

CARACTERES SPECIFIQUES DES VOYELLES NASALES DU FRANCAIS

Jean-Sylvain LIENARD*
Maria-Gabriella DI BENEDETTO**

* LIMSI-CNRS BP133 91403 ORSAY Cedex

** INFOCOM Università La Sapienza, Via Eudossiana 18, 00184 ROMA

Résumé

Dans cette étude on cherche à caractériser les voyelles nasales du français dans la perspective d'aider à l'apprentissage de leur prononciation par un non-francophone. Dans un premier temps on dresse un panorama des connaissances disponibles sur la nasalité vocalique, selon divers points de vue : production, acoustique, perception et synthèse. Dans un deuxième temps on décrit la base de données utilisée, comportant des voyelles isolées émises par divers locuteurs. Puis on décrit les traitements permettant d'associer à chaque segment une enveloppe spectrale. Chaque nasale est ensuite comparée visuellement avec l'orale correspondante, en vue de faire ressortir leurs ressemblances et différences. Ces observations sont confrontées aux connaissances publiées. Enfin on décrit une expérience de reconnaissance automatique, dans laquelle la nasale produite par chaque locuteur dans diverses conditions d'élocution est comparée à sa reconstruction à partir des orales de la même série.

INTRODUCTION

Cette étude préliminaire a été menée dans le but de permettre au système SPELL d'aider un élève à apprendre la prononciation des voyelles nasales du français¹. On suppose qu'il a acquis auparavant la prononciation des voyelles orales. Le système doit décider si la nasale produite par l'élève est correcte, sans pouvoir se référer à un modèle provenant du même locuteur. Cette question suppose que les connaissances concernant la nasalité vocalique pour le français soient disponibles et utilisables en reconnaissance automatique. Nous commencerons par faire le point sur la nasalité vocalique, selon divers points de vue : production, analyse, perception/synthèse, reconnaissance. Puis nous décrirons nos données et l'analyse qui en a été faite afin de mettre en évidence l'enveloppe spectrale. Nous montrerons que, s'il est difficile de dégager des indices

¹ Ce travail a été effectué en partie dans le cadre du contrat ESPRIT SPELL-II (projet 7153), dont le partenaire principal est la société AGORA CONSEIL, en partie dans le cadre d'une coopération entre le CNRS et le CNR (Consiglio Nazionale delle Ricerche).

de la nasalité dans l'absolu, par contre on peut caractériser acoustiquement chaque nasale par rapport à certaines voyelles orales du même locuteur dans les mêmes conditions. Enfin nous montrerons comment ces indices peuvent être validés dans une expérience de reconnaissance automatique.

LA NASALITE VOCALIQUE

Production

La nasalité est produite par la mise en parallèle du conduit nasal avec le conduit oral, par abaissement du velum. Le couplage avec le conduit nasal entraîne l'apparition de plusieurs résonances et antirésonances (pour une discussion nous renvoyons le lecteur à Maeda [1] et Castelli [2]). Mentionnons ici les principaux effets de la nasalisation établis par calcul :

- Premier formant nasal, F_{n1} , aux alentours de 250 Hz. Cette résonance notée par de nombreux auteurs, semble se comporter comme une résonance de Helmholtz et non comme une fréquence propre d'un tuyau acoustique.
- Série de formants aux environs de 1000, 2000, 3000 et 4000 Hz. Ces formants correspondent aux fréquences propres du conduit nasopharyngal considéré comme un tuyau 1/2 onde.
- Paire pôle-zéro dans la région de F_1 : en particulier les simulations montrent qu'une paire pôle-zéro apparaît entre 300 et 1000 Hz dès l'ouverture du velum, même pour de petites aires de couplage.
- Amortissement général du spectre : les pôles du conduit nasopharyngal, les paires pôle-zéro dues au couplage et la dissipation d'énergie dans les fosses nasales ont pour effet d'amortir le spectre dans son ensemble.

Analyse

Si l'on se place du point de vue de l'analyse, on part de matériaux réels (voyelles émises par des sujets bien caractérisés, au lieu de calculs sur un conduit vocal idéalisé), dans lesquels on cherche des indices significatifs du point de vue de la phonétique, de la production ou de la perception.

Le travail de F. Lonchamp en 1978 [4] a été mené dans cette perspective. Deux locuteurs masculins (phonéticiens spécialistes) ont enregistré chacun 6

exemplaires des quatre voyelles nasales du français, dans des conditions de laboratoire (voyelles parfaitement stables, F_0 fixé à 120 Hz). Chaque exemplaire a fait l'objet d'une analyse spectrographique, et l'auteur a déterminé manuellement la position des divers maxima spectraux à partir des "sections" sonographiques (spectre en bande étroite, 45 Hz). Il a ensuite confronté la position de ces divers maxima d'une part avec les mesures publiées par d'autres auteurs, et d'autre part avec les prédictions des divers spécialistes de la production. Il a complété son étude par des synthèses au moyen d'un synthétiseur à formants de type série. Voici quelques-unes de ses conclusions :

- présence d'un premier formant nasal F_{n1} vers 280 Hz,
- présence d'un second formant nasal F_{n2} vers 2000 Hz,
- les voyelles nasales semblent caractérisées par la présence de deux "formants oraux décalés" F_1 et F_2 , sans qu'il soit besoin de faire appel à un indice général de nasalité,
- la voyelle orale française la plus voisine de [ã] est plutôt [ɔ] que [a].

Perception

La plupart des études sur la perception de la nasalité vocalique ont été menées au moyen de la synthèse. Dès 1954 P. Delattre [5] a signalé plusieurs indices utiles pour synthétiser une voyelle nasale à partir de la voyelle orale correspondante, soit par ordre d'importance :

- l'affaiblissement du premier formant F_1 ,
- l'accentuation de la résonance basse à 250 Hz (premier formant nasal),
- la mise en place d'un second formant nasal vers 2000 Hz.

Mais le travail le plus significatif dans la méthodologie synthèse-perception est sans doute celui de S. Hawkins et K. Stevens en 1985 [6]. Utilisant un synthétiseur de Klatt en configuration série ils ont synthétisé divers intermédiaires entre une voyelle orale et la nasale correspondante, en faisant varier les caractéristiques d'un pôle et d'un zéro placés dans la région de F_1 . Ce traitement a été appliqué à cinq voyelles différentes ([i], [e], [a], [o], [u]). Les synthèses ont été évaluées par divers groupes d'auditeurs, notamment par un groupe d'Américains pour qui l'opposition oral / nasal n'est pas phonologiquement pertinente, et divers groupes de locuteurs dont la langue maternelle possède cette opposition pour toutes ou plusieurs des cinq voyelles étudiées. Pour chacune des voyelles traitées les auteurs ont pu trouver un positionnement optimum du pôle et du zéro additionnels, tel que la nasalité a été correctement perçue par tous les groupes d'auditeurs.

Reconnaissance

Il existe peu d'études sur la reconnaissance automatique de la nasalité vocalique. Mentionnons les

travaux de Maeda [1], portant sur des voyelles nasales synthétiques produites par son modèle phonatoire, et de Glass [3], portant sur la nasalisation vocalique en anglais américain. Ce dernier a obtenu un taux d'erreur de l'ordre de 26% sur un corpus multilocuteur, la décision à prendre étant binaire (la voyelle étudiée était soit purement orale, soit nasalisée au contact d'une consonne nasale : le choix au hasard donnerait un taux d'erreur de 50%).

CORPUS ET ANALYSES

Le corpus Corenc a été conçu avec l'idée d'inclure plusieurs sources de variabilité dans les données et d'étudier leurs interactions. Il s'agit d'une petite base de données, comprenant 60 séries de 12 voyelles françaises, prononcées par 13 locuteurs en plusieurs sessions, avec diverses forces de voix. Dans cette phase préliminaire on a choisi de traiter des voyelles isolées et d'utiliser des locuteurs ordinaires, non-spécialistes de phonétique.

Dans l'inventaire des voyelles utilisées on a laissé de côté le [ɔ] qui, à l'inverse des autres voyelles du français, ne peut être noté sous forme d'un mot lexical, ainsi que le [ã], qui n'est plus phonologiquement distinct de [ã] et le [ə] qui n'a pas d'existence autonome. On a aussi considéré un seul [a], le [a] moyen qui correspond aujourd'hui à la majorité des réalisations en français standard. On a demandé au locuteur de répéter les voyelles prononcées par l'opérateur, avec une force de voix en rapport avec la distance à laquelle se trouve ce dernier (30 cm, 1,50 m et 7m). On a évité de se placer dans les cas extrêmes (voix chuchotée et voix criée), pour rester dans le registre des variations habituelles et presque inconscientes de la voix. Les locuteurs sont 6 hommes et 7 femmes, adultes, originaires pour la moitié d'entre eux de la région de Grenoble. Les enregistrements ont été faits en deux sessions distantes de six mois ; certains locuteurs ont participé aux deux sessions.

Les 720 voyelles ont été évaluées auditivement par un groupe de 5 auditeurs majoritairement parisiens et distinct du groupe des locuteurs [7]. Les conclusions en ce qui concerne le problème traité ici sont les suivantes :

- les voyelles isolées sont très largement identifiées par les auditeurs (environ 9% d'erreur, ce qui est relativement faible),
- les voyelles nasales sont moins bien reconnues que les voyelles orales (environ 15% d'erreur, contre 7%, la nasale la plus difficile à identifier étant "on", avec 23% d'erreur),
- les confusions sont fortement liées au locuteur : certains locuteurs semblent avoir plus de mal que d'autres à articuler correctement les nasales ou certaines nasales.

L'analyse a été faite dans la perspective d'une reconstitution ultérieure du signal par une méthode dérivée de l'analyse sinusoidale [8][9][10]. Dans un premier temps on recherche pour chaque segment l'instant qui concilie le plus fort niveau sonore avec la plus forte stabilité spectrale. Dans la suite chaque voyelle sera représentée par le spectre prélevé à cet instant du segment, avec les paramètres suivants : durée 51,2 ms, fenêtre de Hamming, amplitudes en dB. Ce spectre est ensuite réduit à ses pics, après une correction d'amplitude (interpolation parabolique).

Le spectre brut fait apparaître de nombreux pics parasites, qui très souvent ne correspondent pas à des détails spectraux audibles. Pour les éliminer on calcule en chaque point du spectre brut une "fonction de masquage", parabole convexe centrée sur chaque pic et de largeur (évaluée à -20 dB) variant entre 120 Hz dans

le bas du spectre et 330 Hz dans le haut. En chaque point du spectre on ne retient que le maximum de toutes les fonctions de masquage concernant ce point, ce qui constitue le "spectre masqué", auquel on fait ensuite subir un léger lissage. Le résultat de cette opération est un élargissement des raies spectrales et un déplacement d'ensemble du spectre vers les valeurs maxima ; dans ce processus les pointes de faible intensité par rapport à leurs voisines sont masquées. Pour obtenir l'enveloppe il suffit alors de ne retenir que les pics du spectre brut dépassant la valeur du spectre masqué lissé, et d'interpoler linéairement entre pics successifs.

L'échelle linéaire de fréquence donne une trop grande importance à la partie haute du spectre. L'utilisation d'une échelle Bark selon une formulation due à Schroeder [11] apporte une correction à cette distorsion.

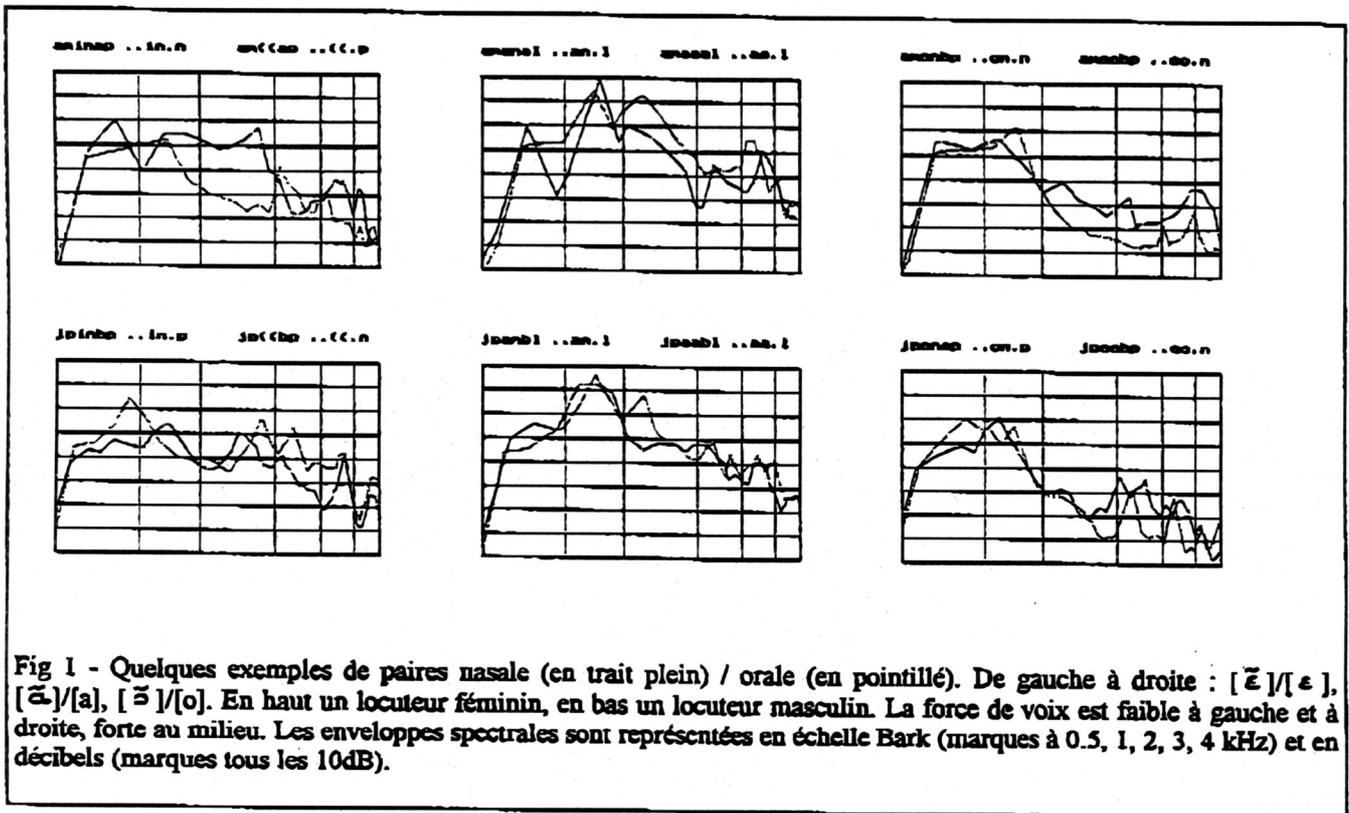


Fig 1 - Quelques exemples de paires nasale (en trait plein) / orale (en pointillé). De gauche à droite : [ẽ]/[ɛ], [ẽ]/[a], [õ]/[o]. En haut un locuteur féminin, en bas un locuteur masculin. La force de voix est faible à gauche et à droite, forte au milieu. Les enveloppes spectrales sont représentées en échelle Bark (marques à 0.5, 1, 2, 3, 4 kHz) et en décibels (marques tous les 10dB).

RESULTATS QUALITATIFS

Sur une partie de notre corpus (3 locuteurs et 3 locutrices, 3 forces de voix, deux sessions) nous avons apparié chaque voyelle nasale avec son homologue orale, toutes deux validées perceptivement en ce qui concerne l'identification vocalique. Quelques exemples de ces paires sont représentés sur la fig 1. On rapporte dans ce qui suit ce que les deux éléments de chaque

paire ont en commun, ce qui les différencie, en quoi les indices observés confirment ou infirment les résultats publiés à ce jour.

La paire [ẽ]-[ɛ]

La caractéristique commune est l'existence d'un creux situé entre 1000 et 1500 Hz, entouré de deux bosses l'une dans la région 400 - 800 Hz (F1 ou F'1), l'autre dans la région 1500 - 2200 Hz (F2 ou F'2).

En ce qui concerne les différences on constate que F1 monte d'environ 150 Hz, F2 descend d'environ 300 Hz, soit pour chaque formant un mouvement de 1 à 2 Bark dans le sens d'une compacité croissante. Le creux qui les sépare est toujours mieux rempli pour la nasale que pour l'orale. Pour certains locuteurs l'effet d'un zéro de nasalité vers 500 Hz est manifeste. Par contre le renforcement de F_{n1} à 250 Hz n'apparaît qu'occasionnellement.

Pour transformer un [e] synthétique en son équivalent nasal Hawkins et Stevens mettent un zéro à 500 Hz et un pôle additionnel à 700 Hz. Ceci a pour effet d'augmenter notablement F1 qui passe de 400 à 700 Hz. F2 reste inchangé à 2000 Hz. Même si les fréquences sont différentes (car la voyelle en question est plus proche du [e] français que du [ɛ] qui nous sert de comparaison), le mécanisme semble bien expliquer l'évolution de F1. Mais l'hypothèse selon laquelle F2 n'est pas affecté par la nasalisation semble ici tout-à-fait irréaliste. D'après Lonchamp on passe de [ɛ] à [ɛ̃] en faisant évoluer F1 de 450 à 680 Hz, F2 de 1800 à 1480 Hz, et en ajoutant le premier formant nasal à 280 Hz. Mis à part le dernier point, cette description est en accord avec nos relevés.

La paire [a]-[ã]

A quelques exceptions près, ces enveloppes des deux voyelles ont en commun un maximum global situé entre 500 et 1200 Hz, et un creux relatif aux alentours de 1,5 à 2 kHz. On ne distingue pas toujours clairement les deux formants du [a]. La partie haute du spectre (au delà de 2 kHz) est souvent renforcée, mais présente une atténuation d'au moins 20 dB par rapport à la région dominante. La partie basse (entre F₀ et 500 Hz) est essentiellement plate ou concave, avec une atténuation moindre que la partie haute (entre 5 et 15 dB).

La différence se trouve essentiellement dans l'atténuation de la région 1200 Hz par rapport à la région dominante : cette atténuation est inférieure à 15 dB pour [a], supérieure à 15 dB pour [ã]. Si l'on prend comme référence l'amplitude du fondamental, l'ensemble du spectre est légèrement atténué pour [ã] par rapport à [a]. Ceci est plus net en voix faible ou normale qu'en voix forte.

Dans l'ensemble on est assez loin des résultats publiés par Hawkins et Stevens, qui nasalisent le [a] de synthèse en mettant un pôle additionnel à 800 Hz et un zéro à 600 Hz, ce qui a pour effet d'augmenter F1 (qui passe de 700 à 800 Hz, soit environ 1 Bark) tout en laissant F2 inchangé (vers 1200 Hz). Par contre Lonchamp considère que F1 reste inchangé (vers 600 Hz) et que F2 baisse de 1150 à 950 Hz (environ 2 Bark), ce qui correspond mieux à nos observations, avec cependant une assez grande variabilité propre au locuteur et à la force de voix. On retrouve ici une observation de Castelli selon laquelle la résonance F_{n1} à 250 Hz apparaît aussi dans les voyelles orales, en parti-

culier dans [a]. Quant au zéro de nasalité vers 500 Hz, il n'est pas observé chez tous les locuteurs.

La paire [ɔ]-[ɔ̃]

Dans tous les cas on observe un maximum global entre F₀ et 800 Hz. La partie basse du spectre est en forme de plateau, et la partie haute (au delà de 1000 Hz) est atténuée d'au moins 20 dB pour la voix forte, 25 à 30 dB pour la voix faible.

En ce qui concerne les différences on observe dans la plupart des cas (soit environ 2 fois sur 3) deux maxima à environ 400 et 700 Hz pour [o] ; pour [ɔ̃] on observe un maximum dans la zone 500-600 Hz, et un creux ou un plateau à 400 Hz environ. On trouve des formes dégénérées, en particulier en voix forte, où n'apparaît qu'un seul maximum dans la zone F₀-800 Hz. Dans ce cas on a affaire à [o] si le maximum est au dessous de 500 Hz, et à [ɔ̃] s'il est au-dessus de 600 Hz. Il subsiste quelques cas ambigus. D'une manière générale [ɔ̃] possède plus d'énergie que [o] dans la bande haute.

D'après Hawkins et Stevens on passe de [o] à son équivalent nasal en plaçant un zéro à 500 Hz, ce qui annule le F1 oral défini à 450 Hz, et un pôle à 650 Hz, F2 restant inchangé à 800 Hz. Comme précédemment ceci correspond à nos observations en ce qui concerne F1, mais peut difficilement être confirmé en ce qui concerne F2, très difficile à identifier dans cette région. D'après Lonchamp le F1 oral disparaît et se transforme en formant de nasalité F_{n1} à 280 Hz, tandis que F2 passe de 800 Hz pour [o] à 650 Hz pour [ɔ̃]. Ceci correspond bien à nos observations, compte tenu des variations dues au locuteur et à la force de voix.

RESULTATS QUANTITATIFS : SPECTRES DIFFERENTIELS MOYENS

On trouvera fig 2 le "spectre différentiel moyen" (sdm), constitué par la différence entre le spectre de la nasale et celui de l'orale correspondante, calculé sur les 18 séries d'un sous-corpus, soit environ 1/3 de la base de données. Malgré l'aspect un peu sommaire de ce moyennage (aucun appareillage manuel ; aucun ajustement des échelles fréquentielles dont on sait qu'elles diffèrent notablement entre voix masculines et voix féminines ; les moyennes peuvent inclure des exemplaires non validés perceptivement), on retrouve dans les sdm certaines des observations rapportées ci-dessus. Pour interpréter les détails du spectre différentiel il faut observer qu'un passage à zéro dans le sens croissant indique un décalage spectral vers l'aigu. On remarquera en particulier la "bosse de compacité" traduisant le rapprochement de F1 et F2 quand on passe de [ɛ] à [ɛ̃], le creux à 1200 Hz caractérisant [ã] par rapport à [a] et dû à l'abaissement de F2, la paire pôle (à 600 Hz) -

zéro (à 400 Hz) et l'augmentation d'énergie au-dessus de 1 kHz qui caractérisent le passage de [o] vers [ɔ̃].

VALIDATION PAR RECONNAISSANCE AUTOMATIQUE

Pour évaluer la pertinence des connaissances implicitement codées dans les sdm nous avons mené une expérience de reconnaissance automatique. Notre corpus est divisé en 3 sous-corpus a, b et c. Dans chaque sous-corpus on trouve 18 séries de 12 voyelles, provenant de 6 locuteurs (3 hommes et 3 femmes) utilisant les 3 forces de voix. Les locuteurs des corpus a et b sont les mêmes, enregistrés à 6 mois d'intervalle. Les locuteurs du sous-corpus c sont différents mais équilibrés de la même manière. On choisit un sous-corpus d'apprentissage, sur lequel on calcule pour chacune des 3 nasales 3 sdm avec 3 voyelles orales "de base", en l'occurrence [ɛ], [a] et [o]. On choisit ensuite un sous-corpus de test, qui est examiné série par série. Pour chaque série on constitue un ensemble de 12 spectres de référence, soit les 9 orales de la série, et trois "nasales synthétiques", constituées à partir des 3 orales de base et des sdm appris précédemment. La raison pour utiliser 3 orales de base alors qu'en principe une seule devrait suffire (la plus proche de la nasale considérée), est que même à l'intérieur d'une série le locuteur peut avoir introduit une variabilité notable : en moyennant trois reconstructions de la même nasale à partir de 3 orales différentes on espère compenser dans une certaine mesure ces variations. La reconnaissance proprement dite consiste à comparer chaque nasale "vraie" à l'ensemble des 12 spectres de référence. La distance de deux spectres est calculée par le biais de 12 coefficients cepstraux, selon une technique décrite dans [12] et lar-

gement utilisée aujourd'hui dans les systèmes de reconnaissance. Les résultats obtenus en permutant les sous-corpus d'apprentissage et de test sont reportés dans le tableau ci-dessous (valeurs données en pourcentages d'erreur arrondis à l'entier le plus proche, 54 essais par cas):

ref-->	s/corp a	s/corp b	s/corp c
tst --> a	22	26	48
tst --> b	20	15	43
tst --> c	43	44	20

Pour interpréter ces chiffres il faut se souvenir des conditions expérimentales. D'un côté un choix aléatoire (parmi 12 possibilités) mènerait à une valeur de 91%. De l'autre les évaluations perceptives ont montré un taux d'erreur de 15% sur les nasales, au dessous duquel il est peu probable qu'un processus automatique puisse descendre. Les résultats sont donc très bons en auto-cohérence (corpus d'apprentissage identique au corpus de test), encore bons d'une session à l'autre avec les mêmes locuteurs, et nettement moins bons quand on change l'ensemble des locuteurs. Cependant, même dans ce dernier cas, le pourcentage d'erreur moyen qui est de l'ordre de 45% reste très prometteur. D'une part on peut penser qu'environ un tiers de ces erreurs sont dues à des données mal articulées. D'autre part on a déjà mentionné le fait que la procédure de calcul des sdm est très simple et ne tient pas compte de toutes les informations disponibles dans le corpus d'apprentissage. En tout état de cause ces résultats permettent de valider les observations faites plus haut et illustrées par la fig 2 ci-dessous.

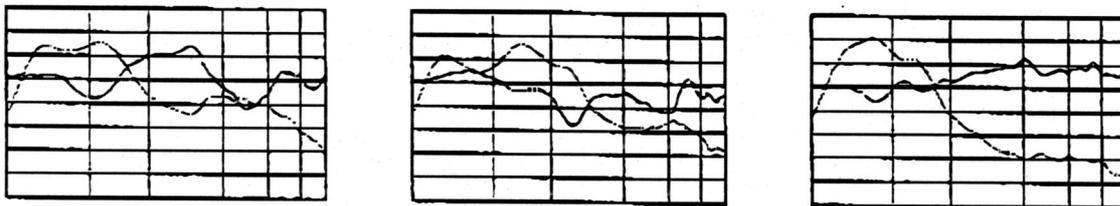


Fig 2 - Spectres différentiels moyens (en trait plein) et voyelles orales moyennes (en pointillé), obtenues sur le sous-corpus b (18 séries : 6 locuteurs et 3 forces de voix). De gauche à droite : sdm [ɛ̃]-[ɛ] et orale moyenne [ɛ], sdm [ɔ̃]-[a] et orale moyenne [a], sdm [ɔ̃]-[o] et orale moyenne [o]. Les sdm sont centrés sur l'ordonnée 50 dB.

CONCLUSION

Cette étude nous a permis de faire le point sur le problème de la nasalité vocalique en français, et sur sa détection par des moyens automatiques dans des conditions peu contraintes (locuteurs non-professionnels, conditions d'élocution relativement libres). Nous nous sommes placés dans la perspective particulière de l'évaluation des nasales par rapport à une ou plusieurs voyelles orales du même locuteur. Les premiers résultats semblent indiquer la possibilité d'identifier les nasales du français dans ces conditions. On notera le fait qu'une étude phonétique de la nasalité fait habituellement appel à une description spectrale en termes de formants ; une telle description est extrêmement difficile à mettre en oeuvre dans un processus automatique, même pour des voyelles orales, dès lors que l'on sort des conditions de laboratoire. C'est pourquoi nous avons préféré rechercher des indices purement spectraux, en l'occurrence les différences spectrales caractérisant chaque nasale par rapport à l'orale correspondante. Dans cette approche le passage du spectre brut à une enveloppe phonétiquement significative, indépendante du fondamental et du bruit de fond, constitue une phase critique du travail d'analyse. L'hypothèse générale de notre travail, qui consiste à ne considérer que le spectre de stabilité, est évidemment provisoire dans la mesure où il est établi que la reconnaissance d'une voyelle est aussi liée à son évolution temporelle [13].

Nous n'avons pas manqué, tout au long de cette étude, de nous poser la question de l'existence d'indices absolus de la nasalité vocalique, indépendants de la voyelle nasale considérée. Il est sans doute trop tôt pour pouvoir répondre à cette question. Nos relevés, confrontés aux connaissances publiées, montrent notamment que les répercussions spectrales de la paire pôle-zéro dans la région de F1 contribuent notablement à distinguer la voyelle nasale de sa correspondante orale. Ceci corrobore partiellement les expériences de Hawkins et Stevens. Pourtant il semble bien que l'hypothèse d'invariance de F2 ne soit pas vérifiée dans notre corpus, en particulier pour les paires []-[a] et []-[], ce qui rejoint certaines des conclusions de Lonchamp. On peut douter de la valeur universelle, du point de vue perceptif, de F1, de la faible valeur ou de l'élargissement de F1, de l'amortissement général du spectre, voire même de la perception d'un zéro en tant que tel dans la région 400-600 Hz. Par contre pour le français il semble que la perception des voyelles nasales ne soit pas fondamentalement différente de celle des voyelles orales : nous rejoignons Lonchamp dans l'idée que les unes comme les autres peuvent trouver leur place spécifique dans une classification bidimensionnelle, dont les dimensions perceptives, qui restent à préciser, ne sont sans doute pas exactement identiques aux formants pris en considération par la théorie de la production.

REFERENCES

- 1 - Maeda S. : "Acoustics of vowel nasalization and articulatory shifts in French nasal vowels", in Phonetics and Phonology, vol 5, "Nasals, Nasalization, and the Velum", eds M.K.Huffman and R.A.Krakow, Ac. Press Inc., 1993.
- 2 - Castelli E. : "Caractérisation acoustique des voyelles nasales du français", thèse Inst. Nat. Polyt. de Grenoble, juin 1989.
- 3 - Glass J.R. and Zue V.W. : "Detection of nasalized vowels in American English", IEEE ICASSP, Tampa, 1985.
- 4 - Lonchamp F. : "Recherche sur les indices perceptifs des voyelles orales et nasales : application à la structure du système vocalique français et de diverses autres langues", thèse 3e cycle, univ. Nancy II, mai 1978.
- 5 - Delattre P. : "Les attributs acoustiques de la nasalité vocalique et consonantique", Studia Linguistica 8, 103-109, 1954.
- 6 - Hawkins S. and Stevens K.N. : "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels", J. Acoust. Soc. Am 77(4), April 1985.
- 7 - Liénard J.S. et Di Benedetto, M.G. : "Evaluation perceptive d'un corpus de voyelles françaises émises isolément par plusieurs locuteurs selon diverses forces de voix", 19e Journées d'Etude sur la Parole de la SFA, Bruxelles, mai 1992.
- 8 - Serra X. : "A system for sound analysis, transformation and synthesis based on a deterministic plus stochastic decomposition", Ph.D. dissertation, Stanford University, Oct. 1989.
- 9 - McAulay R.J. and Quatieri T.F. : "Speech analysis/synthesis based on a sinusoidal representation", IEEE Tr. on ASSP, vol ASSP 34, 744-754, 1986.
- 10 - Grau-Grovel S. : "Utilisation de la représentation par formes d'ondes formantiques et de la représentation sinusoïdale dans l'élaboration d'un synthétiseur paramétrique de parole", thèse, univ. Paris XI, nov. 1992.
- 11 - Schroeder M.R., Atal B.S. and Hall J.L. : "Optimizing digital speech coders by exploiting masking properties of the human ear", J. Acoust. Soc. Am 66(6), Dec. 1979.
- 12 - Davis S.B and Mermelstein P. : "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", IEEE Tr. on Acoustics, Speech and Signal Processing, vol ASSP-28 n° 4, Aug. 1980.
- 13 - Di Benedetto M.G. : "Vowel representation : some observations on temporal and spectral properties of the first formant frequency", J. Acoust. Soc. Am. 86 (1), July 1989.