la plus proéminente détermine la position du formant, ce dernier va changer de position en fonction de F_0 . Il aurait convenu de vérifier la cohérence des prédictions de ce dernier point avec les données empiriques.

L'hypothèse de SCOTT est séduisante à première vue. Mais il faudrait vérifier si elle rend compte des modifications de timbre pour les voyelles plus ouvertes, où la forme de l'onde temporelle ne dépend plus seulement du rapport entre la fréquence fondamentale et la fréquence de F1. Si F2 ne contribue qu'une oscillation rapide de faible amplitude pour les voyelles très fermées, ce n'est plus le cas pour les voyelles ouvertes.

On peut dégager quelques indices intéressants des données de SCOTT. Les résultats de l'expérience 2 (figure 2.6) indiquent que la position de la seule harmonique la plus intense ne rend pas compte des résultats. A 115 Hz, les stimuli 4 et 5, dont les premiers formants sont à 330 et 350 Hz n'ont pas exactement le même timbre, bien que l'harmonique à 345 Hz, très proche de la fréquence formantique, soit de loin la plus intense. Les harmoniques adjacentes participent également au positionnement du formant auditif. Mais il paraît impossible de rendre compte à la fois de la légère variation de timbre qui se produit à 120 Hz, et de la stabilité aux autres fréquences fondamentales, en considérant l'amplitude des

composantes. Il est plus clair de le montrer sur les résultats de l'expérience 3. Les stimuli 4 des séries 2 et 3, à la même fréquence formantique, ont des timbres identiques, donc le même formant auditif. Or ils sont à des fréquences fondamentales légèrement différentes, ce qui implique que les harmoniques adjacentes n'ont pas exactement le même niveau. Le formant auditif devrait donc être à des fréquences différentes. Pour sortir de ce cercle vicieux, la forme de l'onde temporelle, qui montre un changement de part et d'autre de 115 Hz à l'expérience 2, vient à point pour nous tirer d'affaire, en introduisant un degré de liberté supplémentaire. Mais la variation de la forme de l'onde temporelle n'est pas la seule différence qui pourrait expliquer la discontinuité entre 115 Hz et 120 Hz. On sait que la phase d'une harmonique subit une rotation de 180° de part et d'autre de la fréquence formantique (FLANAGAN 1970 p. 61). Si on fait l'hypothèse que le système de détermination du timbre est sensible à la phase des composantes, on peut songer à décrire phénoménologiquement le résultat de l'expérience 2 comme suit : lorsque l'harmonique la plus intense est à une fréquence plus basse que la position physique du formant le timbre est légèrement plus fermé, impliquant un formant auditif un peu plus bas. Lorsque l'harmonique la plus proéminente est plus élevée que le formant, le timbre est plus ouvert. Les données sont un peu plus complexes et nous ne posons ici qu'une hypothèse, sans prétendre rendre compte du détail des expériences. D'autres effets de phase seront décrits plus bas au § 2.14.

Une discussion sur des modifications si ténues du timbre ne doit pas faire oublier que les données d'identification dépendent pour une part d'effets de contraste liés à l'ordre de présentation des stimuli. Il est tout à fait plausible que des biais non contrôlés interdisent toute analyse fine des données. Seule la répétition des effets dans des expériences indépendantes permettra de trancher.

2.9 - FLOREN (1978 / 1979)

La contribution de FLOREN, peu connue, a été élaborée dans le laboratoire de LINDBLOM et mérite notre attention par l'intérêt des résultats bruts. L'expérience décrite avait pour but d'étudier la manière dont le timbre vocalique est affecté par l'interaction entre la position du pic formantique et l'amplitude des harmoniques. Elle porte sur des sons synthétiques mono-formantiques identifiables comme [a,)] ou [o]. Pour des stimuli dont les fréquences formantiques sont expérimentalement choisies pour produire chacun de ces timbres, elle produit des séries qui diffèrent uniquement par la fréquence fondamentale employée, de manière à ce que l'harmonique la plus intense soit proche, ou éloignée, de la fréquence formantique. FLOREN propose trois hypothèses sur le mécanisme de positionnement du formant auditif :

- a - l'auditeur peut reconstruire la position physique exacte du formant à partir de

- l'amplitude de toutes les harmoniques sous l'enveloppe formantique, sans qu'il y en ait nécessairement une à la fréquence exacte du formant physique.
- b - l'auditeur peut utiliser l'harmonique la plus intense, en ignorant le niveau des autres.
 - c - l'auditeur peut procéder à une somme pondérée d'un nombre variable d'harmoniques.

On retrouve dans cette troisième hypothèse la conclusion de CARLSON & al. Ces hypothèses conduisent à des prédictions différentes en ce qui concerne l'évolution du timbre. Si l'auditeur reconstruit la fréquence physique exacte, le timbre sera invariant car la fréquence formantique ne varie pas. Les résultats antérieurs montrent déjà que cette hypothèse est contredite par les faits de perception, et elle peut être éliminée sans plus attendre. Si l'auditeur n'emploie que la fréquence de l'harmonique la plus intense, on s'attend à ce que le timbre fluctue. Un exemple le montre facilement. Soit un formant à 400 Hz. Pour $F_0 = 100$ Hz, supposons un timbre [o] défini par l'harmonique la plus intense, à 400 Hz. Si F_0 est placé à 120 Hz, l'harmonique la plus intense sera à 360 Hz. L'harmonique à 480 Hz, plus éloignée de la fréquence formantique sera plus faible. Le timbre sera donc identique à celui d'une voyelle dont le premier formant est à 360 Hz, donc plus fermée. En revanche, si F_0 s'élève à 150 Hz, l'harmonique la plus intense est à 450 Hz, celles à 300 et 600 étant beaucoup plus éloignées de la fréquence formantique. Le timbre sera celui d'une voyelle plus ouverte. C'est cette fluctuation potentielle du timbre à F_1 constant sous l'effet d'une variation de la fréquence fondamentale que FLOREN surnomme l'effet "zigzag". L'hypothèse 3, tant que les détails de la pondération ne sont pas précisés, ne permet pas de prédiction détaillée.

Les résultats sont reproduits à la figure 2.8. La conclusion de FLOREN est qu'ils ne permettent pas de conclure ! Les zigzags ne se produisent pas à l'endroit où l'hypothèse d'une utilisation de l'harmonique la plus intense les placerait (harmoniques équidistantes de la fréquence formantique). Seuls les résultats de la voyelle ouverte ($F_1 = 720$ Hz) pourraient montrer la relation linéaire entre F_1 et F_0 prévue par FANT (1957). Aucune tentative n'est faite pour évaluer plus avant la proposition de CARLSON & al. Nous remarquons qu'il est possible de donner une interprétation des données dans cet esprit, qui sera reprise plus loin dans la discussion de notre propre expérience et de celle de CARRE (1988).

Pour le deuxième stimulus ($F_1 = 550$ Hz), placer F_0 à 275 Hz produit une deuxième harmonique très intense à la fréquence formantique. On obtient 50% de réponses [o]. Pour $F_0 = 228$ Hz, l'harmonique la plus intense est à 456 Hz, 94 Hz sous F_1 . Mais la troisième à 684 Hz n'est que 134 Hz plus haut que le formant. En tenant compte de la pente du spectre de sonie (+ 6 dB jusqu'à 600 Hz) qui peut amplifier la représentation auditive de cette composante au détriment de la précédente, on peut concevoir que la moyenne pondérée

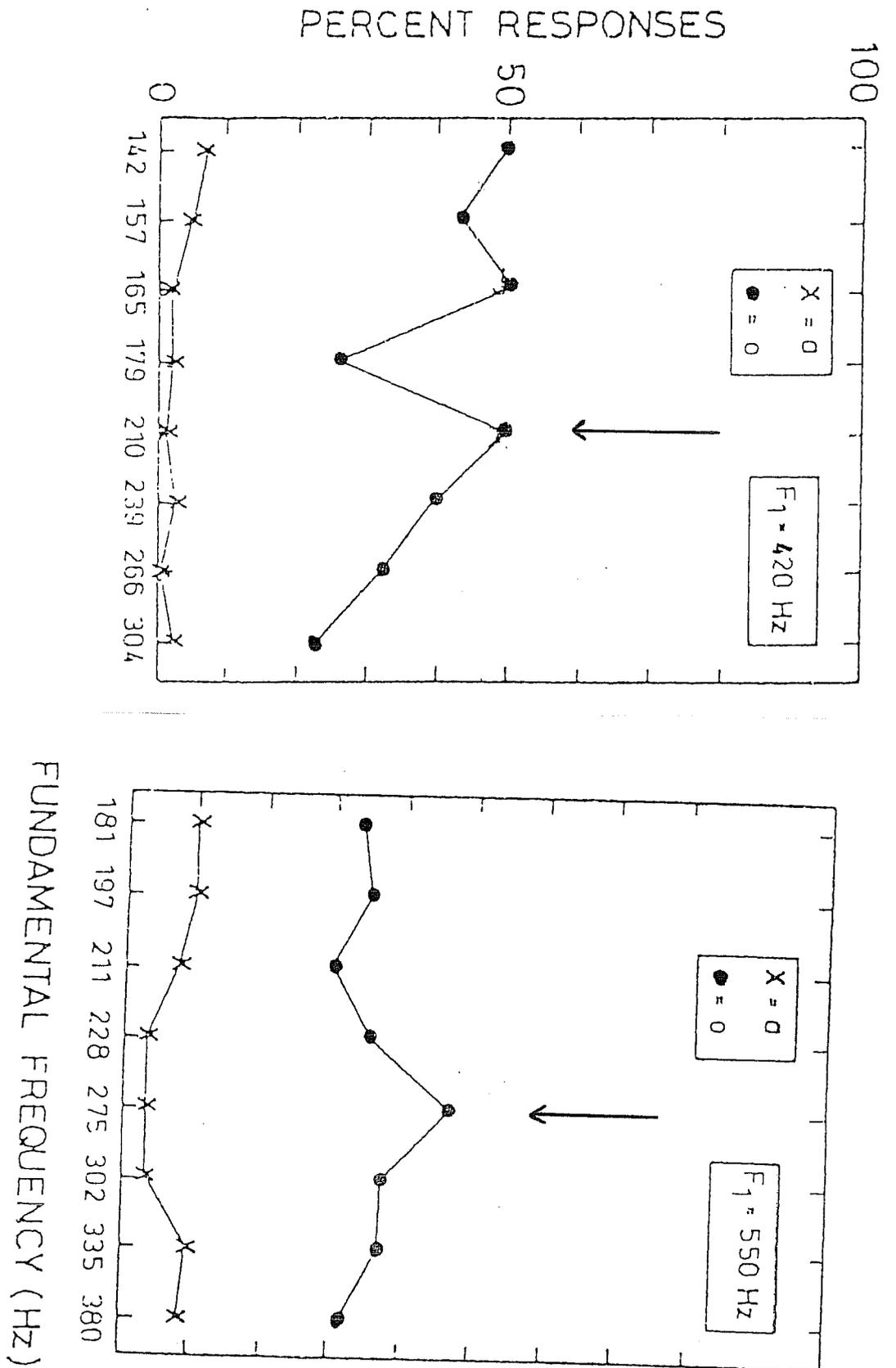


Figure 2.8 - Identification de stimuli mono-formantiques en fonction de F_0 : la flèche indique la fréquence de F_0 où une harmonique coïncide avec la fréquence formantique (d'après FLOREN 1978/9).

des fréquences soit un peu au-dessus de 550 Hz, donnant un timbre plus ouvert et donc un plus petit nombre de réponses [o]. Les paramètres de synthèse n'étant pas connus en détail, on ne peut faire d'analyse plus quantitative. Lorsque F_0 est à 211 Hz, l'harmonique de 633 Hz est plus proche du formant (83 Hz) et donc plus intense que celle de 422 Hz (128 Hz d'écart). La moyenne peut donc être là encore supérieure à 550 Hz, d'où un plus grand nombre de réponses [)]. A $F_0 = 302$ Hz, l'harmonique la plus intense est à 602 Hz, d'où plus de réponses [)]. On peut interpréter de la même manière les réponses à la première voyelle ($F_1 = 420$ Hz). En particulier, lorsque $F_0 = 179$ Hz, l'harmonique la plus intense est sous la fréquence formantique à 358 Hz (-62 Hz), mais la troisième est seulement 83 Hz plus haut que le formant. On pourrait également faire l'hypothèse que lorsque deux harmoniques ont approximativement le même écart par rapport à la fréquence formantique, celle de plus haute fréquence pèse plus "lourd" dans la pondération.

Globalement, on obtiendrait l'effet suivant. Lorsqu'une harmonique coïncide avec la fréquence du formant, le timbre résultant serait plus fermé que lorsque le formant auditif est constitué de la contribution de deux harmoniques.

Le phénomène de suppression donnerait une base physiologique à cette dernière hypothèse. Rappelons l'aspect essentiel de cette notion classique de psychoacoustique. On constate expérimentalement que le pouvoir masquant d'un son A sur un son B peut être diminué par la présence d'un son C. L'interprétation psychoauditive, approximativement conforme à ce que l'on observe sur les données neurophysiologiques, est que l'activité du son C "supprime" une partie de l'activité de A. Son pouvoir masquant sur B est donc plus faible. En règle générale, un son exerce un effet de suppression au-dessous et au-dessus de sa propre position fréquentielle. La magnitude de l'effet de suppression dépend de manière complexe de l'écart fréquentiel entre les deux sons, de l'amplitude relative, et absolue, des deux, de la technique de mesure, et du sujet testé ! (cf. par ex. DUIFHUIS 1980). On peut simplement remarquer que la suppression induite est plus forte sur les sons de fréquence inférieure, mais sur une zone fréquentielle plus réduite. La variabilité individuelle est très importante. Pour des niveaux SPL de l'ordre de 50 à 60 dB, un son de 1.2 kHz à 60 dB ne provoque aucune suppression sur un son de 56 dB à 1 kHz pour l'un des sujets de DUIFHUIS, alors que la suppression atteint 23 dB pour un autre dans les mêmes conditions. Ces résultats montrent qu'un son faible peut exercer une influence sur un son plus intense placé plus bas en fréquence. En règle générale, l'effet de suppression causé par un son au dessus de sa fréquence n'est sensible que s'il domine en énergie. La présence d'un bruit blanc constant, d'intensité modérée, réduit ou supprime les effets de suppression. On peut se demander si la présence, dans une voyelle, d'autres harmoniques ne provoque pas la même disparition. L'effet de suppression ne se produit pas lorsque les deux fréquences sont trop proches. La séparation doit être au moins d'une bande critique, d'après l'examen des données disponibles.

Il aurait été intéressant de vérifier sur les données de FLOREN l'influence de la forme d'onde temporelle, comme le suggère SCOTT. Enfin, il est curieux de constater que les deux voyelles à $F_1 = 550$ Hz, $F_0 = 275$ Hz et $F_1 = 440$ Hz, $F_0 = 210$ Hz, ont exactement le même pourcentage de réponses [o] et [ɔ], ce qui jette un doute sur les analyses précédentes.

2.10 - TRAUNMÜLLER (1981, 1985)

Dans une série de publications, TRAUNMÜLLER a soumis à investigation expérimentale, parmi d'autres facteurs, l'interaction de la fréquence fondamentale et de F_1 dans la qualité du timbre vocalique. Un article majeur, en 1981, souligne que "la perception du degré d'aperture d'une voyelle ne dépend pas, de toute évidence, de la seule valeur de F_1 . Il a été montré que F_0 joue également un certain rôle" (p. 1465). Après avoir attribué à POTTER & STEINBERG (1950) la première formulation de l'hypothèse selon laquelle "... dans une certaine limite, une forme particulière de stimulation présente sur la membrane basilaire sera identifiée comme le même son, quelle que soit sa position exacte sur celle-ci" (op. cit. p. 812), il rappelle que l'unité naturelle de fréquence est le Bark, correspondant à une taille fixe sur la membrane basilaire humaine (1.3 mm, soit l'étendue de 150 cellules ciliées, ZWICKER & FELDKELLER (1967)). L'échelle Bark, d'après le nom du psychophysicien allemand BARKHAUSEN, peut être établie à partir d'expériences déterminant la largeur de bande critique. Lorsqu'on fait estimer subjectivement la sonie d'une bande de bruit de largeur variable, on constate que celle-ci reste constante jusqu'à une certaine largeur qui dépend de la fréquence. Puis elle croît au delà de cette largeur de bande critique, qui vaut par exemple 100 Hz pour une fréquence centrale de la bande de bruit de 100 Hz et 350 Hz à 2000 Hz (SCHARF 1970). L'interprétation est la suivante. La sensation totale de sonie est fonction de la sonie dans chacune des bandes spectrales couvertes par le stimulus. L'oreille est donc en première approximation considérée comme un banc de filtres dont les largeurs sont égales à la largeur de bande critique. Lorsque le stimulus comporte de l'énergie dans plus d'une bande, la sonie totale augmente. La valeur en Bark de la largeur de bande critique est empiriquement donnée par la relation

$$Z_{\text{Bark}} = (26.81 F_{\text{Hz}} / (1960 + F_{\text{Hz}})) - 0.53 \quad (\text{TRAUNMÜLLER 1983})$$

Les travaux plus récents (MOORE & GLASBERG 1983 a), fondés sur la mesure directe, par des techniques de masquage, de la largeur de bande des filtres auditifs conduisent à des valeurs nettement inférieures : 40 Hz de large pour une fréquence centrale de 100 Hz, 200 Hz à 2 kHz. Dans la suite de ce travail, nous avons vérifié que nos conclusions ne

dépendaient jamais de la différence entre les estimations de la largeur de bande critique.

La première expérience porte sur 13 sons mono-formantiques, soumis à 23 écoliers autrichiens qui devaient transcrire le timbre perçu à l'aide de nombreux symboles phonétiques correspondant au riche inventaire phonétique de leur dialecte. L'hypothèse testée est que la distance entre la fréquence fondamentale et le formant sur une échelle Bark est le critère essentiel permettant d'apprécier le degré d'ouverture phonétique. Les résultats sont reproduits à la figure 2.9. Les lettres A à E ont été ajoutées pour faciliter une discussion ultérieure. Si la distance $F1 - F_0$ est le critère essentiel du degré d'aperture phonétique, les limites entre les niveaux d'aperture devraient être parallèles à l'abscisse. Ceci est approximativement vrai pour F_0 en-dessous de 350-400 Hz. Les expériences suivantes portent sur des voyelles multi-formantiques. L'expérience 2 étudie l'effet d'un décalage parallèle de $F1$ et F_0 en échelle Bark sur le timbre de 4 voyelles antérieures [i, e, ε, ø], sans changement pour les formants supérieurs. Le degré d'aperture perçu reste quasiment constant pour F_0 sous 400 Hz, mais les voyelles [e] et [ε] sont perçues comme de plus en plus arrondies. L'expérience 3 confirme un résultat classique. Un déplacement de $F1$ seul provoque une variation du degré d'aperture perçu, qui se produit à la même fréquence pour les voyelles arrondies et non arrondies. Les résultats de l'expérience 4 où F_0 seul est modifié pour les voyelles [e, ε, ae] indiquent que le degré d'aperture diminue lorsque la distance $F1 - F_0$ devient plus faible. Cette relation est beaucoup moins nette pour la voyelle [ae], où un certain nombre de sujets ne semblent utiliser que la fréquence de $F1$. L'auteur suggère finalement dans la discussion qu'une intégration spectrale à large bande (3 - 3.5 Bark) est un modèle rendant compte des résultats. Dans ce cas, la zone de F_0 et $F1$ ne produit qu'un seul pic auditif. Ce pic, et les pics suivants, sont comparés à des références stockées en mémoire sous réserve d'un "glissement" potentiel conservant les distances entre les pics.

Il existe un problème commun à toutes ces expériences. En dessous de 160 - 200 Hz, une modification de la distance $F1 - F_0$ ne joue aucun rôle. Seul la fréquence de $F1$ est pertinente. TRAUNMÜLLER propose une modification ad hoc de l'échelle Bark ("end-of-scale" effect) rendant la valeur Bark constante pour $F < 160$ Hz. Cet effet est décrit comme une conséquence de l'intégration spectrale à large bande qui ne peut s'effectuer pleinement que si les F_0 et $F1$ ne sont pas trop proche de l'extrémité basse de l'échelle.

On notera que cette étude ne concerne que le degré d'aperture perçu. Il y a bien modification du timbre à $F1 - F_0$ constant si l'on tient compte de la variation du degré d'arrondissement perçu.

Une évaluation précise des résultats est rendue difficile en raison de l'hypothèse implicite de TRAUNMÜLLER : La fréquence de $F1$ correspond à la valeur du paramètre de synthèse. Plaider que la distance $F1 - F_0$ reste constante suppose que l'on connaisse la fréquence de $F1$. Comment définir par exemple la fréquence d'un $F1$ à 600 Hz lorsque F_0 vaut 400 Hz ?

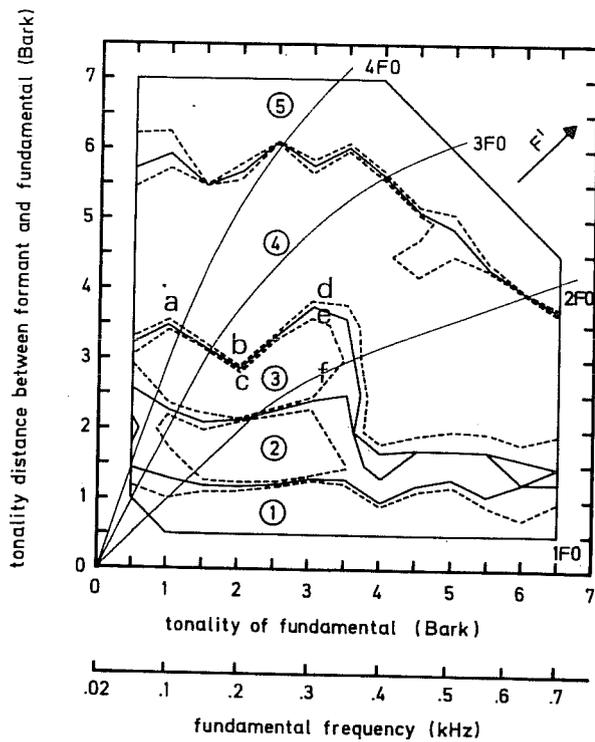


FIG. 4. One-formant stimuli: Dominantly perceived degrees of openness (thick lines) and regions of at least 50% agreement among subjects (dashed lines), shown together with positions of first 4 partials (thin lines). Encircled figures designate degree of openness (cf. Table II and Table III).

Figure 2.9 - Degré d'aperture perçu en fonction de la distance entre F1 et Fo (d'après TRAUNMULLER 1981; pour les lettres 'a' à 'f', voir texte).

Les expériences de psychoacoustique montrent que la fréquence fondamentale et les harmoniques suivantes sont séparées perceptivement. Mais de la prééminence des harmoniques nous ne voyons pas comment le système auditif peut déduire directement la fréquence physique de F1 lorsque Fo est très élevé, l'échantillonnage des points de mesure fréquentiels étant beaucoup trop faible. Si l'on suppose une intégration à large bande, la position de F1 et Fo n'est plus disponible. Il faut donc conclure logiquement que la règle F1 - Fo ne décrit qu'une règle devant être respectée par les paramètres d'un synthétiseur. Nous verrons plus loin, dans notre propre expérience, que l'hypothèse d'une intégration à large bande ne rend pas compte de nos résultats. Le problème de l'extraction des paramètres pertinents du timbre lorsque Fo est très élevé reste donc intact. Précisons à nouveau que notre travail ne concerne que le cas où les harmoniques sont en nombre suffisant pour qu'on puisse raisonnablement penser que la position du formant auditif est estimée à partir de la prééminence de plusieurs harmoniques. Pour les fréquences fondamentales élevées, il faut imaginer un modèle permettant de mettre en correspondance des spectres dont le degré de définition spectrale varie de manière très importante. Ce redoutable problème ne fait ici l'objet d'aucune étude. Aucune solution satisfaisante n'a été proposée à notre connaissance à ce jour.

Comme hypothèse de travail, nous considérerons ici, pour la gamme limitée de variations de Fo qui ont été testées, que le timbre d'une voyelle est défini par la position absolue des formants. Les résultats des études sur la modification du timbre sous l'effet d'une variation de Fo ne sont pas qualitativement différents pour des fréquences fondamentales entre 100 et 300 Hz. On peut donc tenter de proposer un mécanisme commun, indépendant d'autres hypothèses concernant le mécanisme d'attribution du timbre.

TRAUNMÜLLER (1985) réaffirme les conclusions présentées ci-dessus. L'identité phonétique est principalement déterminée par la position relative des formants sur la membrane basilaire de la cochlée. Pour une fréquence fondamentale supérieure à 160 Hz, et inférieure à 350 Hz environ, la distance tonotopique entre le premier formant et la fréquence fondamentale est le corrélat invariant du degré d'aperture. Il présente ensuite une série d'expériences informelles. Si, pour une voyelle dont les formants supérieurs sont ceux d'un [i] masculin, on augmente la fréquence de Fo et F1 en conservant entre eux une distance de 1 bark, sans modifier la fréquence des formants supérieurs, le timbre reste approximativement le même, même lorsque F1 atteint 840 Hz et Fo 700 Hz. Mais la voix est sentie comme criée. La même augmentation de F1, sans changement de Fo, modifie le timbre jusqu'à [ae]. Si les formants supérieurs sont également déplacés de manière à conserver une distance F1 - Fo constante en Bark, le timbre reste le même, proche de [i]. Mais l'impression est que la voyelle est émise par un conduit vocal de plus en plus petit. Des expériences identiques sur la voyelle [u] montrent qu'il est impératif dans ce cas de conserver la distance relative entre le premier et le deuxième formant si on ne veut pas que le

timbre change. Bien que ces résultats soient très intéressants, ils concernent principalement des fréquences où on ne peut plus parler de formants, surtout en ce qui concerne le premier d'entre eux. Selon nous, ils concernent plus la manière dont l'auditeur rapproche des sons humains certains sons présents dans l'environnement, que les mécanismes perceptifs d'extraction des maxima du spectre. On sait par exemple que des sons purs évoquent des timbres vocaliques différents selon leur fréquence. Cette correspondance approximative ne révèle que peu de choses sur le mécanisme de mesure du formant auditif, mais beaucoup sur le traitement ultérieur de catégorisation phonétique.

Il est possible de tester l'hypothèse d'invariance de la distance $F1 - F_0$ comme paramètre essentiel du timbre sur les données des travaux précédents. Sous cette forme, il ne s'agit pas de l'hypothèse exacte de TRAUMÜLLER, car elle n'est pas limitée à la seule perception du degré d'aperture, mais concerne la totalité du timbre. Ces tests ne sont d'ailleurs pas très puissants, car, dans plusieurs cas, $F2$ et / ou les formants supérieurs covarient avec $F1$. Mais ils conservent une valeur indicative. Des données de FUJISAKI & KAWASHIMA, on peut déduire le tableau suivant, qui présente en Bark, la variation nécessaire pour ramener un stimulus sur la frontière phonétique.

	Variation F_0 (Hz)	Valeur en Bk	Décalage $F1$ (Bk)	Décalage $F2$ (Bk)
[e] - [u]	130 > 290	1.64	0.44	0.61
[a] - [o]	130 > 290	1.64	0.88	0.70

La variation de F_0 est bien plus importante que la variation de $F1$ ou $F2$, même lorsqu'on utilise comme ici les anciennes valeurs de SCHARF (1970) de la correspondance Hz / Bark. Pour les expériences de AINSWORTH sur le couple [U - ae], la pente maximale du changement de $F1$ en fonction de F_0 est de 66 Hz pour une variation de 100 Hz de F_0 . Ceci se traduit au plus par une variation en Bark de $F1$, inférieure de moitié à celle de F_0 . Les valeurs moyennes sont bien inférieures, le rapport n'étant que de 1 à 3. Chez SLAWSON, une variation de F_0 de 1.4 Bark se trouve compensée par une variation de $F1$ et de $F2$ d'au plus 0.8 et 0.9 Bark, respectivement, pour les voyelles ouvertes. La modification est bien moindre pour les voyelles fermées. Il est clair que, bien que le test soit imparfait, l'hypothèse de TRAUMÜLLER étendue à toute modification de timbre n'est confirmée dans aucun de ces cas. Notre propre expérience permettra un test plus fin, également négatif.

Il existe d'ailleurs une explication concurrente pour certains des résultats de TRAUMÜLLER, qui fait appel à l'hypothèse d'une position du formant auditif différente de la fréquence du paramètre de synthèse. A l'expérience 4, les résultats pour la voyelle dont le premier formant est à 350 Hz montrent que faire passer F_0 de 150 à 280 Hz ramène le score de [e] de 60 à 0 %. Celui de [i] passe dans le même temps de 40 à 100 %. Si pour

$F_0 = 150$ Hz, on fait l'hypothèse que le formant auditif a une fréquence voisine de sa valeur physique, il ne peut en être de même pour $F_0 = 280$ Hz. Dans ce cas, la deuxième harmonique est à 560 Hz. Un formant à 350 Hz donne une fréquence fondamentale très intense, et une très faible deuxième harmonique car elle est largement en dehors de l'enveloppe du formant. Le pic auditif est vraisemblablement proche de la fréquence fondamentale, c'est à dire plus bas que pour $F_0 = 150$ Hz. La chute du premier formant auditif pourrait bien, dans ce cas, expliquer la différence de timbre. De la même manière, certains des résultats de la première expérience ne se laisse pas facilement expliquer par l'invariance de $F_1 - F_0$. Pour ces stimuli mono-formantiques (cf. figure 2.9), la frontière entre le troisième niveau d'aperture ([o, ø, e]) et le quatrième [ɔ, oe, ε] se place d'abord de manière constante entre 450 et 475 Hz : cf. notre ligne A-B sur la figure 2.9. Puis, pour F_0 supérieur à 200 Hz, elle change brutalement pour suivre la ligne médiane entre la 2ème et la 3ème harmonique (notre ligne C-D). la frontière passe donc de 475 à 700 Hz quand F_0 varie de 200 à 300 Hz. Ceci suggère la règle de décision suivante. Quand la fréquence formantique est plus proche de la deuxième harmonique, celle-ci sera plus proéminente, et déterminera la position du maximum auditif. L'inverse sera vrai lorsque le formant est plus proche de la troisième harmonique. Considérons le point marqué E sur la figure. Un tel stimulus possède une fréquence fondamentale de 300 Hz et un formant à 700 Hz. Mais comme cette dernière valeur est plus proche de la seconde harmonique (600 Hz) que de la troisième (900 Hz), on peut inférer que sa position auditive est celle du point F, si le formant auditif est principalement défini par la position de l'harmonique la plus intense. Une valeur de 600 Hz réduit, mais ne supprime pas, l'écart entre la position de la frontière aux fréquences fondamentales hautes et basses. Le décalage de la frontière est dans ce cas inférieur à la variation de F_0 exprimée en Bark. On peut suggérer que le changement dans la règle de décision se produit lorsque l'espacement entre les harmoniques dépasse deux bandes critiques.

Deux conclusions se dégagent de cette discussion. Tout d'abord, le critère d'une distance constante en Bark entre F_0 et F_1 ne rend pas compte des données disponibles, du moins lorsqu'on l'applique à une différence quelconque de timbre. Deuxièmement, une discussion de la modification du timbre par F_0 ne peut se contenter des valeurs formantiques utilisées comme paramètres de synthèse. Le faire reviendrait à supposer que le formant auditif est à la même valeur que le formant physique, ce qui est une hypothèse très contraignante, et peu plausible lorsque les harmoniques sont écartées.

Ces considérations ont été de celles qui nous ont conduit à réaliser l'expérience de mesure psychophysique de la position du formant auditif qui sera décrite plus loin (LONCHAMP 1985), puis une nouvelle expérience de détermination de la variation de timbre en fonction de la valeur de la fréquence fondamentale.

2.11 - SUNDBERG & GAUFFIN (1982)

Le problème soulevé dans ce travail est celui du rôle potentiel de l'amplitude de la fréquence fondamentale sur le timbre de voyelles chantées à des fréquences fondamentales très élevées. A partir de résultats d'identification obtenus sur des voyelles synthétiques, SUNDBERG & GAUFFIN concluent que l'augmentation de l'amplitude de F_0 , qui est souvent aussi un corrélat important du premier formant des voyelles les plus fermées, correspond perceptivement à un abaissement du formant. C'est le cas de F_1 pour les voyelles fermées, et également de F_2 lorsque celui-ci est bas, dans les voyelles postérieures. Ce résultat est compatible avec l'hypothèse d'un calcul de la position du formant auditif à partir de plusieurs harmoniques. Un F_0 plus intense abaisse la position du formant définie par F_0 et l'harmonique suivante. Le point délicat est que la taille de cet effet dépend des modalités de présentation du test : il est plus fort si le test contient une plus grande diversité d'amplitude de F_0 . La contribution d'une comparaison auditive consciente, d'origine non perceptive, ou d'un mécanisme d'appariement du timbre phonétique faisant appel à d'autres informations que la fréquence des maxima auditifs, n'est donc pas exclue.

2.12 - MOORE & GLASBERG (1983 b)

Cette étude des spectres psychophysiques des voyelles [ae] et [i] reste la plus complète à ce jour quant au nombre de sujets étudiés à des fréquences proches de F_1 et mérite d'être scrutée attentivement, bien qu'elle ait été conduite à une seule fréquence fondamentale (100 Hz). Une "représentation interne" ou "spectre auditif" peut être mesurée par une technique de masquage si on fait l'hypothèse que le seuil de perception d'un signal test ("probe") est déterminé par le niveau d'un son masquant auquel on donne le spectre d'une voyelle. L'évolution, en fonction de la fréquence, du niveau de seuil du son-test sera le reflet du spectre interne de la voyelle.

HOUTGAST (1974) a été le premier à déterminer le spectre auditif de sons vocaliques par la méthode du seuil des pulsations. Cette méthode consiste à insérer un son sinusoïdal de fréquence déterminée dans l'intervalle séparant les (nombreuses) répétitions d'une même voyelle servant de masque. Les durées du son à tester et de l'intervalle entièrement rempli par la sinusoïde sont d'ordinaire égales à 125 ms. On constate que dans une plage relativement étroite d'intensité, la sinusoïde paraît continue, "traversant" sans discontinuité audible la voyelle dont on veut mesurer le spectre. En dehors de cette plage, la sinusoïde semble soumise à une pulsation d'intensité, d'où le nom de la méthode. L'interprétation classique est que l'absence de pulsation correspond à l'établissement d'un même degré

d'excitation auditive par les composantes spectrales du son testé et par la sinusoïde. C'est cette absence de variation dans le degré d'excitation auditive à une fréquence donnée qui provoquerait la continuité auditive de la composante spectrale. Déterminer le spectre auditif consiste donc à mesurer, pour des fréquences rapprochées d'une bande donnée, l'énergie de la sinusoïde correspondant à l'absence de pulsations. Les remarquables travaux d'ESCUDIER & SCHWARTZ à Grenoble (cf SCHWARTZ 1987, ESCUDIER & SCHWARTZ 1985 a, 1985 b) sur les fondements théoriques et méthodologiques de la méthode du seuil de pulsation en font une méthode certes lourde à mettre en oeuvre, mais riche de promesses. Leurs études ont surtout porté sur la zone des formants supérieurs, qui nous intéresse moins ici. Notons pourtant qu'ils ont mis en évidence qu'un petit pic de 2 dB placé à 1300 Hz sur le flanc droit d'un formant à 400 Hz était parfaitement détectable par la méthode, et donc par le système auditif. L'effet de suppression des fréquences immédiatement inférieures (800 - 1100 Hz) est minime, si tant est qu'il existe, et ne provoque donc pas une mise en relief de l'irrégularité spectrale. Le pic auditif apparaît un peu plus large, et moins intense que le pic acoustique. Ils montrent enfin que l'amplitude d'un formant de basse fréquence dépend peu de la pente spectrale. Toutes ces caractéristiques peuvent trouver, selon eux, une explication dans un équilibre entre l'effet de suppression du premier formant sur le petit pic, la suppression causée par le petit pic sur les fréquences plus basses et un léger lissage du spectre due à la sélectivité fréquentielle limitée du système auditif périphérique. Ce dernier facteur jouerait d'ailleurs le rôle le plus important.

Les spectres auditifs d'HOUTGAST (1974) sont reproduits à la figure 2.10. On remarque une très bonne correspondance entre les maxima spectraux auditifs et acoustiques, ainsi que la présence d'effets de suppression, à gauche du deuxième formant de [i] par exemple. Vers 1800 Hz, le spectre auditif est "creusé" relativement au spectre acoustique. Dans la zone de F1, on ne note aucune divergence entre les deux courbes pour les deux voyelles non ouvertes. Avec la même méthode, TYLER & LINDBLUM (1982) ont étudié, chez un unique sujet, des voyelles [i], [e] et [y] synthétiques (figure 2.11 : cercles et ligne pointillée). On note que les premiers formants sont peu visibles : ils semblent soumis à une forte suppression venue de la fréquence fondamentale ou du formant glottal. BACON & BRANDT (1982) présentent également des spectres auditifs mesurés par le seuil de pulsation pour des sujets à audition normale et pathologique (figure 2.12). Pour la voyelle [a], les sujets S1 et S2 montrent un pic auditif pour F1 et F2. Ce n'est pas le cas des deux autres sujets. Deux sujets également (S1 et S4) ont des spectres auditifs conformes au spectre acoustique pour la voyelle [ae]. Dans aucun des cas, le pic auditif de F1 ne semble décalé par rapport à la fréquence acoustique.

Les résultats d'une expérience préliminaire conduite à Grenoble, où nous servions d'unique sujet, sont présentés à l'annexe 2. Ils montrent que, pour les fréquences fondamentales

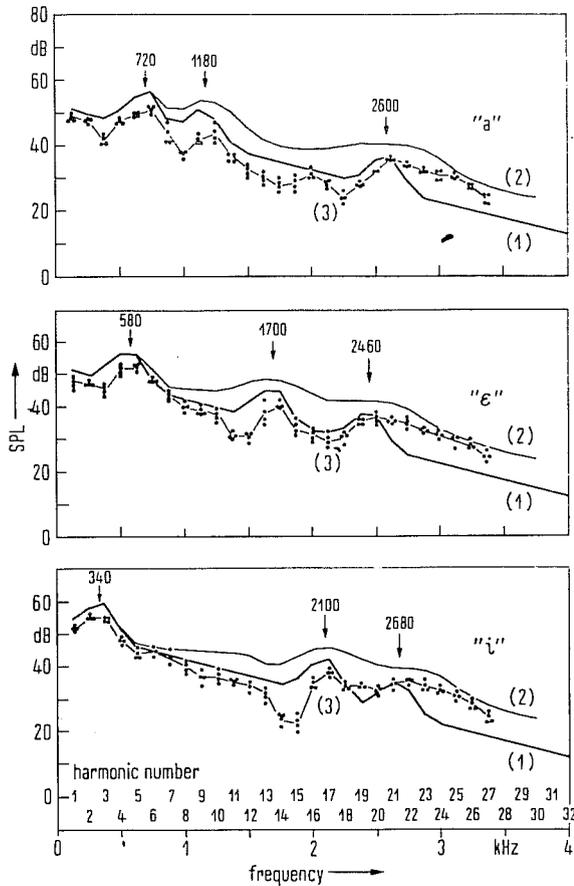


Fig. 3. For each of the three vowel-like sounds three curves are given:
 (1) the level of each of the 32 harmonics (note the formant-like structure),
 (2) the output level of 1/3 octave-band filter centered at the first 30 harmonics and
 (3) the pulsation-threshold levels for a test tone centered at the first 27 harmonics (one subject).

Figure 2.10

Figure 2.10 - Spectres auditifs de 3 voyelles (courbe 3):
 méthode du seuil de pulsation (in HOUTGAST 1974)

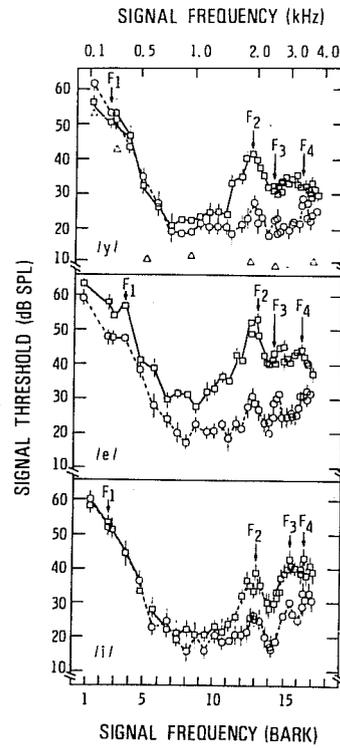


FIG. 1. Vowel masking for the vowels /y/, /e/, and /i/. Signal thresholds (in dB SPL) are shown as a function of signal frequency in the presence of the vowel for simultaneous-masking (squares) and pulsation-threshold (circles) techniques. Each datum point represents the average of five threshold estimates, with the vertical bars representing ± 1 standard deviation. The frequency location of the harmonic with the largest amplitude near the four formants is shown by F_1 , F_2 , F_3 , and F_4 and associated inverted arrows. In the top panel, the open triangles represent signals thresholds in quiet.

Figure 2.11

Figure 2.11 - Spectres auditifs de 3 voyelles (cercles: méthode du seuil de pulsation, in TYLER & LINDBLOM 1982).

basses, la position du formant auditif est conforme à ce que le spectre physique permet de prévoir. En revanche, ni le spectre physique, ni le spectre auditif d'un [a] émis à un F_0 de 350 Hz ne possèdent de formants caractérisés. Seuls les harmoniques individuelles sont repérables.

Revenons aux travaux de MOORE & GLASBERG. Les résultats obtenus en masquage simultané peuvent être contaminés par des interactions entre le son-test et le masque tels que battements et tons de combinaison. En masquage proactif ("forward masking") le seuil n'évolue pas en proportion du niveau de présentation du masque. Il faut donc une correction dont le détail n'a pas sa place ici. Les masques vocaliques étaient des sons synthétiques formés par l'addition de sinusoïdes multiples de 100 Hz en phase. Remarquons qu'un tel choix, dicté apparemment par la volonté d'avoir de manière simple un son-test de même phase que l'harmonique testée, n'est pas le plus judicieux, car l'enveloppe temporelle à grande excursion qui en résulte détériore la qualité naturelle du son, qui est déjà faible sans cela. Une expérience de contrôle ultérieure a montré qu'en haute fréquence (au delà de 1.5 kHz), l'amplitude des fluctuations temporelles n'a pas d'importance, mais les fréquences autour de F_1 n'ont pas été testées. Les amplitudes des harmoniques étaient copiées sur celles d'une voyelle naturelle (figure 2.13). Le premier formant du [i] n'est pas clairement visible. Les stimuli étaient présentés à trois niveaux d'intensité (50, 70 et 90 dB SPL) à trois sujets. Ne considérant que la représentation interne dans la zone du premier formant, on note en masquage simultané (figure 2.14) l'absence de F_1 pour [i], comme sur le spectre acoustique, mais également une pente descendante d'environ -12 dB / octave jusqu'à 800 Hz. L'harmonique à 200 Hz semble de ce fait provoquer un effet de masque supérieur à celle de 300 Hz, dont l'intensité physique est pourtant identique. Mais cette interprétation est incorrecte. Pour obtenir le spectre auditif réel, il ne faut pas oublier d'inclure la pente de + 6 dB / octave de la fonction de transfert de l'oreille moyenne et externe (cf. BASTET 1978 par ex.). Les mesures des spectres auditifs sont en effet relatives, l'effet spectral des éléments placés avant l'oreille interne étant le même pour le son-test et le masque. Pour [ae] (figure 2.15), le pic auditif coïncide avec la fréquence du formant. A la plus forte intensité, ce pic s'étend à l'harmonique inférieure pour deux sujets. En masquage proactif, le spectre auditif du [i] présente la même forme qu'en masquage simultané, avec une pente encore plus raide (15 - 20 dB / octave) dans certains cas. Pour [ae], le premier "formant auditif" est bien marqué, mais, pour deux sujets (B.M. et H.M.), il s'étend au delà de la fréquence physique au trois niveaux d'intensité testés. Pour le sujet B.G, l'inverse est vrai à 90 dB SPL, mais le niveau absolu bien plus élevé indique des difficultés de détection. La vallée creusée à gauche de F_1 , plus profonde dans le spectre auditif que dans le spectre acoustique, est classiquement interprétée comme la trace d'un effet de suppression psychophysique. Les harmoniques intenses affaiblissent les harmoniques plus faibles de fréquence inférieure. Le spectre auditif résultant est ainsi

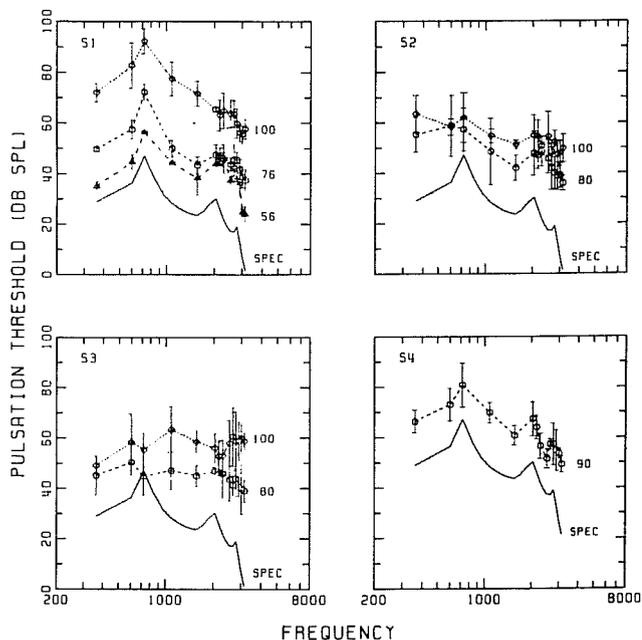


FIGURE 3. As in Figure 2, for the vowel masker /æ/.

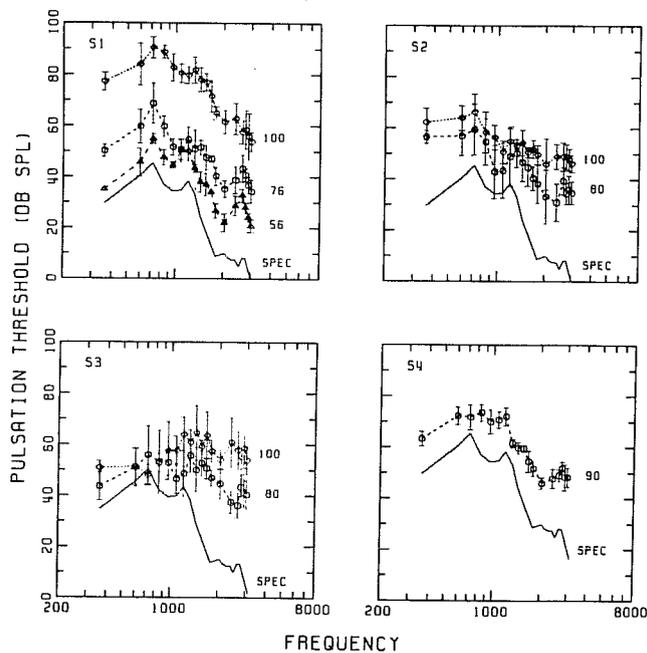


FIGURE 2. /a/ vowel pulsation patterns for the four normal-hearing listeners, with masker level (dB SPL) as the parameter. The acoustic spectrum is displaced downward by an arbitrary amount.

Figure 2.12 - Spectres auditifs des voyelles |a| et |æ| (in BACON & BRANDT 1982). Méthode du seuil de pulation.

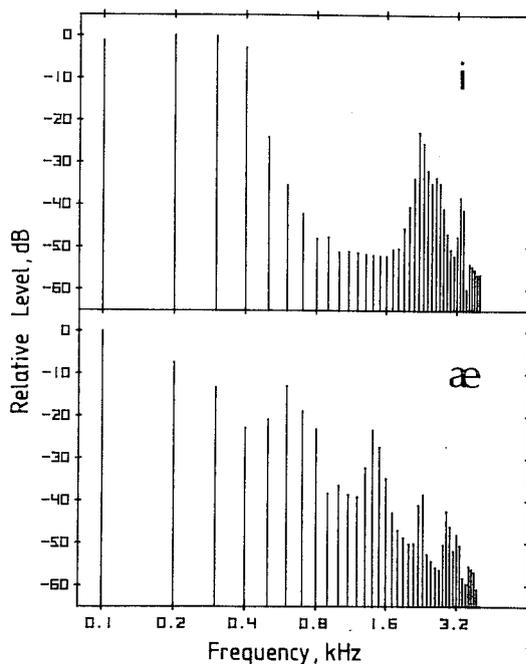
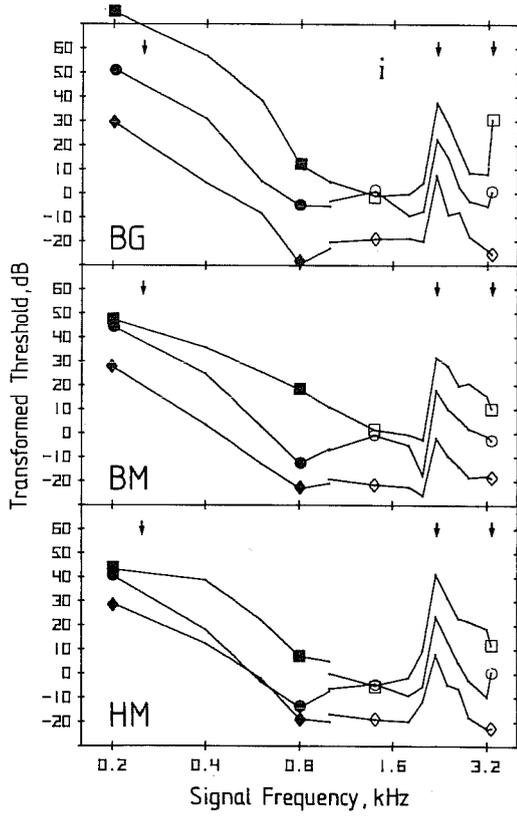


FIG. 1. Spectra of the two synthetic vowels as generated by the computer, measured with a Hewlett-Packard 3582A analyzer. The level of each harmonic is expressed relative to the level of the most intense harmonic.

Figure 2.13 - Spectres physiques des stimuli synthétiques utilisés par MOORE & GLASBERG (1983 b).



Masking patterns in forward masking for the vowel /i/.

Figure 2.14

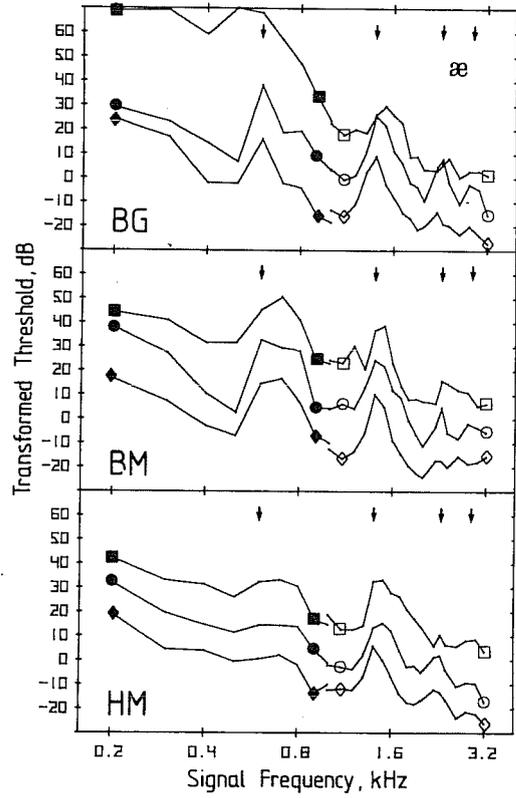


Figure 2.15

Figure 2.14 - Spectres auditifs de la voyelle [i] : masquage proactif (MOORE & GLASBERG 1983 b)

Figure 2.15 - d° : voyelle [æ].

plus contrasté que l'original. Il n'est pas facile d'invoquer la suppression pour expliquer le décalage du premier formant auditif. Si le flanc gauche est creusé, celui de droite n'a aucune raison d'être rehaussé. Pour plaider que la suppression change le "centre de gravité" du formant en "mordant" sur son flanc gauche, il faudrait admettre que celui-ci a une largeur de bande auditive très importante, s'étendant sur trois harmoniques (250 Hz) au minimum. Le même effet existe pour F2 au niveau le plus intense. MOORE & GLASBERG ne commentent pas ce détail et n'offrent donc aucune explication. Phonétiquement, on pourrait penser qu'une voyelle ouverte doit apparaître légèrement plus ouverte (F1 plus élevé) que son analyse physique ne l'indique. Mais si l'effet est constant, il sera bien sûr indétectable en raison du manque de référence absolue.

2.13 - LONCHAMP (1985)

Notre travail préliminaire n'avait pas pour but immédiat de proposer un modèle de l'interaction timbre - Fo, mais d'accumuler des informations sur la forme du spectre auditif dans la région du premier formant, en vue de tester certaines hypothèses de CARLSON & al. (1975) qui ont été décrites en détail plus haut. Dans cette présentation, nous corrigerons ce qui nous apparaît aujourd'hui comme une erreur d'interprétation.

Un modèle adéquat de perception de la parole doit décrire à la fois comment les indices acoustiques sont transformés en indices perceptifs, et la manière exacte dont ces derniers sont combinés et pondérés dans le processus de prise de décision phonétique. Un exemple de cette démarche est l'hypothèse de CARLSON & al., selon laquelle le corrélât perceptif du premier formant est "calculé" comme une somme pondérée des deux harmoniques les plus intenses en échelle de sonie. Le test est la capacité de ce "calcul" à prédire l'évolution de la frontière entre [i] et [e] en fonction de la fréquence fondamentale.

Les spectres auditifs peuvent être mesurés par deux techniques psychophysiques : la méthode du seuil de pulsation, employée par TYLER & LINDLOM (1982) et ESCUDIER & SCHWARTZ (1985) par exemple, et celle du masquage proactif (ou postérieur : "forward masking" : le terme "forward", "antérieur" en français, indique que le masque précède le son-test). Cette technique a été utilisée par MOORE & GLASBERG (1983) ou SORIN (1987). Un bref son test sinusoïdal ou de bruit à bande étroite suit un son masque dont on veut étudier le spectre auditif. On constate expérimentalement en neurophysiologie que le son servant de masque produit une adaptation se traduisant par une diminution de la cadence de décharge moyenne dans les fibres nerveuses excitées par le masque. Cette adaptation se poursuit quelques dizaines de ms après la fin du masque. Si le son-test est placé dans cet intervalle, son seuil d'audibilité est fonction de l'intensité du son masquant dans la même bande de fréquence. Si on modélise en première approximation le système auditif comme une batterie de filtres à fort taux de recouvrement, dont les largeurs de bande

sont égales aux valeurs des bandes critiques déterminée psycho-expérimentalement (SCHARF 1970), les harmoniques de basse fréquence doivent être présentes individuellement dans le spectre auditif. Les estimations récentes de la largeur rectangulaire équivalente des filtres auditifs (MOORE & GLASBERG 1983 a) sont en fait plus basses que les valeurs anciennement acceptées (par exemple 38 Hz à 100 Hz, 77 Hz à 500 Hz, contre 100 Hz à 150 Hz, et 110 Hz à 450 Hz pour SCHARF). On peut donc prévoir que les harmoniques de basse fréquence seront toujours résolues, même pour les fréquences fondamentales les plus basses. C'est bien ce qu'on constate dans les mesures par masquage proactif de MOORE & GLASBERG (1983 c) sur des sons complexes contenant les 10 premières harmoniques de 100, 200 ou 400 Hz. Les trois ou quatre premières harmoniques sont visibles. Le contraste spectral minimal que l'on peut détecter, en raison de la variance des résultats, est d'environ 3 dB. Il est clair que le masquage proactif sous-estime la résolution auditive. Notre choix s'est porté sur le masquage proactif en raison des critiques émises par MOORE sur la méthode du seuil de pulsation. ESCUDIER & SCHWARTZ (1985 a, b) ont depuis montré que le seuil de pulsation fournissait des informations fiables, sous réserve de se prémunir contre deux biais expérimentaux. La durée du son-test était fixée à 20 ms. Sa largeur de bande intrinsèque ne le rend pas optimal pour la détection des harmoniques. Mais elle présente l'avantage de donner une valeur plus favorable à l'accroissement relatif du seuil de détection relativement à l'accroissement de l'intensité locale du spectre. Les formants seront donc plus nettement marqués sur le spectre auditif. L'auteur comme seul sujet a testé, entre 200 et 600 Hz par pas de 50 Hz, trois voyelles synthétiques de timbre [e] : V1 et V2, à une fréquence fondamentale de 101 Hz et avec F1 placé à 400 ou 450 Hz; V3 possédant un Fo plus élevé (150 Hz) et un F1 à 450 Hz. Les autres paramètres de la synthèse série étaient les suivants : F2 = 1800 Hz; F3 = 2750 Hz, F4 = 3600 Hz, largeur de bande de F1 (Bd 1) = 50 Hz, Bd2 = 70, Bd3 = 100, Bd4 = 120 Hz ; onde glottale polynomiale de ROSENBERG : coefficient d'ouverture = 0.40, coefficient de fermeture = 0.25 ; durée 190 ms ; attaque adoucie par une fenêtre demi-cosinus de 5 ms.

Le choix des paramètres principaux avait pour but de tester les points suivants :

- 1 - Le pic auditif est-il plus élancé lorsque le pic spectral coïncide avec une harmonique (V1 et V3) que quand il tombe entre deux harmoniques (V2) ?
- 2 - Le pic auditif est-il mieux marqué lorsque les harmoniques adjacentes, plus éloignées, sont d'intensité plus faible (V3 / V2) ?
- 3 - Le phénomène de suppression provoque-t-il un fossé sur le flanc gauche du formant, le mettant ainsi en relief, comme dans les données de MOORE & GLASBERG (1983 a) sur le premier formant d'un [ae] vers 800 Hz ?
- 4 - La forme globale du pic est-elle conforme à ce que la transformation de sonie implique ?

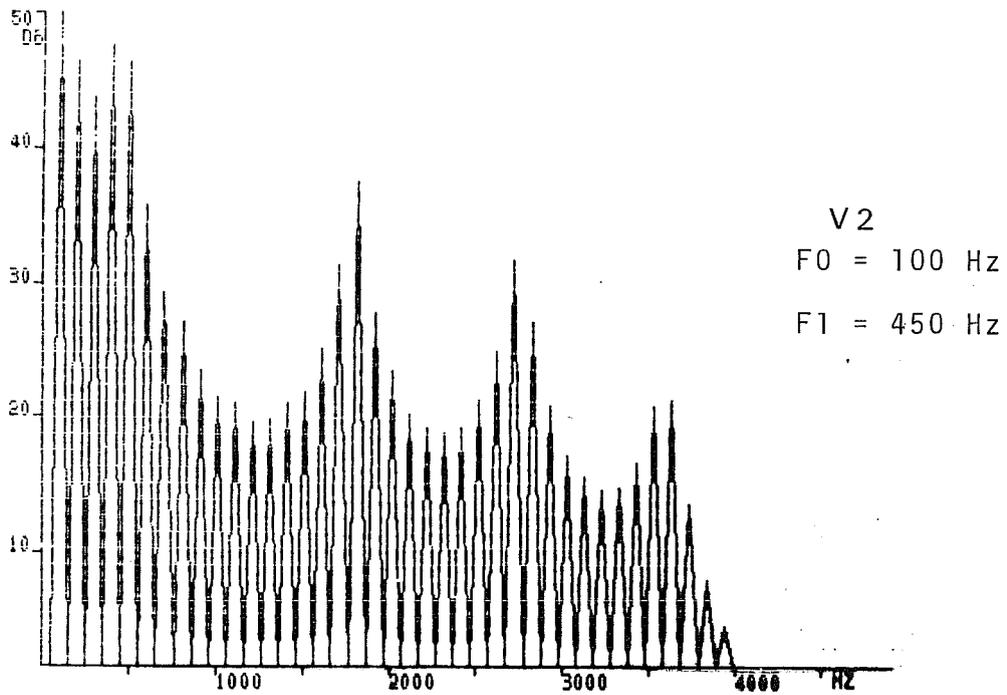
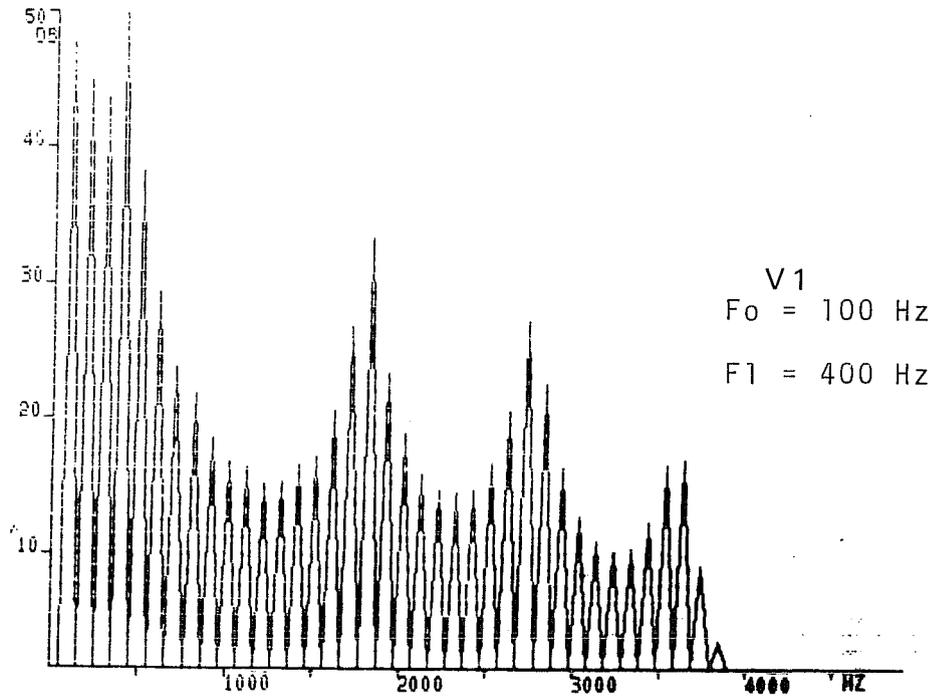
Les spectres F.F.T. des trois voyelles sont présentés à la figure 2.16. On notera que la 4ème et 5ème harmoniques de V2 ($F_1 = 450$ Hz) sont presque au même niveau, 4 dB au-dessus de la 3ème, alors que la 4ème harmoniques de V1 ($F_1 = 400$ Hz) est 6.5 dB au-dessus de la 3ème. La troisième harmonique de V3, à la fréquence formantique, domine la troisième de 8.5 dB. L'origine de l'échelle en dB est arbitraire.

La procédure de test était copiée sur celle de MOORE & GLASBERG. Les paramètres du son-test étaient : durée 20 ms ; temps de montée / descente 5 ms ; intervalle masque - test de 5ms mesuré à - 6 dB ; convertisseur A/D 12 bits à 10 kHz ; casque AKG K 242 ; écoute dans un studio d'enregistrement insonorisé ; niveau 72 SPL. La procédure adaptative avec feedback de LEVITT, avec une augmentation de + 2dB après une erreur et une chute de -2 dB après 2 réponses correctes, étaient pilotée par l'ordinateur. Les résultats graphiques (figure 2.17) portent sur les 5 dernières séries sur 10 à 12, et donnent les seuils bruts, sans la transformation de compensation de la compression du seuil de MOORE & GLASBERG. Les premières séries ont fourni l'entraînement indispensable. D'après leurs résultats moyens, on peut estimer la compression à un facteur 0.5. Une variation de l'énergie du masque de + 10 dB donne donc une élévation du seuil d'environ 5 dB. L'origine de l'échelle en dB est arbitraire. Les valeurs SPL sont inférieures d'environ - 19 dB.

La variance des seuils paraît grande de prime abord quand on considère les moyennes des 5 séries. Mais la variance des estimations à l'intérieur d'une série (10 déterminations du seuil pour chaque fréquence) est plus faible. Les moyennes et écarts-type de V1 à 400 Hz s'établissent par exemple à : 63.4 (0.7); 66.6 (1.3); 69.6 (1.3); 70.6 (1.4); 66.4 (0.7). L'écart-type des valeurs moyennes à cette fréquence est de 2.8 dB, mais il ne dépasse pas 1.4 dB pour les mesures d'une même série. MOORE & GLASBERG ont également noté une augmentation de la variance aux basses fréquences. Leur écart-type moyen (1.5 dB) est très semblable au nôtre (1.7 dB). On n'observe aucune corrélation entre la variance et le niveau d'excitation.

Pour la voyelle V1 ($F_0 = 101$ Hz; $F_1 = 400$ Hz) , il n'y a pas de pic clairement visible à la fréquence formantique, et les écarts à 400 Hz sont importants. Les résultats semblent plus stables au-dessus et en dessous de cette fréquence (300, 500 Hz). Les harmoniques ne sont pas visibles individuellement. La forme globale du spectre peut être décrite comme une épaulement de 100 à 150 Hz de large, sur une pente fortement descendante.

La voyelle V2 ($F_0 = 101$ Hz; $F_1 = 450$ Hz) possède les mêmes caractéristiques. Les écarts restent importants, sauf à 250 Hz. La plus grande stabilité des résultats au-dessous et au-dessus de la fréquence du formant n'est plus évidente ici. On note même, surtout si l'on prend en compte les résultats de deux estimations antérieures du niveau spectral à 350 Hz notés par des croix (57.7, 61.5 dB), une petite dépression sous la fréquence formantique. L'épaulement décrit plus haut semble plus large, s'étendant à une plus haute fréquence que pour V1, ce qui correspond bien à une fréquence formantique plus élevée.



• Figure 2.16 - Spectres physiques des 3 voyelles utilisées dans l'expérience de masquage proactif.

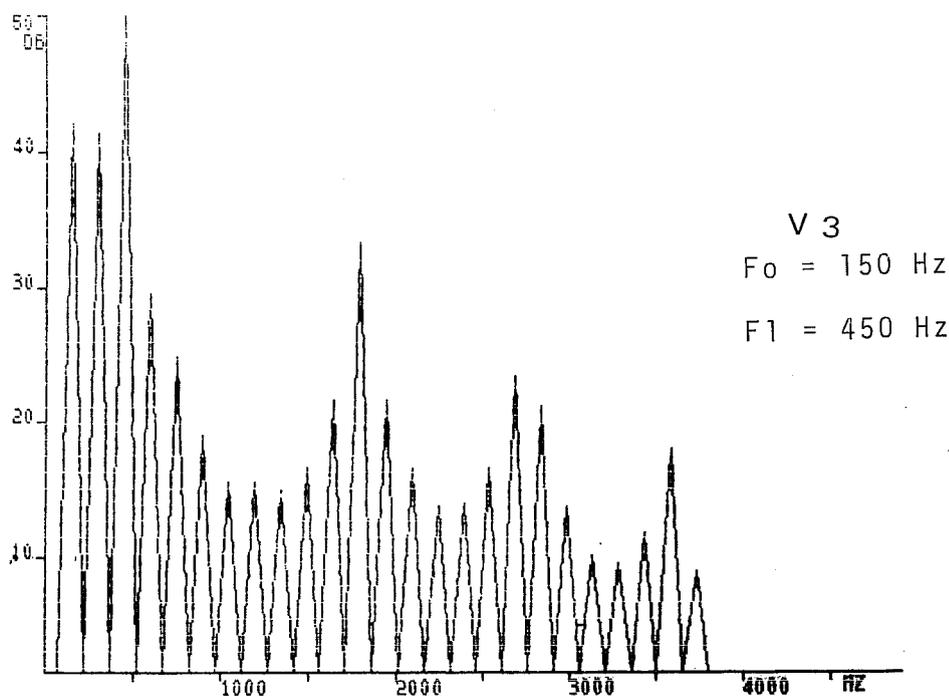


Figure 2.16 (Suite) - Spectres physiques.

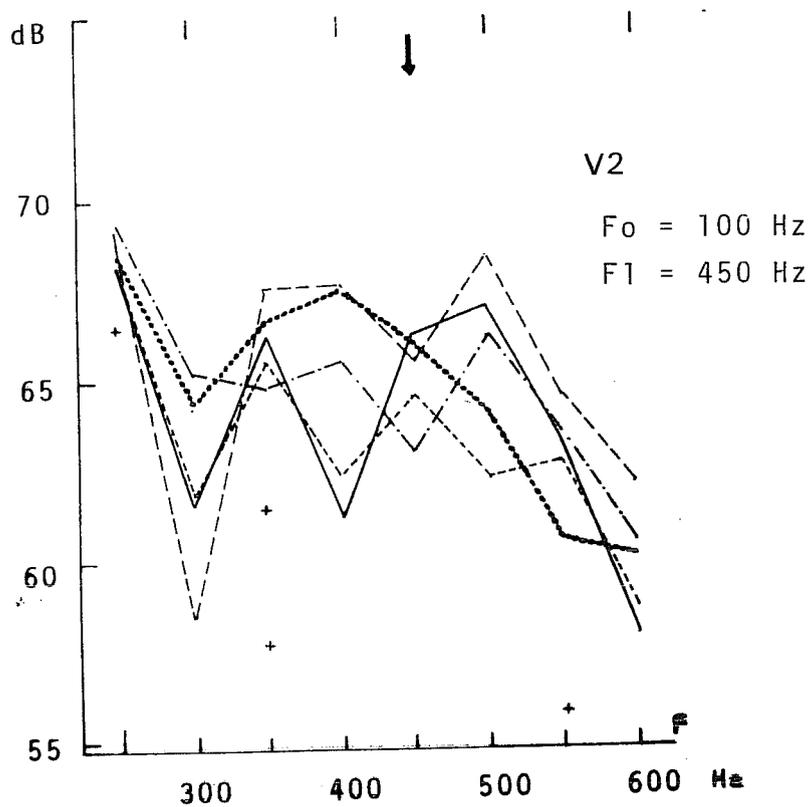


Figure 2.17 - Seuil de masquage proactif pour le stimulus V2 : 5 estimations; La flèche est à la fréquence harmonique, les lignes à celles des harmoniques; Le niveau SPL est inférieur de 19 dB.

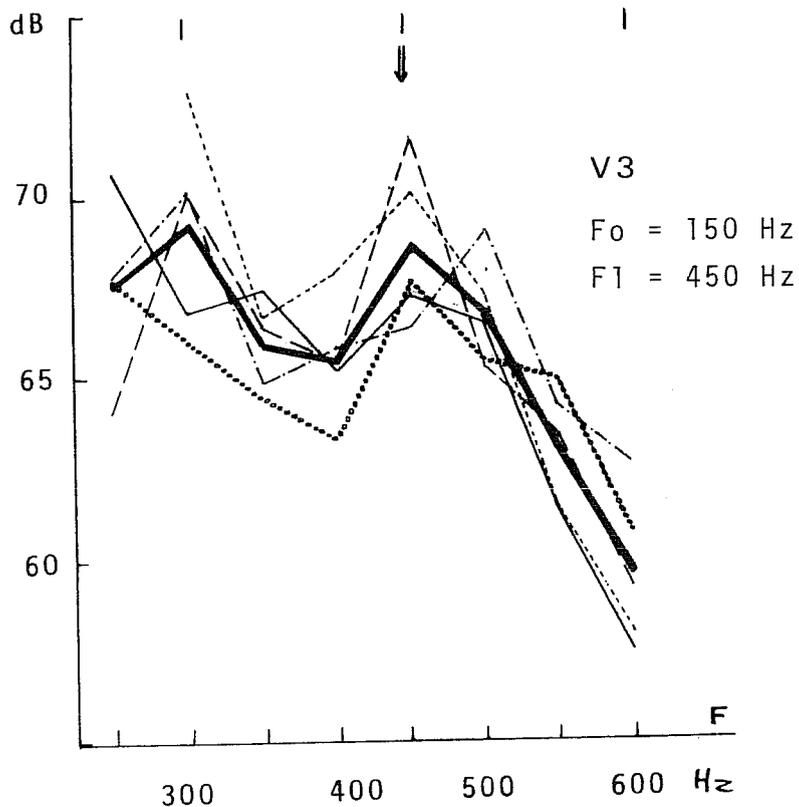
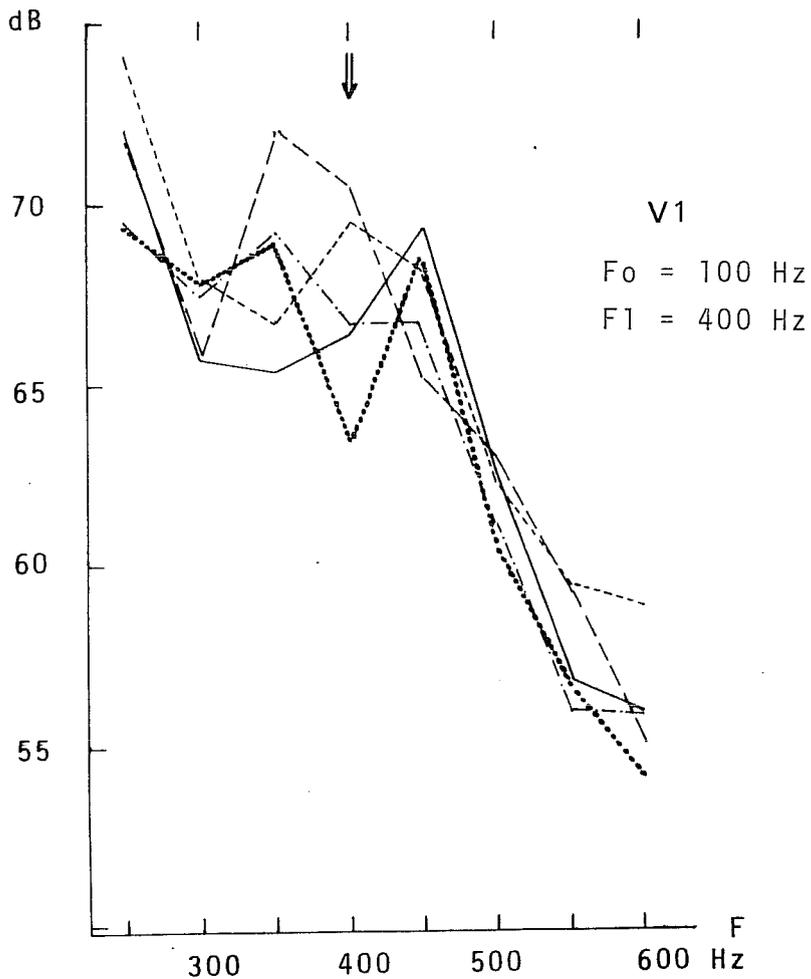


Figure 2.17 (suite). Seuil de masquage proactif : pour V3, la moyenne est indiquée par la ligne noire plus épaisse.

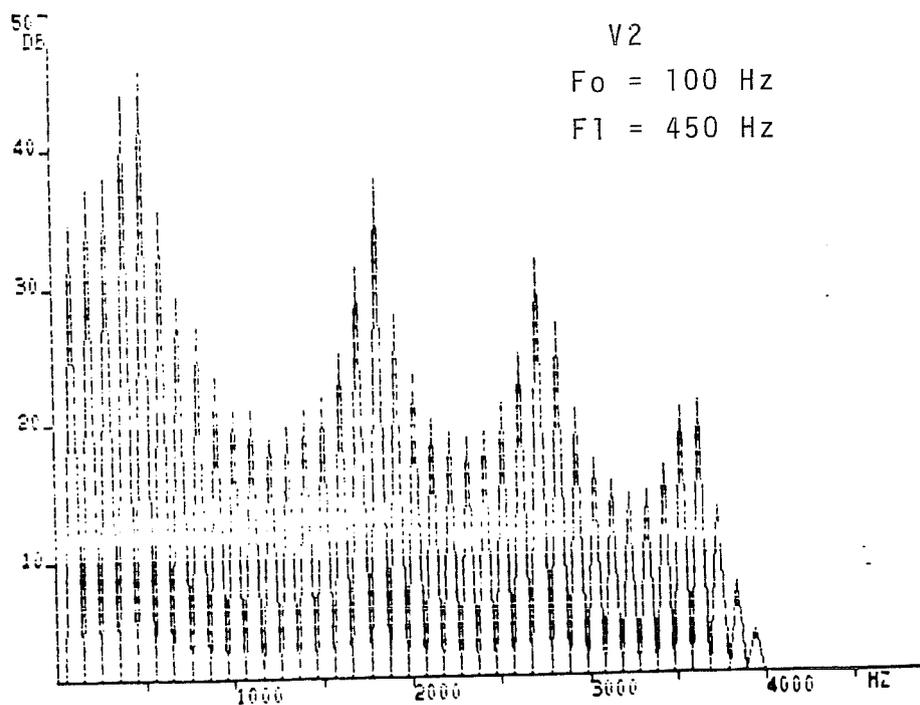
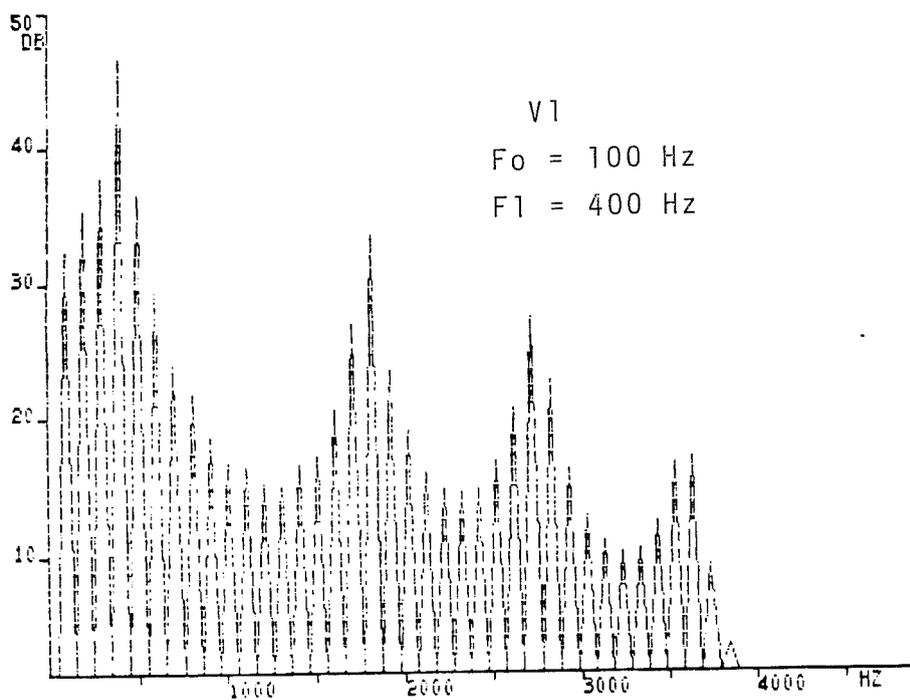


Figure 2.18 - Spectres de sonie des stimuli V1 à V3 : une correction de -6 dB / octave est appliquée au spectre physique sous 600 Hz .

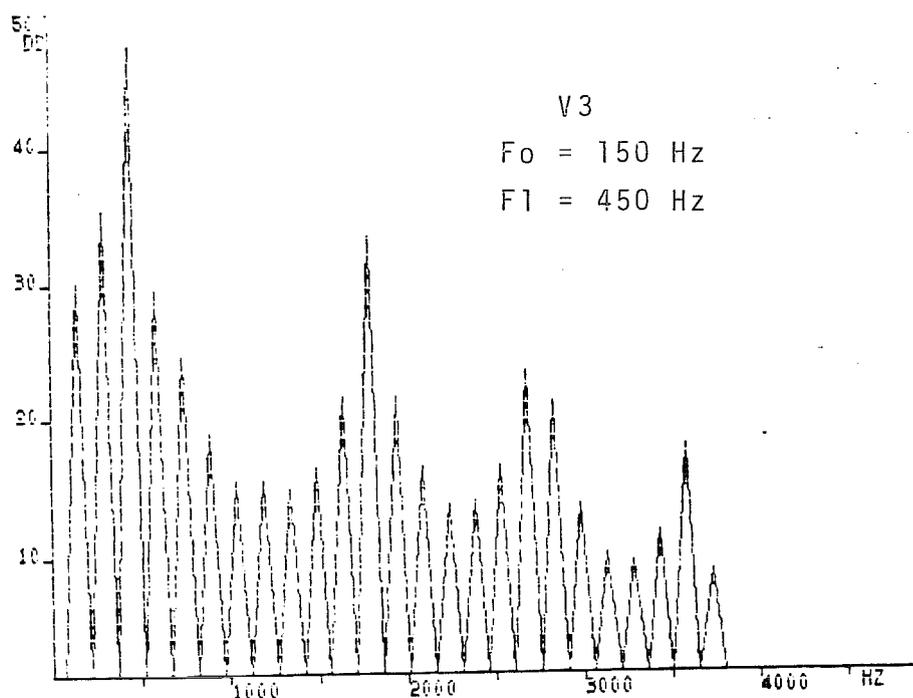


Figure 2.18 (Suite).

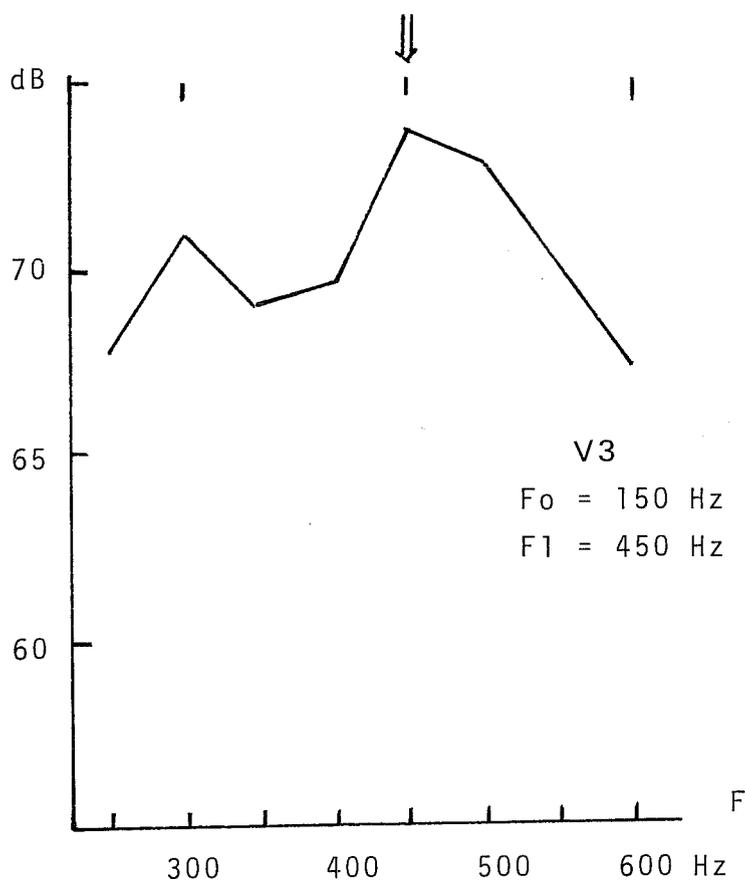


Figure 2.19 - Seuil de masquage moyen de la voyelle V3 avec correction de la pr emphase de l'oreille externe et moyenne.

Dans le cas de la voyelle V3 ($F_0 = 150$, $F_1 = 150$ Hz), on peut distinguer un petit pic à la fréquence formantique qui coïncide avec une harmonique intense, à la fois sur les résultats individuels et sur la courbe moyenne en trait gras. Il n'est pas possible de dire s'il y a également un pic à la deuxième harmonique, mais on remarque une dépression vers 350 - 400 Hz.

Dans notre présentation de 1985, nous insistions dans la discussion sur la différence nette de pente spectrale entre le spectre auditif et le spectre physique. Nous ne tenions pas compte, et c'est l'une des erreurs que nous voulons corriger dans cette discussion, de l'effet correctif de + 6 dB / octave jusqu'à 1000 Hz apporté par l'oreille externe et moyenne (cf. DALLOS (1970); BASTET 1978 pour une revue). Ce filtrage passe-haut a déjà été mentionné dans notre présentation des résultats de MOORE & GLASER (1983 a). Nous remarquons par exemple que pour la voyelle V3, le pic à 450 Hz n'était pas plus élevé que l'harmonique à 300 Hz, alors que ce dernier est physiquement plus bas de 8.5 dB. La différence était encore plus frappante dans la comparaison avec les spectres de sonie dont CARLSON & al (1975) suggèrent qu'ils permettent le calcul du pic auditif de basse fréquence. Ces spectres de sonie incorporent notamment une pente de + 6 dB / octave en basse fréquence (jusqu'à 600 Hz). Les spectres physiques des voyelles ainsi modifiés sont présentés à la figure 2.18. La différence entre la deuxième et la troisième harmonique (formant) de V3 est alors de 12.1 dB. Devant cette différence entre les spectres auditifs et les spectres de sonie, nous concluons que la suggestion de CARLSON & al. n'était pas conforme à nos résultats. Nous suggérons que nos données étaient plus en accord avec la relation d'"égale importance psychophysique" ("equal significance relation") déterminée par MUSHNIKOV & CHISTOVICH (1972). Rappelons que pour des stimuli de timbre intermédiaire entre [e] et [i] ne comportant que deux harmoniques dans la région de F_1 (vers 400 Hz), dont on pouvait modifier individuellement les fréquences, les chercheurs de l'Institut Pavlov de Leningrad montraient que le niveau de l'harmonique basse requis pour garder le stimulus sur la frontière phonétique devait décroître si sa fréquence était abaissée. A niveau sonore constant, une harmonique de basse fréquence devait donc être plus "proéminente", en terme d'effet sur le timbre, qu'une harmonique de fréquence plus élevée, bien que sa sonie soit moindre. Ils ont baptisé "equal significance (curve)", la pente spectrale de - 6 dB / octave sous 600 Hz nécessaire pour rendre compte de l'effet, sur le timbre, de la modification de la fréquence d'une harmonique. Appliquée aux spectres physiques de nos voyelles, cette pondération rapprochait les spectres physiques et auditifs. Pour V3, la différence entre la deuxième et la troisième harmonique était ramenée à 5 dB, proche du minimum détectable par la méthode de masquage proactif.

Mais si l'on tient compte des effets de l'oreille externe et moyenne (figure 2.19), la voyelle V3 présente un pic net à la fréquence du premier formant. Il domine d'environ 6 à 7 dB l'harmonique 2, si l'on se souvient du fait que nos spectres auditifs sous-estiment d'un

facteur 2 les différences spectrales. Le spectre auditif présente donc un pic dont l'écart avec l'harmonique placée en dessous est à une valeur proche de celle du spectre physique brut (8.5 dB). Cette valeur est inférieure à celle que prévoit un spectre de sonie (12 dB). Il est clair cependant que la variabilité des estimations rend toute conclusion tranchée impossible. L'hypothèse de CARLSON & al. n'est pas confirmée, ni infirmée, en raison de la variabilité des mesures subjectives. Au contraire, on peut plus facilement rejeter notre conclusion de 1985 quant à la bonne valeur prédictive d'un spectre d'"égale importance". Il ne nous paraît pas possible de suggérer que cette discussion est fondée sur une erreur de méthode parce que nous négligeons les différences de pente spectrale dans leur contribution à l'amplitude du pic auditif. Nous avons rappelé plus haut que ESCUDIER & SCHWARTZ (1985 a) ont montré que l'amplitude d'un pic de fréquence basse était relativement indépendante de la pente de ses flancs.

On notera enfin que le formant de V3 semble légèrement décalé vers une fréquence plus élevée, comme chez MOORE & GLASBERG. La pente spectrale à droite semble relativement faible, et l'allure générale par sa dissymétrie est voisine des courbes de masque psychophysiques. Pour les voyelles V1 et V2, le plateau formantique devrait être remplacé par un pic discret, mais la variabilité des mesures interdit une représentation exacte. Pour les quatre interrogations qui motivaient ce travail, on peut conclure que la variabilité des mesures est trop forte pour permettre de savoir si un pic centré sur une harmonique est plus large qu'un pic spectral entre deux harmoniques. En revanche, il est vrai qu'un pic spectral est d'autant plus distinct que les harmoniques droite et gauche sont de niveau plus faible. La suppression ne provoque pas un creux très marqué à la gauche du formant, mais sa trace est peut-être visible sur les spectres auditifs de V2, vers 350 Hz. Enfin, il paraît impossible de savoir si un spectre de sonie rend compte au mieux des résultats. On peut ajouter un cinquième aspect. La position du maximum formantique pourrait être légèrement décalée vers une fréquence légèrement plus élevée. Mais les spectres auditifs mesurés par le seuil de pulsation ne montrent jamais ce biais.

Nous n'avons pas poursuivi ce type d'investigation car la variabilité constatée dans toutes les études, même pour des sujets entraînés, rend trop aléatoire le test d'hypothèses précises sur la forme et la position du maxima auditif. On nous permettra peut-être de souligner que l'obtention de ces quelques résultats a nécessité plus de 30 h d'écoute au casque, et que la tâche exige la plus extrême concentration, le moindre relâchement ou perturbation provoquant une élévation des seuils.

2.14 - DARWIN & GARDNER (1985, 1986)

On se souvient que CARLSON & al. ont montré que l'utilisation des deux harmoniques les plus proéminentes fournissait les meilleures prédictions de l'évolution du timbre en

fonction de la fréquence fondamentale. DARWIN & GARDNER (1985) reposit le problème du nombre d'harmoniques participant à la fixation de la fréquence du formant auditif. Leur expérience montre que la contribution des autres, plus lointaines, n'est pas nulle. La fréquence de la frontière phonétique, entre la 3ème et la 4ème harmonique, étant déterminée à partir des données d'identification, ils démontrent qu'une augmentation de 6 dB de l'harmonique 2 (250 Hz) ou 6 (750 Hz) provoque un déplacement de la frontière de voyelles de timbre [I] ou [e]. L'effet est plus important pour un accroissement de + 12 dB. Les résultats sont asymétriques dans la mesure où la modification d'une harmonique plus basse que le formant provoque un déplacement de frontière plus faible que celle d'une harmonique plus élevée. L'analyse L.P.C. donne une estimation de F1 très proche de la frontière phonétique des sujets, car elle prend en compte l'énergie spectrale dans une large bande autour de F1. Selon nous, quelques facteurs sont susceptibles d'expliquer les conclusions divergentes des deux expériences. L'usage d'une échelle de fréquence en Bark montrerait que les harmoniques hautes et basses, équidistantes sur une échelle en Hz, ne le sont plus dans ce cas, les fréquences basses étant "dilatées", c'est à dire plus "lointaines" que les fréquences élevées du formant auditif. On pourrait y chercher la raison de la moindre contribution de l'harmonique basse. La différence est cependant faible. Une plus forte pondération des harmoniques élevées semble une constante dans la plupart des expériences. Le facteur suivant paraît plus important. L'examen des spectres des voyelles utilisées montre que les harmoniques éloignées de la fréquence formantique possédaient une forte amplitude en raison de l'emploi d'une largeur de bande de F1 élevée (90 Hz). Un surcroît d'énergie de 12 dB place l'harmonique de 750 Hz à - 8 dB de la plus proéminente, à 375 Hz, et celle de 250 Hz largement au dessus. Il n'est pas exclu que les harmoniques lointaines des voyelles de CARLSON & al. aient été d'amplitude beaucoup plus faible. Un autre détail de la synthèse pourrait expliquer la différence entre ces données et celles de CARLSON & al.(1975). Les voyelles étaient de durée brève (56 ms), et possédaient un front de montée très raide (16 ms seulement). Nous discuterons plus loin les implications de cette remarque à partir de données voisines de CARRE (1988).

Les expériences décrites par DARWIN & GARDNER (1986) concernent la relation entre les mécanismes de perception de la hauteur tonale et ceux destinés à la perception du timbre. Les diverses théories de la perception de la hauteur tonale ("pitch") s'accordent pour affirmer que celle-ci est calculée à partir des harmoniques de basse fréquence. Tout se passe comme si les composantes fréquentielles d'un son complexe passaient au préalable dans un tamis n'autorisant la sortie que des seules composantes harmoniques. MOORE & al (1985 a, b) ont montré qu'en dessous de 3% d'inharmonicité, une composante participe pleinement à la formation de la hauteur, et que sa contribution diminue graduellement ensuite pour s'éteindre vers 8%. On sait que la fréquence du premier formant auditif est vraisemblablement estimée à partir de l'amplitude des deux harmoniques les plus

proéminentes. Si le même tamis est utilisé dans la perception du timbre, il est vraisemblable que l'inharmonicité d'une composante doit modifier le timbre phonétique. La première expérience porte sur des voyelles synthétiques classiques dont la fréquence fondamentale est fixée à 125 Hz et dont l'harmonique de 500 Hz, au-dessus de la frontière phonétique et donc du formant auditif, était affectée d'un désaccord harmonique de ± 0.2 à ± 8 %. Les résultats (figure 2.20) montrent d'abord un effet prévisible pour les valeurs élevées d'inharmonicité : la frontière est décalée vers les fréquences élevées, ce qui indique que le formant auditif est plus bas comme il convient si la contribution de l'harmonique de 500 Hz est plus faible. Mais l'effet se fait également sentir pour des degrés d'inharmonicité très faibles. En valeur absolue, les différences de frontière phonétique sont minimales (10 Hz au maximum; mais voir infra). Les auteurs rappellent qu'un décaccord harmonique léger se traduit également par une variation dans le temps de la phase de la composante relativement à celles des autres composantes. La seconde expérience confirme les résultats de la première pour des voyelles en synthèse harmonique à différence de phase nulle pour toutes les composantes sauf celle de 500 Hz. Ouvrons une parenthèse : il est rassurant pour les tenants d'une théorie spectrale, et non temporelle comme SCOTT (1976), de la perception formantique de constater que la frontière phonétique change de 5 Hz seulement pour une voyelle synthétique ayant une forme d'onde temporelle si éloignée, par son facteur de pic énorme, de celle d'une voyelle naturelle ou synthétique classique. Si l'interprétation de DARWIN & GARDNER est correcte, alors une simple différence dans la phase d'une composante parfaitement harmonique doit permettre de reproduire cet effet. La confirmation est excellente, bien que la discussion soit trop longue et complexe pour être présentée ici. DARWIN & GARDNER notent qu'une expérience pilote a indiqué que l'oreille ne peut exploiter la différence de phase de 180° qui accompagne les harmoniques lorsqu'elles passent de part et d'autre d'un formant, en raison sans doute des distorsions aléatoires de phase exercées par tout environnement normal, et, ajoutons-nous, par tous les moyens de reproduction sonore. Mais les indications sont trop succinctes pour déterminer si les conditions étaient celles discutées dans l'analyse des résultats de SCOTT. Une dernière expérience montre que l'effet de phase sur le timbre est moins important pour une composante de fréquence plus basse, sous F1.

PALMER & al. (1987) étendent ces résultats en montrant que l'augmentation maximale de F1 due à une rotation de 90° de la phase de l'harmonique au-dessus du formant est équivalente à une augmentation de son niveau de 4 dB. Le débattement de la frontière est de 20 Hz pour l'ensemble des variations de phase. D'un changement de phase par rapport à celle de la synthèse classique, il résulte toujours une diminution de la fréquence de la frontière phonétique, c'est-à-dire une élévation du formant auditif. Les résultats neurophysiologiques du taux de décharge synchrone de fibres du noyau cochléaire dont les fréquences centrales sont proches de la fréquence formantique révèlent qu'une variation de

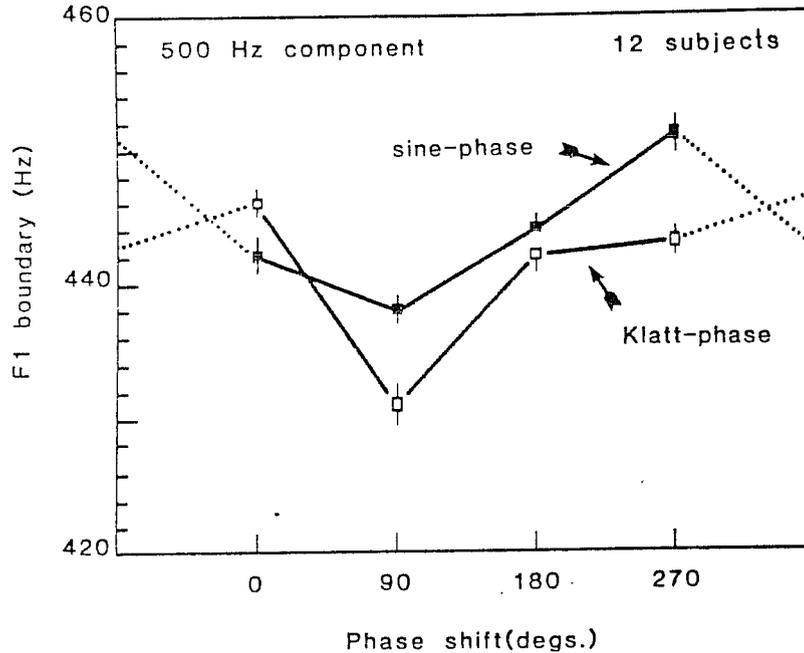


FIG. 4. Phoneme boundaries of vowel continua as a function of the phase shift in degrees of the 500-Hz fourth harmonic relative to the 0 deg point. Vowels were synthesized on a fundamental of 125 Hz with amplitude values of the individual harmonics derived from the transfer function of the Klatt synthesizer (Klatt, 1980). The open symbols give values for vowels with the starting phases of the harmonics derived from the Klatt transfer function. The filled symbols give values for vowels with all harmonics starting at a phase of 180 deg. Vertical bars are standard errors across 12 subjects.

Figure 2.20 - Evolution de la frontière phonétique en fonction de la phase de l'harmonique au-dessus de la fréquence formantique (DARWIN & GARDNER 1986).

phase correspond toujours à une augmentation de la fréquence du maximum de synchronisation d'au moins 40 Hz. Les fibres dont les fréquences centrales sont proches de F1 montrent une synchronisation des décharges plus forte sur l'harmonique haute dans le cas où la phase est modifiée. Ce n'est pas le cas du taux de décharge moyen, qui reste pratiquement constant. La cohérence des résultats psychoacoustiques et neurophysiologiques montre que la comparaison des valeurs d'une frontière phonétique dans deux conditions doit prendre en compte l'incertitude liée aux différences de phase d'une condition à l'autre. SHAMMA (1985) fait remarquer que les réponses des fibres auditives montrent un déphasage important pour les fibres dont fréquences centrales sont de part et d'autre de la fréquence d'une composante spectrale intense, et il propose un modèle de détection des maxima spectraux fondé sur la détection de ces changements de phase par inhibition latérale.

La condition où nous avons suggéré un effet de phase, c'est-à-dire le cas d'une seule harmonique proéminente très proche par excès ou par défaut de la fréquence formantique, n'a pas été testée dans ces expériences .

Il conviendrait, je crois, de reprendre également ces expériences pour des fréquences fondamentales plus élevées. Ce que l'on sait de l'audibilité monorale des variations de phase permet de faire l'hypothèse que les effets de phase décroissent avec l'augmentation de la fréquence fondamentale.

2.15 - HOLMES (1986)

La contribution de HOLMES au symposium sur "L'Invariance et la Variabilité des Processus de Parole" concerne principalement le problème de la normalisation. L'hypothèse que le timbre phonétique global est fonction de la seule différence en Bark entre les formants ne reçoit pas un appui clair. Les résultats proviennent d'une sélection par des sujets de la voyelle la plus "naturelle" dans une synthèse de haute qualité faisant varier Fo et F3 de diverses manières. Le problème est, nous le savons, dans le fait que pour les voyelles fermées et même ouvertes, la variation nécessaire de F1 et F2 pour conserver le timbre n'est pas proportionnelle à celle de Fo. Nous avons figuré à la figure 2.21, le déplacement de la position la plus "naturelle" des voyelles dans le cas d'une simple augmentation de Fo d'un Bark (environ 100 Hz). Cette variation fait passer Fo d'une valeur typique pour les hommes à une valeur typique pour les femmes. Dans la synthèse, Fo chutait d'une demi-octave dans les deux cas. On constate que, comme chez SLAWSON, le déplacement pour les voyelles fermées est très faible pour F1 et F2 de [i], plus important pour le deuxième formant de [u] (presque 1 Bark). Pour les voyelles ouvertes, le déplacement est voisin pour F1 et F2, mais de taille le plus souvent largement inférieure à 1 Bark.

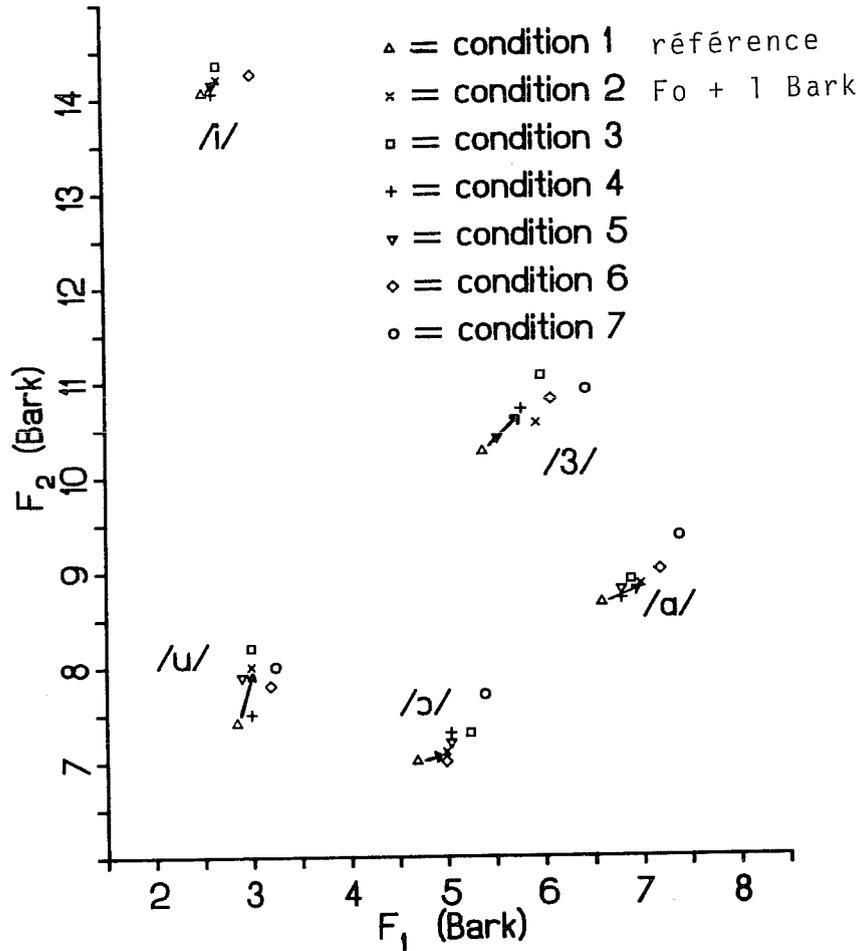


FIG. 16.1. Mean F_1 and F_2 values judged acceptable for five British vowels. The conditions are described in the text.

Figure 2.21 - Déplacement du centroïde de voyelles en fonction d'une augmentation de F_0 de 1 Bark : flèche entre condition 1 et 2 (in HOLMES 1986).

2.16 - VAN TASSEL, FABRY & THIBODEAU (1987)

Certaines figures et commentaires de cet article consacré aux spectres auditifs de sujets malentendants concernent un sujet à l'audition normale. Les spectres auditifs de 4 voyelles synthétiques à une fréquence fondamentale de 100 Hz, obtenus par masquage proactif, sont reproduits à la figure 2.22. Les spectres physiques sont également reproduits. En nous limitant à la bande contenant F1, on constate que le spectre de la voyelle [ʌ] montre un pic qui correspond exactement à la position des deux harmoniques proéminentes (600 et 700 Hz). La profonde vallée à droite, qui n'existe pas dans le spectre physique, pourrait être due à la suppression. Pour la voyelle [i], le pic auditif forme un large plateau qui s'étend jusqu'à 500 Hz, alors que le spectre physique à cette fréquence montre une harmonique de -20 dB inférieure à celle de 300 Hz. Il est difficile d'imaginer comment le système perceptif code la fréquence du premier formant si tel est bien le spectre auditif. En revanche, le flanc droit du premier formant de la voyelle [u], qui est acoustiquement très proche de celui de [i], montre une pente forte. Le premier formant de la rétroflexe [ɤ] est, quant à lui, parfaitement défini, alors que le second formant de [u] a disparu. Devant ce qui apparaît comme une situation incohérente, on peut chercher des explications pour chaque cas particulier, faisant un large appel à la suppression, ou conclure que la tâche était trop difficile pour le sujet, et / ou enfin que la technique doit être soumise à des analyses théoriques et expérimentales plus poussées, comme celles menées par SCHWARTZ & ESCUDIER pour le seuil de pulsation, avant que l'on puisse déduire des informations fiables sur la représentation auditive des formants. Une cause possible de la variabilité des spectres acoustiques est le niveau élevé de présentation (92 dB).

2.17 - ASSMANN & NEAREY (1987)

Ce travail très récent examine le problème de l'estimation perceptive du premier formant à partir des harmoniques. Il rappelle tout d'abord les conclusions opposées de MUSHNIKOV & CHISTOVICH et de CARLSON & al. La première expérience teste l'appariement entre un ensemble de voyelles synthétiques composées de 32 harmoniques de 125 Hz et dont le premier formant couvre les fréquences 175 à 850 Hz par pas de 25 Hz, et une voyelle possédant une bande d'harmoniques d'égale intensité vers F1. Cette voyelle est modifiée par l'affaiblissement ou la suppression d'une ou plusieurs harmoniques. Pour une voyelle ne possédant qu'une harmonique sous celle de 500 Hz, une réduction d'intensité de l'harmonique de 375 Hz (de 0 à -15 dB sous le niveau de celle de 500 Hz) ne modifie presque pas le timbre : la variation de timbre correspond à celle causée par un abaissement de F1 de 40 Hz. Il en est de même pour des stimuli à 3 ou 4 harmoniques lorsqu'on

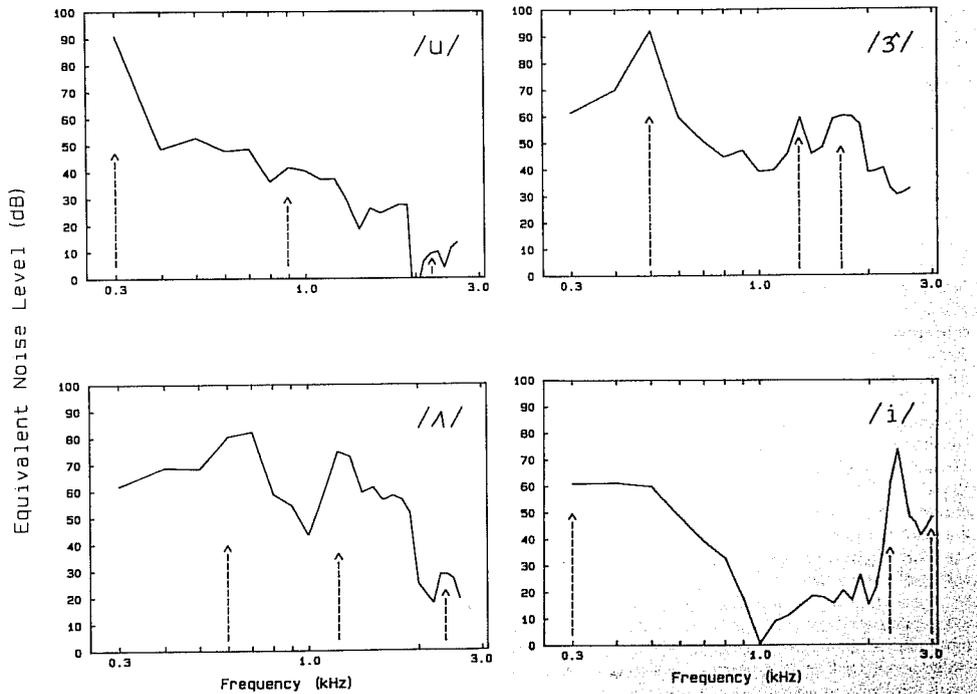


FIG. 3. Transformed thresholds as a function of probe frequency for normal subject DF. Vowel maskers were presented at 95 dB SPL. Vertical dashed arrows indicate frequencies and relative amplitudes of the vowel F_1 - F_3 , displaced downward from the pattern by an arbitrary amount.

Figure 2.22 - Seuil de masquage proactif de 4 voyelles
(in VAN TASSEL & al. 1987).

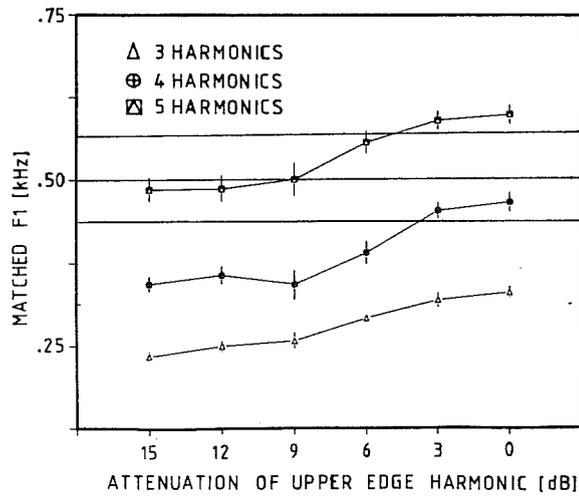


FIG. 2. Means and standard deviations of matched F_1 values from experiment 1. Top panel: lower-edge condition; bottom panel: upper-edge condition.

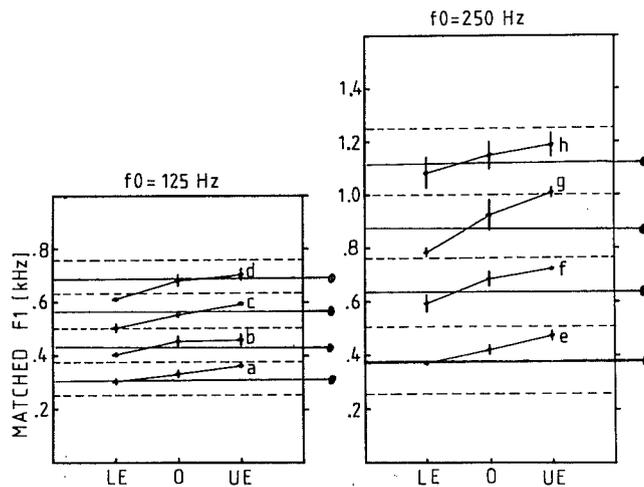


FIG. 4. Means and standard deviations of matched F_1 values from experiment 2. Low f_0 condition is shown in the left panel, high f_0 condition is shown in the right. Horizontal dashed lines indicate harmonic frequencies. LE: low-frequency harmonic of pair boosted by 9 dB; UE: high-frequency harmonic of pair boosted by 9 dB; 0: neither harmonic boosted. Harmonic pairs: a-250 + 375 Hz; b-375 + 500 Hz; c-500 + 625 Hz; d-625 + 750 Hz; e-250 + 500 Hz; f-500 + 750 Hz; g-750 + 1000 Hz; h-1000 + 1250 Hz.

Figure 2.23 - in ASSMANN & NEAREY (1987) : voir texte pour les détails.

modifie le niveau des harmoniques à 250 ou 125 Hz, les autres conservant une amplitude égale à celle de 500 Hz. La différence de timbre est négligeable. En revanche lorsqu'on modifie l'harmonique la plus élevée, le timbre se modifie notablement. A partir de -12 dB, cette harmonique ne joue plus qu'un rôle négligeable (figure 2.23). La fréquence d'une harmonique unique à 500 Hz correspond à un formant à 500 Hz. Pour une voyelle à 4 harmoniques (125, 250, 375 Hz : même intensité; 500 Hz : - 6 dB) la variation de timbre est équivalente à une baisse de F1 de 75 Hz (465 à 390 Hz). ASSMANN & NEAREY testent 4 modèles de prédiction : harmonique la plus intense, somme pondérée des deux, ou trois, harmoniques les plus intenses avec pondération spectrale de + 6 dB / octave jusqu'à 600 Hz, comme suggèrent CARLSON & al.; pic LPC avec la même préemphasis. Ce sont les pics LPC qui fournissent les meilleures prédictions, puis le modèle à deux harmoniques. L'article ne précise pas si la pondération est basée sur les amplitudes en sonnes ou en dB. On peut regretter aussi qu'une échelle Bark n'ait pas été employée. Les auteurs ne formulent pas non plus un modèle d'estimation plus précis. De l'examen des résultats, on peut tirer les conclusions suivantes :

- a - Lorsque plus de deux harmoniques ont la même intensité, seules les deux plus hautes contribuent à l'établissement du formant auditif.
- b - Lorsque deux harmoniques ont la même intensité physique, le formant auditif est plus proche de l'harmonique la plus haute. Cet effet semble plus marqué quand le spectre comprend plus de deux harmoniques dans la zone de F1.

On pourrait songer à expliquer cette deuxième conclusion, que nous avons déjà suggérée pour d'autres expériences, par la préemphasis de + 6 dB / octave due à l'oreille externe et moyenne. Mais les modèles de prédiction qui comporte cette préemphasis, comme le notent ASSMANN & NEAREY, ont néanmoins tendance à sous-évaluer la fréquence du formant auditif. L'usage d'une échelle Bark, qui dilate les fréquences basses et donc comprime les fréquences plus élevées, ne fait qu'aggraver ce décalage. Dans l'expérience 1, cet effet semble plus important pour les harmoniques élevées. Le décalage par rapport à la moyenne des deux harmoniques s'élève à 35 Hz pour les harmoniques de 625 et 500 Hz, et à 15 Hz seulement pour les harmoniques de 250 et 375 Hz. Ce fait est d'autant plus surprenant que la proéminence de l'harmonique élevée après préemphasis est plus faible en haute qu'en basse fréquence (1.5 contre 3.5 dB). On s'attendrait donc à un décalage d'autant plus important que les harmoniques concernées sont plus basses. C'est d'ailleurs ce que l'on semble obtenir pour la deuxième expérience dans des conditions identiques sauf en ce qui concerne le nombre d'harmoniques, limité à deux. Une analyse fine de l'expérience 1 montre que la présence d'une troisième harmonique sous les deux concernées contribue à élever légèrement le formant auditif (+ 30 Hz environ). Ce décalage est bien visible aussi à l'expérience 2 pour des harmoniques espacées de 250 Hz. Il existe même au delà de 600 Hz, à des fréquences où, la préemphasis ne jouant plus de rôle, intensité physique et sonie

sont proportionnelles. On peut le vérifier sur les figures, où nous avons tracé (trait plein) la moyenne des harmoniques. Pour la condition notée 0 (harmoniques d'égale intensité), la moyenne des appariements ("matched F1") est au-dessus de cette ligne. Si le modèle à deux harmoniques et l'analyse L.P.C prévoient bien les résultats pour $F_0 = 125$ Hz, ils présentent des sous-estimations non négligeables pour les harmoniques plus espacées.

Étudions, en faisant l'hypothèse que le formant auditif est légèrement plus élevé que la moyenne de deux harmoniques d'égale intensité, la position de ce dernier quand la fréquence formantique coïncide avec une harmonique, et quand le formant est équidistant de deux harmoniques. Ces deux cas se présenteront par exemple pour $F_1 = 480$ Hz, et F_0 valant 120 ou 137 Hz. Si l'harmonique coïncide avec le formant, on a vu que le formant auditif est à la fréquence de l'harmonique. En revanche, si deux harmoniques encadrent la fréquence formantique, le formant auditif sera à une fréquence un peu plus élevée car plus proche de l'harmonique de haute fréquence. Le timbre sera un peu plus ouvert. On devrait donc obtenir une petite fluctuation de timbre lorsque la fréquence fondamentale change. Le timbre devrait être un peu plus fermé quand une harmonique coïncide avec la fréquence formantique. La frontière phonétique, de ce fait, devrait se placer à une fréquence formantique plus élevée à proximité de la zone où un formant coïncide avec une harmonique. C'est le résultat que nous avons observé dans les données de FLOREN, par exemple. Cette modification du formant auditif est compatible avec l'hypothèse d'un effet de suppression de l'harmonique haute sur l'harmonique (ou les harmoniques) basse(s), qui diminue le niveau de cette dernière, décalant le formant auditif vers le haut. ASSMANN & NEAREY présente une hypothèse alternative, dans la lignée des travaux de DARWIN (1984). L'auditeur aurait implicitement conscience de contraintes sur la forme du spectre. Une modification de la fréquence de F_1 conduit à une plus grande variation dans le niveau des harmoniques proche du flanc droit, de haute fréquence, du formant, dont la pente spectrale est supérieure. Une pondération plus forte de l'harmonique de plus haute fréquence compense en quelque sorte l'asymétrie de niveau des harmoniques du premier formant des voyelles fermées.

La troisième expérience d'ASSMANN & NEAREY compare l'identification de voyelles synthétiques en fonction de la valeur de F_1 (250 - 850 Hz) dans plusieurs conditions : spectre basse fréquence complet; deux harmoniques seulement dans la zone de F_1 ; suppression de toutes les harmoniques sous, ou au-dessus, des deux plus proéminentes de la zone de F_1 . Les courbes d'identification, couvrant les timbres [i] à [ae], sont pratiquement identiques, montrant que le timbre est largement défini par les deux harmoniques les plus proéminentes. La suppression des harmoniques sous le formant est la seule qui conduise à une très légère augmentation de la valeur de la frontière. Cet abaissement de la valeur du formant auditif pourrait résulter, selon ASSMANN & NEAREY, d'un changement du centre de gravité global du spectre. Ils rappellent les

expériences de LINDQVIST & PAULI (1968) où un abaissement du niveau de F1 de 12 dB dans des voyelles fermées conduisait à un déplacement de la frontière vers des valeurs plus élevées de F2.

2.18 - CONCLUSIONS

Il n'est pas facile de fournir un résumé à la fois clair et complet de cette revue interprétative de la masse disparate et foisonnante des études antérieures. Les points essentiels, qui ont pesé sur la conception de l'expérience qui sera décrite dans les paragraphes suivants, nous semblent les suivants :

- 1 - La modification du timbre sous l'effet d'une variation de F_0 est plus sensible pour les voyelles ouvertes que pour les voyelles fermées.
- 2 - Cette modification ne peut pas être décrite par une translation de F1 et/ou F2 égale à celle de F_0 , sur une échelle en Hz ou en Bark.
- 3 - Certains sujets ne détectent pas de changement de timbre lorsque le front de montée des stimuli est très raide.
- 4 - La position du formant auditif est déterminée en première approximation par une somme pondérée de la sonie des deux harmoniques les plus intenses lorsque leur différence est inférieure à 10 - 15 dB. Outre un léger effet de lissage spectral, qui ne remet pas en cause la présence d'harmoniques dans le spectre auditif, un spectre de sonie présente une préemphasis de + 6 dB / octave jusqu'à 600 Hz par rapport au spectre physique.
- 5 - Une modification de la forme d'onde temporelle, notamment la disparition d'un passage par zéro positif, pourrait également jouer un rôle dans la modification du timbre.
- 6 - Lorsqu'un pic spectral est physiquement composé de deux harmoniques d'égale intensité, le formant auditif pourrait se trouver plus près de l'harmonique élevée, même si l'on tient compte de la préemphasis due à l'oreille externe et moyenne. Des harmoniques intenses sous les deux plus proéminentes auraient le même effet. La suppression du pic de haute fréquence sur le pic de basse fréquence est une explication potentielle.
- 7 - La détermination de la fréquence du pic auditif par masquage proactif ("forward masking") montre une légère élévation par rapport à la fréquence formantique, mais ce n'est pas le cas pour la technique du seuil de pulsation.
- 8 - Une manipulation de la phase des harmoniques de part et d'autre du formant conduit dans tous les cas à une légère élévation du pic auditif, dont on trouve trace

dans les résultats neurophysiologiques. D'autre part, le formant auditif pourrait être à une fréquence plus basse lorsque l'harmonique la plus proéminente est à une fréquence légèrement inférieure à celle du formant physique.

C'est une partie de ce pot-pourri d'hypothèses et de faits expérimentaux qui nous a incité à réaliser l'expérience décrite ci-dessous.

3 - ETUDE EXPERIMENTALE DE LA VARIATION DE TIMBRE INDUITE PAR UNE MODIFICATION DE F_0 POUR DES VOYELLES OUVERTES

Les paragraphes suivants décrivent une expérience portant sur la modification de timbre liée à un changement de F_0 pour des voyelles synthétiques de timbre compris entre [oe] et [a]. Nous présenterons d'abord une justification du choix des paramètres et de la méthode retenue, puis une description des stimuli et des conditions de présentation. Après une description des résultats, nous procéderons à une discussion de différents facteurs pouvant rendre compte des résultats. Le travail très récent de CARRE (1988) fournit la matière à une intéressante comparaison.

3.1 - LE CHOIX DE LA METHODE D'IDENTIFICATION ET DES PARAMETRES DE SYNTHESE.

Le premier but de l'expérience qui va être décrite est de fournir des informations quantitatives sur les modifications de timbre qui découlent d'une variation de la fréquence fondamentale. Le but ultime est la découverte, par l'examen des spectres acoustiques et des simulations des spectres auditifs, de règles au moins qualitatives permettant la prédiction de ces modifications de timbre.

La méthode retenue pour l'analyse du timbre est la méthode d'identification phonétique de stimuli. Elle consiste à demander au sujet de noter le timbre perçu à l'aide d'un symbole phonétique. La raison essentielle de ce choix est la simplicité du dispositif expérimental, le caractère naturel de la tâche et la possibilité d'accumuler rapidement des résultats pour un nombre important de sujets. Une technique différente a été retenue par ASSMANN & NEAREY. Le sujet doit procéder à l'appariement des éléments d'une série invariante de stimuli avec des stimuli modifiés. La série invariante constitue donc un repère invariant, et on peut ainsi mesurer quantitativement l'effet d'une modification. L'avantage essentiel est que le sujet n'est pas obligé d'associer le timbre perçu à un symbole déterminé. Pour notre étude, cependant, cette technique ne répond pas au but fixé. Il n'est pas possible de savoir si la modification de la position du formant auditif correspond effectivement à une modification du timbre, ni de connaître son ampleur.

La méthode d'identification phonétique souffre de deux désavantages. Les scores d'identification sont soumis à un effet de contraste provenant de la succession à bref interval de stimuli de timbre différent. Bien qu'aucune théorie des effets séquentiels sur l'identification ne soit universellement acceptée à ce jour, un examen d'expériences et de commentaires publiés (CROWDER & REPP (1984), FOX 1985, REPP & al. 1979 par ex.) indique qu'un stimulus de timbre bien défini exerce un puissant effet de contraste sur un stimulus ambigu qui le suit. Cet effet est gênant lorsqu'on souhaite comparer les résultats de plusieurs séries. Un choix aléatoire de l'ordre de présentation pour des séries différentes peut faire précéder les voyelles de timbre ambigu de voyelles de timbres différents dans chaque série. Afin de minimiser l'importance de ce facteur, nous avons soigneusement égalisé l'étendue de chaque série et retenu le même ordre pour chaque série. Les stimuli ambigus sont donc précédés de voyelles ayant des timbres de nature voisine d'une série à l'autre. L'effet de contraste sera donc approximativement le même dans chaque série. Les voyelles sont présentées deux fois, séparées de 750 ms. Cette double présentation permet au sujet d'utiliser la première comme un signal d'attention et surtout de "gommer" dans la mémoire auditive la trace du couple précédent, présenté 4 s plus tôt. La deuxième présentation doit ainsi être plus libre d'effets de contraste.

Le deuxième inconvénient est que les réponses d'identification phonétique ne révèlent vraisemblablement qu'une partie seulement des informations que possède le sujet sur le timbre. Un timbre ambigu ne peut être reconnu qu'en procédant à l'analyse de la moyenne des réponses. Les réponses individuelles ne peuvent être analysées que si le même stimulus est présenté un nombre important de fois, ce qui allonge la durée de la session. Il est néanmoins possible d'améliorer la technique d'identification en proposant au sujet une échelle dont les extrémités correspondent à des timbres francs. On donne ainsi au sujet la possibilité de spécifier de façon précise les nuances du timbre. Lorsque les différences attendues sont de faible ampleur, comme ici, une échelle de 5 à 9 positions est empiriquement adéquate. Nous avons retenu une échelle à 7 positions. La distance entre [a] et [oe] est donc égale à 7 unités arbitraires, et la valeur 4 caractérise un stimulus ambigu.

Le choix des paramètres de synthèse a été dicté par les considérations suivantes :

- a - Les modifications de timbre paraissent plus importantes pour les voyelles ouvertes que pour les voyelles fermées.
- b - Pour simplifier l'analyse et tester la règle $F_1 - F_0$, F_2 et F_3 devaient être à une fréquence constante, et correspondre si possible à une harmonique .
- c - La fréquence fondamentale devait couvrir la gamme 100 - 300 Hz environ. Les résultats de TRAUMÜLLER (1981) suggèrent en effet un changement dans les réponses au delà de 350 Hz.
- d - Il paraissait enfin souhaitable d'inclure une fréquence fondamentale égale au double d'une autre pour créer une condition de similarité spectrale maximale.

L'auteur a procédé à l'écoute informelle d'un très grand nombre de voyelles synthétiques. Il est apparu que la frontière entre [oe] et [a] permettait de satisfaire cet ensemble de contraintes. Les écoutes ultérieures ont notamment eu pour objet d'égaler subjectivement l'étendue des variations de timbre à diverses fréquences fondamentales.

3.2 - DESCRIPTION DES STIMULI.

Nous avons synthétisé 4 séries de 7 stimuli à l'aide d'un synthétiseur numérique de structure série. Les fréquences fondamentales de chaque série étaient constantes et sans vibration ("jitter") à 110, 188, 220 ou 270 Hz. Ces chiffres sont en fait approximatifs : il serait plus exact de préciser les périodes : 9,1 ms (109.9 Hz) ; 5,3 ms (188.7 Hz) ; 4.5 ms (222.2 Hz) et 3.7 ms (270.3 Hz). Les paramètres communs étaient une durée de 275 ms, incluant un front de montée et de descente relativement long de 35 ms de forme exponentielle. La fréquence d'échantillonnage était de 10 KHz, et le codage sur 11 bits + signe. L'onde glottale était la forme polynomiale de ROSENBERG, avec une phase montante de 40% de la durée totale de la période, une phase descendante de 25%, et une phase fermée de 35 %. Les fréquences (et largeurs de bande) de F2, F3 et F4 étaient fixes, à 1320 (70), 2420 (100) et 3300 (120) Hz respectivement. Notons que F2 coïncide avec la fréquence d'une harmonique à 110, 188 et 220 Hz. A 270 Hz, le formant est seulement 30 Hz au dessus de l'harmonique. Il en est de même pour F3 à 110, 220 et 270 Hz.

Le front de montée a été choisi long en raison des résultats de SLAWSON(1968) montrant, pour certains sujets, un moindre effet de la fréquence fondamentale sur le timbre lorsque l'attaque était brutale.

La gamme de variation de F1 était ajustée pour chacune des séries à des fréquences fondamentales différentes. Le pas d'incrémentatation était de 28 Hz, ce qui est proche du seuil de détection pour les voyelles de spectre constant (cf. les "difference limens (DL)" de MERMELSTEIN (1978) : 33 Hz pour F1 près de la frontière [ɛ - ae]).

110 Hz : 550 - 715 Hz

188 Hz : 578 - 743 Hz

220 Hz : 605 - 830 Hz (en omettant les stimuli à 633 et 800 Hz)

270 Hz : 605 - 770 Hz

La largeur de bande de F1 était constante à 50 Hz. Quelques spectres physiques sont présentés à la figure 2.24.

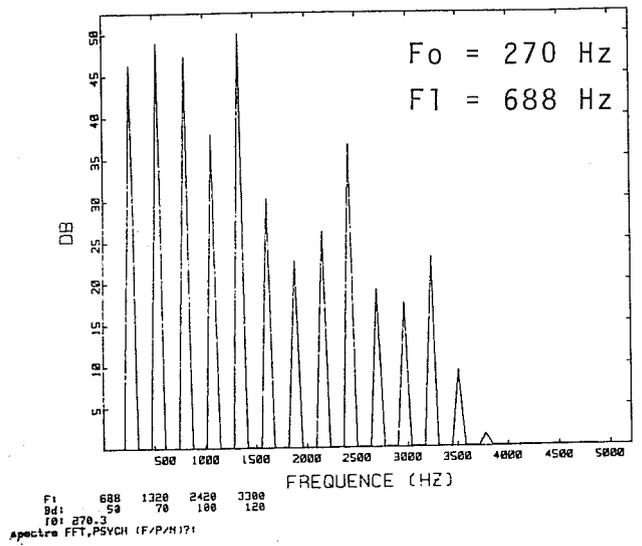
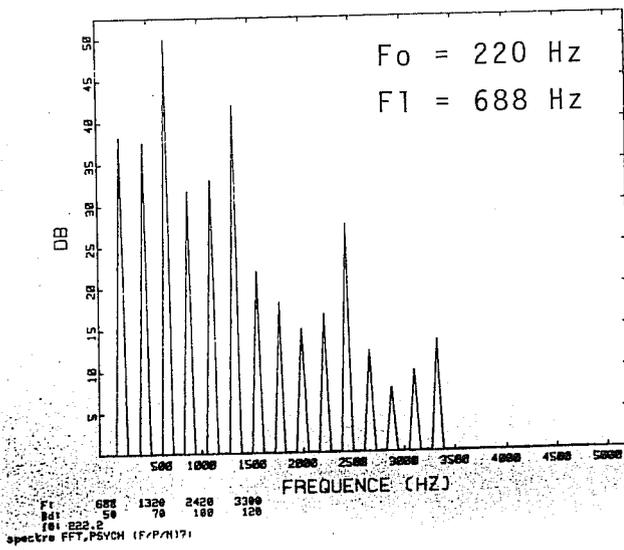
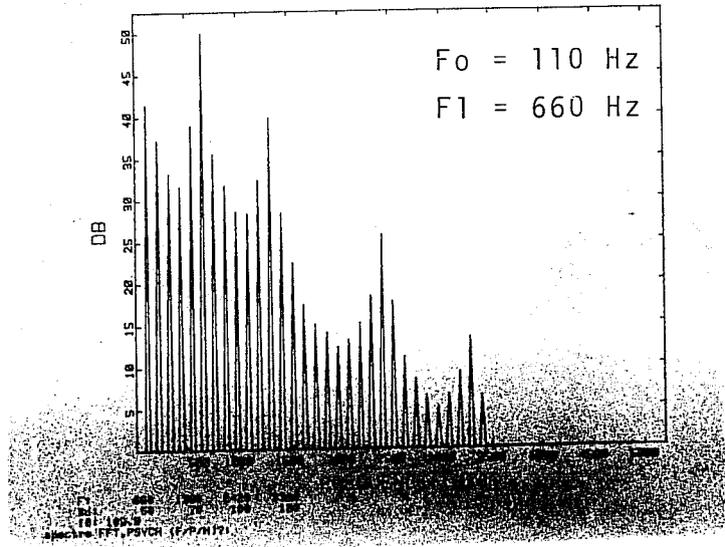
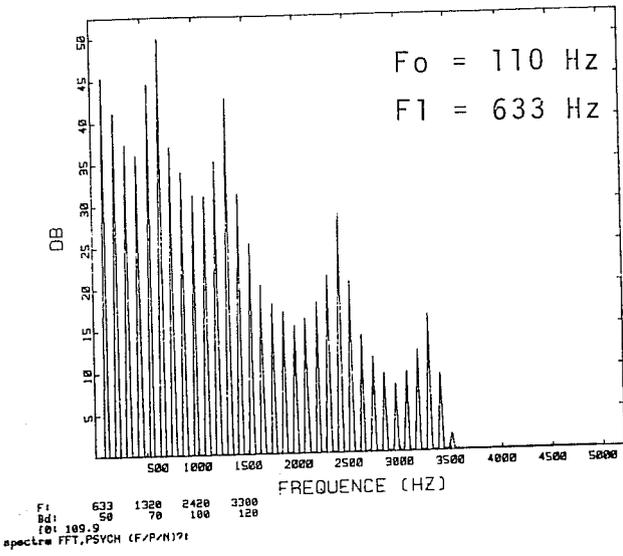


Figure 2.24 - Spectres physiques de stimuli utilisés dans l'expérience d'identification.

3.3 - CALCUL DES SPECTRES AUDITIFS

Les spectres physiques ne se prêtent pas bien à une discussion de la position du formant auditif. Les raisons essentielles sont qu'ils ne tiennent pas compte de la fonction de transfert de l'oreille externe et moyenne ni de la résolution fréquentielle variable de l'oreille interne. MOORE & GLASBERG (1983 a) ont proposé une méthode de calcul de la forme de l'excitation interne ("excitation pattern") provoquée par un stimulus sonore. Le système auditif est modélisé comme une batterie de filtres qui se recouvrent largement. La largeur des filtres est celle déterminée par les expériences de masquage où l'on détermine le seuil d'audibilité d'une sinusoïde flanquée de deux bandes de bruit à une distance variable. Comme la forme des flancs des filtres change avec la fréquence, il est plus commode d'exprimer la largeur de bande comme celle d'un filtre rectangulaire transmettant la même énergie ("equivalent rectangular bandwidth : ERB"). Une estimation aux moindres carrés, fondée sur les résultats de 6 expériences, de la largeur de bande équivalente en fonction de la fréquence est donnée par

$$LB_{ERB} = 6.23 F^2 + 93.4 F + 28.5 \text{ Hz} , \text{ la fréquence centrale } F \text{ du filtre étant en kHz.}$$

Ces valeurs sont largement inférieures, comme nous l'avons déjà souligné, aux valeurs des bandes critiques (cf. § 2.10). On peut également définir une échelle fréquentielle proportionnelle aux largeurs de bande, à laquelle MOORE & GLASBERG donnent le nom d'échelle ERB, à l'instar de l'échelle Bark ou échelle de bande critique. Pour la présentation de nos résultats nous avons conservé l'échelle classique en Hz, plus lisible. La seule difficulté est que la moyenne entre deux fréquences n'est pas exactement la même sur les deux échelles. Mais la différence est négligeable en pratique. On peut dériver la forme de l'excitation provoquée par un stimulus en réalisant la convolution de chacun des filtres avec le spectre du stimulus. La forme des filtres, déterminée par les expériences de masquage, est approximée par $W(g) = (1 + p g)e^{-p g}$, où g est la valeur relative de la déviation fréquentielle d'avec la fréquence centrale F , et p un paramètre de forme égal à $4 F / LB_{ERB}$. La simulation de la forme d'excitation est calculée à l'aide de filtres dont les fréquences centrales sont espacées de 10 Hz. On corrige ensuite le résultat de la pente donnée par la fonction de transfert de l'oreille externe et moyenne. Nous appellerons par commodité "spectres auditifs" les formes d'excitation ainsi calculées. MOORE & GLASBERG signalent avec honnêteté les limites et approximations d'une telle approche. Les spectres auditifs s'appliquent à des sujets jeunes à des intensités modérées. Les largeurs de bande ont tendance à s'accroître lorsque le niveau sonore d'une composante spectrale dépasse 40 dB, ce qui est le cas pour la parole. La résolution spectrale sera donc plus faible. La forme des filtres change également avec le niveau. Le flanc de haute fréquence devient plus raide, ce qui implique que la pente de la forme d'excitation devient au contraire plus douce. Pour une composante spectrale donnée, le masquage des fréquences plus élevées s'accroît donc.

Enfin, et surtout, ces spectres auditifs ne modélisent pas les effets de suppression. Les maxima des spectres auditifs pourraient s'en trouver affinés et les niveaux relatifs changés. La présentation de nos résultats fait appel à une échelle en dB et non en sone. L'intensité maximale est arbitrairement fixée à 50 dB. La raison essentielle est que nous ne connaissons pas le niveau absolu de présentation des stimuli pour les sujets, ce qui interdit le calcul de l'échelle phone, puis de l'échelle sone. La prise en compte de la fonction de transfert de l'oreille moyenne incorpore cependant un aspect essentiel de l'échelle phone. La pente spectrale sous 600 Hz dans nos figures (figure 2.25) est un peu trop forte. Dans la discussion, les écarts en dB des niveaux des harmoniques ont été recalculés avec plus de précision.

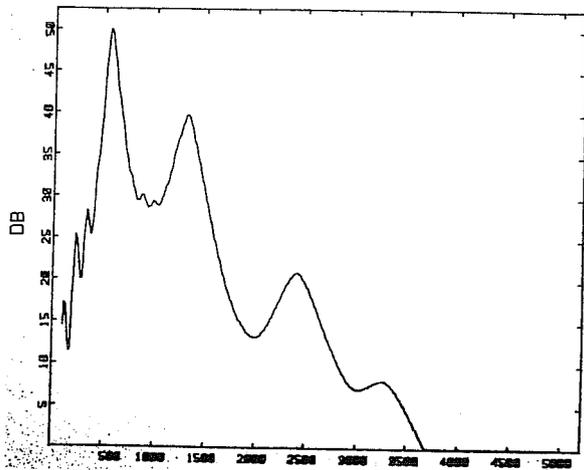
3.4 - SUJETS ET CONDITIONS DE PRESENTATION.

Nous avons recueilli les réponses de 16 étudiants francophones (11 femmes et 5 hommes), préparant une licence ou une maîtrise des Sciences du Langage à l'Université de Nancy II. Ils possédaient tous une bonne pratique de la transcription phonétique et n'étaient pas surpris d'entendre des voyelles de timbre ambigu. Les voyelles synthétiques ont été présentées à deux groupes de 8 sujets par haut-parleurs, en champ libre dans un grand studio d'enregistrement à l'insonorisation de très haute qualité. Les séries à 110 et 220 Hz ont été présentées dans une première session de 40', et les deux autres séries (180 et 270 Hz) la semaine suivante. Chaque série débutait par l'écoute de 42 stimuli en ordre aléatoire afin de familiariser les sujets avec la gamme des timbres. Simultanément, l'auteur illustre au tableau l'utilisation de la feuille de réponse. Dans chaque série, chaque stimulus était présenté 11 fois en ordre (pseudo) aléatoire : aucune voyelle ne pouvait être présentée plus de deux fois successivement et nous avons légèrement modifié l'ordre aléatoire de manière que le groupe de 3 voyelles du milieu de la gamme soit précédé d'un nombre équilibré de voyelles extrêmes. Les 7 premières réponses n'ont pas été prises en compte, ce qui donne 10 réponses par stimuli et par sujet, pour un total général de 4880 réponses.

3.5 - RESULTATS.

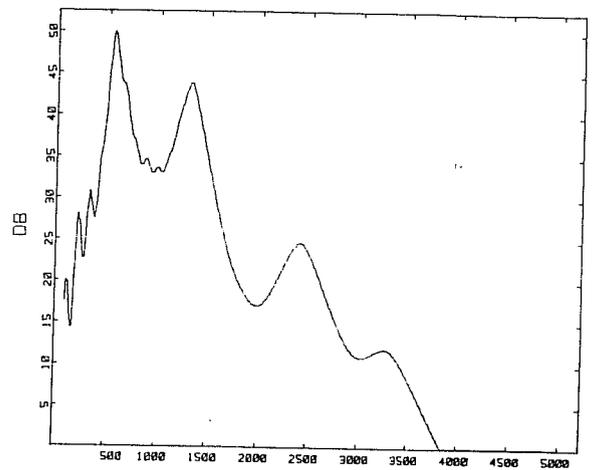
Les fonctions d'identification moyenne sont présentées à la figure 2.26. Nous avons renoncé à égaliser l'étendue des réponses des sujets en normalisant par sujet et par série la distance entre les stimuli extrêmes. Cette normalisation aurait masqué le fait que l'étendue individuelle était très voisine de l'étendue moyenne, elle-même proche de l'étendue maximale. Ceci montre que la variation de F1 pour chaque série était adéquate. Mais surtout une normalisation aurait pu décaler les voyelles considérées comme ambiguës par les sujets. Dans les quatre séries, c'est au 5ème stimulus que les sujets attribuent le score le plus

Fo = 110 F1 = 550 Hz

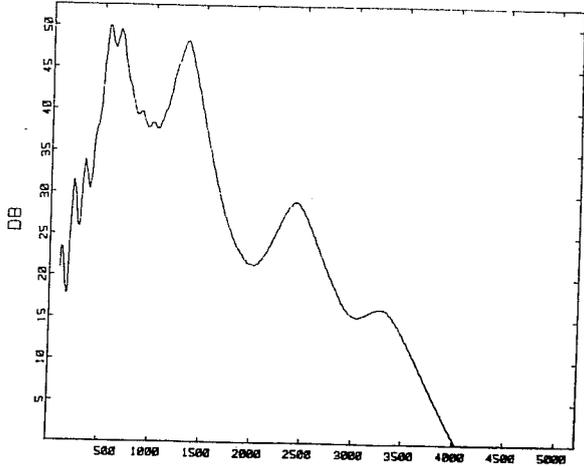


271

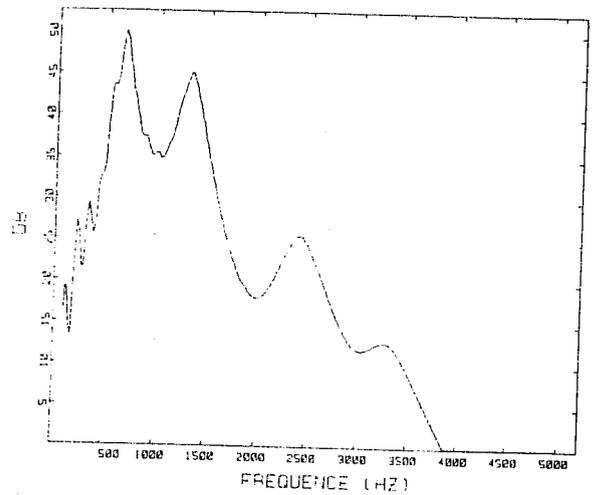
Fo = 110 F1 = 578 Hz



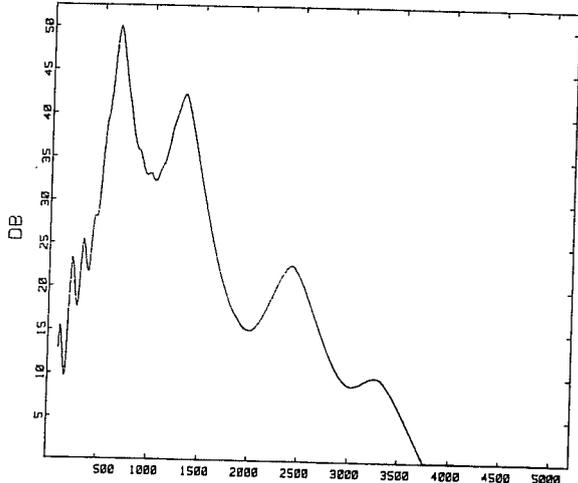
Fo = 110 Hz F1 = 605 Hz



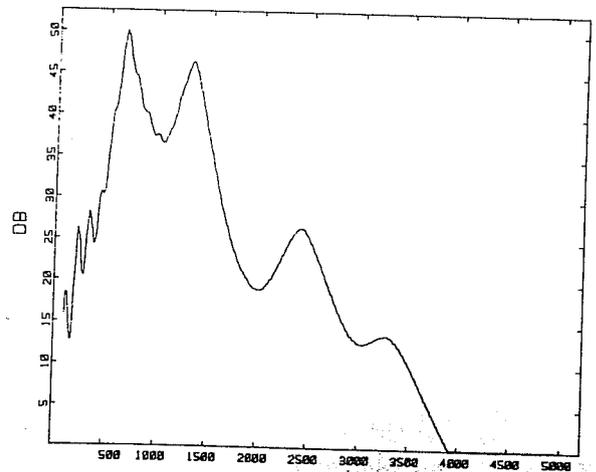
Fo = 110 Hz F1 = 633 Hz



Fo = 110 F1 = 660 Hz



Fo = 110 F1 = 688 Hz



Fo = 110 Hz F1 = 715 Hz

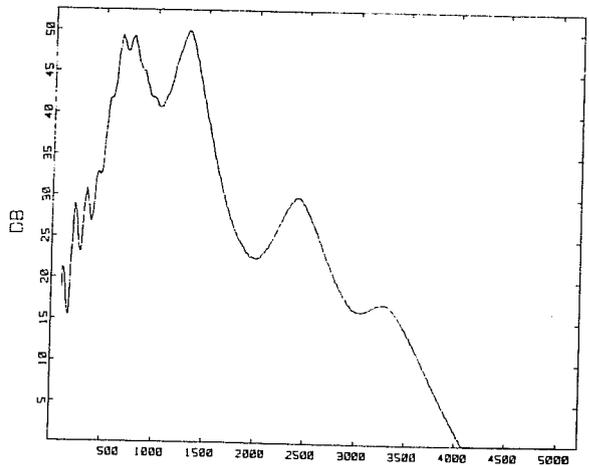
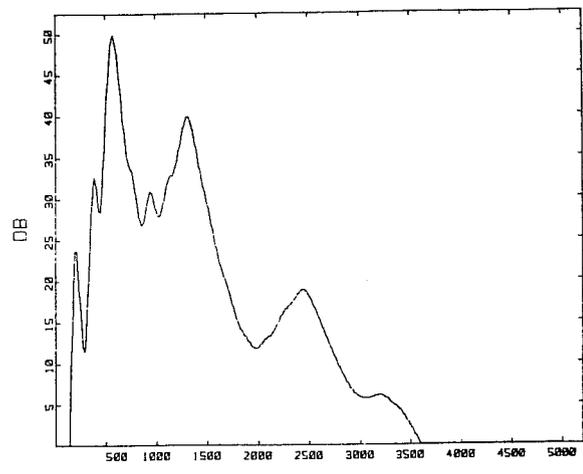
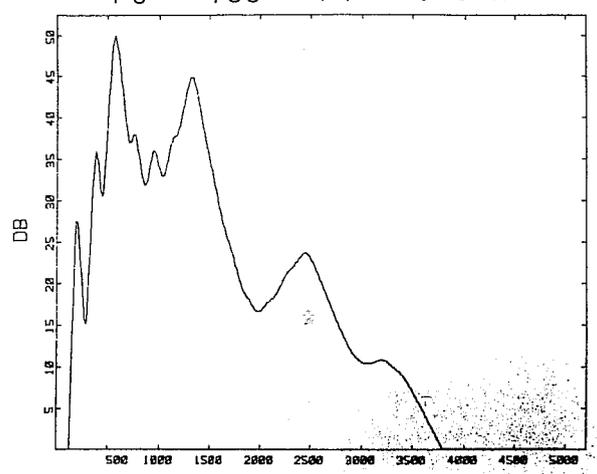


Figure 2.25 A - Spectres auditifs de la série de 110 Hz.

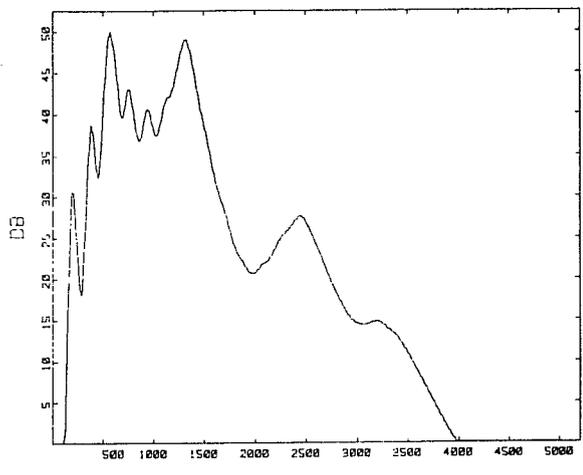
F0 = 188 F1 = 578 Hz



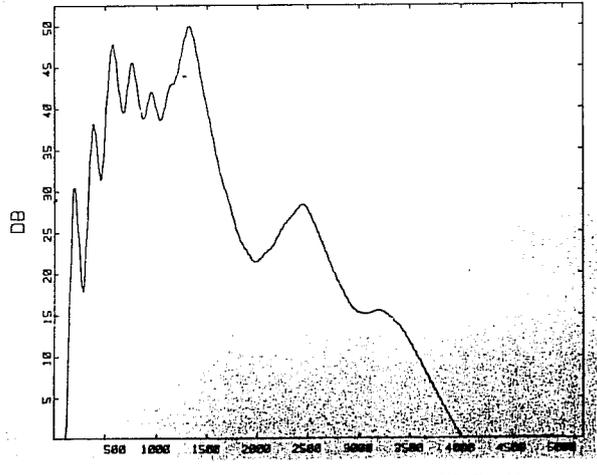
273



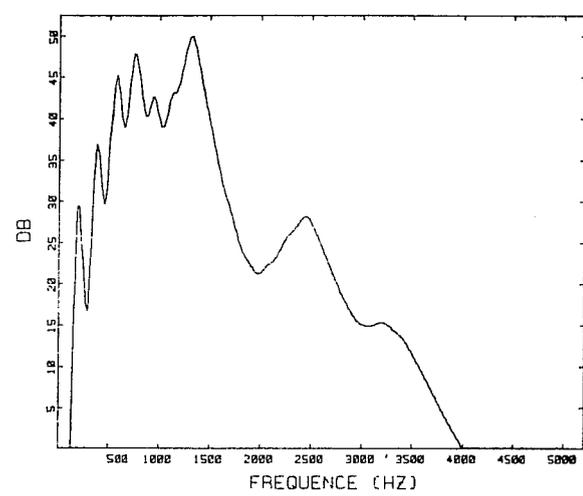
F0 = 188 F1 = 633 Hz



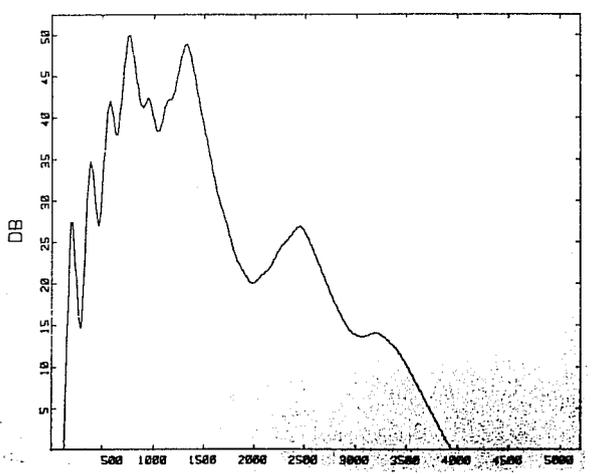
F0 = 188 F1 = 660 Hz



F0 = 188 F1 = 688 Hz



F0 = 188 F1 = 715 Hz



F0 = 188 F1 = 743 Hz

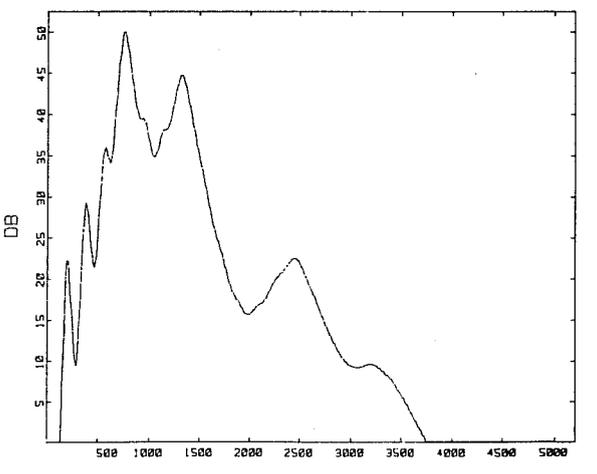
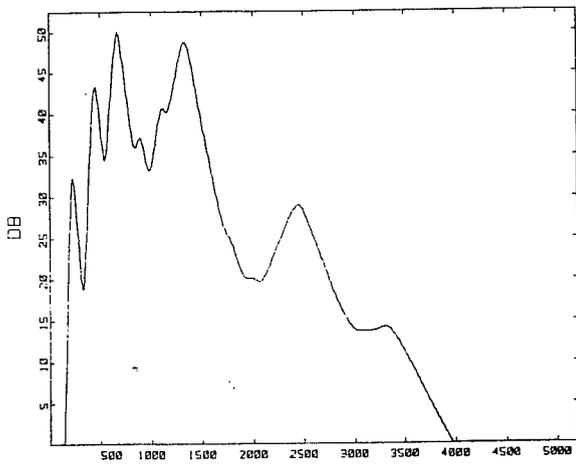
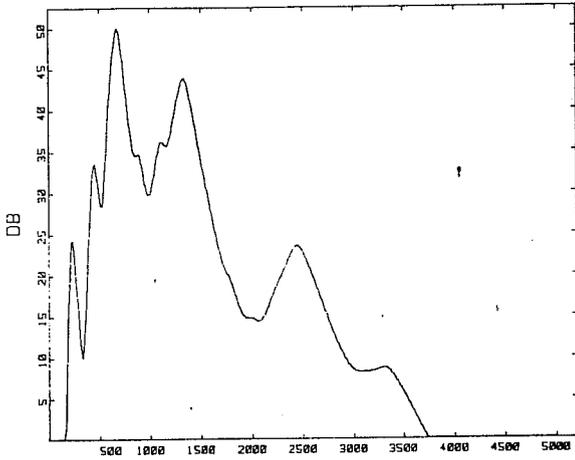


Figure 2.25 - Spectres auditifs de la série de 188 Hz.

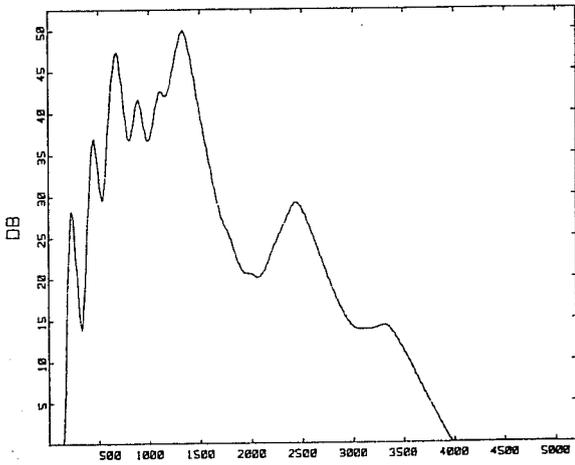
Fo = 220 F1 = 605 Hz



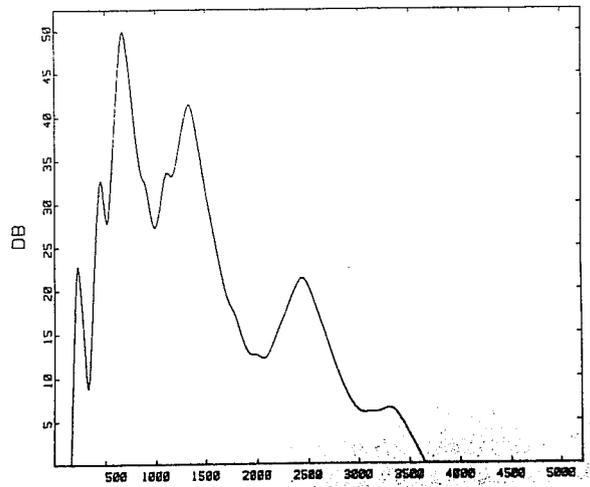
Fo = 220 F1 = 688 Hz



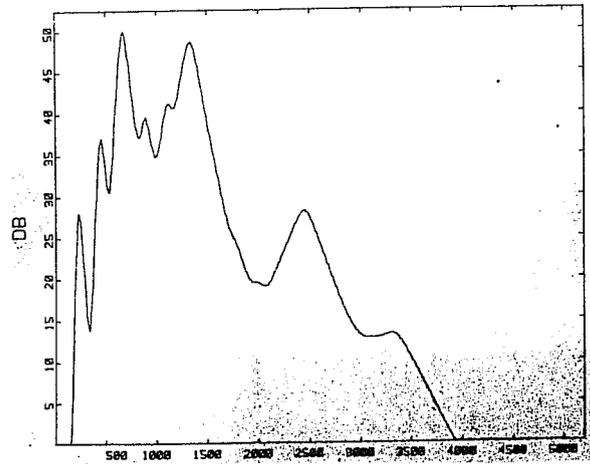
Fo = 220 F1 = 743 Hz



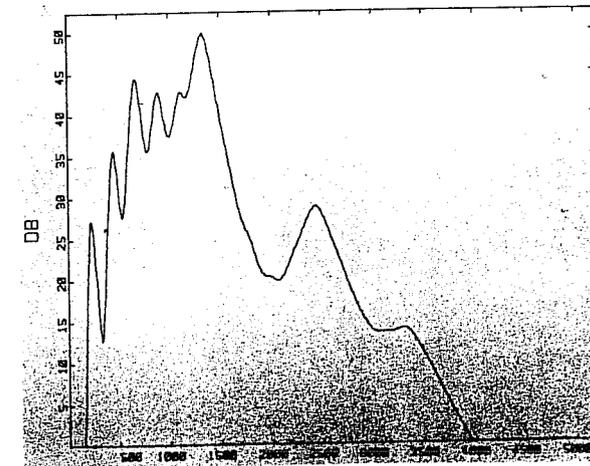
275



Fo = 220 F1 = 715 Hz



Fo = 220 Hz F1 = 770 Hz



Fo = 220 Hz F1 = 830 Hz

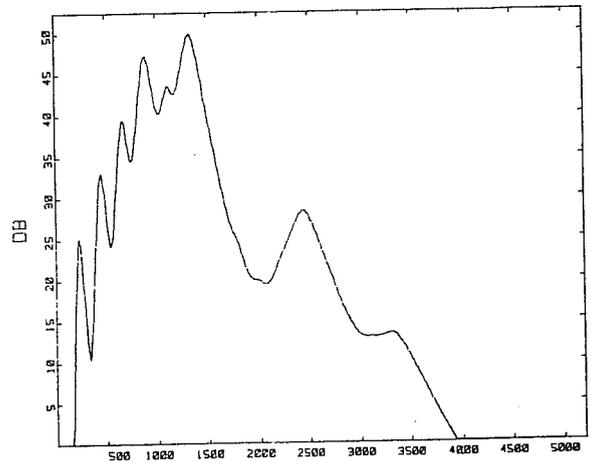
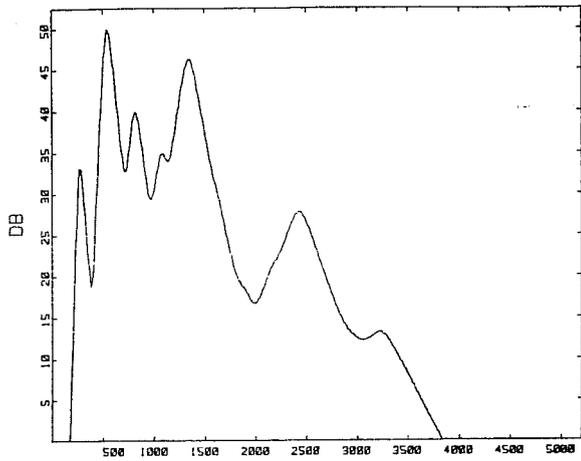
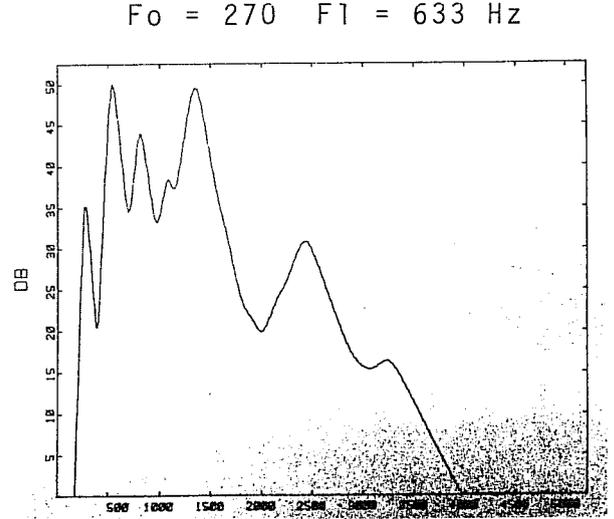


Figure 2.25 C - Spectres auditifs de la série de 220 Hz.

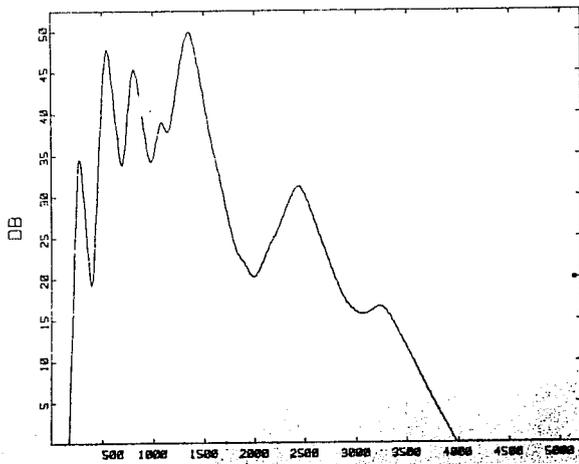
Fo = 270 F1 = 605 Hz



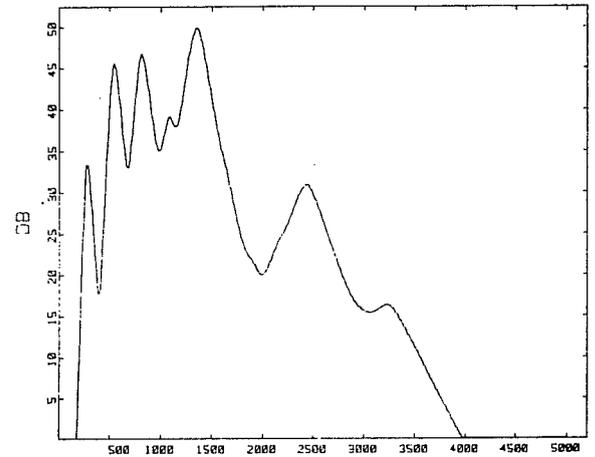
276



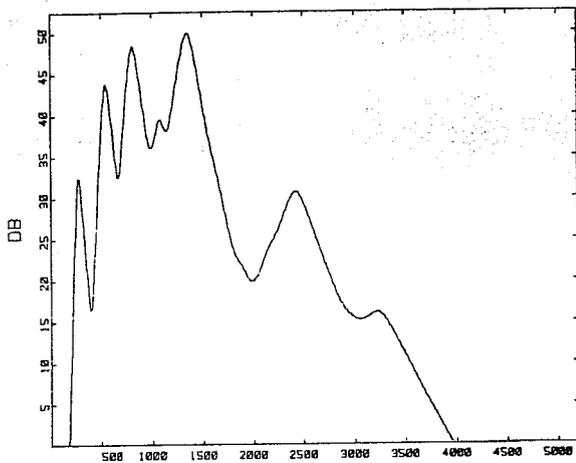
Fo = 270 F1 = 660 Hz



Fo = 270 F1 = 688 Hz



Fo = 270 F1 = 715 Hz



Fo = 270 F1 = 743 Hz

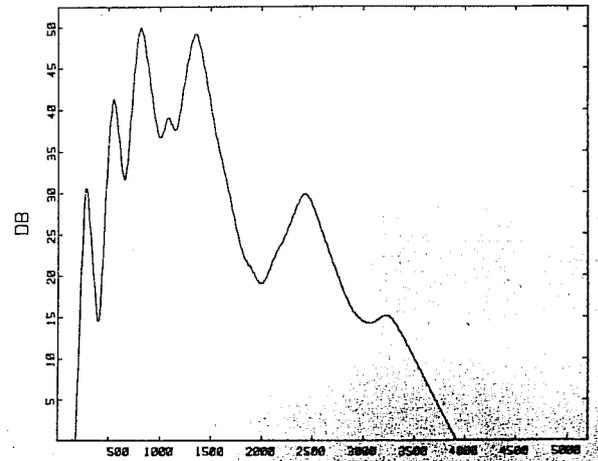
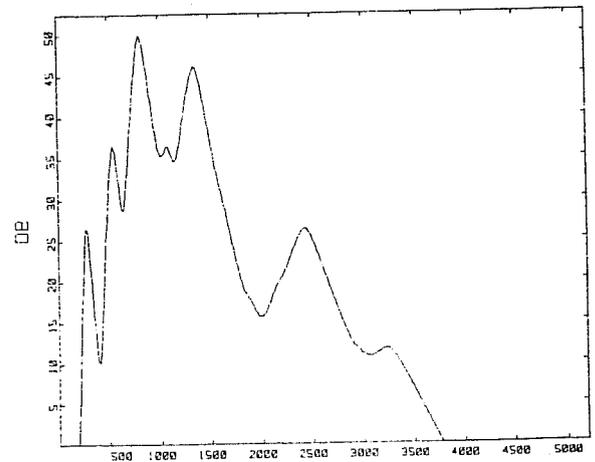


Figure 2.25 D - Spectres auditifs de la série de 280 Hz.

Fo = 270 F1 = 770 Hz



proche de 4, signe d'un timbre ambigu. Ceci implique que la frontière phonétique de l'auteur, qui a servi à déterminer l'étendue des variations, est placée environ un pas fréquentiel plus bas (28 Hz).

Le fait le plus frappant est que la courbe d'identification de la série de fréquence fondamentale la plus élevée (270 Hz) est en dessous de celle de la série à 220 Hz. Il n'y a donc pas de déplacement de la frontière proportionnel à la variation de F_0 . L'examen des résultats individuels, présentés à l'annexe 3, confirme que ce "point de rebroussement" existe pour chacun des 16 sujets. Un test non paramétrique de signe indique que la différence est hautement significative ($p < 0.001$). Un calcul approximatif de la position de la frontière des données individuelles, à partir d'une interpolation linéaire des deux estimations les plus proches de 4, montre que la courbe d'identification de la série à 220 Hz est plus élevée de 31 Hz en moyenne que celle de 270 Hz. L'écart-type de la différence est de 15 Hz, avec une étendue de 13 à 68 Hz. Les autres différences sont également significatives. Le calcul pour les séries les plus proches (110 et 188 Hz) révèle une séparation de 22 Hz, avec un écart-type de 14 Hz pour une étendue de 1 à 51 Hz. Le test de signe confirme que cette différence est hautement significative. Les écarts-type des frontières phonétiques s'établissent à 18, 15, 14 et 20 Hz pour les séries synthétisées à 110, 188, 220 et 270 Hz, respectivement. La fonction d'identification de la série de 220 Hz est moins régulière et de pente plus faible que les autres, autant en moyenne que sur les courbes individuelles.

3.6 - DISCUSSION.

3.6.1 - La série de 110 Hz

Examinons tout d'abord les spectres auditifs de la série de fréquence fondamentale égale à 110 Hz (figure 2.25). Pour F_1 variant de 550 à 605 Hz, l'harmonique dominante de 550 Hz est rejointe par l'harmonique de 660 Hz. Au delà de 605 Hz, la 6ème harmonique (660 Hz) domine le spectre jusqu'à 685 Hz. A 715 Hz, la 6ème et 7ème harmoniques sont d'égale intensité. Une élévation de F_1 de 633 à 688 Hz provoque un changement de timbre d'un peu plus du tiers de la "distance" phonétique entre [oe] et [a], si l'on considère que sur l'échelle subjective des timbres, dont l'étendue est de 6 unités, le timbre change de 2.2 unités. La comparaison des spectres auditifs correspondants est instructive (figure 2.27). La 5ème harmonique (550 Hz) passe de 45.7 à 42 dB, alors que la 7ème (770 Hz) gagne 4.2 dB (39.2 à 43.4 dB). On ne peut pas prédire le changement de timbre si l'on tient compte uniquement de la position de l'harmonique la plus proéminente, ici la 6ème à 660 Hz. La détermination de la position du formant auditif est vraisemblablement réalisée par une moyenne pondérée des niveaux des harmoniques les plus proéminentes, comme les travaux

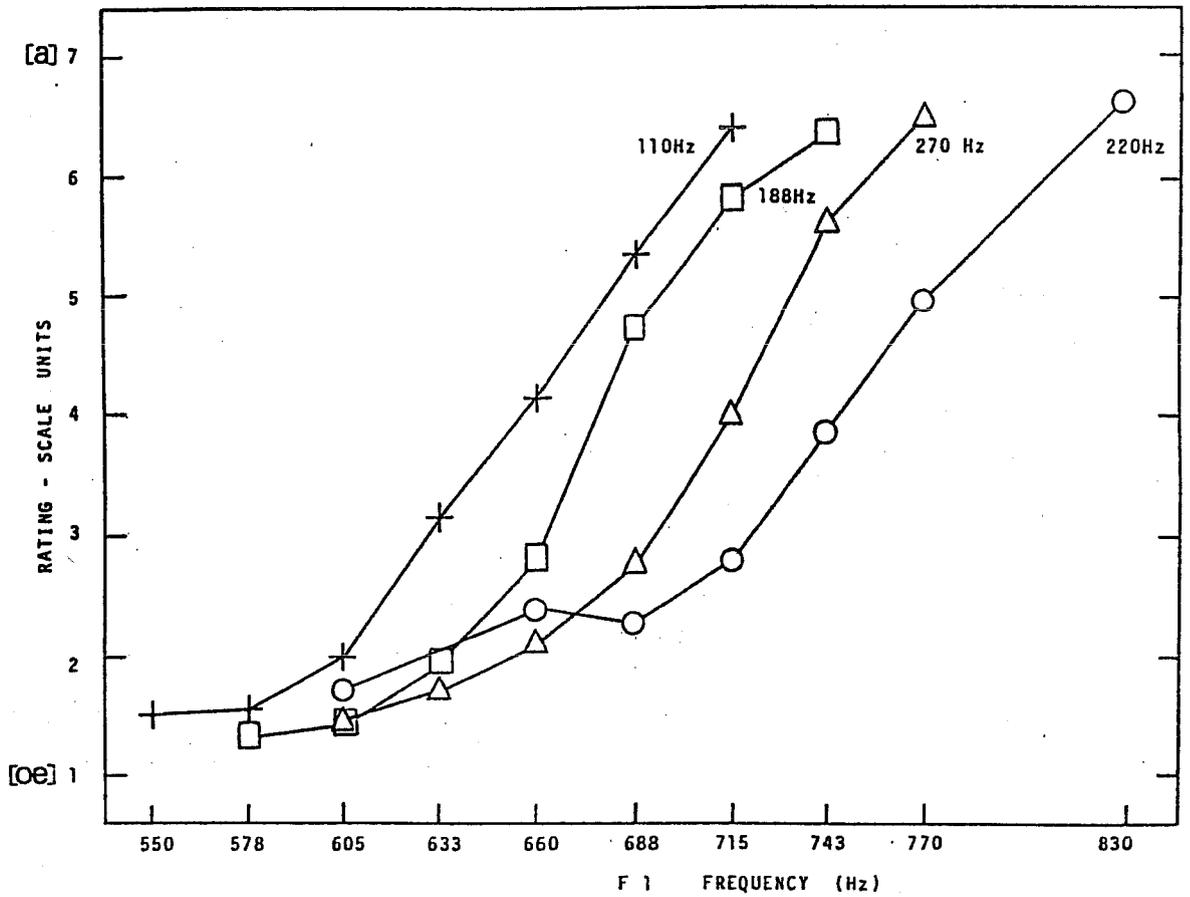


Figure 2.26 - Fonctions d'identification des 4 séries de stimuli. L'ordonnée est graduée en unités de l'échelle subjective de timbre.

de CARLSON & al. (1975) l'ont montré.

3.6.2 - Echec de la prédiction par un filtrage large bande.

La comparaison entre ces deux voyelles permet également de montrer que la valeur du formant auditif ne peut pas être la conséquence automatique d'un filtrage ultérieur à large bande. Nous avons simulé ce filtrage à large bande en calculant le spectre auditif avec des filtres dont les largeurs de bande étaient trois fois plus fortes que les valeurs de MOORE & GLASBERG (1983a) (figure 2.28). Une valeur plus faible ne gomme pas les pics des harmoniques et ne résoud donc pas le problème de la détermination des maxima du spectre. Nous avons vérifié que notre conclusion ne dépend pas de la taille exacte de la largeur des filtres. On peut ajuster les niveaux relatifs des deux spectres de manière que le flanc basse fréquence et le pic de F1 coïncident presque parfaitement. La différence essentielle est alors dans le niveau de F2. Si on ajuste les niveaux de F2, la différence est alors le niveau de F1. On sait que l'amplitude relative des formants ne joue qu'un rôle mineur dans la formation du timbre. Il est donc peu vraisemblable que ces différences soient à l'origine de la variation de timbre mesurée chez nos sujets. Nous avons déterminé la position exacte du maximum spectral dans plusieurs cas pour les voyelles de Fo égal à 110 Hz. Le timbre est donné par sa position moyenne sur l'échelle à 7 positions : 1 [oe] - 7 [a].

Fréquence physique de F1	Maximum large bande	Timbre
605	640	1.99
633	670	3.16
660	670	4.14
(660 à Fo = 220 Hz)	(690)	(2.42)
688	680	5.36
715	770	6.39

On voit que les variations de timbre ne correspondent pas en magnitude à ce que prévoit la position du maximum du spectre large bande, et surtout que le timbre peut être différent (voyelles à 633 et 660 Hz) pour une même position du maximum. En anticipant sur la discussion, remarquons que la voyelle de 220 Hz dont le formant est à 660 Hz possède un maximum plus élevé que celle de 110 Hz, alors que son timbre est plus proche de [oe]. La largeur des filtres est dans cet exemple égale à 4 fois, et non 3, celle déterminée par masquage. Cet échec du mécanisme de filtrage large bande, qui est une première conclusion importante de ce travail, est dû au fait que les harmoniques de plus faible niveau, de part et d'autre des harmoniques intenses, sont "gommées" par le lissage de large bande, alors que leur importance perceptive semble réelle.

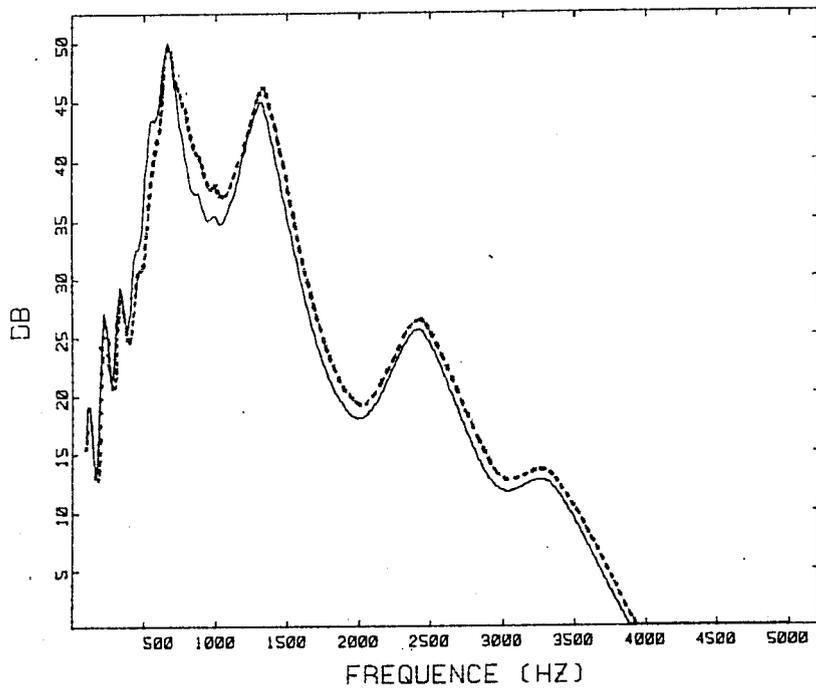


Figure 2.67 - Comparaison des spectres auditifs de deux voyelles à la même fréquence fondamentale (110 Hz) dont les premiers formants sont à 633 et 688 Hz.

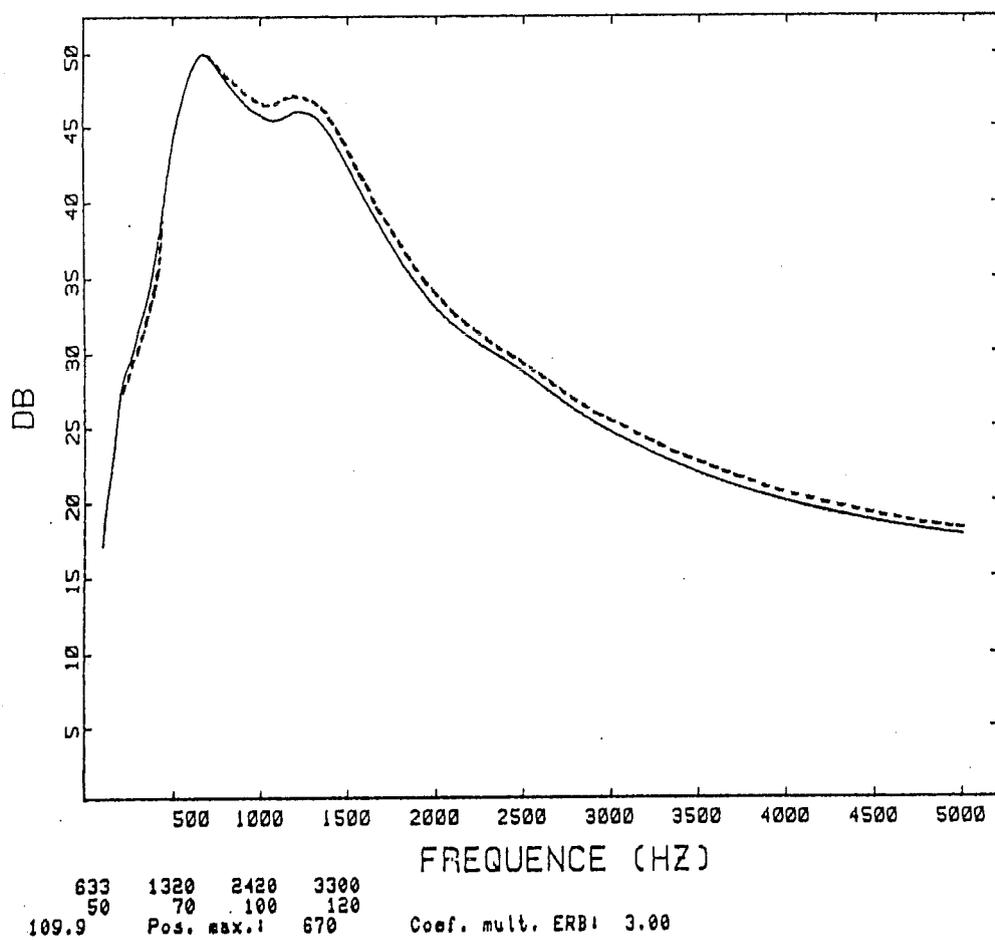


Figure 2.28 - Spectres large-bande de deux voyelles physiquement identiques : $F_1 = 660$ Hz, $F_0 = 110$ et 220 Hz.

3.6.3 - Le rôle des harmoniques sous la plus proéminente.

La position du formant auditif est sensible à de faibles variations du niveau d'une harmonique. Tous les sujets, sans exception, sont sensibles à la différence de timbre entre les voyelles dont les formants sont à 660 et 688 Hz. La seule différence entre ces deux voyelles est le niveau de la 7^{ème} harmonique qui passe de 39.4 à 43.4 dB, soit un gain de 4 dB. Ceci confirme que le niveau des harmoniques à moins de 10 dB sous l'harmonique principale est important. On voit que le seuil différentiel de déplacement formantique (DL) est inférieur à 28 Hz pour les voyelles stables, ce qui est dans la partie basse de la fourchette donnée par MERMELSTEIN (1978). Notre pas d'échantillonnage fréquentiel n'est donc pas trop fin. Cette forte sensibilité est obtenue pour un son dont le timbre est intermédiaire entre deux voyelles. Si l'hypothèse d'une estimation du formant auditif à partir de deux harmoniques est retenue, il est clair que le DL n'est pas une constante du système perceptif, mais une mesure liée au seuil de perception du niveau relatif de deux composantes spectrales.

Si l'on considère que la position du formant auditif est déterminée par la position de l'harmonique la plus proéminente, dans le cas où les autres sont de niveau faible, le formant auditif de la voyelle ambiguë de la série de 110 Hz doit être très proche de 660 Hz, c'est-à-dire de la position du formant physique. Les harmoniques à 550 et 770 Hz sont -8.1 et -10.6 dB plus bas. Pour F1 placé à 688 Hz, l'harmonique haute (770 Hz) gagne 4 dB, - 6.6 dB sous la plus proéminente, l'harmonique basse (550 Hz) restant au même niveau. A 633 Hz, l'harmonique de 550 Hz est de niveau élevé, - 4.3 dB seulement sous celle de 660 Hz, alors que l'harmonique haute est - 10.8 dB en dessous. Il paraît donc possible de rendre compte de la modification du timbre en tenant compte des harmoniques voisines de la plus proéminente, si leur intensité n'est pas inférieure à - 10 dB relativement à celle de la plus proéminente. Cet ordre de grandeur est voisin de celui que l'on peut déduire des données de ASSMANN & NEAREY (1987).

3.6.4 - La série de 188 Hz.

Examinons dans cette optique les résultats de la série de fréquence fondamentale égale à 188 Hz. Pour F1 à 660 Hz, la 3^{ème} harmonique à 566 Hz domine la quatrième (755 Hz) de 2.1 dB. La moyenne entre ces deux harmoniques est de 660 Hz. Le formant auditif étant un peu plus proche de la troisième harmonique plus intense, il est vraisemblablement un peu en dessous de cette fréquence, ce qui est parfaitement compatible avec un timbre un peu plus proche de [oe]. Pour le formant à 688 Hz, l'harmonique haute (755 Hz) domine la troisième de 2.5 dB. Le formant auditif passe légèrement au dessus de la fréquence moyenne, qui est aussi la fréquence de la frontière phonétique, et le timbre est un peu plus proche de [a] que

de [oe].

Pour ces deux séries, l'évolution du timbre est compatible avec la variation présumée de la position du formant auditif. Pour un timbre donné, le glissement vers des fréquences plus élevées des courbes d'identification en fonction de la position du formant physique ne paraît refléter que l'ajustement nécessaire de l'amplitude des harmoniques pour amener le formant auditif à des fréquences semblables.

3.6.5 - La série de 270 Hz.

Si l'on considère maintenant la fréquence fondamentale la plus élevée (270 Hz), la voyelle de timbre ambigu est celle dont le premier formant est à 715 Hz. Dans ce cas, la deuxième harmonique à 540 Hz est 1.7 dB sous la troisième à 810 Hz. La moyenne des harmoniques est de 675 Hz sur une échelle en Hz, mais de 665 Hz sur l'échelle ERB de MOORE & GLASBERG. Comme l'harmonique supérieure est légèrement plus haute que la seconde, le formant auditif devrait se placer au dessus. Le timbre devrait donc être plus proche de [a]. Pour F1 à 688 Hz, la seconde harmonique domine la troisième de 1.7 dB. Le formant auditif doit logiquement se trouver sous la moyenne (665 Hz) et peut-être sous la frontière, comme dans nos résultats. Il y a donc un léger décalage entre les prédictions et les données. Ce décalage correspond à une élévation de la frontière phonétique d'environ 30 Hz. Il pourrait cependant être plus important si l'on se souvient que ASSMAN & NEAREY ont mis en évidence un biais dans l'estimation des formants auditifs : l'harmonique de haute fréquence "pèse" plus lourd que l'harmonique de basse fréquence.

On peut passer en revue plusieurs hypothèses pour tenter de rendre compte du décalage dans nos données. Tout d'abord, la différence de phase : les résultats de SCOTT montrent peut-être un timbre plus ouvert lorsque l'harmonique proéminente est au dessus de la fréquence formantique. Ce cas ne nous concerne pas ici car les harmoniques sont de niveau voisin, de part et d'autre du formant physique. Nous avons par ailleurs vérifié (figure 2.29) que l'onde temporelle ne présente pas de changement d'allure marqué entre 660 et 743 Hz. Ce fait est d'ailleurs prévisible. Les formants F1 et F2 sont trop proches, et la durée de la période trop faible, pour que l'évolution temporelle soit sensible à une petite variation de l'un d'eux. Reste les effets de suppression. Si le biais en faveur de l'harmonique élevée est dû à un effet de suppression sur l'harmonique inférieure, on peut imaginer un biais contraire qu'exercerait le deuxième formant sur l'harmonique élevée du premier. L'examen des spectres auditifs montre que le pic du deuxième formant domine les harmoniques du premier de 1.3 dB à 688 Hz et 1.5 dB à 715 Hz. Le problème est que cela est vrai également pour les voyelles synthétisées à 188 Hz, où nous n'avons pas eu besoin de cette hypothèse. Mais il est vrai que les harmoniques dans les deux cas ne sont pas exactement aux mêmes fréquences. En fait, dans l'état actuel de nos connaissances, un effet de

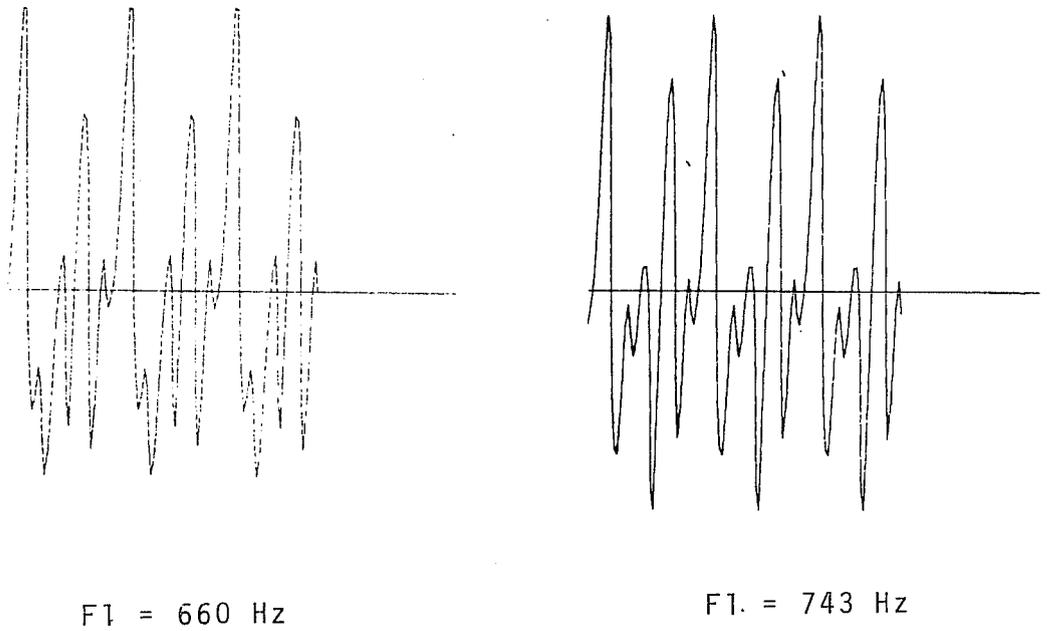


Figure 2.29 - Formes d'onde temporelle : le nombre d'oscillations ou de passages par zéro reste constant pour $F1 = 660$ et $F1 = 743$ Hz dans la série de 270 Hz.

suppression n'est pas une hypothèse testable. Par chance, les données pour les voyelles de 220 Hz montreront que l'effet de suppression ne peut pas être le seul "coupable".

Dans notre présentation de 1986, nous faisons grand cas des travaux sur la suppression au niveau du nerf auditif (ABBAS & SACHS 1976, JAVEL & al. 1983, DELGUTTE 1986). Ils montrent en effet qu'une sinusoïde sous la fréquence caractéristique (ou centrale : CF) d'une fibre possède un pouvoir de suppression supérieur à une sinusoïde de fréquence supérieure. Par analogie, on pourrait considérer que l'harmonique de basse fréquence exerce un effet de suppression sur l'harmonique haute, même si le niveau de cette dernière est légèrement plus fort. Le décalage constaté aurait donc une base auditive. L'examen de la figure 10 de JAVEL & al. (figure 2.30) montre par exemple que pour une fibre de fréquence centrale de 1 kHz, une sinusoïde de 800 Hz à 70 dB exerce un effet de suppression à 1 kHz bien supérieur à celui d'une sinusoïde de 80 dB à 1200 Hz. Plusieurs spécialistes, dont B. DELGUTTE et E. YOUNG nous ont fait remarquer que les données physiologiques au niveau du nerf auditif n'étaient pas en accord avec les données psychoacoustiques, qui reflètent d'évidence un traitement beaucoup plus élaboré de l'information (cf. par ex. DUIFHUIS 1980). Les résultats de ASSMANN & NEAREY pointent d'ailleurs d'avantage vers un effet de suppression de l'harmonique haute sur l'harmonique basse, ou vers un biais attentionnel sur l'harmonique haute.

3.6.6 - La série de 220 Hz.

Les voyelles dont la fréquence fondamentale est de 220 Hz fournissent par construction un test puissant de l'influence de la fréquence fondamentale sur le timbre. Les spectres physiques sont voisins de ceux de la série de 110 Hz, dont ils ne se distinguent que par l'absence d'une harmonique sur deux. Pour F1 à 660 Hz, les harmoniques voisines, à 440 et 880 Hz, sont -12.5 et -17.7 dB sous la plus proéminente. Le formant auditif devrait être proche de 660 Hz. Or la fonction d'identification montre que le timbre de cette voyelle est proche de [oe]. L'hypothèse d'une frontière phonétique fixe est donc mise en défaut. Une discussion complète sera présentée au paragraphe suivant. La figure 2.31 compare les spectres auditifs des deux voyelles à F1 placé à 660 Hz et confirme la très grande similitude des spectres auditifs. Il faut pousser F1 jusqu'à 743 Hz pour obtenir un changement de timbre net lorsque Fo est à 220 Hz. L'examen de l'amplitude des harmoniques est révélateur. A 688 et 715 Hz, les harmoniques voisines sont à plus de 10 dB sous l'harmonique la plus intense (- 12.2 , -15.3 dB et -10.0, -10.8 dB respectivement). La légère inflexion de la courbe d'identification à 688 Hz, qui apparaît pour 6 sujets sur 16, ne trouve pas son explication dans l'amplitude des harmoniques. Nous la considérerons comme une fluctuation statistique. A 743 et 770 Hz, l'harmonique de haute fréquence (880 Hz) émerge rapidement (- 6.4 et - 2.5 dB). On peut considérer que la hausse plus modérée

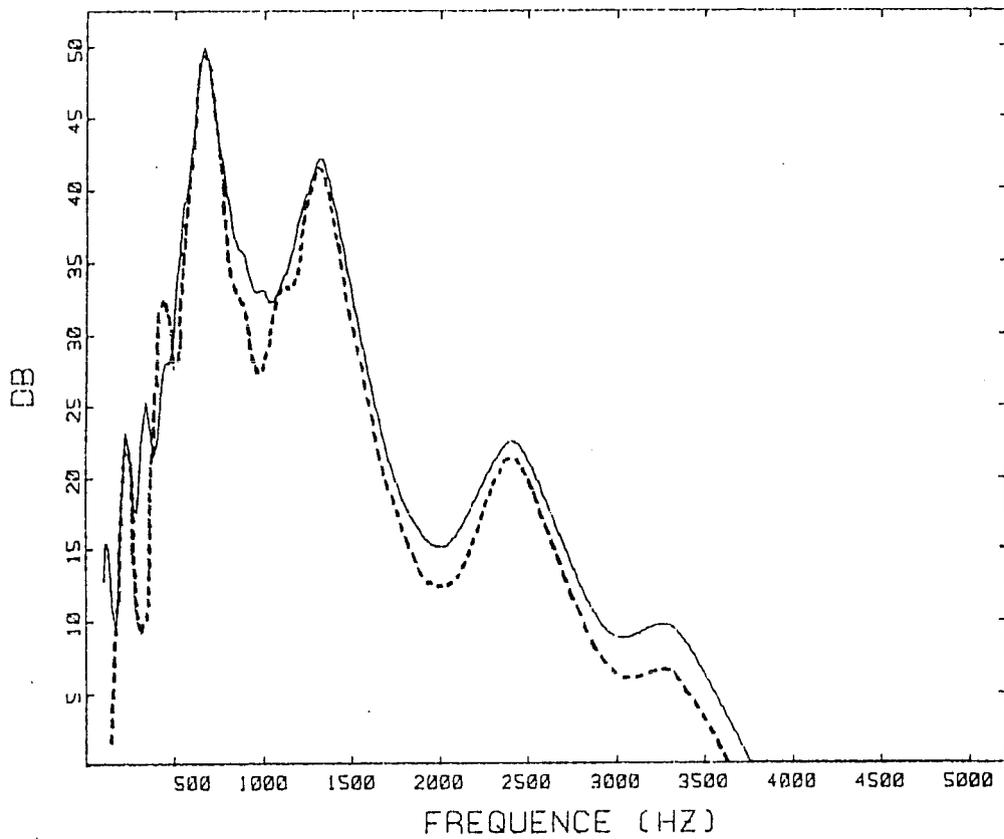


Figure 2.31 - Comparaison des spectres auditifs de deux voyelles physiquement identiques ($F_1 = 660$ Hz) à deux fréquences fondamentales différentes : 110 et 220 Hz. L'égalité des spectres auditifs n'implique pas celle du timbre.

de l'harmonique de basse fréquence (440 Hz : -8.0 et - 6.5 dB) ne peut s'opposer au déplacement du formant auditif, qui atteint et dépasse la frontière phonétique. Le décalage important de la fonction d'identification sur l'axe des fréquences formantiques physiques et la pente plus douce semblent, d'une manière évidente, liés au fait que seul un accroissement conséquent de la fréquence formantique permet à l'harmonique supérieure d'atteindre et de dépasser son seuil de contribution au positionnement du formant auditif. L'évolution parallèle de l'amplitude de l'harmonique haute et du timbre renforce notre conviction que la position du formant auditif est déterminée par la position de l'harmonique la plus proéminente et de celle de l'harmonique voisine si son amplitude est suffisante. Un seuil relatif de 10 dB paraît à nouveau rendre compte de nos résultats. Ceci constitue une seconde conclusion majeure de notre travail.

3.6.7 - Le problème du déplacement de la frontière phonétique.

Mais il reste à rendre compte du fait que pour les séries de fréquences fondamentales de 220 et 270 Hz, l'hypothèse d'une frontière auditive fixe à 660 Hz ne permet pas une prédiction du timbre perçu. La série de 220 Hz permet d'éliminer comme cause unique un effet de suppression. En effet, l'harmonique proéminente de 660 Hz domine le spectre, et il est probable qu'elle supprime plus qu'elle n'est supprimée. De plus, un effet de suppression ne conduirait qu'à une diminution de son amplitude, et non à une variation de la position du formant auditif. Aucune des conclusions que l'on peut tirer du travail d'ASSMANN & NEAREY ne s'applique au cas d'une harmonique intense isolée. Il est très difficile d'imaginer un biais d'origine acoustique qui déplacerait la position du formant auditif dans ce cas. Un biais de réponse dû à l'étendue du timbre ("range effect") n'est guère plausible. On sait que les sujets ont tendance à répartir leurs réponses de manière équilibrée entre les deux extrémités d'une échelle de réponses. La distribution des réponses n'est pas parfaitement équilibrée ici car les timbres [oe] dominent dans les réponses. Mais l'effet d'étendue devrait conduire à une augmentation du nombre de réponses [a], un biais jouant en sens inverse de celui constaté.

Trois possibilités demeurent pour rendre compte du déplacement de la frontière. On peut songer tout d'abord à un effet de la qualité de la synthèse. Les voyelles synthétisées sans inflexion de F_0 et avec un spectre rigoureusement constant sont peu naturelles et leur timbre, surtout quand F_0 est élevé, pourrait être senti comme moins distinct, ou moins extrême. Il se produirait ainsi un biais vers la voyelle neutre [oe]. On pourrait tester cette hypothèse en étudiant le couple [ø - oe]. Si le biais est vers la voyelle de timbre moins extrême, il devrait s'exercer dans le sens opposé à celui trouvé ici. Les résultats de CARRE & QUACH-TUAN (1986), qui montrent également un déplacement de la frontière dans le cas des voyelles [i] et [e], ne vont pas dans ce sens si [e] est pris comme une voyelle moins

extrême que [i].

La deuxième hypothèse s'appuie, par analogie, sur la modification de la fréquence centrale des fibres auditives qui se produit quand le niveau sonore s'accroît (cf. par ex. MØLLER 1983, p. 202 - 205). Celle-ci diminue pour les fibres de basse fréquence (3 kHz pour le rat par exemple), et augmente pour les fibres de haute fréquence. Il faut alors supposer que le codage de la fréquence du pic auditif s'effectue en tenant compte d'un facteur autre que la position tonotopique des fibres, déterminée par la position des cellules ciliées cochléaires, ou de la fréquence synchronisée des impulsions nerveuses, hypothèses généralement admises. On trouvera à l'annexe 3 un rappel de résultats sur le codage neurophysiologique des informations spectrales. Nous suggérons avec prudence que cette hypothèse rend compte de la modification du timbre des voyelles aux niveaux sonores élevés, dépassant 90 dB (DORMAN & al. 1986, 1987). Celui-ci est décrit comme un abaissement de F1, et une élévation de F2. La difficulté est que le niveau de présentation des voyelles, que nous n'avons pas mesuré, n'atteignait certainement pas une valeur aussi élevée. Mais il reste envisageable qu'à des niveaux sonores inférieurs se produisent de légères modifications qui ne peuvent être détectées que dans des tâches de forte sensibilité comme la nôtre, et non par une tâche d'identification d'un petit nombre de voyelles synthétiques couvrant l'ensemble du plan F1 / F2 comme dans DORMAN & al. (1986). Notons qu'à niveau sonore global identique, le pic proéminent d'une voyelle de fréquence fondamentale élevée sera plus intense que celui d'une voyelle de fréquence fondamentale plus basse. L'énergie de la voyelle est concentrée sur un plus petit nombre d'harmoniques aux fréquences fondamentales élevées. La modification auditive du pic pourrait donc être plus importante pour la voyelle de 220 Hz possédant le même spectre physique que la voyelle de 110 Hz, car le niveau absolu de l'harmonique la plus intense est supérieure dans le cas de la fréquence fondamentale élevée.

Mais ces deux hypothèses ne peuvent faire oublier la possibilité d'un changement effectif de la frontière auditive en fonction de la fréquence fondamentale. La cohérence des séries de 110 et 188 Hz permet d'éliminer la possibilité d'une erreur grave dans notre estimation de la fréquence du formant auditif pour les fréquences fondamentales basses. Les harmoniques inférieure et supérieure de la série de 110 Hz sont en effet juste au dessus, et en dessous, du seuil de 10 dB pour F1 = 660 Hz. Le formant auditif pourrait donc à la rigueur être un peu plus haut, ou plus bas, que 660 Hz. Mais une erreur importante ne permettrait pas de prédire le timbre de la série de 188 Hz, à moins de supposer que l'augmentation de F0 ramène le formant auditif vers 660 Hz. Cette éventualité nous paraît peu plausible. Il est impossible de donner une estimation précise du décalage de la frontière pour les fréquences fondamentales élevées car il faudrait connaître la fonction de pondération s'appliquant à l'harmonique moins intense. Pour la série de 220 Hz dont l'harmonique la plus proéminente est à 660 Hz, on constate qu'une augmentation du niveau de l'harmonique haute (880 Hz)

de -10.8 (F1 = 715 Hz) à -6.4 dB (F1 = 743 Hz) suffit à replacer la voyelle sur la frontière, ce qui suggère que celle-ci n'est pas très éloignée de 660 Hz. 50 Hz nous paraît un ordre de grandeur plausible, et une valeur inférieure ne peut être exclue. Plus haut, nous avons estimé, pour la série de 270 Hz, la variation de la frontière à 30 Hz, mais des effets plausibles, comme la suppression ou une pondération plus forte de l'harmonique haute, impliqueraient un décalage supérieur. L'imprécision est telle qu'il est impossible de dire si le décalage de la frontière phonétique est proportionnel à la fréquence fondamentale. Il n'est certainement pas égal à la variation de la fréquence fondamentale, quelle que soit l'échelle, comme nous le montrerons au paragraphe suivant. Par son ampleur limitée, il apparaît comme un effet de second ordre.

Il n'est pas difficile de suggérer une cause possible, d'origine non auditive. Les formants féminins, pour des fréquences fondamentales voisines de celles envisagées ici, montrent un décalage dépassant 15%, c'est-à-dire de plus de 100 Hz. Notons qu'une frontière auditive décalée de 15%, vers 760 Hz donc, ne serait atteinte pour la série de 220 Hz qu'avec un formant physique de 800 Hz environ. Or nos résultats montrent par interpolation que cette voyelle n'est pas intermédiaire entre [oe] et [a] mais possède un timbre proche de [a]. L'absence de correspondance entre le léger décalage de la frontière, et la forte différence entre les réalisations masculines et féminines ne doit pas conduire à rejeter immédiatement cette explication.

On peut supposer que les sujets ont "appris" d'expérience que les formants féminins et enfantins sont placés plus haut que les formants masculins. Il paraît d'ailleurs obligatoire d'augmenter la fréquence de la frontière, sinon à geste articuloire homologue, les formants des [oe] féminins se trouveront à des fréquences proches de ceux des [a] masculins. Les simulations de TRAUMÜLLER (1983, 1984) démontrent qu'il est possible de prévoir, avec une bonne précision, les différences entre les fréquences formantiques des voyelles masculines et féminines en ne tenant compte que des différences anatomiques : longueur du conduit vocal, variation du rapport des longueurs du conduit buccal et pharyngal. Les gestes articulatoires vocaliques sont donc très vraisemblablement homologues. Ils ne visent donc pas à la production de formes acoustiques soumises à des contraintes spécifiques et différentes pour chaque sexe. En d'autres termes, la possibilité de confusions auditives ne paraît pas être un facteur exigeant une modification des positions articulatoires féminines relativement à celles des hommes.

On peut ne pas comprendre dans un premier temps, pourquoi les frontières phonétiques ne sont pas décalées d'une différence auditive égale à la différence physique des réalisations vocaliques. Il semble que le petit décalage de la frontière qui ressort de nos données soit insuffisant pour assurer une perception conforme au timbre de la voyelle prévue par le geste articuloire. Concrètement, la fréquence moyenne du premier formant des [oe] féminins (LONCHAMP, sous presse) est voisine de 650 Hz. Cette moyenne est très proche de celle

de la frontière phonétique. Il ne paraît donc pas exclu, l'écart-type étant de 58 Hz, qu'un nombre non négligeable de [oe] féminins aient un premier formant auditif au-dessus de la frontière, même si l'on tient compte du fait que celle-ci est légèrement plus élevée que 660 Hz en raison de la fréquence fondamentale haute des voix féminines. Or on ne constate pas d'ordinaire de confusions entre les [a] masculins et certains [oe] féminins.

Le raisonnement précédent est inexact car il ne tient pas compte de la fréquence du deuxième formant. Lorsque celui-ci s'élève, comme dans les productions féminines, il est probable que la frontière auditive de F1 s'élève également. Nous n'avons pas directement testé ce point, mais il est visible dans les résultats de CARLSON, GRANSTRÖM & FANT (1970) : sur l'axe F2, la limite entre les stimuli identifiés comme Ø et A est d'autant plus élevée que F2 est plus haut. Les confusions prévues ne se produiraient que si la fréquence de F2 restait identique. Mais un geste articulatoire féminin homologue du geste masculin conduit automatiquement à une fréquence de F2 plus élevée. La fréquence de F2 dans cette expérience est de 1320 Hz. Or la moyenne des F2 féminins pour la voyelle [oe] est de 1690 Hz, celle des réalisations masculines étant à 1390 Hz. Pour ce qui concerne F1, la frontière ne se trouvera pas vers 660 Hz dans les réalisations féminines, mais à une fréquence plus élevée. Dans ces conditions, aucun [oe] féminin ne devrait être perçu comme [a].

Ces mêmes remarques valent pour [a]. D'après nos mesures, la moyenne du premier formant masculin s'établit à 684 Hz seulement, avec un écart-type de 47 Hz. Avec une frontière phonétique à 660 Hz, beaucoup de [a] masculins devraient être perçus comme [oe]. Mais la fréquence moyenne de F2 n'est que de 1256 Hz, plus basse que celles des voyelles de notre expérience (1320 Hz). La frontière est donc elle-même plus basse pour ces valeurs plus faibles de F2. La marge de sécurité des voyelles féminines paraît néanmoins supérieure. Il est réconfortant enfin de constater que les résultats de l'expérience perceptive, utilisant des voyelles synthétiques, sont au moins qualitativement cohérents avec les résultats de mesures acoustiques.

Cette discussion incite au test d'une nouvelle hypothèse pour rendre compte du décalage de la frontière phonétique. L'harmonique la plus intense de la zone de F1 provoquerait un effet de suppression sur F2, qui est souvent de moindre amplitude. Si cette suppression "mord" sur le flanc gauche de F2, sa fréquence pourrait être décalée vers le haut, et la frontière serait elle-même plus haute. On peut rejeter cette hypothèse en remarquant que pour F1 = 660 Hz, les amplitudes de F1 et F2 diffèrent peu lorsque les fréquences fondamentales sont de 110 et 220 Hz. Un effet de suppression éventuel serait sans doute de même ampleur dans les deux cas, et la différence de timbre n'est pas expliquée.

3.6.8 - Le caractère 'masculin' et 'féminin' des voyelles

D'un faible décalage de la frontière phonétique, on peut déduire une conséquence

importante. La différence des fréquences formantiques entre les productions masculines et féminines, pour ne pas parler des productions enfantines, est supérieure à celle qui semble nécessaire pour conserver le timbre des voyelles à des fréquences formantiques différentes. En admettant que les séries de 220 et 270 Hz soient représentatives des variations physiques nécessaires, la moyenne du décalage physique de F1 est de l'ordre de 10%. Cette valeur est voisine de celle mesurée par SLAWSON (1968). Or le décalage mesuré dans les productions naturelles est de 25 et 15 % pour le premier formant de [oe] et [a] (LONCHAMP, sous presse). Ceci implique que les voyelles féminines sont plus extrêmes, c'est-à-dire plus loin de la frontière que strictement nécessaire. Bien que la forme du mécanisme d'assignation d'un timbre particulier ne soit pas connue avec précision, on peut supposer que ces valeurs formantiques plus extrêmes sont un corrélat de la qualité "féminine" de la voix, au même titre que la forme différente de l'onde glottale ou le niveau plus élevé de bruit de haute fréquence présent dans le spectre glottal. Il est bien sûr possible d'inverser cette conclusion en disant que les voyelles masculines sont moins extrêmes, si on refuse de prendre les productions masculines comme norme ! On sait que la seule élévation de la fréquence fondamentale ne procure pas à une voyelle synthétique un timbre "féminin". Une enquête à posteriori a révélé que 3 seulement de nos sujets sur 16 considéraient les voyelles présentées aux deux fréquences fondamentales élevées comme ayant un timbre "féminin". 6 ont associé toutes les voyelles à une production artificielle, par ordinateur. 7 sujets ont déclaré que toutes les voyelles possédaient plutôt un timbre "masculin".

3.6.9 - Une distance F1-Fo constante ne conserve pas le timbre invariant.

L'hypothèse d'une distance constante entre Fo et F1, dérivée de celle de TRAUNMÜLLER (1981), n'est pas validée par notre expérience. Pour Fo = 270 Hz, la fréquence de F1 qui place la voyelle sur la frontière doit être inférieure ou égale à la fréquence de la troisième harmonique (810 Hz), qui est la plus intense. Si on admet qu'à Fo = 110 Hz le formant physique et auditif de la voyelle de timbre analogue est à 660 Hz, une augmentation de Fo de 1.8 Bark est compensée par une élévation de F1 d'au plus 1.1 Bark. (660 à 810 Hz). Si l'on tient compte du fait que la seconde harmonique de la voyelle à 270 Hz n'est que 4.5 dB sous la plus proéminente, et qu'elle abaisse en conséquence le formant auditif, l'erreur de prédiction sera encore plus grande. Pour la série de 220 Hz, le formant auditif de la voyelle ambiguë devrait se trouver à 835 Hz. Cette valeur est obtenue en ajoutant l'écart en Bark entre les fréquences fondamentales (1.28 Bark) à la fréquence du formant auditif (660 Hz = 6.22 Bark), ce qui donne 7.50 Bark, c'est-à-dire 835 Hz. Or, d'après notre discussion ci-dessus, il ne peut se trouver plus haut que 710 - 730 Hz. Nous utilisons les fréquences des formants auditifs pour cette discussion, car nous ne voyons aucune raison de supposer

que les sujets ont accès aux fréquences formantiques physiques dans cette expérience. L'utilisation des écarts calculés sur l'échelle ERB est aussi défavorable : la différence entre les fréquences fondamentales atteint 3.47 unités (110 - 270 Hz), alors que la variation du formant auditif n'est, dans le cas le plus favorable à l'hypothèse (660 - 810 Hz), que de 1.5 unités. Pour la voyelle de 220 Hz, le formant auditif se trouverait à 925 Hz. La conclusion est que, pour la gamme de F_0 que nous avons utilisée, la règle $F1_{\text{auditif}} - F_0$ ne prédit pas la conservation du timbre vocalique.

3.6.10 - Analyse des résultats de CARRE (1988).

Il ne paraît possible de donner une interprétation des résultats de CARRE (1988) identique à celle des nôtres que pour un seul sujet sur 3. CARRE étudie l'identification de voyelles synthétiques de timbre variant entre [i] et [e] à des fréquences fondamentales entre 140 et 200 Hz (figure 2.32). Les formants supérieurs sont constants (2100, 2850, 3400 Hz) et la fréquence de $F1$ varie de 250 à 350 Hz par pas de 25 Hz. Ces paramètres sont ceux de CARLSON & al. (1975). La durée des voyelles est de 200 ms dans l'expérience principale, et de 50 ms dans une expérience complémentaire. En utilisant une source "raisonnable", nous avons calculé l'amplitude des harmoniques importantes en tenant compte de la fonction de transfert de l'oreille externe et moyenne. Une connaissance exacte du spectre glottal ne serait utile que si l'on connaissait précisément les valeurs de pondération à appliquer aux harmoniques en fonction de leur niveau relatif. Faisons l'hypothèse que deux facteurs déterminent la position du formant auditif et le timbre perçu. Le premier est la position de l'harmonique la plus intense et des harmoniques proches à moins de 10 dB sous le niveau de la plus proéminente, et le deuxième la position de la frontière phonétique, qui peut dépendre du sujet. Rappelons que la valeur de -10 dB est compatible avec nos données et celles de ASSMANN & NEAREY.

Pour le sujet RC, la frontière phonétique est à 300 Hz : pour $F_0 = 150$ Hz et $F1$ à 300 Hz, la voyelle est ambiguë. Les autres harmoniques sont plus de 13 dB sous celle de 300 Hz, et ne contribuent pas au placement du formant auditif. Pousser le formant à 325 Hz ne change pas de façon significative le niveau des harmoniques (-13.3 à -11.5 dB à 150 Hz, -19.4 à -14.7 dB à 450 Hz), et le timbre reste approximativement le même. En revanche, pour $F1 = 350$ Hz, l'harmonique de 450 Hz n'est plus qu'à 8.8 dB sous celle de 300 Hz, et décale le formant auditif vers [e], si on admet que l'amplitude plus élevée de F_0 (-8.8 dB) n'est pas pertinente. Cette dernière valeur est approximative car elle dépend de la source glottale réellement employée. A $F_0 = 140$ Hz, l'harmonique proéminente est à 280 Hz. Le niveau de F_0 est relativement faible (-10.5 dB). Mais l'harmonique suivante (420 Hz) est à -10.7 dB sous la deuxième dès que $F1$ est égal à 325 Hz, ce qui fait pourrait déplacer le timbre vers [e], bien que le niveau de F_0 soit identique. Pour F_0 à 160 Hz, l'harmonique de haute

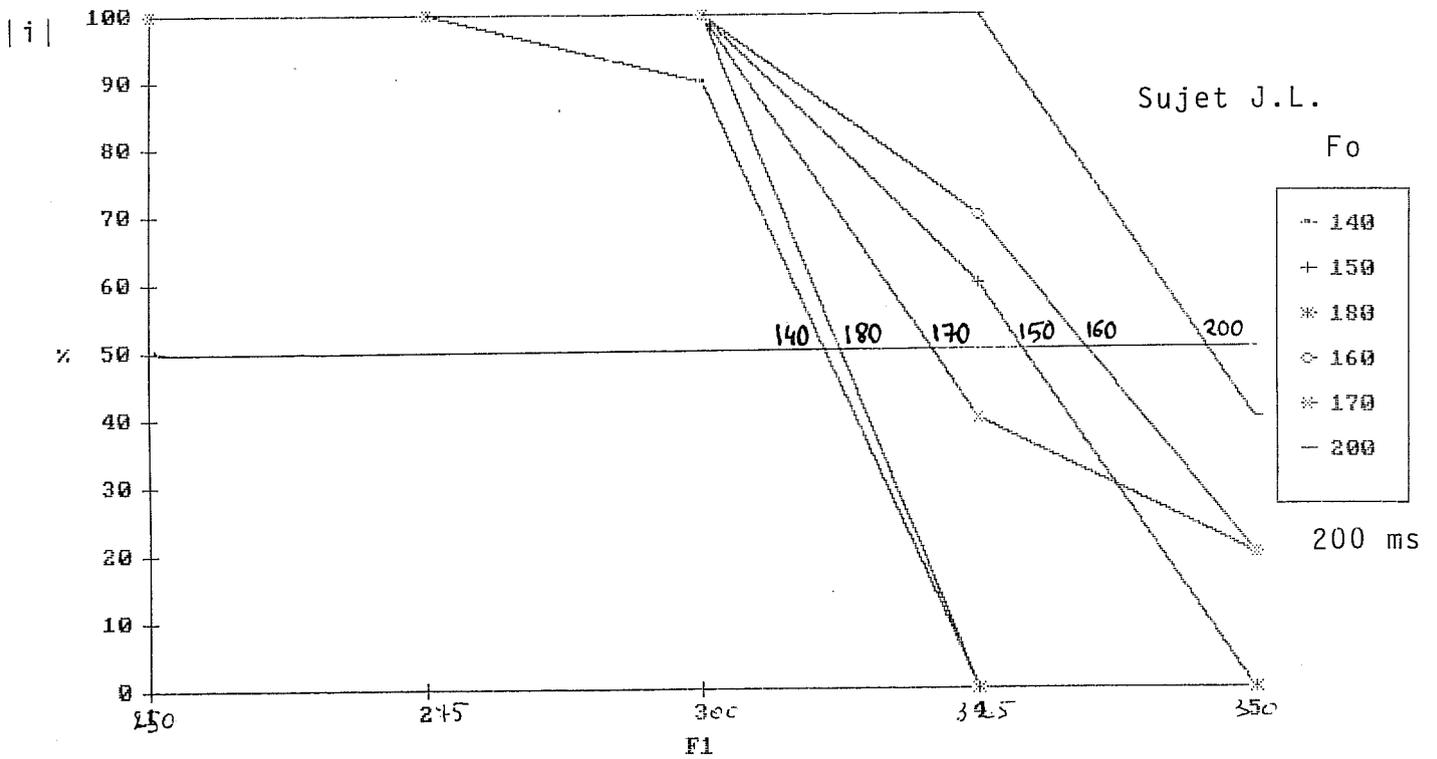
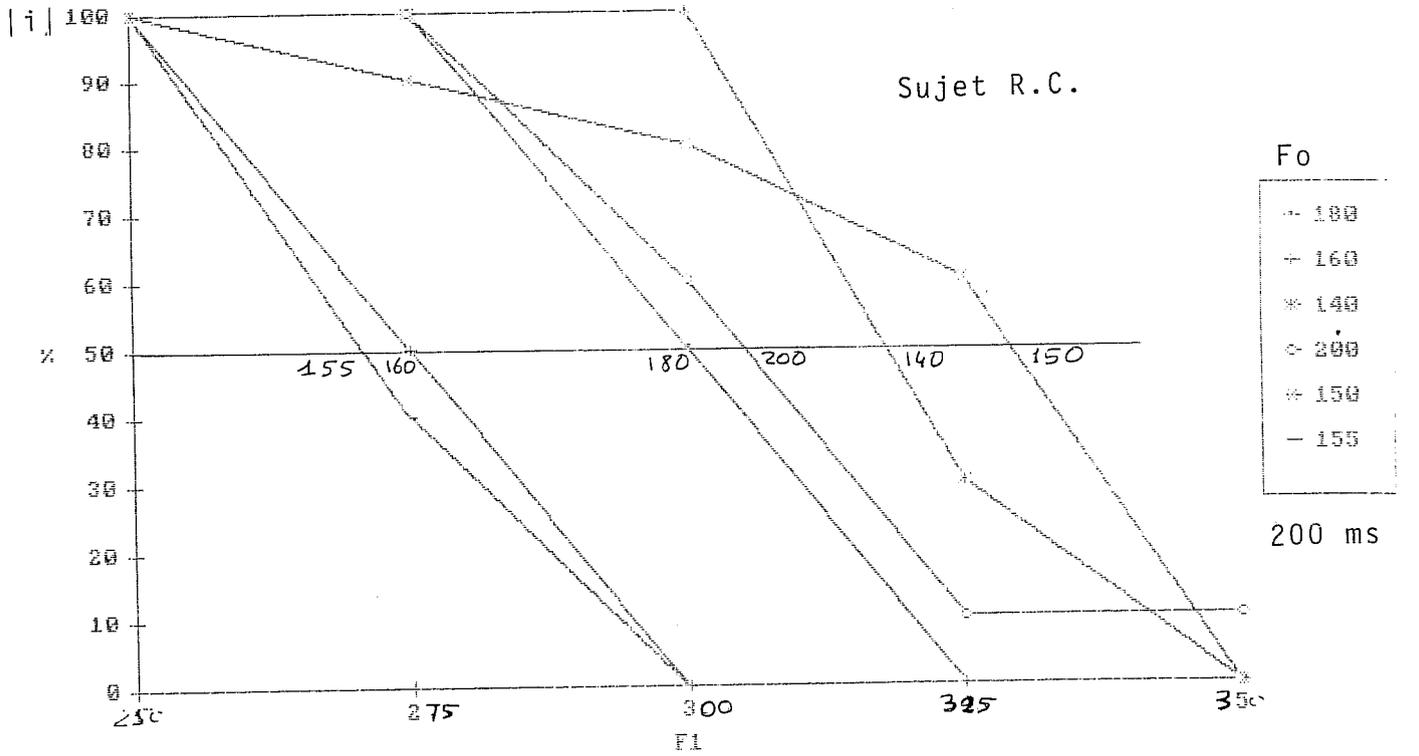


Figure 2.32 - Résultats individuels d'identification de CARRE (1988).

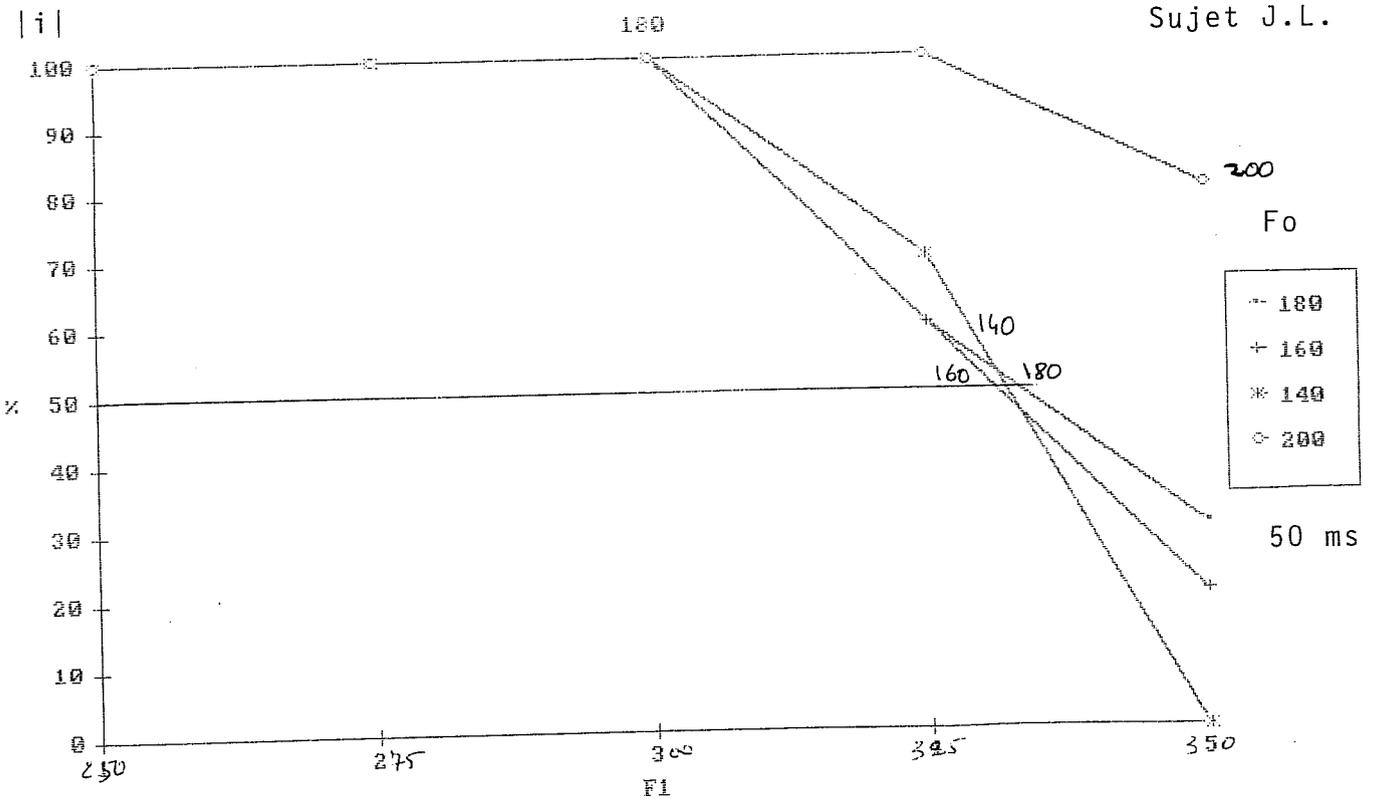
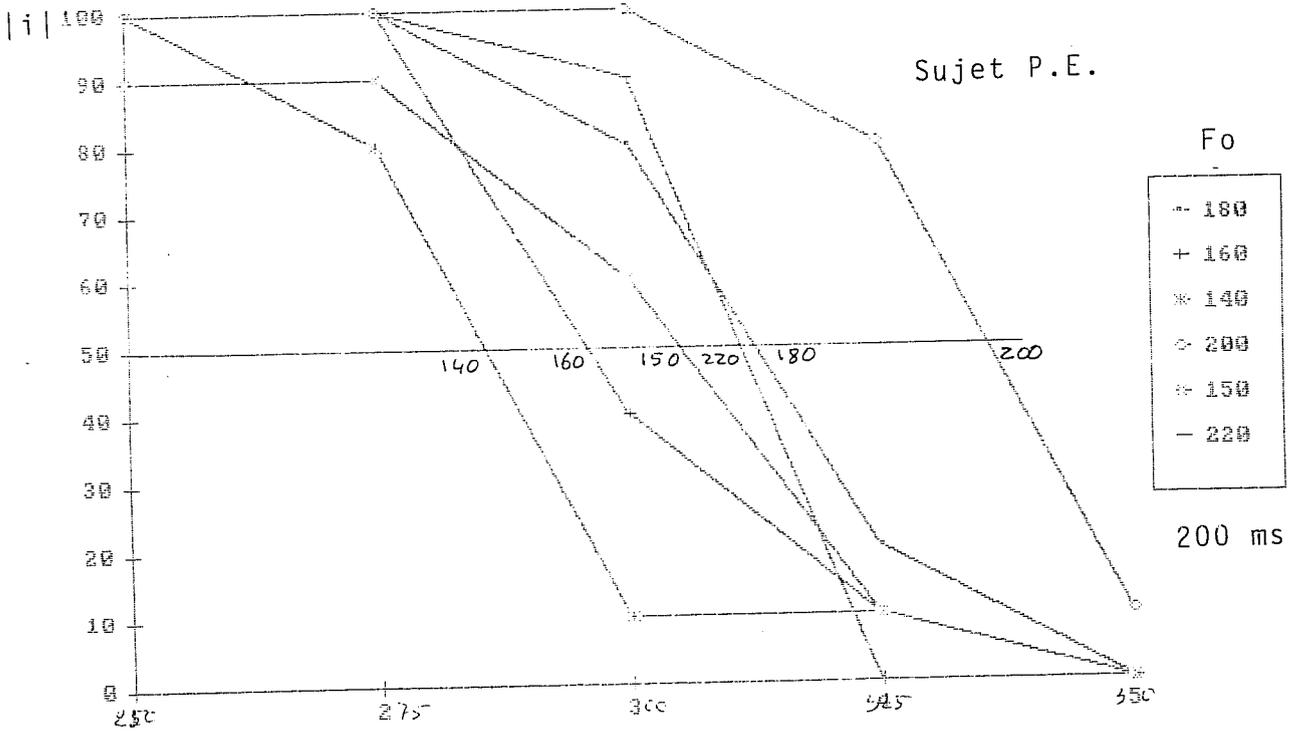


Figure 2.32 (Suite).

fréquence (480 Hz) reste à des niveaux faibles (-19.8 à -14.2 dB). En revanche, pour F1 à 275 Hz, la fréquence fondamentale est seulement - 5.4 dB sous celle de 320 Hz. Pour placer le pic auditif vers 310 Hz, il faut admettre que l'amplitude de la fréquence fondamentale n'a que peu d'influence. A 300 Hz, au contraire, le niveau de Fo descend à -11.3 dB. La fréquence du pic est celle de la fréquence de l'harmonique proéminente (320 Hz), au dessus de la frontière phonétique, d'où le timbre [e]. Cette analyse est plus difficile pour Fo à 155 Hz, où l'harmonique n'est qu'à 310 Hz. L'énorme distance entre les fonctions d'identification des voyelles à 150 et 155 Hz (60 Hz environ) est surprenante. Notre analyse suppose une frontière très précise, et une règle de décision très stable. On ne peut évoquer un effet de phase dans les conditions que nous avons décrites précédemment. En effet, pour F1 = 325 Hz, l'harmonique proéminente est à gauche du formant pour Fo de 150 à 160 Hz. On ne voit pas bien pourquoi la suppression aurait des effets si différents sur des spectres très voisins. Il faudrait certainement vérifier les effets de contraste potentiel dans l'ordre de présentation des voyelles et la forme de l'onde temporelle. A 180 et 200 Hz, l'amplitude de la fréquence fondamentale est supérieure ou égale (+ 7.5 à - 2.5 dB) à celle de la seconde harmonique pour F1 égal à 275 et 300 Hz. Le formant auditif est donc à une valeur intermédiaire, entre Fo et 2 x Fo. Il faut que F1 atteigne 325 Hz pour que la seconde harmonique à 360 ou 400 Hz devienne proéminente (+ 8.7 et +2 dB) et pousse ainsi le formant auditif au delà de la frontière phonétique. La proéminence nécessaire est moindre quand Fo est à 200 Hz, car l'harmonique de 400 Hz est plus loin de la frontière et la moyenne est plus élevée.

Pour le sujet RC, la prise en compte du niveau des harmoniques à moins de 10 dB de la plus proéminente conduit à une interprétation plausible de l'évolution du timbre en fonction des fréquences formantiques et de la fréquence fondamentale. Mais il faut admettre qu'une harmonique de fréquence supérieure à celle de la plus proéminente peut avoir un poids supérieur à une harmonique basse d'intensité supérieure.

Pour le sujet PE, en revanche, les résultats ne sont pas compatibles avec l'interprétation précédente. Pour Fo à 150 Hz et F1 à 300 Hz, l'harmonique de 300 Hz est seule proéminente (-13.3 et -19.4 à 150 et 450 Hz). Le timbre est ambigu, ce qui implique que la frontière est très proche de 300 Hz. C'est également le cas à 160 Hz, pour l'harmonique de 320 Hz (-11.3 et -19.2 dB à 160 et 480 Hz) dont la fréquence un peu plus haute que précédemment rend compte du timbre un peu plus ouvert. Le changement de timbre pour F1 égal à 325 Hz n'est compréhensible que si des harmoniques de niveau faible jouent un rôle. L'harmonique de 480 Hz reste exactement au même niveau (-19.2 dB) et celle de 180 Hz décroît de 2.8 dB. On peut plaider que la baisse du niveau de la fréquence fondamentale (160 Hz) décale le formant auditif vers une fréquence plus élevée, mais il faut alors admettre que cette dernière joue un rôle non négligeable à -11.3 dB sous la plus proéminente, qui disparaît à -14.1 dB. Pour Fo à 150 Hz, cette analyse ne tient plus, car l'harmonique basse

augmente de 1.8 dB. Il est vrai que l'harmonique haute augmente de 4.7 dB mais elle reste très faible (-14.7 dB). Les problèmes paraissent insurmontables pour la fréquence fondamentale la plus basse. Pour F1 à 300 Hz, le timbre [e] provient d'une harmonique proéminente à 280 Hz, et de deux harmoniques à -12.3 (140 Hz) et -16.9 dB (440 Hz). Il n'existe aucun argument pour "pousser" le formant auditif au delà de la frontière phonétique vers 300 Hz. L'hypothèse d'un déplacement de la frontière ne résoud pas le problème. Le niveau des harmoniques pour F1 à 275 et 300 Hz est très proche : moins de 0.1 dB d'écart à 140 Hz, et 3 dB à 440 Hz, mais -16.9 dB sous le pic. Les résultats aux fréquences fondamentales élevées sont de moindre intérêt car la fréquence fondamentale peut dominer l'harmonique suivante et le pic auditif est à une valeur intermédiaire qui doit dépendre à la fois de la moyenne entre les composantes et du rapport des amplitudes.

Nous reviendrons bientôt sur les problèmes posés par les résultats de ce sujet.

Penchons nous vers les données du dernier sujet qui ne montrent qu'une faible influence de la fréquence fondamentale sur le timbre. On peut songer à placer la frontière vers 330 Hz, en remarquant que pour les fréquences fondamentales de 150 à 170 Hz, les harmoniques proéminentes (300 à 340 Hz) sont plus de 11 dB au dessus des autres quand F1 est égal à 325 Hz. Mais pour expliquer le timbre [i] à 300 Hz il faut supposer que l'amplitude de la fréquence fondamentale joue un rôle. Si elle est de -6.5 dB pour Fo égal à 170 Hz, elle n'est que de -13.3 dB pour Fo = 150 Hz, ce qui est trop faible pour lui assigner le moindre rôle. La fréquence fondamentale la plus basse (140 Hz) pose également un problème insurmontable. Pour F1 à 325 Hz, le timbre est [e] alors que l'harmonique proéminente est à 280 Hz, loin de la frontière phonétique (330 Hz), et, comme pour le sujet PE, rien dans l'amplitude de l'harmonique supérieure (-10.7 dB) ne permet de supposer que le formant auditif est décalé vers le haut de 50 Hz.

Pour ce sujet également, il paraît impossible de prévoir la position du formant auditif avec les hypothèses énoncées plus haut : rôle majeur de l'harmonique proéminente et de l'harmonique placée plus haut si son amplitude n'est pas trop faible, d'environ -10 dB au plus.

CARRE (1988) remarque, en faisant la synthèse de la plupart des données disponibles (CARLSON & al 1975, CARRE & QUACH-TUAN 1986, LONCHAMP 1986, CARRE 1988), que la discontinuité dans la position estimée de la frontière phonétique se produit lorsqu'une harmonique coïncide avec la fréquence formantique proche de la frontière phonétique (figure 2.33). Cet effet est peu visible pour les sujets PE et JL. Il suggère deux hypothèses. Dans cette configuration, la voyelle est plus facilement identifiable comme l'une ou l'autre des voyelles. Nous avouons ne pas comprendre exactement pourquoi ceci conduirait à une discontinuité de la frontière. Ou bien, les frontières entre deux voyelles se situent dans les parties du plan Fo / F1 où le premier formant ne comprend pas d'harmoniques en son milieu. Une raison serait que la définition du timbre serait meilleure

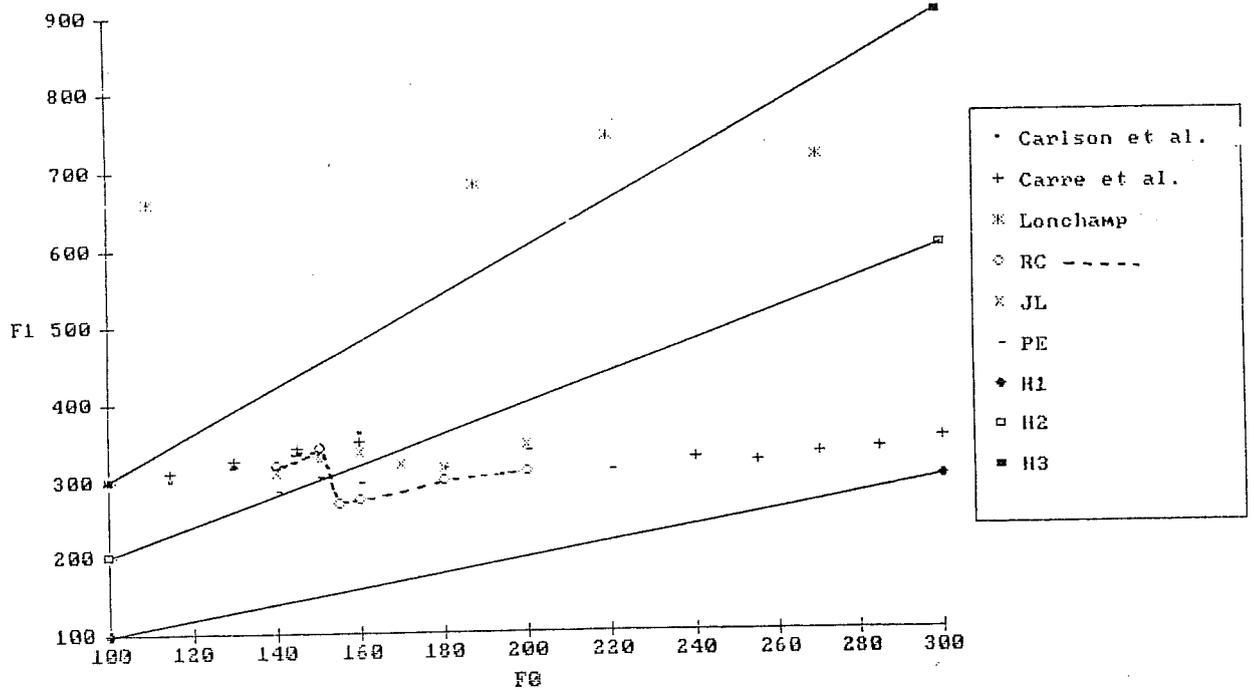


Figure 2.33 - Evolution de la frontière phonétique dans plusieurs expériences d'identification. (in CARRE 1988).

lorsque une harmonique coïncide avec la fréquence formantique.

Nous croyons que la discontinuité de la frontière, lorsqu'elle existe, reflète exactement les mouvements du formant auditif. Un exemple théorique le montrera clairement (figure 2.34). Supposons que la frontière phonétique corresponde à une fréquence F du formant auditif. Plaçons le formant physique à cette fréquence. Si la fréquence fondamentale est telle que l'harmonique la plus proéminente est légèrement inférieure à la fréquence du formant physique, le formant auditif sera à la fréquence de l'harmonique la plus proéminente puisque les autres harmoniques seront d'amplitude faible. Le formant auditif est donc sous la frontière. Pour le faire passer au dessus de la frontière, à F_0 constant, il faudra augmenter la fréquence du formant physique de manière importante pour amener l'harmonique supérieure à un niveau élevé. Le graphe de la position du formant physique (en fonction de F_0) qui produit une voyelle ambiguë, c'est-à-dire sur la frontière, montrera donc l'inflexion vers les hautes fréquences visible sur la figure de CARRE (figure 2.32) : un formant physique élevé est nécessaire pour augmenter suffisamment l'amplitude de l'harmonique supérieure lorsque l'harmonique la plus proéminente est à une valeur juste inférieure à la fréquence du formant. Si, en revanche, l'harmonique la plus proéminente est au dessus de la fréquence du formant physique, et donc le formant auditif au dessus de la frontière, il faudra largement baisser le formant physique (à F_0 constant) pour que l'harmonique inférieure devienne assez intense pour abaisser le formant auditif. Mais tout ceci n'explique pas les données des sujets PE et JL.

3.6.11 - Le rôle de l'attaque.

L'examen des courbes d'identification du sujet JL lorsque la durée des voyelles est réduite à 50 ms suggère une explication. Les fonctions d'identification pour les séries de 140, 160 et 180 Hz se superposent exactement. Tout se passe comme si ce sujet était capable de recouvrer la fréquence physique du formant et employait une frontière constante à 325 Hz. En revanche, pour $F_0 = 200$ Hz, la frontière est supérieure à 350 Hz. Pour les voyelles de 200 ms, nous avons déjà noté une grande proximité des courbes d'identification de ce sujet, sauf pour $F_0 = 200$ Hz. Lorsque F_1 est à 325 Hz, la fréquence fondamentale de 200 Hz et l'harmonique suivante sont à quelques dB l'une de l'autre (49 et 50 dB dans notre simulation). On comprend que la fréquence formantique soit mal définie. Une augmentation de la fréquence de F_1 abaisse rapidement l'amplitude de la fréquence fondamentale. Ces résultats rappellent ceux de SLAWSON (1968) sur les voyelles à attaque brutale, où le timbre était, pour certains sujets, indépendant de la fréquence fondamentale. Nous faisons l'hypothèse que ce sujet concentre consciemment ou inconsciemment son attention sur l'attaque vocalique. De plus, nous croyons qu'un sujet peut mettre en oeuvre deux mécanismes de codage distincts. Les indices expérimentaux sont analysés à l'annexe 3.

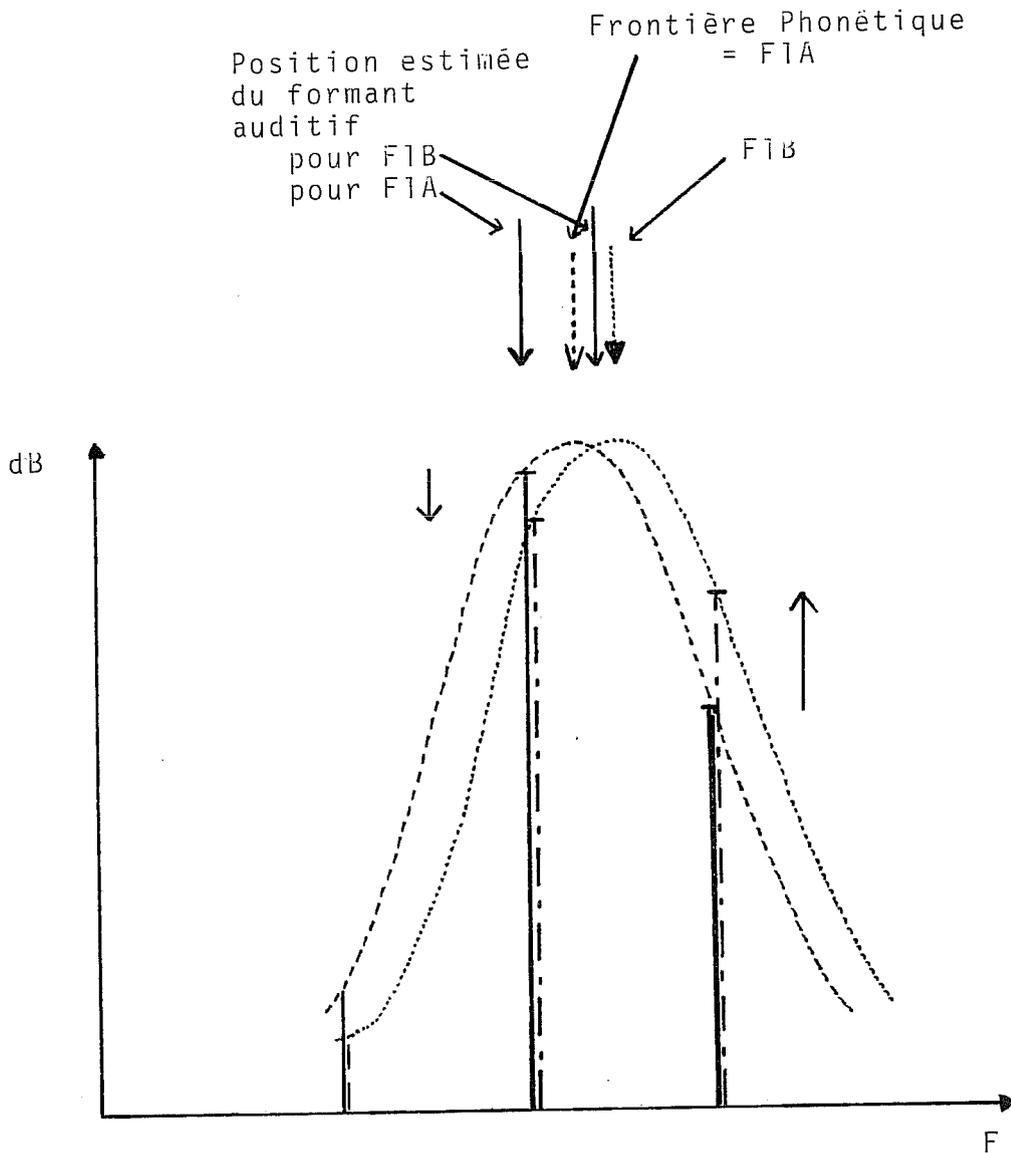


Figure 2.34 - Représentation schématique de la position du formant auditif en fonction de l'amplitude des harmoniques: Soit un formant physique sur la frontière phonétique F1A. Lorsque l'harmonique inférieure domine, le formant auditif est sous la frontière. Il faut une élévation importante du formant physique (F1B) pour amener le formant auditif sur la frontière.

Pour le sujet PE, et RC dans le cas des voyelles brèves, nous suggérons que les sujets utilisent les informations provenant des deux mécanismes de codage. Nous ne présenterons pas une analyse détaillée, car l'hypothèse de deux mécanismes augmente le risque d'explications ad hoc. On remarquera cependant que dans les deux cas, la frontière la plus élevée est obtenue pour une fréquence fondamentale de 200 Hz, et que l'écart entre les fonctions d'identification est réduit.

Les résultats de CARRE & QUACH-TUAN (1986), qui portent sur des voyelles synthétiques de 70 ms identiques à celles de CARLSON & al. (1975), montrent que la frontière phonétique entre [i] et [e], mesurée par la position formantique physique donnant un timbre ambigu, fluctue en fonction de la fréquence fondamentale. Le petit tableau suivant est extrait de l'article cité.

Fo	100	115	130	145	160	240 Hz
Position de la frontière	275	285	315	290	300	290 Hz
Fo	255	270	285	300 Hz		
Position de la frontière	285	315	315	335		

Le premier point de rebroussement se produit pour Fo entre 130 et 145 Hz. Il peut s'expliquer si la frontière phonétique se place vers 290 - 300 Hz. En effet, à 130 Hz, l'harmonique la plus intense est à 260 Hz. Pour un formant proche, à 300 Hz, les harmoniques adjacentes sont d'intensité faible, et le formant auditif reste sous la frontière. Ce n'est que lorsque F1 atteint 325 Hz que l'harmonique haute est suffisamment intense pour attirer le formant auditif au delà de la frontière. Pour Fo = 145 Hz et F1 = 300 Hz, la seule harmonique intense est à 290 Hz, très proche de la frontière, d'où le timbre ambigu. On notera que les fréquences fondamentales élevées semblent révéler un déplacement de la frontière. Pour Fo et F1 à 300 Hz, le timbre est encore [i]. Le formant doit être placé plus haut, entre 325 et 350 Hz pour que le formant auditif, qui coïncide ici avec la fréquence fondamentale, soit sur la frontière phonétique.

Il est curieux de constater que les résultats présentés sont ceux de la synthèse produite en contrôlant l'amplitude de la source avant son "injection" dans les filtres résonants. Le spectre de l'attaque, calculé sur 12.8 ms de signal pondéré par une fenêtre de Hamming, présente physiquement un petit maximum d'énergie à la fréquence physique de F1. Ce détail spectral n'est pas présent lorsque l'amplitude de l'attaque est contrôlée après le passage de l'onde glottale dans le filtre. Or l'évolution du timbre décrite plus haut semble bien résulter de la mesure du niveau des harmoniques : le bref surcroît d'énergie à la fréquence de F1 ne paraît pas perturber significativement l'évolution du timbre. La mesure de la position physique du premier formant ne semble pas une hypothèse tenable. Une telle

mesure conduirait à une frontière phonétique stable, comme celle du sujet JL chez CARRE (1988). Les frontières phonétiques des voyelles synthétisées en contrôlant l'amplitude après le filtrage formantique sont systématiquement plus hautes. De plus, la frontière physique ne montre pas de point de rebroussement pour $F_0 = 145$ Hz. Elle s'accroît encore à $F_0 = 160$ Hz, pour atteindre 350 Hz. Ce comportement est compréhensible si l'on suppose que le formant auditif doit être au delà de 320 Hz pour que le timbre passe de [i] à [e]. Mais la raison du changement de critère n'apparaît pas. Il faudrait peut-être la chercher dans l'amplitude plus forte des formants supérieurs dans l'attaque, lorsque la transition d'amplitude est appliquée à la source glottale.

Les résultats obtenus par CARRE peuvent se résumer ainsi. Un sujet, RC, utilise vraisemblablement le même mécanisme de mesure du formant auditif que nos sujets. Mais, dans le cas d'une voyelle brève, il s'appuie en partie sur un second mécanisme, qui semble largement employé par le sujet JL. Ce second mécanisme, que nous décrivons plus en détail dans l'annexe 3, permet à l'auditeur de recouvrer la fréquence physique du premier formant avec une bonne précision en s'appuyant sur un traitement particulier de l'attaque de la voyelle. Mais cette stratégie ne semble efficace que lorsque le formant n'est pas défini par deux harmoniques d'égale amplitude. Le comportement du troisième sujet est intermédiaire. La frontière auditive entre la voyelle [i] et [e] est à une fréquence plus élevée lorsque F_0 est lui-même élevé. Il reste un point obscur. Lorsque l'attaque de la voyelle est déterminé par une variation appliquée à la partie stable de la voyelle, la frontière auditive semble plus élevée que dans le cas où l'amplitude de la source glottale est responsable de l'attaque, comme dans la parole naturelle. Il est clair que dans cette deuxième configuration de synthèse, le spectre initial comporte des informations supplémentaires sur la fonction de transfert de la voyelle, mais l'origine du biais de la frontière auditive reste mystérieux. En tout état de cause, le mécanisme utilisé par le sujet JL ne paraît pas en être responsable.

3.6.12 - Un problème potentiel : La non linéarité des vibrations de la membrane basilaire aux très basses fréquences.

Avant de conclure, nous voulons suggérer que les résultats obtenus par M. H. HOLMES (1980, 1987) sur un modèle hydroélastique des vibrations de la membrane basilaire de la cochlée peuvent éventuellement remettre en cause les interprétations discutées ici. Il ne nous a pas été possible, dans le cadre de ce travail, de procéder à des simulations précises. Nous nous contenterons donc de présenter une interrogation. HOLMES rappelle que la vibration de la membrane basilaire résulte théoriquement de la superposition de deux ondes progressives de directions opposées. L'onde de retour provient d'une réflexion de l'onde initiale à l'hélicotrème de la cochlée. En règle générale, l'affaiblissement de l'onde initiale au delà de la position du maximum de l'enveloppe des vibrations est si fort que la contribution

de l'onde réfléchiée est nulle. Mais ce n'est plus vrai en très basse fréquence, lorsque l'amplitude maximale des vibrations se produit à l'extrémité distale de la cochlée. Pour l'oreille humaine, la contribution de l'onde réfléchiée est sensible jusqu'à 350 Hz, et elle est d'autant plus forte que la fréquence d'excitation est plus basse (figure 2.35). Elle provoque l'apparition d'un deuxième maximum de l'amplitude des vibrations à une fréquence plus élevée, vers 500 Hz pour une excitation à 100 Hz. Une voyelle contient toujours au moins une composante intense entre 100 et 350 Hz : la fréquence fondamentale. On peut s'interroger sur le rôle de ce deuxième maximum. Ne provoque-t-il pas une modification légère de l'amplitude d'harmoniques plus élevées, et un changement de timbre ? Peut-être est-il masqué par la contribution des autres composantes spectrales. Il nous paraît qu'il y a là matière à une intéressante simulation, sans parler d'un travail d'analyse des résultats psychophysiques portant sur les sons de combinaison et la perception de la sonie des sons simples.

4 - CONCLUSIONS

A la fin de notre présentation des travaux de nos prédécesseurs (§ 2.18), nous avons énoncé des conclusions dont les plus importantes sont rappelées, discutées et enrichies ci-dessous à la lumière de l'analyse de nos résultats et de ceux de CARRE (1988). Nous terminerons par plusieurs conclusions nouvelles. Nous n'avons pas rassemblé toutes les remarques faites, mais seulement les plus importantes. Hors de leur contexte, certaines seraient apparues trop disparates, problématiques, ou même contradictoires, incitant trop à une répétition de la discussion complète.

- 1 - La variation de F1 nécessaire pour conserver un timbre constant lorsque Fo est modifié ne peut pas être décrite par une simple translation de F1 et/ou F2 égale à celle de Fo, sur une échelle ERB ou Bark.

Pour la première fois, cette hypothèse a pu être testée sur la position du formant auditif, et elle doit être rejetée. La variation du formant auditif est plus faible que la variation fréquentielle de la fréquence fondamentale, quelle que soit l'échelle retenue. On ne peut songer à utiliser les fréquences physiques, même dans le cas où un sujet se fonde sur l'analyse de l'attaque car, dans ce cas, la frontière se placerait à une fréquence presque fixe.

- 2 - La position du formant auditif est déterminée en première approximation par une somme pondérée de la sonie des deux harmoniques les plus intenses lorsque leur différence est inférieure à 10 - 15 dB. Outre un léger effet de lissage spectral, qui ne remet pas en cause la présence d'harmoniques dans le spectre auditif, un

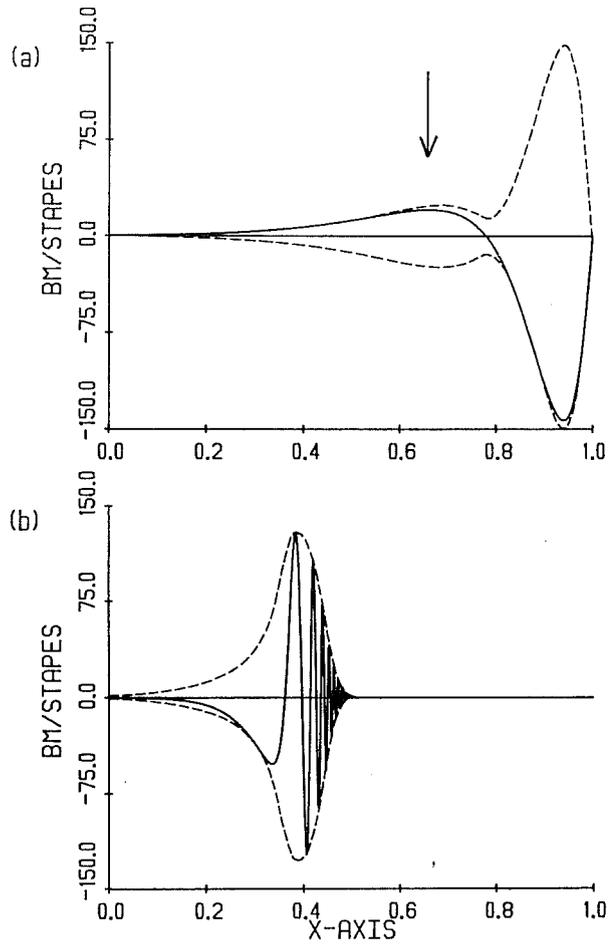


FIG. 2. Wave on the basilar membrane of the human ear as determined by (4), but normalized by the amplitude of the stapes, at one instant of time for a driving frequency of (a) 100 Hz and (b) 2000 Hz. Also shown, by the dashed curves, is the corresponding envelope.

Figure 2.35 - Enveloppe des vibrations de la membrane basilaire pour une fréquence d'excitation de 100 et de 2000 Hz. Notez le second maximum (flèche) pour l'excitation de basse fréquence. (in HOLMES 1987).

spectre de sonie présente une préemphasis de + 6 dB / octave jusqu'à 600 Hz par rapport au spectre physique.

Notre travail offre une bonne confirmation de la conclusion de CARLSON & al. (1975), en l'étendant à des fréquences fondamentales beaucoup plus élevées. Un seuil de - 10 dB permet de déduire de la position du formant auditif une bonne prédiction qualitative de l'évolution du timbre, si la frontière phonétique est adéquatement choisie. Cette stratégie semble utilisée lorsqu'une attaque vocalique douce ne permet pas de recouvrer la fréquence physique de F1. L'hypothèse de l'utilisation de la seule harmonique la plus proéminente ne rend pas compte de l'évolution du timbre pour les faibles variations de la fréquence formantique.

- 3 - Certains sujets ne détectent pas de changement de timbre lorsque le front de montée des stimuli est très raide.

Les résultats de CARRE (1988) pour l'un des sujets confirment les résultats de SLAWSON (1968) et mettent clairement en évidence qu'une attaque brutale permet vraisemblablement à certains sujets de placer le formant auditif très près de la fréquence physique du formant, quelle que soit la fréquence fondamentale. La position du formant auditif à F1 constant reste donc stable, quelle que soit la fréquence fondamentale. Il est probable qu'un auditeur peut utiliser ce deuxième mécanisme de positionnement du formant auditif parallèlement au premier.

- 4 - Lorsque un pic spectral est physiquement composé de deux harmoniques d'égale intensité, le formant auditif pourrait se trouver plus près de l'harmonique élevée, même si l'on tient compte de la préemphasis due à l'oreille externe et moyenne. Des harmoniques intenses sous les deux plus proéminentes auraient le même effet. La suppression du pic de haute fréquence sur le pic de basse fréquence est une explication potentielle.

Aucune de ces conclusions n'a été évoquée longuement dans notre discussion, sans doute parce que la forme spectrale des voyelles ne s'y prêtait pas. Il n'est pas impossible en revanche que, lorsque les deux harmoniques de part et d'autre de la plus proéminente sont intenses, celle de haute fréquence soit perceptivement plus importante.

- 5 - Une modification de la forme d'onde temporelle, notamment la disparition d'un passage par zéro positif, pourrait également jouer un rôle dans la modification du

timbre.

L'étude de la forme de l'onde temporelle faite dans l'analyse de la série de 270 Hz indique que, dans ce cas, la forme de l'onde n'est pas responsable du décalage de la frontière. Le pouvoir explicatif de cet indice est donc limité, mais il demeure une explication potentielle.

- 6 - Une manipulation de la phase des harmoniques de part et d'autre du formant conduit dans tous les cas à une légère élévation du pic auditif, dont on trouve trace dans les résultats neurophysiologiques. D'autre part, le formant auditif pourrait être à une fréquence plus basse lorsque l'harmonique la plus proéminente est à une fréquence légèrement inférieure à celle du formant physique.

Chaque fois que nous avons cherché à expliquer une variation de timbre qui ne paraissait pas pouvoir être expliquée par la sonie des deux harmoniques les plus proéminentes, le changement de phase de 180 ° que subissent les harmoniques de part et d'autre d'un pic formantique ne pouvait être une explication plausible.

Les conclusions suivantes sont la contribution spécifique qu'apporte l'interprétation de notre expérience.

- 7 - La position de la frontière phonétique mesurée à partir de la position présumée du formant auditif semble se décaler légèrement vers les hautes fréquences lorsque la fréquence fondamentale augmente.

Si ce phénomène paraît bien caractérisé, car on le retrouve chez CARRE & QUACH-TUAN, sa magnitude, certainement inférieure à la variation de la fréquence fondamentale, et surtout son origine ne sont pas établies. Nous avons pu éliminer les causes suivantes : un effet de suppression sur une harmonique de la zone de F1, car le décalage se produit même lorsque le formant auditif n'est formé que d'une harmonique très proéminente; un biais de réponse lié au choix de l'étendue des timbres; un effet de la qualité moindre du timbre aux fréquences fondamentales élevées privilégiant les timbres moins extrêmes. Nous avons envisagé avec prudence une explication auditive, qui, par analogie avec la baisse de la fréquence caractéristique des fibres auditives, prévoit une baisse du formant auditif lorsque l'amplitude d'une composante est élevée. L'autre explication possible est de nature différente. Elle procède d'un apprentissage de la différence de fréquences formantiques entre les réalisations masculines, féminines et enfantines. Ce décalage de la frontière phonétique aurait pour effet d'empêcher les confusions entre les réalisations de gestes articulatoires homologues. L'ampleur limitée du décalage de la frontière n'est pas un argument à charge. Nous avons montré un accord satisfaisant de cette interprétation avec les mesures physiques de réalisations naturelles, lorsque l'on tient

compte de la fréquence du deuxième formant.

- 8 - La position du formant auditif n'est pas celle que prévoit une représentation fondée sur un filtrage du spectre auditif par un second filtre large-bande.

Cette conclusion vaut pour le mécanisme de positionnement du formant auditif qui prévaut pour nos sujets. L'amplitude des harmoniques adjacentes de l'harmonique la plus proéminente joue un rôle, alors qu'un filtrage large-bande noie leur contribution dans les flancs d'un large maximum. La position du formant auditif ne résulte donc pas automatiquement d'un processus physiologique de bas niveau.

Nous écrivions dans la présentation de ce travail que nous ne serions pas en mesure de décrire un modèle complet de la représentation auditive du premier formant. On voit maintenant la portée exacte de cette conclusion. S'il paraît possible de déterminer la position du formant auditif, du moins lorsque les auditeurs utilisent l'un des deux mécanismes de mesure spectrale suggérés, la simple connaissance de la position est insuffisante pour prévoir le timbre car la frontière phonétique semble évoluer dans le même sens, mais proportionnellement plus lentement, que la fréquence fondamentale. Si, au sens strict, on entrevoit une réponse au problème de représentation auditive du premier pic spectral, le problème de l'assignation du timbre phonétique reste posé.

D'autres études sont nécessaires pour tester expérimentalement les causes potentielles de la variation de la frontière phonétique en fonction de la fréquence fondamentale. Mais il ne faut pas perdre de vue que, même si le succès couronne ces travaux, il restera à comprendre comment notre oreille et notre cerveau procèdent pour sélectionner un timbre aux fréquences fondamentales très élevées, au delà de 350 Hz, quand le spectre physique ne montre que quelques harmoniques dans la bande de fréquence du premier et du deuxième formant.

ANNEXE I : DISCUSSION DE SCHWARTZ & ESCUDIER " A direct demonstration of a large-scale integrated spectral representation in vowel perception" (soumis à 'Speech Communication') dans SCHWARTZ (1987).

Dans sa thèse SCHWARTZ présente ce travail comme "la première preuve directe de l'existence d'un (ou de plusieurs ?) mécanisme(s) d'intégration spectrale à large bande" (p.279). Nous voudrions discuter cette affirmation.

Si on se place dans l'hypothèse où l'une des représentations spectrales disponibles est fondée sur le "calcul" de F'2, pic unique de haute fréquence résultant d'une intégration large bande, il ne coûte pas beaucoup plus cher de faire l'hypothèse que cette représentation est, dans certaines conditions, utilisée dans les tâches de discrimination. Mais si les prédictions qui en découlent sont infirmées, on pourra seulement conclure que les conditions requises pour l'utilisation de F'2 n'étaient pas, en fin de compte, remplies. Etant donné que, d'après SCHWARTZ, il n'existe pas d'autre preuve de l'existence de F'2, ce mécanisme demeurera comme une explication possible, parallèle à d'autres, de certains résultats d'identification et de discrimination. Pour parler franc, nous ne croyons pas à l'existence psychologique de F'2. Il nous paraît que dans les expériences de mise en correspondance auditive ("matching") de spectres à 2 et 4 formants, la tâche est aisée lorsque l'un des formants supérieurs (souvent F2, quelquefois le bloc F2-F3 comme pour [y] ou [ø], ou F3 - F4 pour [i]) domine le spectre. Dans ce cas la fréquence de F'2 est très proche de celle de ce formant dominant. Mais alors un simple calcul à la KLATT (1982), ou bien l'utilisation d'une distance de Minkowski, sur les seules fréquences formantiques, suffit. Notons qu'un modèle de spectre auditif à la BLADON & LINDBLOM (1981) échoue car il est trop sensible aux différences d'amplitude des parties spectrales éloignées des formants. Dans le cas, rare, où aucun formant ne domine le spectre (spectres intermédiaires entre [i] et [y], ou [e] et [ø]), la variabilité des estimations de F'2, et l'impression des sujets (cf. LONCHAMP (1978)) indiquent qu'une voyelle à 2 formants n'est pas comparable en timbre à une voyelle à 4 formants. Je souligne que cette analyse implique que les pics spectraux proches sont représentés par un pic unique dans la représentation auditive, ce qui nécessite un léger lissage spectral. La fréquence du pic auditif unique correspondant à deux formants pourrait être une somme pondérée des fréquences de chacun d'eux.

SCHWARTZ suppose logiquement que l'usage éventuel de F'2 apparaîtra lorsque la mémoire auditive du spectre "brut" sera dégradée par le passage du temps, mais avant que la

comparaison ne se fasse à l'aide des seuls symboles phonétiques. Il obtient d'abord les scores d'identification pour des stimuli à F1 (450 Hz), F2 (2000 Hz) et F4 (3350 Hz) fixes et F3 variable (2300 - 3000 Hz), dont le timbre est très majoritairement [e]. La valeur de F'2, proche des valeurs de F3 aux extrémités, varie plus rapidement que F3 pour les stimuli centraux. L'auteur prévoit donc une meilleure discrimination des stimuli centraux que ne laisse prévoir le spectre acoustique. Les prédictions liées à la mémorisation feront l'objet d'un court commentaire plus loin. Il compare trois tests de discrimination qui, pour éviter un effet plafond ou plancher, portent sur des stimuli d'écart fréquentiel croissant, les couples étant choisis aux deux extrémités et au centre de la gamme. L'intervalle entre les stimuli à discriminer est bref (200 ms) pour le premier test, long (3 s) pour le deuxième, long (6 s) et rempli de 6 courtes bouffées tonales à 1 kHz pour le troisième. C'est le test 2 qui doit révéler l'effet d'accroissement de la discriminabilité des stimuli centraux dû à l'utilisation de F'2. Les résultats (table III : p. 297) sont les suivants. Au test 2, la paire centrale est discriminée dans 82 et 75% des cas selon l'ordre, la paire de fréquence F3 élevée dans 100 et 82% et la paire 1 (F3 bas) dans 57 et 71% des cas. La prise en compte du taux de fausse alarme dans le calcul des d' ne change pas les conclusions. La paire de haute fréquence est plus discriminable. L'examen du spectre des stimuli (figure 2.36) suggère une explication simple. Beaucoup de résultats électrophysiologiques et psychophysiques suggèrent qu'un pic intense de fréquence inférieure à un pic plus petit a tendance à exercer sur ce dernier un effet de suppression. On trouvera une illustration dans les données de TYLER & LINDBLOM (1982) : le deuxième formant de leur voyelle [y] supprime le troisième formant proche et vraisemblablement moins intense acoustiquement ; pour [i] et [e], en revanche, F3 plus lointain et plus intense est bien visible (cf. aussi l'annexe 2 ci-dessous). Faisons de plus l'hypothèse que deux pics proches (200 - 400 Hz) n'en forment qu'un dans la représentation auditive, à une fréquence égale à la somme des deux composantes pondérée par l'amplitude. Dans ce cas, les stimuli proches de 7 (cf. figure 2.36) ont un pic auditif dominé par F2 fixe, F3 variable étant réduit en amplitude par F2 fixe, et contribuant seulement modestement à la variation de la fréquence du formant auditif. Pour les stimuli proches de 14, au contraire, le pic auditif, composé de F3 variable et intense et de F4 fixe mais réduit en amplitude, possède une fréquence qui varie largement en fonction de F3. Une situation intermédiaire prévaut pour F3 équidistant de F2 et F4. Une même variation de F3 conduit donc à une faible variation du pic auditif pour F3 bas, et à une variation plus importante pour F3 élevé, d'où la différence de discriminabilité. Cette explication est d'ailleurs au mieux incomplète. Les données montrent clairement, et SCHWARTZ le souligne (pour rejeter ensuite cette interprétation (p. 307) par souci de cohérence avec l'hypothèse d'une action de F'2 !), que la nature de la réponse phonétique, liée à l'emplacement des frontières phonétiques, modifie les scores de discrimination. Osons une nouvelle interprétation d'un détail. Les paires de basse fréquence sont plus

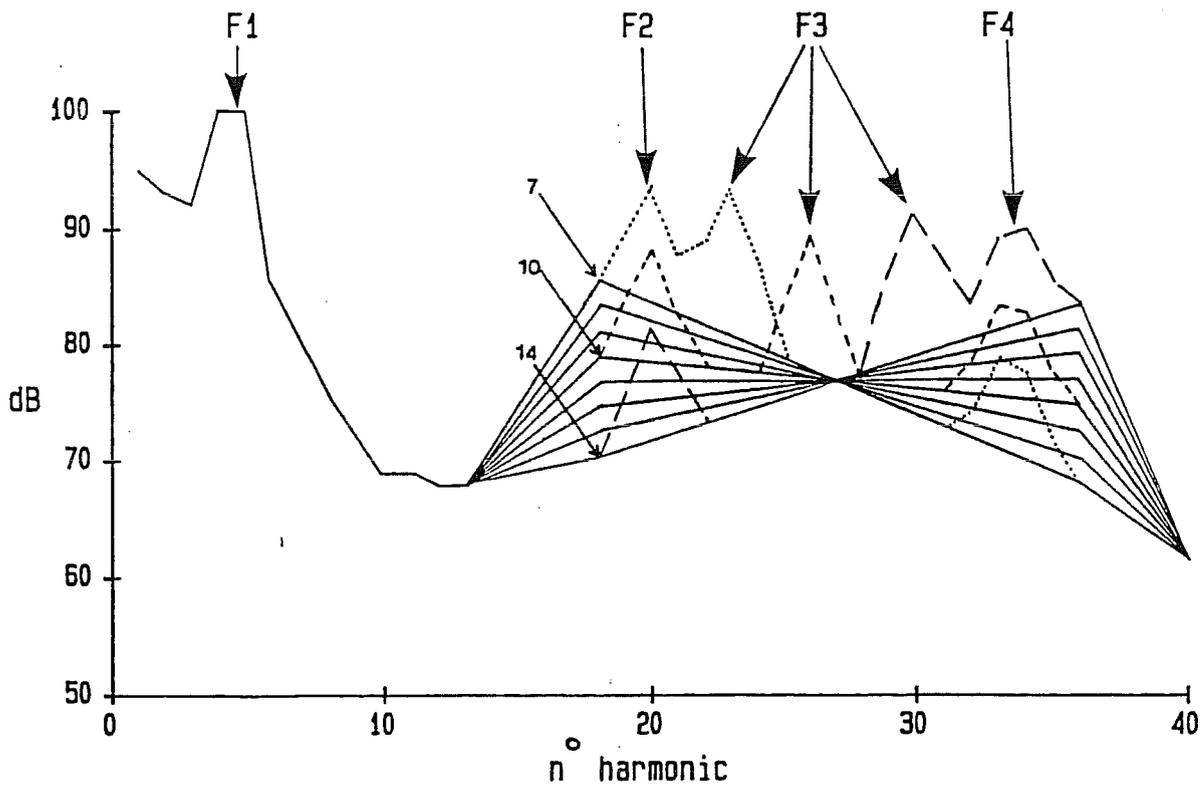


Fig.5

Description of stimuli 7 to 14 of corpus D1 (see text). For better understanding, we have drawn stimuli 7 ($F_3=2300$ Hz, points), 10 ($F_3=2600$ Hz, small dashes) and 14 ($F_3=3000$ Hz, large dashes).

Figure 2.36 - Formes spectrales des voyelles du corpus D1 de SCHWARTZ (1987).

discriminables quand le stimulus le plus près de la frontière [ø] est en premier. Il pourrait s'agir d'un simple effet de contraste. Quand un stimulus de bonne qualité phonétique précède un stimulus de moindre qualité, il exerce sur ce dernier un effet de contraste qui le fait paraître plus différent de timbre. Cette effet est moindre, ou n'existe pas, dans l'ordre de présentation inverse. Cette interprétation fait implicitement appel à un concept de double codage (phonétique et auditif) à la FUJISAKI & KAWASHIMA (1970 par ex.) pour expliquer les scores de discrimination. Une dernière remarque concerne une éventuelle difficulté à mémoriser les stimuli pour lesquels F3 est équidistant de F2 et F4. Le fait à expliquer est le plus grand taux de fausse alarme sur ces stimuli. Ces stimuli sont plus complexes dans la mesure où ils possèdent selon nous trois, et non deux, pics auditifs dans la zone F2 - F4. De ce fait, ils sont plus sensibles à des variations aléatoires dans l'estimation des maxima auditifs par les sujets. Un écart fortuit provoquant une fausse alarme est plus probable.

Comme souvent (toujours ?) dans les expériences de perception, on peut concevoir des explications concurrentes rendant compte des mêmes données. Ici, la prédiction d'une plus grande discriminabilité dans la zone où F'2 varierait le plus vite n'est pas vérifiée. Mais comme nous l'avons fait remarquer au début, cette absence de confirmation ne signifie pas que F'2 n'existe pas. Il reste un candidat possible, mais il n'est pas indispensable à la compréhension des expériences de discrimination.

ANNEXE 2 : SPECTRES AUDITIFS DE TROIS VOYELLES NATURELLES OBTENUS PAR LA METHODE DU SEUIL DE PULSATION.

Les figures 2.37 à 2.39 présentent les spectres auditifs de trois voyelles naturelles, obtenus par la méthode du seuil de pulsation, ainsi que les spectres F.F.T. correspondants. Cette expérience préliminaire, dont l'auteur était le sujet, a été réalisée en 1981 au Laboratoire de la Communication Parlée de l'E.N.S.E.R.G., et nous voulons remercier son directeur, R. CARRE, ainsi que M. BOULOGNE, J.M. DOLMAZON et P. ESCUDIER de l'autorisation d'utiliser le matériel, et de l'aide qu'ils nous ont apportée, sans parler du temps qu'ils nous ont consacré sans compter ni de l'accueil très chaleureux qu'ils nous ont réservé.

Les figures présentent 5 estimations du spectre, ainsi que la médiane des valeurs. Le seuil de pulsation a été exploré à un pas fréquentiel de 50 Hz. La présentation des voyelles s'est faite à 60 dB SPL. L'audiogramme de l'oreille gauche du sujet, qui a été utilisée, est acceptable en basse fréquence (+10 dB à 500 Hz par rapport aux normes ISO-1964 et ANSI-1969), excellent jusqu'à 2000 Hz (0 et -5 dB à 1000 et 2000 Hz), acceptable (+10 dB) à 3000 Hz. Au delà, bien que cette zone ne nous concerne pas, l'audiogramme montre des pertes de 45 dB, 40 dB et 35 dB à 4000, 6000 et 8000 Hz, liées à un tinnitus (acouphèmes) consécutif à une exposition à des bruits d'armes à feu pendant son service militaire. Elles sont similaires mais de moindre amplitude (-5 à -15 dB) pour l'oreille droite, sauf à 500 Hz, où l'audition est excellente.

Pour la voyelle [y], dont la fréquence fondamentale est de 135 Hz, le spectre auditif moyen de la région de F1 semble montrer, malgré la variabilité des résultats, la contribution auditive de la fréquence fondamentale et de l'harmonique suivante. Le deuxième pic est à 300 Hz, proche de la fréquence de l'harmonique de 270 Hz. Si l'on tient compte de la préemphasis spectrale de l'oreille externe et moyenne, qui abaissera le premier de 6 dB, le maximum auditif sera vers 300 Hz, très proche de la fréquence de la seconde harmonique. La troisième composante spectrale est juste discernable à 500 Hz. Le spectre auditif est de pente globale différente du spectre physique. La fonction de transfert de l'oreille externe et moyenne n'est pas en cause, car elle n'intervient évidemment pas dans le spectre physique, ni dans les mesures brutes du spectre auditif présentées ici qui ne dépendent que de l'égalité du son-test et de la voyelle-masque. La préemphasis s'appliquant au son-test comme au masque vocalique, elle ne change pas le niveau mesuré. Les formants supérieurs sont plus intenses que dans le spectre physique. Dans le spectre auditif réel, après prise en compte de la préemphasis de +6 dB / octave jusqu'à 600 Hz, la dominance des formants supérieurs serait encore plus marquée. La réduction notable du 3ème formant pourrait résulter d'un

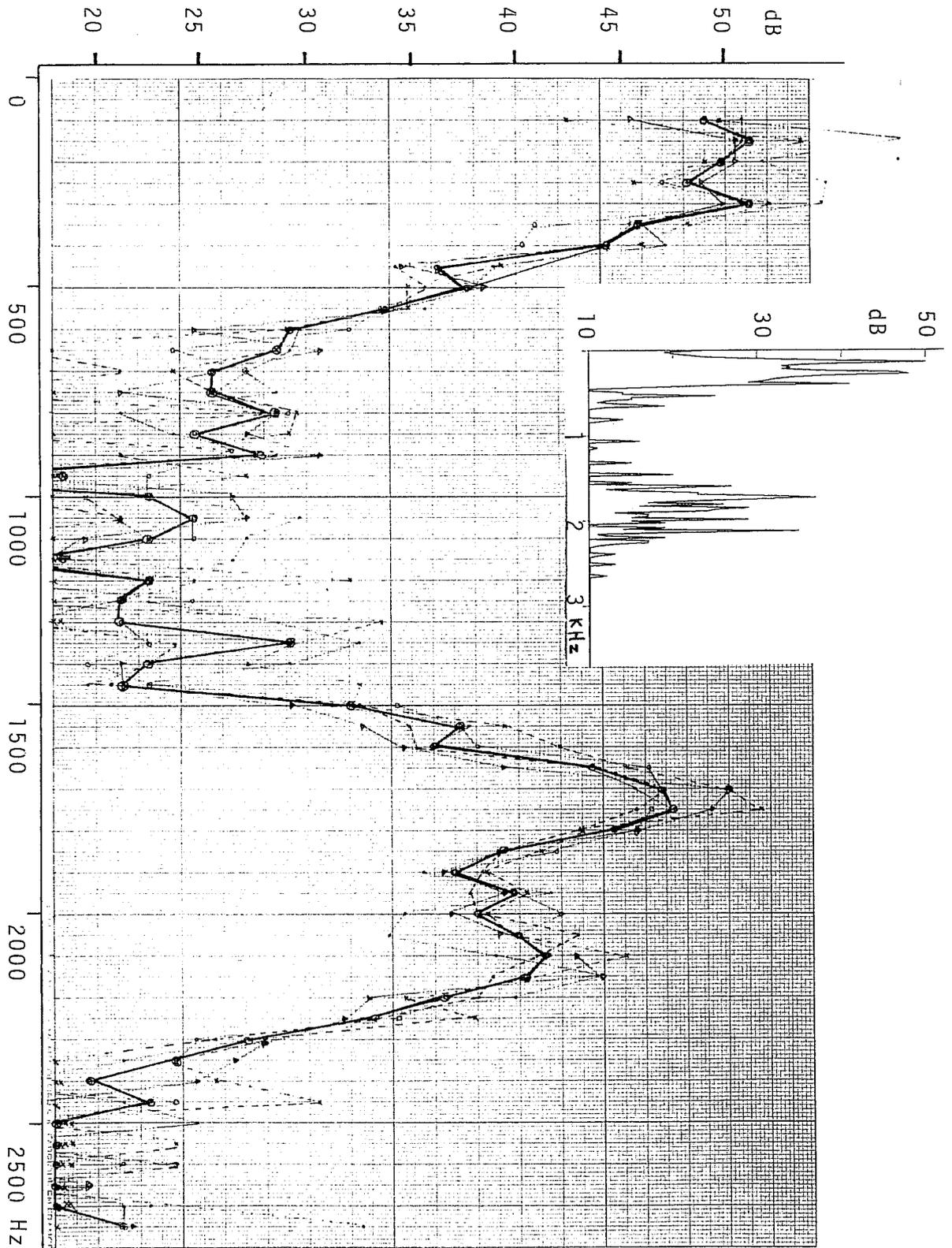


Figure 2.37 - Spectre auditif d'une voyelle [y] : méthode du seuil de pulsation.
Le cartouche montre le spectre physique

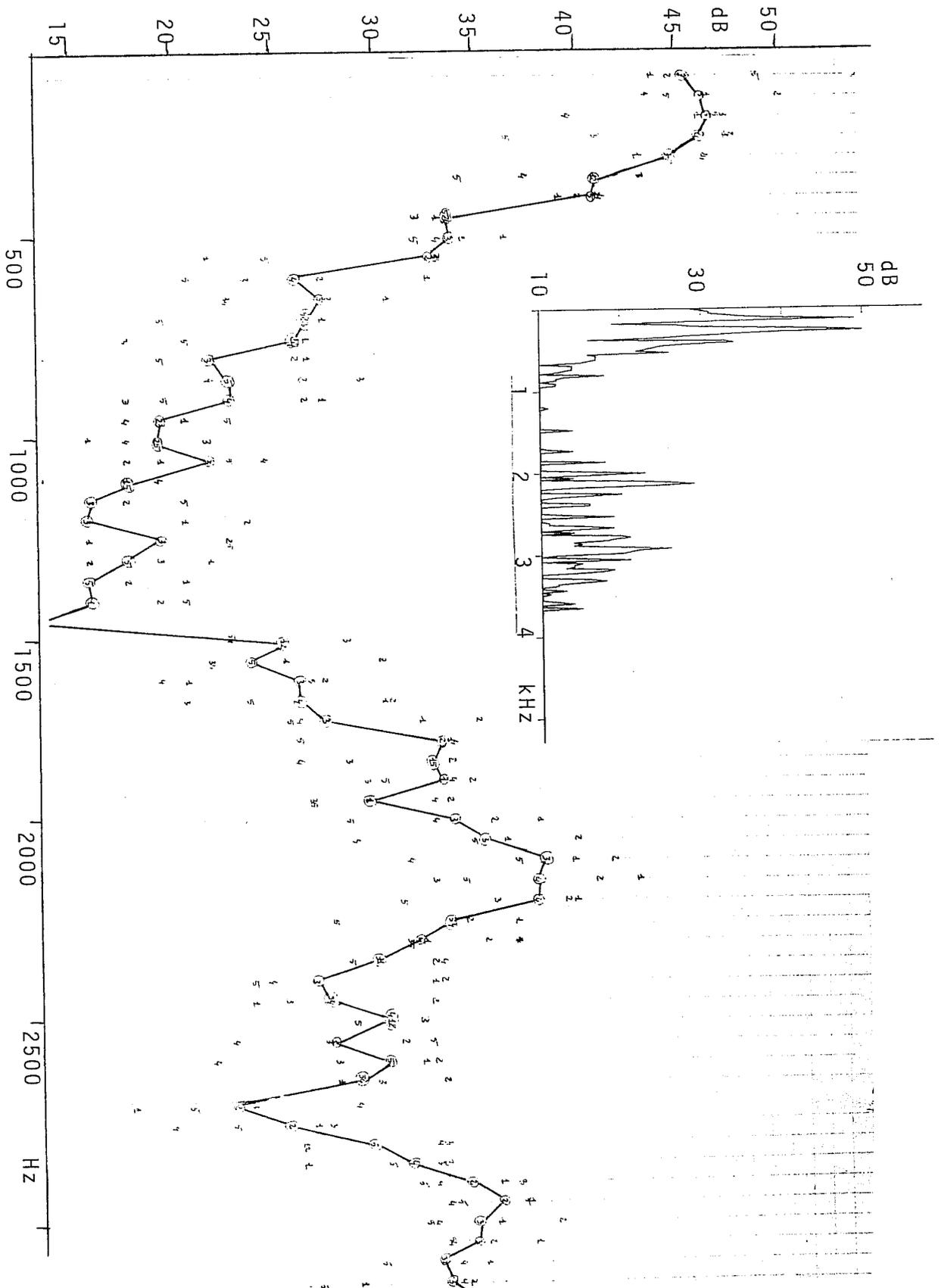


Figure 2.38 - Spectre auditif d'une voyelle [i] : le cartouche montre le spectre physique.

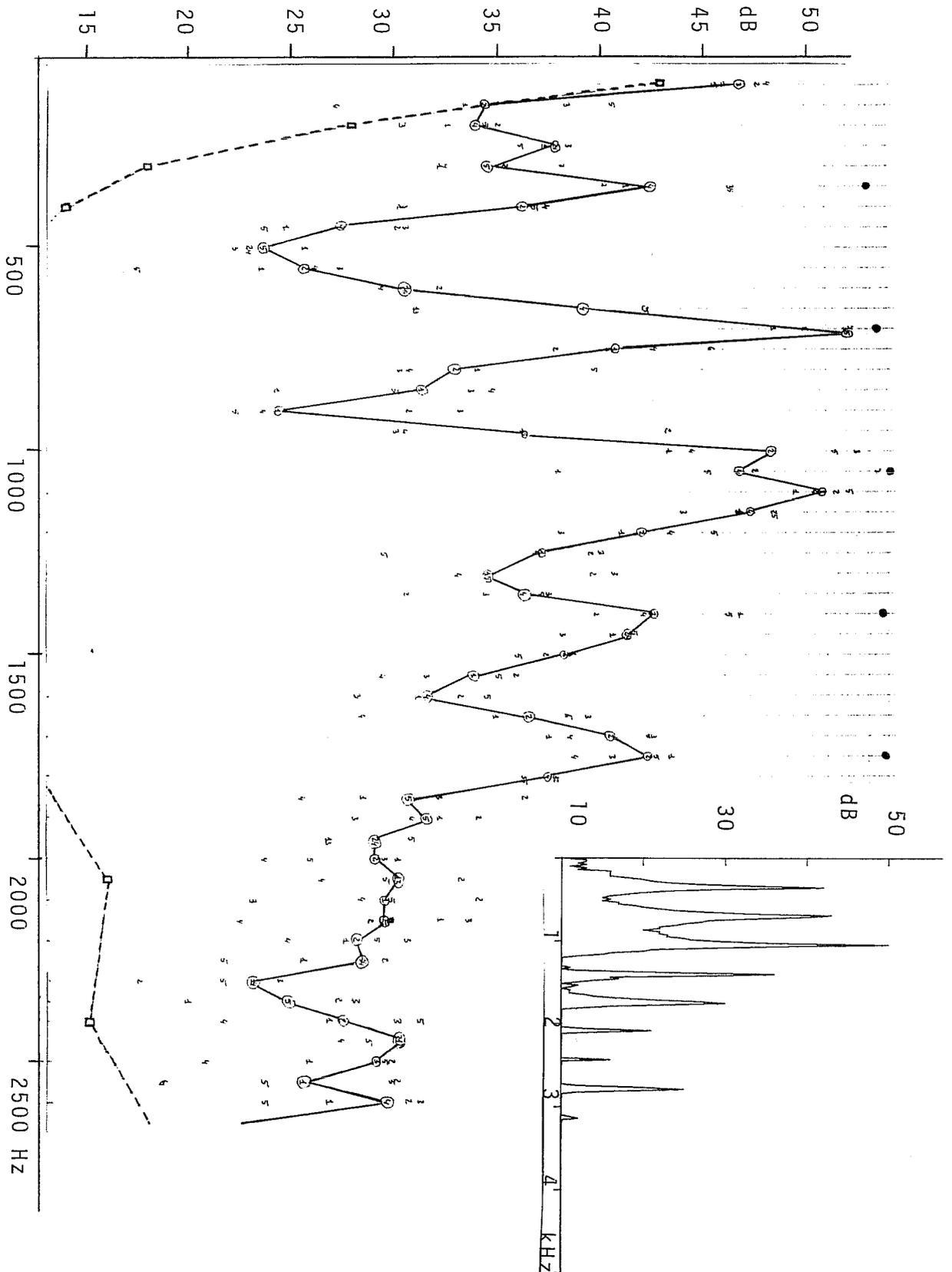


Figure 2.39 - Spectre auditif d'une voyelle [a] : Fréquence fondamentale de 350 Hz; le cartouche montre le spectre physique; la ligne pointillée indique le seuil d'audibilité absolu; La fréquence des harmoniques est marquée d'un point noir.

effet de suppression provenant du 2ème formant qui le domine.

La voyelle [i] montre une bosse centrée sur 200 Hz comme représentation du premier formant. Les deux premières harmoniques d'égale intensité du spectre physique ne sont donc pas séparées comme dans la voyelle précédente. Si l'on tient compte du fait que, dans la réalité auditive, la première harmonique (F_0) sera 6 dB plus basse que la seconde, il est probable que le formant auditif se placera très près de la fréquence de la 2ème harmonique, vers 250 Hz. De petits paliers jusqu'à 800 Hz signalent parfaitement les harmoniques suivantes sur le spectre auditif, malgré la variabilité des estimations. La représentation des formants 2 et 3 est nette, et l'effet de suppression de F_2 sur F_3 paraît négligeable. L'intensité des formants supérieurs est ici aussi plus forte que sur le spectre physique.

La troisième voyelle est un [a] féminin émis à une fréquence fondamentale aiguë de 350 Hz. C'est à notre connaissance la première mesure d'une voyelle à une fréquence élevée. Il est impossible de déterminer les deux premières fréquences formantiques sur le spectre physique. Le spectre auditif met clairement en évidence les 5 premières harmoniques, jusqu'à 1750 Hz (cercles noirs sur le spectre auditif). En très basse fréquence, la ligne pointillée indique le seuil absolu d'audition du sujet. L'amplitude de la fréquence fondamentale semble réduite par un effet de suppression qui pourrait provenir de la seconde harmonique. L'amplitude de la 3ème harmonique paraît très faible en comparaison de sa valeur physique, et la variabilité des estimations très importante. Mais il faut remarquer que la valeur du seuil est déjà supérieure à toutes les valeurs mesurées pour les spectres précédents. Nous soupçonnons que cet effet résulte d'une présentation trop intense, eu égard à la prééminence de la 3ème harmonique, causant une saturation du seuil de pulsation. Il est également possible que la gamme d'intensité du son-test offerte par le dispositif expérimental ait été insuffisante dans ce cas. Une chose est claire cependant. Ce n'est pas le niveau où est élaboré la représentation du spectre auditif qui fournit au mécanisme d'identification phonétique une estimation précise de la fréquence des deux premiers formants !

En conclusion, les deux spectres de voyelles à fréquence fondamentale basse montrent une position du premier formant auditif conforme à ce que les spectres F.F.T. permettent de prévoir. Pour ces voyelles fermées, les formants supérieurs sont auditivement très proéminents, et F_3 est d'amplitude faible lorsqu'il est proche de F_2 , vraisemblablement en raison d'un effet de suppression. Pour les fréquences formantiques élevées, le spectre auditif ne montre pas plus de formants que le spectre physique.

ANNEXE 3 : LE DOUBLE CODAGE DE L'INFORMATION SPECTRALE - TEMOIGNAGES DU DOMAINE DE LA PERCEPTION DE LA PAROLE.

Cette annexe rassemble quelques résultats d'expériences de perception de la parole qui s'interprètent de manière satisfaisante lorsque l'hypothèse d'un double codage des informations spectrales est envisagée. Elle développe les quelques remarques faites à ce sujet dans un article de synthèse (LONCHAMP (1984)).

1 - BASES NEUROPHYSIOLOGIQUE DU CODAGE SPECTRAL PAR L'OREILLE.

Les lignes suivantes sont un rappel très succinct des connaissances actuelles sur la neurophysiologie du nerf auditif. On sait que pour les fréquences inférieures à 4 - 5 kHz une fibre nerveuse auditive répond à une stimulation sonore en manifestant deux caractéristiques. On observe d'une part une augmentation de son taux de décharge moyen d'impulsions nerveuses au delà du taux spontané. Cette augmentation se produit pour des sons d'autant plus faibles qu'ils sont proches de la fréquence de plus grande sensibilité de la fibre (CF : fréquence caractéristique ou centrale). D'autre part, ces impulsions nerveuses sont produites à des instants particuliers, correspondant à une configuration déterminée de la forme de la membrane basilaire en vibration. On parle alors de verrouillage de phase ("phase locking") pour cette analyse du taux de décharge dit instantané, c'est-à-dire à l'échelle d'une période. (cf. par ex. DELGUTTE & KIANG 1984 a).

Un premier aspect capital du taux de décharge moyen est la présence d'une dynamique limitée. Pour chaque fibre, le taux moyen ne varie que pour une gamme limitée d'intensité, de l'ordre de 20 à 30 dB. Au dessous et dessus de deux seuils variables pour chaque fibre, le taux reste constant, présentant donc une saturation non linéaire. DELGUTTE (1987) a néanmoins montré que si l'on tient compte de trois populations de fibres distinguées par leur taux de décharge spontané, qui corrèle avec le seuil de déclenchement et de saturation, on pouvait, sous certaines hypothèses, prévoir les caractéristiques auditives de la discrimination d'intensité. Cette démonstration rend moins probante l'interprétation de SACHS & YOUNG (1979) et YOUNG & SACHS (1979) : seule l'utilisation de la réponse verrouillée en phase, ou taux de décharge synchrone, permet au système auditif de disposer d'un spectre interne possédant des maxima spectraux aux fréquences formantiques quelle que soit l'intensité des stimuli sonores. En effet, à l'échelle d'une période, les fibres nerveuses répondent vigoureusement dans une large plage d'intensité à une composante intense de fréquence proche de leur fréquence caractéristique. Il faut néanmoins noter une détérioration pour les intensités les plus fortes. Cette réponse se produit à un moment déterminé du battement de la membrane basilaire. On constate expérimentalement que plus

l'intensité d'une composante de fréquence proche de la fréquence caractéristique d'une fibre est grande, plus l'intervalle entre les impulsions est proche d'un multiple entier de la période de la composante. Le degré de synchronisation est donc une mesure indirecte de l'amplitude d'une composante sonore.

Mais le taux de décharge moyen possède une autre caractéristique essentielle. La plage d'intensité capable de moduler ce taux est bien plus grande dans les 10 - 30 ms qui suivent l'apparition d'une composante spectrale (SMITH 1979, SMITH & BRACHMAN 1980, DELGUTTE 1980, par ex.). Le taux présente ensuite une adaptation, qui se traduit par un écrasement des taux de décharge en fonction des différences d'amplitude, et par une saturation réduisant la gamme utile d'analyse de l'intensité. DELGUTTE (1980) montre en particulier que la présence du formant de basse fréquence d'une consonne nasale adapte fortement la réponse des fibres au spectre de la voyelle suivante. Cette adaptation est plus forte pour la zone de fréquence qui a été stimulée par le formant nasal.

2 - CODAGE D'INTENSITE ET CODAGE TEMPOREL.

Intuitivement, les deux mécanismes de codage spectral possèdent des caractéristiques qui les rendent efficaces dans des situations différentes. Le codage par le taux de décharge moyen, que nous appellerons codage d'intensité, semble bien adapté au codage d'informations brèves, non périodiques et qui ne nécessitent pas une grande précision d'analyse. Ce pourrait être le cas des barres d'explosion des occlusives et des bruits fricatifs. Son efficacité est maximale lorsque le stimulus sonore présente une forte discontinuité d'amplitude, bien qu'il le codage d'intensité soit sensible à la saturation si l'intensité atteinte est trop forte. Au contraire, le codage que nous appelons temporel, fondé sur la capacité de la fibre à répondre en synchronie avec l'évolution temporelle de la composante, paraît plus efficace pour coder la fréquence exacte et l'amplitude des composantes spectrales périodiques, quelle que soit l'intensité globale. Ce codage sera d'autant plus efficace que la durée d'accumulation des informations sera plus longue. Ce mode de codage paraît indispensable pour atteindre la précision de mesure que révèlent les expériences de discrimination de la hauteur tonale et très utile pour le positionnement exact du premier formant auditif à partir du niveau précis des deux harmoniques proéminentes. Il n'est bien sûr pas possible de savoir si ces deux mécanismes peuvent être utilisés simultanément.

3 - POLS & SCHOUTEN (1978, 1981) : LE ROLE D'UN PREFIXE DE BRUIT DANS LA PERCEPTION DES OCCLUSIVES.

Le modèle de perception qui va être discuté a été conçu à l'origine pour rendre compte des

résultats de OHDE & SCHARF (1977, 1981) et POLS & SCHOUTEN (1978, 1981). Ces études, ainsi que plusieurs travaux antérieurs, montraient que l'identification des syllabes voyelle + occlusive était meilleure que l'identification des syllabes occlusive + voyelle. OHDE & SCHARF (1981) et POLS & SCHOUTEN (1981) présentent un résumé complet. Pour rendre les conditions comparables, la barre d'explosion et le bruit de friction étaient supprimés, et la supériorité des groupes VC ne pouvait être due qu'à un traitement plus efficace des transitions vocaliques. Or les transitions sont spectralement au moins aussi marquées dans les groupes CV que dans les groupes VC, ce qui rend difficile une interprétation de cette asymétrie.

Un autre enseignement de ces travaux, que nous ne développerons pas, est que le rôle perceptif de la barre d'explosion des consonnes sourdes est considérable (cf. aussi LARIVIERE & al. 1975). Ceci explique pourquoi la simple copie des transitions naturelles en synthèse ne permet pas la création de consonnes de bonne qualité, même pour les occlusives sonores (ROSSI & MARTIN-LAVAL 1985). Il nous semble que les transitions formantiques de pente exagérée nécessaires en synthèse en l'absence de barre d'explosion, ont pour but de créer un spectre voisin de celui qui existerait si la barre d'explosion était présente.

POLS & SCHOUTEN (1978), à partir de l'idée intuitive que la brusque transition d'amplitude des groupes CV peut produire une sorte de "clic" auditif, d'ailleurs inaudible objectivement, font précéder le stimulus d'un bruit relativement faible. Les scores d'identification augmentent alors fortement. OHDE & SCHARF (1981) obtiennent également une légère amélioration en utilisant une fenêtre initiale de 6ms. SUMMERFIELD & ASSMANN (1987) rapportent qu'ils ont également observé d'une manière informelle une augmentation de l'intelligibilité de groupes CV dans du bruit blanc lorsque celui-ci débute quelques centaines de ms avant la syllabe. Cette avantage n'existe pas lorsque le bruit commence en même temps que le début de la syllabe. SCHEFFERS (1983) a mis en évidence le même avantage pour des voyelles.

On peut tenter de rendre compte de ces résultats dans le cadre d'un modèle à double codage. Dans le cas d'une brusque discontinuité d'amplitude, le système auditif privilégie, peut-être de façon automatique, le codage d'intensité. Mais les résultats de cette analyse ne sont pas très utiles dans la mesure où le spectre ne possède plus la barre d'explosion caractéristique. Ce sont les transitions seules qui sont soumises à cette analyse dont les résultats peuvent de plus être trop grossiers pour assurer une bonne identification. Si l'on admet que les deux modes de codage ne peuvent être menés en parallèle, on peut faire l'hypothèse que le codage temporel plus précis s'exercera trop tard, sur une partie du stimulus où les déflexions fréquentielles des transitions sont déjà trop réduites. La supériorité des groupes VC s'explique par le fait que le suivi des transitions par le codage temporel peut se faire sans interruption jusqu'à son terme, où réside l'information la plus pertinente. Toute

l'information disponible est donc utilisée dans ce cas. La présence d'un bruit précédant la discontinuité d'amplitude ou d'une fenêtre évite peut-être que le codage d'intensité soit utilisé exclusivement, ou décale moins le point de départ de l'analyse temporelle. L'information portée par les transitions est mieux exploitée, et les scores d'identification sont meilleurs.

Nous voulons maintenant présenter quelques études qui confortent, ou précisent, certains aspects de ce modèle très fruste.

CHISTOVICH (1971) a décrit très brièvement une expérience qui s'interprète aisément si l'on fait appel à l'adaptation constatée pour le taux de décharge moyen. Elle concerne une voyelle à deux harmoniques proéminentes dans la zone de F1. Pour maintenir la voyelle sur une frontière phonétique lorsqu'on introduit l'une des harmoniques de F1 avec un retard croissant, il faut diminuer son amplitude. Tout se passe comme si une composante qui apparaît tardivement possédait une sonie supérieure, ou réciproquement, comme si la contribution d'une harmonique présente dès le début de la voyelle diminuait graduellement avec le temps. Comme il n'y a pas d'énergie à la fréquence de l'harmonique manquante, les fibres de fréquence voisine sont dans un état non adapté au moment où celle-ci est introduite, et elles répondent de façon vigoureuse. Les travaux de SUMMERFIELD & ASSMANN (1987) confirment que le système auditif privilégie les variations spectrales sur les zones stables.

SLAWSON (1968), comme il a été vu plus haut, démontre que certains auditeurs ne traitent pas de la même façon une voyelle synthétique selon que l'attaque est douce ou brutale. Comme le contrôle de l'amplitude était placé avant le calcul des résonances, il est probable que les stimuli n'étaient pas physiquement identiques (cf. CARRE & QUACH-TUAN (1986)), ce qui ne permet pas d'affirmer avec certitude qu'un auditeur analyse l'attaque avec un mécanisme différent de celui utilisé pour la partie stable. Il peut simplement être sensible aux différences spectrales entre les stimuli à attaque douce et brutale quel que soit le mécanisme mis en jeu. Mais CARRE (1988) montre que certains sujets répondent de manière différente à des stimuli physiquement identiques, ce qui suggère fortement la présence de deux mécanismes distincts. Cette tendance est plus accusée lorsque la voyelle est brève (50 ms), c'est-à-dire lorsque le codage temporel est moins efficace. Que le timbre des voyelles à attaque brutale soit moins sensible à une variation de la fréquence fondamentale concorde avec l'idée que le codage d'intensité fournit une représentation plus proche du spectre physique, comparable à celle fournie par un filtrage large-bande, ou une analyse LPC. Il est vraisemblablement moins précis dans le cas où un formant est physiquement composé de deux harmoniques d'égale intensité, et il n'est pas, ou guère, utilisable pour des variations spectrales sans discontinuité d'amplitude.

4 - UNE FORTE INTENSITE DETERIORE L'IDENTIFICATION DES SONS

De nombreuses études confirment que la précision de l'identification et de la discrimination des sons diminue lorsque le niveau sonore augmente. Mais il n'est pas facile de déduire des résultats lequel des mécanismes est responsable de la détérioration. SYRDAL-LASKY (1978) conclut qu'à 92 dB, l'identification et la discrimination du lieu d'articulation d'occlusives sourdes et sonores en synthèse à 2 formants est de moindre qualité. L'absence de F1 au début des occlusives sourdes ("F1 cutback") permet d'éliminer l'hypothèse d'une extension de masquage important de F1 vers les transitions de F2 ("upward spread of F1 masking"). Il est intéressant de constater que l'addition d'un bruit de fond de 60 dB améliore les performances. Ce résultat doit selon nous être rapproché de celui de POLS & SCHOUTEN. VAN TASSELL & CRUMP (1981) note une réduction générale de l'efficacité de la transition de F3 à séparer [da] de [ga] quand le niveau passe de 60 à 100 dB. La pente d'identification est plus faible et non monotone, et les stimuli extrêmes ne sont plus identifiés qu'à 80%. Les auteurs incriminent l'extension du masquage de F1 et F2 vers les fréquences supérieures, hypothèse rejetée par SYRDAL-LASKY pour ses résultats. DORMAN & DOUGHERTY (1981), pour des syllabes synthétiques à deux formants de forme occlusive sonore+voyelle, confirment que les fonctions d'identification sont très perturbées à 90 dB relativement à celles obtenues à 55 et 70 dB. Les détails des modifications ne correspondent pas à ceux des études précédentes, ce qui est pour nous le signe d'une simple détérioration de la qualité de l'information spectrale conduisant à des réponses aléatoires.

Les études de DORMAN & al. (1986, 1987) confirment que l'intelligibilité de voyelles brèves (50 ms) diminue à partir d'une intensité de 90 - 95 dB SPL. Pour des voyelles plus longue, le réflexe stapédien, d'une latence supérieure à 50 ms, intervient pour réduire au niveau de l'oreille moyenne l'intensité du son transmis, ce qui assure une bonne intelligibilité jusqu'à 110 dB SPL. Les auteurs rappellent que les sujets ayant subi une stapéctomie, c'est-à-dire une section du muscle stapédien qui joue un rôle essentiel dans le réflexe acoustique, montrent également une diminution de l'intelligibilité des voyelles vers 90 dB (BORG & ZAKRISSON 1973). Les voyelles ouvertes sont plus atteintes que les voyelles fermées (DORMAN & al. 1986). Le sens des confusions semble indiquer que F1 décroît et F2 croît avec l'augmentation de l'intensité. CARRE (1988) constate également aux niveaux élevés une distorsion de la courbe d'identification, qu'il attribue à un effet de saturation d'origine neurophysiologique.

Cette diminution de l'intelligibilité à des intensités élevées est compatible avec l'hypothèse que le codage des voyelles est fondé pour une part sur le taux de décharge moyen. Mais il est également probable que le verrouillage de phase devient moins précis aux intensités

élevées. S'il est permis de comparer des expériences différentes, la détérioration de l'identification des occlusives à 92 dB est déjà bien établie, alors que celle des voyelles ne fait que commencer, ce qui pourrait pointer vers une pondération différente des mécanismes de codage dans l'identification, ou la discrimination, des voyelles et des occlusives. Mais ces tâches exigent peut-être des degrés différents de précision pour les voyelles et les consonnes, ce qui rendrait la comparaison invalide.

5 - LE CODAGE DE L'ATTAQUE : LES EXPERIENCES DE LACERDA (1987)

Certaines expériences décrites par LACERDA (1987) sont d'un grand intérêt pour tester le modèle proposé. La première compare la discriminabilité d'une variation de F2 dans des stimuli brefs (50 ms) à attaque brutale ou graduelle. Les formants F1, F3 et F4 sont identiques, et F2 est à une fréquence différente, mais stable, pour chaque stimulus. L'hypothèse testée par LACERDA est que l'attaque graduelle adapte les réponses nerveuses et conduit à une représentation auditive moins différenciée. La discrimination est donc plus difficile. Les résultats confirment cette prédiction, mais il nous semble possible d'en donner une autre explication. Tout d'abord, on peut songer à plaider que dans le cas d'une attaque graduelle, un codage temporel intrinsèquement précis est utilisé, mais que la durée du stimulus est trop faible pour que les résultats auditifs soient fiables. La comparaison avec les résultats de FLANAGAN (1955) pour des stimuli de 750 ms, identiques à ceux obtenus par LACERDA pour les stimuli à attaque graduelle, permet de rejeter cette hypothèse. On peut en revanche considérer que le sujet est capable d'utiliser les deux sources d'informations dans les sons à attaque brutale car la structure formantique est stable. Ce surcroît d'informations assure une meilleure discriminabilité.

La deuxième expérience montre qu'une variation de F2 dans la syllabe [ad] est plus discriminable que dans [da]. Elle confirme les résultats présentés plus haut (POLLS & SCHOUTEN (1978, 1981)). Nous avons déjà présenté notre interprétation. L'auteur propose deux explications complémentaires et complexes : la première est fondée sur la présence d'une excitation résiduelle d'origine interne dans les syllabes VC. Cette hypothèse est testée en faisant suivre le stimulus d'un appendice formantique constant. La prédiction est que l'effet de cette adjonction doit être négligeable, car cette nouvelle information est déjà disponible ! Le résultat est conforme aux prédictions, et l'hypothèse est considérée comme validée. On peut démontrer de la même façon que la lune possède une forte personnalité. La prédiction est que la lune refusera de changer de direction, même quand l'expérimentateur se mettra en colère. L'expérience prouve que cette prédiction est vraie, et on "démontre" ainsi que la lune possède effectivement une volonté inflexible. Précisons que notre ironie ne vise ici que le test expérimental, et non l'hypothèse. Aucun modèle ne peut être mis en difficulté par la prédiction d'un effet nul, car il est toujours possible de plaider

qu'aucune information nouvelle n'est apportée, ou prise en compte par l'auditeur. La deuxième hypothèse est que la représentation auditive est perturbée par la variation du spectre pendant la durée des transitions car l'adaptation n'est pas instantanée. La discriminabilité devrait augmenter quand la transition de F2 est précédée par un "préfixe" formantique. Les résultats le confirment, ce que notre modèle peut prévoir si on remarque que lorsque le préfixe est soumis au codage d'intensité, une plus grande partie de la transition peut être suivie par le codage temporel. L'expérience suivante examine le rôle de la source sonore. SUMMERFIELD & al. (1985) rapportent que dans des groupes VCV de synthèse à source de bruit, la partie CV est plus discriminable de part et d'autre d'une frontière phonétique. LACERDA, pour les stimuli à spectres constants de la première expérience excités par du bruit, démontre que l'attaque graduelle ou brutale ne joue plus de rôle. Notre modèle rend sans difficulté compte de ce résultat. La possibilité d'un codage temporel étant fortement réduite par la nature non périodique du stimulus, le codage d'intensité est seul pertinent. Il apparaît que la définition spectrale ne dépend pas de la forme spécifique de l'attaque pour ce type de codage. Mais, à l'inverse également des résultats de SIDWELL & SUMMERFIELD (1986), qui portent sur des groupes CVC, la partie VC d'un groupe VCV conserve un léger avantage dans l'expérience décrite ensuite. De toute évidence, les résultats sur ces groupes plus complexes dépendent de nombreux paramètres insuffisamment maîtrisés aujourd'hui. Ils ne peuvent donc fournir un test pour un modèle. LACERDA poursuit en montrant que la direction de la pente n'est pas un paramètre pertinent. L'avantage de la séquence [ab] sur [ba] est claire. Notre modèle ne prévoit pas non plus de différence en fonction de la direction des transitions. La dernière expérience que nous mentionnerons porte sur l'effet d'une barre d'explosion. Son spectre est identique à celui de la première période de la transition. La durée de silence (16 ou 33 ms) entre l'explosion et la partie vocalique n'a presque aucun effet. Mais la présence d'une barre d'explosion améliore la discriminabilité. C'est exactement ce que prévoit notre modèle. La présence de l'explosion devant la partie vocalique permet au codage temporel de se produire plus tôt, et donc un suivi plus complet de la transition formantique.

Il apparaît que le modèle perceptif que nous proposons, fondé sur un double mécanisme de codage, ne se heurte à aucune difficulté majeure. Il permet de donner une interprétation cohérente à un ensemble d'expériences.

6 - LES EFFETS DE L'ADAPTATION : REPP (1987)

Un dernier test est fourni par la série d'expériences décrite par REPP (1987) qui portent sur l'effet du murmure nasal sur l'identification des consonnes [m] et [n]. L'hypothèse testée par REPP, et rejetée en conclusion, trouve son origine dans les résultats, déjà mentionnés, de DELGUTTE (1980) montrant que le premier formant d'une consonne nasale produit un

effet d'adaptation sur le spectre neurophysiologique de la voyelle. L'adaptation modifie la partie initiale du spectre de la voyelle et la comparaison avec le spectre de la partie finale de la voyelle fournit, par différence, un indice du lieu d'articulation de la consonne. Une discussion complète de cette hypothèse est hors de propos. Nous noterons cependant avec REPP que la conséquence de l'adaptation due au murmure nasal doit être de réduire l'amplitude spectrale de la voyelle aux fréquences basses. Il peut en résulter une plus grande "audibilité" des fréquences élevées où se concentrent les indices du lieu d'articulation. Concernant notre modèle, notre analyse est la suivante. La présence du murmure nasal, comme le bruit pré-consonantique de POLS & SCHOUTEN (1978) améliore les performances d'identification en permettant un meilleur suivi des transitions vocaliques. Cette prédiction est directement testée par REPP dans la cinquième expérience où il utilise une fenêtre identique à celle de OHDE & SCHARF (1981) sur la partie vocalique de la syllabe. Un effet faible, mais statistiquement significatif, est obtenu. Cette augmentation ne se produit que lorsque la partie vocalique est perçue comme nasale, impliquant la présence d'une consonne nasale éliminée. Elle ne concerne donc que la distinction [m]-[n], et non les stimuli perçus comme étant précédés de [b]-[d]. Nous sommes d'accord avec REPP pour reconnaître que ce fait n'est pas de prime abord en accord avec le rôle attribué à un préfixe par POLS & SCHOUTEN, et surtout par nous même qui prévoyons un suivi de formant plus efficace qui devrait améliorer l'identification du lieu d'articulation dans tous les cas. Mais on peut incriminer un biais de réponse en faveur de [b], comme le fait REPP ailleurs. Ce biais de réponse proviendrait de l'absence de barre d'explosion et non d'une perception différente des transitions. Nous ne sommes pas d'accord avec REPP quand il conclut que l'effet de la fenêtre n'explique pas pourquoi des scores d'identification aléatoires d'un murmure et d'une zone vocalique peuvent s'améliorer quand les deux parties sont réunies. Rappelons notre argumentation. Le murmure nasal peut contenir peu d'information phonétique. La partie vocalique seule est mal identifiée car le suivi des formants est rendu difficile par le brusque saut d'intensité qui déclenche un codage d'intensité. La présence du murmure nasal devant la partie vocalique permet un meilleur suivi des formants.

Dans la première expérience de REPP, inverser l'ordre de la partie vocalique et du murmure nasal détériore les scores d'identification comme dans le cas des occlusives non précédées d'un bruit, ce qui est également conforme aux prédictions de notre modèle. La deuxième expérience présente en écoute dichotique, c'est-à-dire dans une oreille différente, le murmure nasal et la partie vocalique. L'adaptation, d'origine périphérique, doit disparaître, et la fiabilité de l'identification diminuer. Les résultats le confirment. Notre modèle fait la même prédiction si on admet que le déclenchement de l'analyse d'intensité est déterminé par une détection unilatérale, donc périphérique, d'une brusque augmentation d'intensité globale. Si l'on sépare le murmure de la partie vocalique par une durée de silence, les résultats d'identification doivent décroître lorsque la durée du silence est augmentée. Le

modèle de REPP, comme le nôtre, prévoit une diminution plus rapide que celle constatée (5 - 7 % en moyenne), qui reste faible même pour une durée de silence de 240 ms. Pour conserver notre hypothèse, il faut admettre que le codage temporel commencé sur le murmure n'est guère perturbé par un silence de plus de 100 ms suivi d'une brusque variation d'énergie. Si ceci est vrai, la barre d'explosion d'une consonne occlusive intervocalique n'a pas d'effet bénéfique sur le suivi de formant subséquent. Elle ne contribue qu'à la mesure de sa propre distinctivité. L'expérience décrite ensuite est difficile à commenter en quelques lignes. La durée du murmure nasal ne joue qu'un rôle mineur de 10 à 60 ms, ce qui est compatible avec notre modèle, qui prévoit un bon suivi des formants par le codage temporel si le codage d'intensité n'est pas déclenché. Il ne fait aucune prédiction sur le rôle d'un murmure nasal extrait d'un contexte différent de celui de la partie vocalique suivante. De manière complexe, cet effet peut être prévu par l'effet d'adaptation. Les expériences 6 et 7 testent directement l'hypothèse d'adaptation en affaiblissant par filtrage la zone de basse fréquence de la partie vocalique. Cette hypothèse prévoit une amélioration des scores d'identification en rendant plus "audibles" les indices de haute fréquence. Or les résultats vont en sens contraire. La détérioration des scores est légèrement plus marquée pour un filtrage de durée brève (10 ms). Notre modèle ne prévoit clairement ni amélioration, ni détérioration de l'identification, et nous ne souhaitons pas proposer une explication ad hoc. Ces expériences conduisent REPP à rejeter l'hypothèse d'un effet bénéfique de l'adaptation sur le mécanisme d'identification phonétique.

7 - CONCLUSIONS

En conclusion, le modèle de double codage est compatible avec une série de résultats expérimentaux, le plus important étant l'augmentation des scores d'identification quand une information cruciale portée par une évolution formantique n'est pas présentée juste après une brusque variation d'intensité. Le point le plus délicat est le rôle facilitant que l'on doit attribuer à un "préfixe" comme le murmure nasal placé jusqu'à 240 ms devant la partie vocalique à forte variation initiale d'énergie.

ANNEXE 4 : RESULTATS INDIVIDUELS DE L'EXPERIENCE PRINCIPALE

On trouvera ci-dessous les scores d'identification sur une échelle allant de 1 [oe] à 7 [a].
Chaque valeur est la moyenne de 10 estimations.

Fo = 110 Hz

F1	550	578	605	633	660	688	715 Hz
Sujet							
1	1.0	1.1	1.6	2.7	4.2	5.1	6.3
2	1.4	2.0	2.0	3.3	4.9	5.7	6.3
3	1.1	1.2	1.7	5.2	5.5	6.9	7.0
4	1.1	1.1	1.4	3.1	4.5	5.9	6.7
5	1.0	1.3	1.9	3.9	5.1	6.0	6.9
6	2.0	1.4	2.1	3.8	4.3	5.8	6.1
7	1.8	2.6	2.2	3.3	4.0	5.5	6.7
8	1.3	1.4	2.0	2.2	2.3	4.4	6.0
9	1.8	1.8	3.2	4.3	5.2	6.3	6.8
10	1.8	1.9	2.0	2.4	4.4	5.9	6.7
11	2.1	2.2	2.3	2.2	2.9	4.3	5.3
12	1.1	1.0	1.5	2.2	3.6	4.3	6.5
13	2.1	2.0	2.3	2.7	3.6	4.5	6.1
14	2.5	1.5	2.4	3.0	4.2	6.0	6.4
15	1.2	1.3	1.7	3.4	3.7	4.9	5.8
16	1.0	1.0	1.6	2.8	3.9	4.3	6.6

moyenne	1.52	1.55	1.99	3.16	4.14	5.36	6.39
écart-type	0.49	0.48	0.44	0.83	0.84	0.83	0.45

Fo = 188 Hz

F1	578	605	633	660	688	715	743 Hz
sujet							
1	1.2	1.5	2.2	2.8	4.5	5.6	5.9
2	1.6	1.3	2.8	3.9	4.2	5.7	6.5
3	1.3	1.3	2.0	4.4	6.6	6.9	7.0
4	1.6	1.7	2.7	3.9	5.3	6.2	6.5
5	1.5	1.5	1.7	2.8	5.3	6.3	6.9
6	1.0	1.1	1.5	2.3	4.0	4.0	5.2
7	1.6	1.6	1.7	2.0	5.3	6.8	7.0
8	1.4	1.7	1.9	2.0	3.9	5.5	5.6
9	1.7	2.0	3.3	3.5	5.3	6.2	6.7
10	1.0	1.0	1.1	1.6	4.9	6.4	6.8
11	1.0	1.2	1.4	2.3	4.1	5.2	6.2
12	1.0	1.1	1.5	2.5	4.8	5.9	7.0
13	1.3	1.7	1.4	2.1	3.3	4.4	5.3
14	1.4	1.6	2.1	2.3	3.0	4.4	5.5
15	1.1	1.1	2.6	3.2	5.2	6.1	6.6
16	1.8	1.9	1.7	2.6	5.6	6.7	6.9

moyenne	1.33	1.43	1.96	2.82	4.71	5.81	6.38
écart-type	0.27	0.32	0.59	0.82	0.80	0.87	0.63

Fo = 220 Hz

F1	605	660	688	715	743	770	830 Hz
sujet							
1	1.6	1.8	1.7	2.3	4.0	5.0	6.9
2	1.7	2.0	1.8	2.8	4.0	5.7	7.0
3	1.3	2.1	3.0	4.2	6.4	7.0	7.0
4	1.7	3.2	3.1	3.1	5.4	6.0	6.7
5	1.7	2.3	2.4	2.8	3.9	5.0	6.6
6	1.8	2.2	2.4	2.5	4.2	4.8	6.1
7	2.6	2.6	1.5	2.3	3.8	5.0	6.8
8	1.1	2.0	2.0	2.3	2.7	4.6	6.7
9	2.5	2.6	2.8	3.1	3.7	5.7	6.7
10	1.3	2.7	2.3	2.6	4.2	4.9	6.2
11	1.9	2.4	2.2	2.8	3.1	4.6	6.2
12	1.0	1.0	1.2	1.4	2.1	2.4	6.4
13	1.7	2.2	2.2	3.6	3.0	4.3	5.9
14	2.8	3.3	2.6	3.2	4.2	4.8	6.4
15	2.0	2.5	2.7	3.2	4.8	5.7	7.0
16	1.0	2.5	2.5	2.5	2.2	4.1	6.8

moyenne	1.73	2.42	2.27	2.79	3.85	4.97	6.63
écart-type	0.56	0.62	0.55	0.66	1.15	1.03	0.31

Fo = 270 Hz

F1	605	633	660	688	715	743	770 Hz
sujet							
1	1.5	1.7	2.2	2.6	3.8	5.4	6.6
2	1.2	1.5	1.8	2.4	2.9	5.0	6.4
3	1.7	2.0	2.6	4.2	6.6	6.6	6.9
4	2.6	2.9	3.2	3.8	4.8	5.2	5.9
5	1.1	1.6	2.1	2.9	4.5	6.0	6.6
6	1.0	1.2	1.7	2.3	4.0	5.6	6.7
7	1.3	1.7	1.8	2.2	3.6	6.2	7.0
8	1.4	1.9	2.2	2.4	3.2	4.1	5.6
9	2.1	2.3	3.0	3.0	3.9	5.9	6.8
10	1.5	1.7	1.6	2.3	3.5	6.3	6.6
11	1.1	1.3	2.0	2.8	3.9	4.9	6.1
12	1.0	1.0	1.1	1.7	3.0	5.1	6.6
13	2.0	1.8	2.1	2.6	3.7	4.3	5.5
14	2.3	2.3	2.5	3.0	3.8	5.2	6.2
15	1.0	1.1	2.3	3.7	4.4	6.2	7.0
16	1.2	1.6	2.0	2.6	4.1	6.8	7.0

moyenne	1.48	1.71	2.12	2.78	3.99	5.60	6.50
écart-type	0.49	0.48	0.51	0.63	0.84	0.79	0.48

BIBLIOGRAPHIE DE LA DEUXIEME PARTIE

LISTE DES ABBREVIATIONS DES REVUES ET CONFERENCES

- Ann. Rpt. Eng. Res. Inst. - Annual Report of the Engineering Research Institute, University of Tokyo, Japon.
- IEEE - ICASSP - Institute of Electrical and Electronics Engineers - International Congress on Acoustics, Speech and Signal Processing.
- JASA - Journal of the Acoustical Society of America, U.S.A.
- J. Exp. Psych. : HPP - Journal of Experimental Psychology : Human Performance and Perception, U.S.A.
- JSHR - Journal of Speech and Hearing Research, U.S.A.
- PERILUS - Phonetic Experimental Research at the Institute of Linguistics University of Stockholm, Suède.
- Proc. Inst. Acoust. - Proceedings of the Institute of Acoustics, London, G.B.
- Proc. Nth Int. Cong. Acoust. - Proceedings of the Nth International Congress on Acoustics.
- P&P - Perception and Psychophysics, U.S.A.
- STL- QPSR - Speech Transmission Laboratory - Quaterly Progress on Speech Research - Royal Institute of Technology, Stockholm, Suède

- ABBAS P.J. & SACHS M.B. (1976) "Two-tone suppression in auditory nerve fibers : extension of a stimulus-response relationship", JASA, 59, 112 - 122.
- AINSWORTH W.A. (1971) "Perception of synthesized isolated vowels and h-d words as a function of fundamental frequency", JASA, 49, 1323 - 1324.
- AINSWORTH W.A. (1974) "Influence of fundamental frequency on perceived vowel boundaries in English", in Speech Communication, FANT G. (ed.), Almqvist & Wiksell, Stockholm, 123 - 129.
- ASSMANN P.F. & NEAREY T.M. (1987) "Perception of front vowels : the role of harmonics in the first formant region", JASA, 81(2), 520 - 534.
- BACON S.P. & BRANDT J.F. (1982) "Auditory processing of vowels by normal-hearing

- and hearing-impaired listeners", JSHR, 25(3), 339 - 347.
- BASTET L. (1978) Etude et réalisation d'un modèle fonctionnel du système auditif périphérique, Thèse de Docteur - Ingénieur, I.N.P.Grenoble.
- BLADON R.A.W. (1982a) "Problems of normalizing the spectral effects of variations in the fundamental", Proc. Inst. Acoust. 1982, A51 - A55.
- BLADON R.A.W. (1982 b) "Arguments against formants in the auditory representation of speech", in CARLSON R. & GRANSTRÖM B. (eds), The representation of speech in the peripheral auditory system, Elsevier, Amsterdam, 95 - 102.
- BLADON R.A.W. & LINDBLOM B. (1981) "Modeling the judgment of vowel quality differences", JASA, 69, 1414 - 1422.
- BORG E. & ZAKRISSON J. (1973) "Stapedius reflex and speech features", JASA, 54, 525 - 527.
- CARLSON R., FANT G. & GRANSTRÖM B. (1975) "Two-formant models, pitch and vowel perception", in Auditory analysis and Perception of Speech, FANT G. & TATHAM M.A.A. (eds), Academic Press, London, 55 - 82.
- CARLSON R., GRANSTRÖM B. & FANT G. (1970) "Some studies concerning the perception of isolated vowels", STL - QPSR, 2/3-70, 19 - 35.
- CARRE R. & LANCIA R. (1975) "Perception of vowel amplitude transients", in FANT G. & TATHAM M.A.A. (eds.), Auditory analysis and perception of speech, Academic Press, London, 89 - 90.
- CARRE R. (1988) "Rôle de la fréquence fondamentale dans la perception des voyelles de synthèse", communication personnelle, à paraître.
- CARRE R. & QUACH-TUAN (1986) "Effects of non-stationary characteristics on the perception of the vowels", Proc. NATO Advanced Research Workshop on Psychophysics on Speech Perception, Utrecht, 1986 (aussi in Bull. Lab. Comm. Parlée Grenoble, 1, 307 - 318, 1987).
- CHIBA T. & KAJIYAMA M.(1941) The vowel : Its nature and structure, Tokyo-Kaisekan Tokyo.
- CHISTOVICH L. A. (1970) "Problems of speech perception", in HAMMERICH L.L., JAKOBSON R. & ZWIRNER E. (eds.), Form and Substance, Akademik Forlag, Copenhagen,
- CHISTOVICH L.A. (1971) "Auditory processing of speech stimuli - evidence from psychoacoustics and neurophysiology", Proc. 7th Int. Cong. Acoust., Akademiai Kiado, Budapest, 1, 27 - 42.
- CROWDER R.G. & REPP B.H. (1984) "Single formant contrast in vowel identification", P&P, 35, 372 - 378.
- DALLOS P. (1970) "Low-frequency auditory characteristics : species dependence", JASA, 48(2), 489- 499.

- DARWIN C.J. (1984 a) "Perceiving vowels in the presence of another sound : constraints on formant perception", JASA, 1976, 1636 - 1647.
- DARWIN C.J.(1984 b) "Auditory processing and speech perception", in BOUMA H. & BOUWHUIS D.G. (eds.), Attention and performance X : Control of language processes, L.E.A., Hillsdale, U.S.A., 197 - 210.
- DARWIN C.J. & GARDNER R.B. (1985) "Which harmonics contribute to the estimation of first formant frequency ? ", Speech Communication, 4, 231 - 235.
- DARWIN C.J. & GARDNER R.B. (1986) "Mistuning a harmonic of a vowel : grouping and phase effects on vowel quality", JASA, 79(3), 838 - 845.
- DELGUTTE B. (1980) "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers", JASA, 68, 848 - 857.
- DELGUTTE B. (1986) "Two-tone rate suppression in auditory-nerve fibers : variation with suppressor level and frequency", rapport non publié, communication personnelle.
- DELGUTTE B (1987) "Peripheral auditory processing of speech information : implications from a physiological study of intensity discrimination", in SCHOUTEN M.E.H.(ed.), Psychophysics of speech perception, Martinus Nijhoff, Dordrecht, 333 - 353.
- DELGUTTE B. & KIANG N.Y.S. (1984) "Speech coding in the auditory nerve : I. Vowel-like sounds", JASA, 75(3), 866 - 878.
- DORMAN M.F., CEDAR I., HANNLEY M.T., LEEK M.R. & LINDHOLM J.M. (1986) "Influence of the acoustic reflex on vowel recognition", JSHR, 29, 420 - 424
- DORMAN M.F. & DOUGHERTY K. (1981) "Shifts in phonetic identification with changes in signal presentation level", JASA, 69(5), 1439 - 1440
- DORMAN M.F., LINDHOLM J.M., HANNLEY M.T & LEEK M.R (1987)"Vowel intelligibility in the absence of the acoustic reflex : performance-intensity characteristics", JASA, 81(2), 562 - 564.
- DUIFHUIS H. (1980) "Level effects in psychophysical two-tone suppression", JASA, 67, 914 - 927.
- ESCUQUIER P. & SCHWARTZ J.L. (1985 a) "Pulsation threshold patterns of synthetic vowels : study of the second formant emergence and the 'center of gravity' effect", Speech Communication, 4, 189 - 198.
- ESCUQUIER P & SCHWARTZ J.L. (1985 b) "Pulsation threshold test : evidence for the existence of two biases and discussion of their implications", JASA, 78, 463 - 473.
- FANT G. (1957) Den Akustiska Fonetikens Grunder, Tech. Report n° 7, STL - RIT, Stockholm.
- FANT G. (1959) "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics, 1, 3 - 108.

- FANT G., CARLSON R. & GRANSTRÖM B. (1974) "The [e] - [ø] ambiguity", FANT G. (ed.), in Speech Communication, Almqvist & Wiksell, Stockholm, 117 - 121.
- FLANAGAN J.L. (1955) "A difference limen for vowel formant frequency", JASA, 27, 613 - 617.
- FLANAGAN J.L. (1970) Speech analysis, synthesis and perception, Springer Vlg., N.Y., 2ème ed.
- FLOREN Å. (1978/1979) "why does [a] change to [ɔ] when Fo is increased ?", PERILUS, 1, 13 - 23.
- FOX R.A. (1985) "Auditory contrast and speaker quality variations in vowel perception", JASA, 77(4), 1552 - 1559.
- FUJISAKI H. & KAWASHIMA T. (1968) "The roles of pitch and higher formants in the perception of vowels", IEEE-AU, 16, 73 - 77.
- FUJISAKI H. & KAWASHIMA T. (1970) "Some experiments on speech perception and a model for the perceptual mechanisms", Ann. Rpt. Eng. Res. Inst., U. of Tokyo, 29, 207 - 214. (cf. aussi FUJISAKI H. (1979) "On the modes and mechanisms of speech perception, in LINDBLOM B. & ÖHMAN S. (eds.), Frontiers of Speech Communication research, Academic Press, London, 177 - 189).
- HOLMES J.N. (1986) "Normalization in vowel perception", in PERKELL J.S. & KLATT D.H. (eds.), Invariance and Variability in Speech processes, L.E.A., Hillsdale N.J., 346 - 357.
- HOLMES M.H. (1980) "An analysis of a low-frequency model of the cochlea", JASA, 68, 482 - 488.
- HOLMES M.H. (1987) "Frequency discrimination in the mammalian cochlea : theory versus experiment", JASA, 81(1), 103 - 114.
- HOUTGAST T. (1974) "Auditory analysis of vowel-like sounds", Acustica, 31, 320 - 324.
- JAVEL E., MCGEE J., WALSH E.J., FARLET G.R. & GORGA M.P. (1983) "Suppression of auditory nerve responses : II. Suppression threshold and growth, iso-suppression contours", JASA, 74(3), 801 - 813.
- KAKUSHO O., HIRATO H., KATO K. & KOBAYASHI T. (1971) "Some experiments of vowel perception by harmonic synthesizer", Acustica, 53, 132 - 142.
- KARNICKAYA E.G., MUSHNIKOV V.N., SLEPOKUROVA N.A. & ZHUKOV S. J. (1975) "Auditory processing of steady state vowels", in FANT G. & TATHAM M.A.A. (eds.), Auditory Analysis and Perception of Speech, Academic Press, London, 37 - 53.
- KLATT D.H. (1982) "Prediction of perceived phonetic distance from critical-band spectra : a first step", IEEE-ICASSP, 1278 - 1281.
- LACERDA F. (1987) Effects of peripheral auditory adaptation on the discrimination of

- speech sounds, Ph.D, U. de Stockholm, PERILUS, 6.
- LARIVIERE C, WINITZ H. & HERRIMAN E. (1975) "Vocalic transitions in the perception of voiceless initial stop", JASA, 57, 470 - 475.
- LINDQVIST J. & PAULI S. (1968) "The role of relative spectrum levels in vowel perception", Proc. 6th Int. Congress on Acoustics, Tokyo, B91 - B94.
- LONCHAMP F. (1978) Recherches sur les indices perceptifs des voyelles orales et nasales, Thèse de 3ème cycle, U. de Nancy II.
- LONCHAMP F. (1981) "Multidimensional vocalic perceptual space : how many dimensions ?", JASA, 69, S94 ; cf. aussi "Analyse multidimensionnelle de l'espace perceptif des sons vocaliques", 12ème J.E.P. du GALF, Montréal 1981, 406 - 418.
- LONCHAMP F. (1984) "Perceptual cue validation", Speech Communication, 3(1), 17 - 28.
- LONCHAMP F. (1985) "Formants, harmonics and vocalic timbre : experiments and speculations", in GUERIN B. & CARRE R. (eds), Actes du Symposium franco-suédois sur la Parole, GALF - SFA, vol. 1, 189 - 208.
- LONCHAMP F. (1986) "Prediction of Fo-induced phonetic change in vowels with identical spectral envelopes", Proc. NATO Advanced Research Workshop on Psychophysics on Speech Perception, Utrecht, 1986.
- LONCHAMP F.(sous presse) "Acoustique descriptive des sons du français", in CALLIOPE (ed.), La parole et son traitement automatique, Masson.
- MERMELSTEIN P. (1978) "Difference limens for formant frequencies of steady-state and consonant-bound vowels", JASA, 68(2), 572 - 581.
- MILLER R.L. (1953) "Auditory tests with synthetic vowels", JASA, 25, 114 - 121.
- MØLLER A.R. (1983) Auditory physiology, Academic Press, N.Y.
- MOORE B.C.J. & GLASBERG B.R. (1983 a) "Suggested formulae for calculating auditory filter bandwidths and excitation patterns", JASA, 74(3), 750 - 753.
- MOORE B.C.J. & GLASBERG B.R. (1983 b)"Masking patterns for synthetic vowels in simultaneous and forward masking", JASA, 73(3), 906 - 917.
- MOORE B.C.J. & GLASBERG B.R. (1983 c) "Forward masking patterns for harmonic complex tones", JASA, 73(5), 1682 - 1685.
- MOORE B.C.J., GLASBERG B.R. & PETERS R.W. (1985 a) "Relative dominance of individual partials in determining the pitch of complex tones", JASA, 77, 1853 - 1860.
- MOORE B.C.J., PETERS R.W. & GLASBERG B.R. (1985 b) "Thresholds for the detection of inharmonicity in complex tones", JASA, 77, 1861 - 1868.
- MUSHNIKOV V.N & CHISTOVICH L.A. (1972) "Method for the experimental investigation of the role of component loudness in the recognition of a vowel",

- Soviet Physics - Acoustics, 17(3), 339 - 344.
- MUSHNIKOV V.N. & CHISTOVICH L.A. (1973) "Experimental testing of the band hypothesis of vowel perception", Soviet Physics - Acoustics, 19 (3), 250 - 254.
- ODHE R.N. & SCHARF D.J. (1977) "Order effect of acoustic segments of VC and CV syllables on stop and vowel identification", JSHR, 20, 543 - 554.
- ODHE R.N; & SCHARF D.J. (1981) "Stop identification from vocalic transition plus vowel segments of CV and VC syllables : a follow-up study", JASA, 69(1), 297 - 300.
- PALMER A.R., WINTER I.M., GARDNER R.B. & DARWIN C.J. (1987) "Changes in the phonemic quality and neural representation of a vowel by alteration of the relative phase of harmonics near F1", in SCHOUTEN M.E.H. (ed.), Psychophysics of speech perception, Martinus Nijhoff, Dordrecht.
- PETERSON G.E. & BARNEY H.L. "Control methods used in a study of the vowels", JASA, 24, 175 - 184.
- POLS L.C.W & SCHOUTEN M.E.H. (1978) "Identification of deleted consonants", JASA, 64, 1333 - 1337.
- POLS L.C.W. & SCHOUTEN M.E.H. (1981) "Identification of deleted plosives : the effect of adding noise or applying a time window (A reply to OHDE & SCHARF 1981)", JASA, 69(1), 301 - 303.
- POTTER R.K. & STEINBERG J.C. (1950) "Toward the specification of speech", JASA, 22, 807 - 820.
- REPP B.H. (1987) "On the possible role of auditory short-term adaptation in perception of the prevocalic [m]-[n] contrast", JASA, 82(5), 1525 - 1538.
- REPP B.H., HEALY A.F. & CROWDER R.G. (1979) "Categories and context in the perception of isolated steady-state vowels", J. Exp. Psych. : HPP, 5, 129 - 145.
- ROSSI M. & MARTIN-LAVAL E. (1985) "The effect of selective adaptation on the perception of place of articulation in French voiced stops", Speech Communication, 4(1-3), 199 - 212.
- SACHS M.B. & YOUNG E.D. (1979) "Encoding of steady-state vowels in the auditory nerve : representation in terms of discharge rate", JASA, 66, 470 - 479.
- SCHEFFERS M.T.M. (1983), I.P.O., manuscript 450/II (cité in SUMMERFIELD & ASSMANN 1987)
- SCHARF B. (1970) "Critical bands", in TOBIAS J.V. (ed.), Foundations of modern auditory theory, Academic Press, N.Y., vol. 1, 159 - 202.
- SCHWARTZ J.L. (1987) Représentations auditives de spectres vocaliques, Thèse de Doctorat d'Etat, I.N.P.Grenoble.
- SCOTT B.L. (1976) "Temporal factors in vowel perception", JASA, 60(6), 1354 - 1356.
- SCOTT B.L. (1980) "Speech as patterns in time", in COLE R.A. (ed.) Perception and

- production of fluent speech, L.E.A., Hillsdale, USA, 51 - 71.
- SHAMMA S.A. (1985) "Speech processing in the auditory system I : The representation of speech sounds in the response of the auditory nerve", JASA, 78(5), 1612 - 1621.
- SIDWELL A. & SUMMERFIELD Q. (1986) "The auditory representation of symmetrical CVC syllables", Speech Communication, 5, 283 - 297.
- SLAWSON A.W. (1968) "Vowel quality and musical timbre as functions of spectral envelope and fundamental frequency", JASA, 43, 87 - 101.
- SMITH R.L. (1979) "Adaptation, saturation and physiological masking in single auditory-nerve fibers", JASA, 65, 166 - 178.
- SMITH R.L. & BRACHMAN M.L. (1980) "Operating range and maximum response of single auditory-nerve fibers", Brain Research, 184, 499 - 505
- SORIN C. (1987) "Psychophysical representation of stop consonant and temporal masking in speech", in SCHOUTEN M.E.H., Psychophysics of speech perception (Proc. NATO Advanced Research Workshop, Utrecht, 1986), Martinus Nijhoff, Dordrecht, 241 - 247.
- SUMMERFIELD Q. & ASSMANN P. (1987) "Auditory enhancement in speech perception", in SCHOUTEN M.E.H. (ed.), Psychophysics of speech perception, Martinus Nijhoff, Dordrecht, 140 - 150.
- SUMMERFIELD Q., FOSTER J., TYLER R. & BAILEY J. (1985) "Influence of formant bandwidth and auditory frequency selectivity on the identification of place of articulation in stop consonants", Speech Communication, 4, 213 - 229.
- SUNDBERG J. & GAUFFIN J. (1982) "Amplitude of the voice source fundamental and the intelligibility of super pitch vowels", in CARLSON R. & GRANSTRÖM B. (eds), The representation of speech in the peripheral auditory system, Elsevier, Amsterdam, 223 - 228.
- STUMPF C. (1926) Die Sprachlaute, Springer, Berlin.
- SYRDAL-LASKY A. (1978) "Effects of intensity on the categorical perception of stop consonants and isolated second formant transitions", P&P, 23(5), 420 - 432.
- TENENHOLTZ E. (1978 / 1979) "Vowel identification as a function of increasing fundamental frequency", PERILUS, 1, 38 - 48.
- TRAUNMÜLLER H. (1981) "Perceptual dimension of openness in vowel", JASA, 69(5), 1465 - 1475.
- TRAUNMÜLLER H. (1983) On Vowels : Perception of spectral features, related aspects of production and sociophonetic dimensions", Ph. D., University of Stockholm.
- TRAUNMÜLLER H. (1984) "Articulatory and perceptual factors controlling the age - and sex-conditioned variability in formant frequencies of vowels", Speech Communication, 3(1), 49 - 62.
- TRAUNMÜLLER H. (1985) "The role of the fundamental and the higher formants in the

- perception of speaker size, vocal effort and vowel openness", Symposium Franco-Suédois sur la parole, GALF - SFA, Grenoble 1985, p.209 - 219.
- TYLER R.S & LINDBLOM B. (1982) "Preliminary study of simultaneous masking and pulsation threshold patterns for vowels", JASA, 71, 220 - 224.
- VAN TASSELL D.J. & CRUMP E.S.A. (1981) "Effects of stimulus level on perception of two acoustic cues in speech", JASA, 70(5), 1527 - 1529.
- YOUNG E.D. & SACHS M.B. (1979) "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers", JASA, 66, 1381 - 1403.
- ZWICKER E. & FELDTKELLER R. (1967) Das Ohr als Nachrichtenempfänger, S. Hirzel Vlg., Stuttgart. (aussi en traduction française par C. SORIN, L'oreille, récepteur d'information, Masson, 1980)

ENVOI

Banalité, lieu commun, clause de style, habilité éculée que de dire que le travail présenté est imparfait et incomplet, et de jurer, par sous-entendu prudent, de le poursuivre et le parfaire. Ce sont pourtant les mots qui viennent spontanément à l'esprit.

Je les rejette pourtant. Voici le produit de mes efforts, au jour d'aujourd'hui. L'expérience de ma thèse de 3ème cycle me montre que demain, à la lumière d'autres lectures et réflexions, je tirerai peut-être des conclusions différentes. Je reviendrai à ces données, en ajouterai de nouvelles, ou j'explorerai d'autres domaines, laissant alors à d'autres le soin de montrer, s'ils le jugent utile, les fautes et imperfections, ou les mérites de ce travail.

Eo, opus est.

*

TABLE DES MATIERES

0 - PRESENTATION	5
PREMIERE PARTIE : LES INDICES ACOUSTIQUES DE LA NASALITE VOCALIQUE	
1 - INTRODUCTION	7
2 - ORGANISATION	7
3 - ANATOMIE DES FOSSES NASALES	8
3.1 - Introduction	8
3.2 - Le nez	10
3.3 - Les fosses nasales et le rhinopharynx	10
3.4 - Les sinus paranasaux	16
3.5 - Coupes tomographiques horizontales et frontales : le problème de la taille du limen nasi	23
4 - ARTICULATION DES VOYELLES NASALES	30
5 - ACOUSTIQUE THEORIQUE DE LA NASALITE ET MODELE DE SIMULATION	37
5.1 - Analyse graphique par le schéma de réactance	37
5.2 - Déplacements fréquentiels dus au couplage nasal	42
5.3 - Ordre des pôles et des zéros	42
5.4 - Le sinus comme résonateur de HELMHOLTZ	44
5.5 - Le modèle de simulation	48
6 - ANALYSE ACOUSTIQUE : LE CHOIX DE LA METHODE	53
7 - LA CONTRIBUTION SPECTRALE DE LA SOURCE ET DES COUPLAGES NON VELAIRES	54
7.1 - La source glottale	54
7.2 - Le couplage sous-glottique	62
7.3 - Le couplage nasal externe (Merlier 1984)	68
8 - LES INDICES ACOUSTIQUES DE LA NASALITE VOCALIQUE : SYNTHESE INTERPRETATIVE	68
8.1 - FANT (1960)	68

8.2 - HOUSE & STEVENS (1956)	77
8.3 - BJUGGREN & FANT (1964)	78
8.4 - FUJIMURA & LINDQVIST (1971)	78
8.5 - MRAYATI (1975, 1976)	83
8.6 - LINDQVIST & SUNDBERG (1976)	85
8.7 - LONCHAMP (1978)	90
8.8 - MAEDA (1982)	90
8.9 - FENG, ABRY & GUERIN (1985) - FENG (1986)	92
8.10 - HAWKINS & STEVENS (1985)	102
8.11 - Conclusions	105
9 - SIMULATIONS ET ANALYSES	106
9.1 - Sensibilité des fréquences de résonance du conduit nasal à la modification des paramètres articulatoires	106
9.2 - Du rôle de l'abaissement du voile dans les variations formantiques et de la dissymétrie des fosses nasales	111
9.3 - Simulation et analyse de la consonne nasale vélaire	112
10 - IDENTIFICATION DES INDICES DE NASALITE DANS LES VOYELLES NATURELLES DU FRANCAIS	128
10.1 - Le corpus	129
10.2 - La voyelle [ɛ̃]	131
10.2.1 - Description spectrale	131
10.2.2 - Interprétation	139
10.2.3 - Conclusions	148
10.3 - La voyelle [ã]	148
10.3.1 - Description spectrale	148
10.3.2 - Interprétation	155
10.3.3 - Conclusions	161
10.4 - La voyelle [õ]	162
10.4.1 - Description spectrale	162
10.4.2 - Interprétation	169
10.4.3 - Conclusions	174
11 - LES INDICES ACOUSTIQUES DE LA NASALITE VOCALIQUE : UN ESSAI DE SYNTHESE	175
12 - PERCEPTION DES INDICES DE NASALITE VOCALIQUE	178
12.1 - HAWKINS & STEVENS (1985)	178
12.2 - CHENG & GUERIN (1987 b)	182
12.3 - MAEDA (1984)	185
12.4 - BOGNAR & FUJISAKI (1986)	186

12.5 - WRIGHT (1986); KAWASAKI (1986)	187
12.6 - Les indices perceptifs de la nasalité vocaliques : quelques réflexions	189
12.7 - Conclusions	192
- ANNEXE A: Fonctions d'aire orales	194
- ANNEXE B: Fonctions d'aire nasales	196
- ANNEXE C: Les sinus	198
- BIBLIOGRAPHIE DE LA PREMIERE PARTIE	199

DEUXIEME PARTIE : LA MODIFICATION DU TIMBRE PAR LA FREQUENCE FONDAMENTALE

0 - INTRODUCTION : LA REPRESENTATION INTERNE DES FORMANTS	208
0.1 - Formants et fréquence fondamentale	208
0.2 - Le formant comme indice perceptif: discussion des critiques "classiques"	209
0.3 - Objectif de l'étude : La représentation perceptive des maxima spectraux	210
1 - ORGANISATION	210
2 - L'INTERACTION DU TIMBRE VOCALIQUE ET DE LA FREQUENCE FONDAMENTALE : HYPOTHESES ET DONNEES EXPERIMENTALES	211
2.1 - R.L. MILLER (1953)	212
2.2 - FANT (1957); CARLSON & al. (1970, 1974); FANT & al (1974).	212
2.3 - FUJISAKI & KAWASHIMA (1968)	213
2.4 - SLAWSON (1968)	214
2.5 - MUSHNIKOV & CHISTOVICH (1972, 1973)	218
2.6 - AINSWORTH (1974)	220
2.7 - CARLSON, FANT & GRANSTRÖM (1975)	220
2.8 - SCOTT (1976)	223
2.9 - FLOREN (1978/1979)	226
2.10 - TRAUNMÜLLER (1981, 1985)	230
2.11 - SUNDBERG & GAUFFIN (1982)	236
2.12 - MOORE & GLASBERG (1983)	236

2.13 - LONCHAMP (1985)	242
2.14 - DARWIN & GARDNER (1985, 1986)	252
2.15 - HOLMES (1986)	256
2.16 - VAN TASSEL & al. (1987)	258
2.17 - ASSMANN & NEAREY (1987)	258
2.18 - Conclusions	263
3 - ETUDE EXPERIMENTALE DE LA VARIATION DE TIMBRE INDUITE PAR UNE MODIFICATION DE F ₀ POUR DES VOYELLES OUVERTES	264
3.1 - Le choix de la méthode d'identification et des paramètres de synthèse	264
3.2 - Description des stimuli	266
3.3 - Calcul des spectres auditifs	269
3.4 - Sujets et conditions de présentation	270
3.5 - Résultats	270
3.6 - Discussion	278
3.6.1 - La série de 110 Hz	278
3.6.2 - Echec de la prédiction par filtrage large bande	280
3.6.3 - Le rôle des harmoniques sous la plus proéminente	283
3.6.4 - La série de 188 Hz	283
3.6.5 - La série de 270 Hz	284
3.6.6 - La série de 220 Hz	286
3.6.7 - Le problème du déplacement de la frontière phonétique	289
3.6.8 - Le caractère 'masculin' et 'féminin' des voyelles	292
3.6.9 - Une distance F ₁ - F ₀ constante ne conserve pas le timbre invariant	293
3.6.10 - Analyse des résultats de CARRE (1988)	294
3.6.11 - Le rôle de l'attaque	299
3.6.12 - Un problème potentiel : la non linéarité des vibrations de la membrane basilaire aux très basses fréquences	303
4 - CONCLUSIONS	303
ANNEXE 1 : Discussion de SCHWARTZ & ESCUDIER "A direct demonstration of a large-scale integrated spectral representation in vowel perception"	308

ANNEXE 2 : Spectres auditifs de 3 voyelles naturelles obtenus par la méthode du seuil de pulsation	312
ANNEXE 3 : Le double codage de l'information spectrale - Témoignages du domaine de la perception de la parole	318
1 - Bases neurophysiologiques du codage spectral par l'oreille	319
2 - Codage d'intensité et codage temporel	320
3 - POLS & SCHOUTEN (1978, 1981) : le rôle d'un préfixe de bruit dans la perception des occlusives	320
4 - Une forte intensité détériore l'identification des sons	323
5 - Le codage de l'attaque : les expériences de IACERDA (1987)	324
6 - Les effets de l'adaptation : REPP (1987)	325
7 - Conclusions	327
ANNEXE 4 : Résultats individuels de l'expérience principale	328
 - BIBLIOGRAPHIE DE LA DEUXIEME PARTIE	 332

 - ENVOI	 340

 TABLE DES MATIERES	 341
