



# Pertinence des trois premiers formants des voyelles orales dans la caractérisation du locuteur

Odile Mella

## ► To cite this version:

Odile Mella. Pertinence des trois premiers formants des voyelles orales dans la caractérisation du locuteur. JEP 1992 - 19e Journées d'Etude sur la Parole, May 1992, Bruxelles, Belgique. pp.1-6. hal-01737027

**HAL Id: hal-01737027**

**<https://inria.hal.science/hal-01737027>**

Submitted on 19 Mar 2018

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

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

# Pertinence des trois premiers formants des voyelles orales dans la caractérisation du locuteur

Odile Mella

CRIN-CNRS & INRIA lorraine  
B.P. 239 f54506 Vandœuvre-lès-Nancy Cedex

## RESUME

Le travail rapporté dans cet article se situe dans le cadre de la caractérisation automatique du locuteur. Dans cette optique, nous sommes en train de réaliser une étude sur la pertinence de plusieurs paramètres phonétiques et acoustiques dont le premier volet est constitué par l'étude des trois premiers formants de certaines voyelles orales du français. Il s'agit de déterminer les voyelles orales les mieux adaptées à la reconnaissance automatique du locuteur et, pour chacune d'elles de déterminer les formants ou les combinaisons de formants les plus discriminants. Après avoir décrit l'élaboration et l'étiquetage du corpus dont sont issues ces voyelles, nous développerons la méthode de détermination des trois premiers formants des voyelles. Puis, nous présenterons les indicateurs de pertinence utilisés pour classer les voyelles et les combinaisons formantiques avant de terminer par la présentation de quelques résultats.

## INTRODUCTION

La recherche en traitement automatique de la parole a mis en évidence la variabilité du signal de parole en fonction du locuteur. Les études menées dans le cadre de l'identification ou de la vérification du locuteur se classent dans deux catégories selon qu'elles utilisent cette variabilité de façon implicite [ROS, 76], ou qu'au contraire elles essaient de l'extraire du message acoustico-phonétique [CAE, 88]. Notre travail se situe dans cette dernière catégorie. Il a pour but d'étudier la pertinence pour la caractérisation du locuteur d'un certain nombre de paramètres phonologiques, phonétiques et acoustiques spécifiques du français et qui, à terme, pourraient être utilisés dans des systèmes de la première catégorie [GON, 92]. Dans une

première étape, nous avons étudié les trois premiers formants de certaines voyelles orales. Nous allons présenter dans cet article, les trois phases de cette étude : l'élaboration et l'étiquetage du corpus nécessaire à l'étude des paramètres, la détermination des trois premiers formants des voyelles orales retenues et le calcul d'indicateurs de pertinence permettant de classer les voyelles orales et les combinaisons formantiques associées. Puis, nous concluons par la présentation de quelques résultats.

## ELABORATION ET ETIQUETAGE DU CORPUS

### **1 Elaboration du corpus**

Un ensemble de paramètres fréquentiels, temporels et phonologiques susceptibles d'être pertinents pour la reconnaissance du locuteur a été sélectionné avec l'aide de F. Lonchamp, Professeur à l'Institut de Phonétique de Nancy. Parmi les paramètres fréquentiels figurent les formants des voyelles orales, /*ɛ*/, /*e*/, /*œ*/, /*ɔ*/, /*a*/, /*î*/ et /*u*/, précédées d'un contexte neutre au sens de la coarticulation linguale, /*p*/ ou /*b*/, et suivies d'un contexte postérieur allongeant, /*R*/ . Suite à cette sélection, dix-sept phrases ont été construites de façon à minimiser le nombre de phrases à prononcer par rapport au nombre de paramètres retenus. L'emploi de phrases a été jugé préférable à celui de mots isolés. En effet, la lecture de listes de mots hors contexte devient vite fastidieuse pour un locuteur bénévole et reflète peu sa prononciation naturelle. La table 1 présente les phrases du corpus dans lesquelles les occurrences des triplets /*p-voyelle-R*/ et /*b-voyelle-R*/ sont soulignées.

- 1 *Guy a péri bêtement du diabète en Italie.*
- 2 *La porte du garage tomba avec lourdeur.*
- 3 *La partie de belote dura toute la matinée.*
- 4 *Un bateau à vapeur a quitté le port.*
- 5 *Le petit gamin traîne un jouet.*
- 6 *Donne-moi le bocal de cacao !*
- 7 *En ski, la godille permet d'éviter les tournants.*
- 8 *Un coq bien dodu pour demain !*
- 9 *Lequel des bandits guette près du repère ?*
- 10 *Le trappeur commun redoutait le loup-garou.*
- 11 *Douze nains conspirent derrière le bosquet.*
- 12 *Le soldat brisa la baguette de son tambour.*
- 13 *Goûtez-moi ce cake au beurre !*
- 14 *Le rire de la gouvernante est revigorant.*
- 15 *La cousine du nain soupire dans son délire.*
- 16 *Le départ de la course Strasbourg-Paris aura du retard.*
- 17 *Notre guide charmant quitte la jolie route danoise.*

Table 1. Les dix-sept phrases du corpus.

Pour des raisons d'homogénéité de population, dix-sept locuteurs et vingt-et-une locutrices, tous originaires de Lorraine ou y résidant depuis de nombreuses années, ont été choisis pour lire quatre répétitions de chaque phrase. Ces répétitions ont été présentées aux locuteurs sous la forme de listes de phrases rangées aléatoirement afin de ne pas toujours reporter sur les mêmes phrases les phénomènes de fin de liste. L'enregistrement, effectué à partir d'un microphone Shure et d'un magnétophone Revox, ne s'est pas déroulé dans une véritable chambre sourde mais dans une pièce isolée et à des heures calmes.

## 2 Etiquetage du corpus

Une partie du corpus correspondant à dix locuteurs masculins a été numérisée phrase par phrase sur un Masscomp 5600. L'acquisition a été effectuée avec un filtrage passe-bas à 6800 Hz et une conversion A/D sur 12 bits à la fréquence d'échantillonnage de 16 kHz. Puis, les phrases numérisées ont été étiquetées manuellement à l'aide de la version 88 de SNORRI, logiciel interactif d'édition de signal de parole développé au laboratoire [LAP, 88].

La segmentation des phrases, réalisée par l'indication de frontières de segments sur le spectrogramme, a été effectuée suivant un certain nombre de critères. Ces critères ont pour but d'obtenir une segmentation répétitive proche d'une segmentation semi-automatique, de faciliter le calcul des différents paramètres à étudier et de remédier aux hésitations

dues au manque d'expérience de l'"étiqueteuse". Ces règles sont nécessaires lorsqu'il y a hésitation entre plusieurs frontières possibles ou lorsqu'il n'existe pas de frontière évidente. Nous donnons ici les critères qui ont été appliqués dans la segmentation de la voyelle du triplet **/p-voyelle-R/** :

- souvent, le **/R/** est vocalique et donc très difficile à séparer de la voyelle qui le précède. Le positionnement de la frontière s'est surtout fait grâce à l'écoute du segment acoustique, au changement de timbre de la voyelle ;
- lorsque le triplet termine une phrase ou un groupe de phonation, la fin de la voyelle est déterminée par la fin du formant F1 de la voyelle quand ce dernier remonte vers celui du **/R/** et que la barre de voisement disparaît ;
- dans quelques cas, par exemple dans la phrase numéro 1, le phonème **/R/** est en position inter-vocalique. L'ensemble **/voyelle-R-voyelle/** comporte alors des transitions longues comparativement aux parties stables des phonèmes. Les frontières peuvent être positionnées au début, au milieu ou à la fin de ces transitions. Ces triplets servant aussi à l'analyse spectrale du **/R/**, il a été décidé d'inclure les transitions dans les voyelles environnantes et de limiter le **/R/** à sa partie fricative.

La transcription a été effectuée simultanément à la segmentation et par conséquent à partir des mêmes informations. Elle est de nature à la fois phonologique et acoustique et correspond à ce qui a été entendu et vu, sauf dans les cas ambigus où l'étiquette est celle de l'élément attendu [MEL, 89].

Après cette première phase de construction de la base de données nécessaire à l'étude des paramètres sélectionnés, nous avons étudié la pertinence des trois premiers formants des voyelles orales retenues. Pour cela, nous avons conçu une méthode de détermination automatique de ces formants que nous allons développer dans le paragraphe suivant.

## DETERMINATION DES TROIS PREMIERS FORMANTS DES VOYELLES ORALES

Un formant de voyelle orale est déterminé à partir de trois formants intermédiaires calculés à trois emplacements dans la voyelle. Chacun de ces formants intermédiaires est lui-même obtenu grâce à une méthode automatique d'affectation des pôles issus de l'analyse LPC aux formants F1, F2 et F3 de la voyelle orale. Nous allons détailler ces différentes étapes dans

les paragraphes suivants, après avoir défini un formant  $F_i$ , intermédiaire ou final, comme une structure constituée :

- d'une fréquence formantique,  $F_i.fr$  ;
- d'une largeur de bande,  $F_i.bw$  ;
- d'un coefficient de défiance,  $F_i.df$  ;

Le coefficient de défiance  $df$  est composé de trois champs indépendants  $df_3$ ,  $df_2$  et  $df_1$ . Les champs  $df_2$  et  $df_1$  seront mis à jour au cours des différentes phases de calcul des formants finaux alors que  $df_3$  le sera lors du calcul des distances entre locuteurs. Plus le coefficient de défiance est élevé, moins la valeur du formant et la distance calculée à partir de cette valeur sont fiables.

## 1 Calcul et premiers filtrages des racines LPC

Nous effectuons une analyse LPC fondée sur la méthode d'autocorrélation de Durbin [MAR, 76], qui fournit dix-huit coefficients à partir desquels sont extraites toutes les racines LPC ayant une largeur de bande inférieure ou égale à 1000 Hz. Un premier filtrage des racines LPC permet d'éliminer les pôles ayant des fréquences trop voisines. Si l'écart fréquentiel entre deux racines est inférieur à la valeur médiane du pitch calculé sur la phrase contenant la voyelle étudiée, seule est conservée la racine dont la largeur de bande est la plus petite. Ensuite, un deuxième filtrage des racines est effectué afin de ne conserver que les racines dont la fréquence appartient à un intervalle prédéfini (ou domaine de définition). Ces domaines de définition sont les intervalles fréquentiels dans lesquels sont censés se trouver les trois premiers formants des voyelles retenues, précédées des contextes /p/ et /b/ et suivies du contexte /R/ et prononcées par un locuteur masculin. Ces intervalles ont d'abord été établis sur les conseils d'une phonéticienne du laboratoire puis affinés en fonction des premiers résultats obtenus.

## 2 Affectation des racines LPC aux formants intermédiaires F1, F2 et F3

Pour chaque voyelle, les racines restantes, triées dans l'ordre croissant de leurs fréquences formantiques, sont affectées séquentiellement aux trois premiers formants F1, F2 et F3 selon le nombre de racines présentes dans le domaine de définition  $D(F_i)$  de chaque formant :

- si  $D(F_i)$  ne contient aucune racine, le formant intermédiaire  $F_i$  est nul ;
- si  $D(F_i)$  contient une seule racine candidate, elle est affectée au formant  $F_i$  ;

- si  $D(F_i)$  contient deux racines susceptibles d'être affectées au formant, un algorithme sélectionne une des deux racines en utilisant l'une des règles suivantes, classées par ordre de priorité :

- éliminer la racine qui a déjà été affectée à  $F_{i-1}$ ,
- éliminer la racine qui est la seule candidate pour le formant  $F_{i+1}$ ,
- éliminer la racine dont la largeur de bande est supérieure au double de la largeur de bande de l'autre racine,
- éliminer la racine dont la fréquence est la plus éloignée d'une valeur de référence du formant cherché. Nous avons choisi comme valeurs de référence les valeurs médianes des formants F1, F2 et F3 établies par F. Lonchamp dans une étude sur un corpus de triplets /p-Voyelle-R/ [CAL, 89] ;

- si  $D(F_i)$  contient trois racines, un algorithme de choix du même type est appliqué.

La largeur de bande du formant,  $F_i.bw$ , est donnée par la largeur de bande associée à la racine retenue. De plus, le champ  $df_1$  du coefficient de défiance  $F_i.df$  vaut (en Hz)

$$(1) \quad \begin{array}{ll} 1 & \text{si } F_i.bw \geq 250 + \frac{F_i.fr}{10} \\ 0 & \text{sinon} \end{array}$$

## 3 Détermination des formants finaux F1, F2 et F3

L'application de la méthode précédemment décrite permet de calculer, pour chacun des formants d'une voyelle orale, trois valeurs intermédiaires à trois endroits situés dans la partie la plus stable de la voyelle, un emplacement central, un emplacement situé à  $-8$  ms et un autre à  $+8$  ms. L'emplacement central dépend de la durée de la voyelle, variable selon la position syntaxico-sémantique de la voyelle dans la phrase. Si la voyelle est longue (durée  $> 160$  ms), l'emplacement central est pris à 80 ms du début de la voyelle, sinon il correspond au milieu de la voyelle.

Afin d'obtenir des fréquences formantiques robustes à partir desquelles seront calculées des distances entre locuteurs, l'attribution d'une valeur à une fréquence formantique finale dépend de la proximité des valeurs intermédiaires.

En nous fondant sur les résultats de Monsen et Engebretson [MON, 83] sur les erreurs d'estimation des trois premiers formants des voyelles orales en spectrographie comme en analyse par prédiction linéaire, nous avons choisi, comme écart maximal entre fréquences formantiques, 60 Hz pour F1 et 110 Hz pour F2 et F3. Nous avons utilisé ces écarts pour

déterminer la valeur finale du formant et de son coefficient de défiance selon une démarche résumée dans la table 2. On remarque dans cette table que si la fréquence d'un formant n'est pas considérée comme suffisamment fiable, elle est mise à 0.

Une fois les valeurs finales des trois premiers formants des voyelles orales établies, la pertinence pour la caractérisation du locuteur des voyelles orales et de certaines combinaisons de ces formants peut être étudiée. La méthodologie employée pour réaliser cette étude fait l'objet du prochain paragraphe.

## ETUDE DE LA PERTINENCE DES VOYELLES ORALES

Le nombre restreint d'échantillons par triplet (quatre par locuteur) ne nous a pas permis d'utiliser une analyse de type analyse discriminante. Aussi avons-nous décidé de limiter les combinaisons formantiques étudiées aux formants (un, deux ou trois formants par voyelle) et aux écarts entre les formants (un, deux ou trois écarts par voyelle).

Lors de l'étude d'une combinaison formantique, pour chaque répétition  $i$  de la voyelle étudiée, un locuteur est représenté par un vecteur ayant pour composantes les éléments de cette combinaison. Ainsi, lors du test de la pertinence de la combinaison ( $F_1, F_2, F_3$ ), le locuteur  $k$  est représenté par le vecteur  $V_{k_i}(x_1, x_2, x_3) = V_{k_i}(F_1, F_2, F_3)$ .

Les méthodes de classement que nous avons mises en œuvre ont recours au calcul d'une distance euclidienne normalisée entre le vecteur associé à l'occurrence  $i$  d'une voyelle par le locuteur  $k$  et le vecteur associé à l'occurrence  $j$  de la voyelle par un locuteur  $l$ . A cette distance, donnée par la formule 2 ci-dessous, nous avons adjoint un coefficient de défiance  $D_{df}$  composé de trois champs. Les champs  $df_1$  et  $df_2$  sont respectivement les sommes des champs  $df_1$  et  $df_2$  des coefficients de défiance des composantes du vecteur. Le champ  $df_3$  indique qu'un ou plusieurs

termes intervenant dans le calcul de la distance sont nuls en raison de la nullité d'une ou de plusieurs des composantes.

$$(2) \quad D^2(k_i, l_j) = \frac{1}{N} \sum_{n=1}^N \frac{(x_n^{k_i} - x_n^{l_j})^2}{a_n^2}$$

avec :

$N$  : nombre de composantes non nulles de chaque vecteur,

$a_n$  : coefficient normalisateur qui, selon les tests effectués, peut prendre les valeurs suivantes :

- le minimum des deux composantes,
- la valeur de référence du formant de rang  $n$  ou l'écart entre les valeurs de référence, lorsque les composantes sont des écarts entre formants.
- la largeur du domaine de définition du formant de rang  $n$ .

A l'aide de cette distance, nous avons calculé trois indicateurs de pertinence de chacune des voyelles et de chacune des combinaisons formantiques, *Taux* (T), *Score* (S) et *Alpha* (A) :

- *Taux* est le pourcentage de réussite de reconnaissance d'un locuteur parmi dix locuteurs de référence lors de douze expériences de reconnaissance par locuteur. Ces douze expériences ont été obtenues en faisant tourner les différentes répétitions des locuteurs (quatre répétitions de référence et, pour chacune d'elles, trois répétitions "échantillons").

Le locuteur de référence qui minimise la distance au locuteur inconnu est considéré comme reconnu, à condition que l'écart entre cette distance minimale et la deuxième distance la plus faible soit supérieur à un seuil. Sinon, le locuteur reconnu est celui dont la distance a le plus petit coefficient de défiance. Dans les premiers tests, ce seuil a été initialisé à 10% de la distance minimale ;

proximité des formants intermédiaires	formant final fréquence Fi.fr	formant final champ df2	formant final champ df1
3 fréquences proches	moyenne des 3 fréquences	0	somme des 3 $df_1$
3 fréquences proches nulles	0	5	0
2 fréquences proches	moyenne des 2 fréquences	2	somme des 2 $df_1$
2 fréquences proches nulles	0	4	0
2 paires de fréquences proches	moyenne des 3 fréquences ou des 2 meilleures d'entre elles	1	somme des 3 ou des 2 $df_1$
3 fréquences éloignées	0	3	somme des 3 $df_1$

Table 2. Détermination du formant final à partir des trois formants intermédiaires.

- *Score* est le cumul, pour tous les locuteurs et pour les douze expériences de reconnaissance, des rangs auxquels les locuteurs ont été reconnus :
- *Alpha* est un indicateur de type statistique qui estime de façon indirecte le rapport de la variance intralocuteur à la variance interlocuteur. Il est donné par le rapport de la moyenne des distances intralocuteur à la moyenne des distances interlocuteur.

Nous allons présenter dans le paragraphe suivant les valeurs prises par ces trois indicateurs ainsi que le classement des voyelles qui en résulte.

## QUELQUES RESULTATS

### 1 Etude de la combinaison (F1, F2, F3)

La table 3 fournit les valeurs des trois indicateurs de pertinence ainsi que le classement des voyelles dans le cas où un locuteur est représenté par les trois premiers formants des voyelles orales et lorsque le coefficient normalisateur  $a_n$  est donné par le minimum des composantes entrant dans le calcul de la distance.

Ce tableau met en évidence la pertinence de la voyelle / $\epsilon$ / dans la phrase 9, quel que soit l'indicateur. Vient ensuite la voyelle / $\text{œ}$ / dans la phrase 4 qui se

trouve en deuxième position ou en troisième position selon l'indicateur. La voyelle / $\text{ɔ}$ / obtient de bons résultats au sens des indicateurs *Score* et *Alpha* mais réalise un score de reconnaissance très moyen. Nous pouvons noter également la très mauvaise place de / $\text{u}$ / quels que soient l'indicateur, la position de la voyelle dans la phrase ou le contexte antérieur.

Du point de vue général, nous pouvons remarquer que les voyelles situées en fin de phrase ou en fin de groupe de phonation, donc plus accentuées, caractérisent mieux le locuteur que celles situées en début d'un mot lexical ou à l'intérieur d'un mot grammatical ( $\epsilon_{09}$  vs  $\epsilon_{07}$  et  $\text{ɔ}_{04}$  vs  $\text{ɔ}_{02}$ ). En revanche, les rares comparaisons possibles ne nous permettent pas de conclure sur la pertinence du contexte antérieur / $\text{p}$ / par rapport à / $\text{b}$ /.

voyelle	Taux (%)	voyelle	Score	voyelle	Alpha
$\epsilon_{09}$	62	$\epsilon_{09}$	0.78	$\epsilon_{09}$	0.34
$\text{œ}_{04}$	60	$\text{ɔ}_{04}$	0.80	$\text{ɔ}_{04}$	0.35
$\text{œ}_{10}$	56	$\text{œ}_{04}$	0.84	$\text{œ}_{04}$	0.38
$\text{i}_{15}$	55	$\text{œ}_{13}$	0.91	$\text{œ}_{13}$	0.38
$\text{œ}_{13}$	55	$\text{i}_{15}$	0.99	$\text{i}_{15}$	0.40
$\text{ɔ}_{04}$	55	$\text{a}_{03}$	1.17	$\text{œ}_{10}$	0.42
$\text{a}_{03}$	55	$\text{œ}_{10}$	1.24	$\text{ɔ}_{02}$	0.44
$\text{ɔ}_{02}$	53	$\text{ɔ}_{02}$	1.27	$\text{i}_{11}$	0.46
$\text{i}_{11}$	51	$\text{e}_{01}$	1.42	$\text{a}_{03}$	0.47
$\text{u}_{16}$	50	$\text{a}_{16}$	1.53	$\text{e}_{01}$	0.48
$\epsilon_{07}$	47	$\text{i}_{11}$	1.55	$\text{a}_{16}$	0.49
$\text{a}_{16}$	45	$\text{u}_{16}$	1.58	$\text{u}_{16}$	0.53
$\text{e}_{01}$	43	$\text{u}_{12}$	1.84	$\epsilon_{07}$	0.62
$\text{u}_{12}$	43	$\epsilon_{07}$	2.27	$\text{u}_{12}$	0.64
$\text{u}_{08}$	18	$\text{u}_{08}$	3.21	$\text{u}_{08}$	0.72

Table 3. Classement des voyelles pour les trois indicateurs dans un ordre de pertinence décroissante, pour la combinaison (F1, F2, F3).

## 2 Etude toutes combinaisons confondues

La comparaison des indicateurs *Taux* et *Score* entre toutes les combinaisons de formants étudiées montre que la pertinence des voyelles  $\text{\text{E}}_{09}$ ,  $\text{\text{œ}}_{04}$  et  $\text{\text{œ}}_{13}$  pour la caractérisation du locuteur est due aux formants F2 et F3. En effet, comme le montre la table 4, la reconnaissance du locuteur fondée sur la combinaison (F2, F3) est plus performante que celle fondée sur (F1, F2, F3). Les résultats obtenus pour chaque formant (un locuteur étant ainsi représenté par un seul formant) mettent en évidence la prépondérance de F3 sur F2 (24% à 33% pour F2 contre 37% à 49% pour F3).

Pour les combinaisons formées à partir des écarts entre formants, les taux de reconnaissance sont moins bons que ceux obtenus lors de l'utilisation des formants alors que les indicateurs *Score* et *Alpha* conservent à peu près les mêmes valeurs. Bien que  $\text{\text{E}}_{09}$  occupe toujours une bonne place, la voyelle  $\text{\text{œ}}_{04}$  est globalement la plus pertinente quelle que soit sa position dans la phrase. De plus, la table 4 indique un taux de reconnaissance de 70% pour l'occurrence  $\text{\text{E}}_{04}$ . Notons également la relativement bonne pertinence de la voyelle  $\text{\text{e}}$  qui serait due, d'après l'étude individuelle des écarts, à la différence F2-F1.

combinaison	voyelle	Taux (%)	combinaison	voyelle	Taux (%)
F2 F3	$\text{\text{E}}_{09}$	63	F3-F1 F2-F1 F3-F2	$\text{\text{œ}}_{04}$	72
F2 F3	$\text{\text{œ}}_{04}$	62	F2-F1 F3-F2	$\text{\text{œ}}_{04}$	70
F1 F2 F3	$\text{\text{E}}_{09}$	62	F3-F1 F3-F2	$\text{\text{œ}}_{04}$	68
F2 F3	$\text{\text{œ}}_{13}$	60	F3-F1 F2-F1	$\text{\text{œ}}_{13}$	55
F1 F2 F3	$\text{\text{œ}}_{04}$	60	F2-F1 F3-F2	$\text{\text{E}}_{09}$	54
F1 F2 F3	$\text{\text{œ}}_{13}$	56	F3-F1	$\text{\text{u}}_{12}$	53

Table 4. Les meilleures voyelles au sens de l'indicateur *Taux* pour les combinaisons de formants et pour les combinaisons d'écarts entre formants.

## 3 CONCLUSION

Nous sommes en train d'effectuer une étude sur la pertinence de certains paramètres phonétiques et acoustiques pour la caractérisation du locuteur. Nous avons commencé par l'étude des formants des voyelles  $\text{\text{E}}$ ,  $\text{\text{e}}$ ,  $\text{\text{œ}}$ ,  $\text{\text{ɔ}}$ ,  $\text{\text{a}}$ ,  $\text{\text{i}}$  et  $\text{\text{u}}$ , précédées d'un contexte neutre au sens de la coarticulation linguale,  $\text{\text{p}}$  ou  $\text{\text{b}}$  et suivies d'un contexte postérieur allongé,  $\text{\text{R}}$ . Pour cela, nous avons développé une méthode automatique de détermination de formants fiables. Nous avons présenté quelques résultats mettant en évidence la pertinence de certaines voyelles et de certaines combinaisons de formants. Notre objectif est de vérifier ces résultats en faisant varier les conditions d'expérimentation (normalisation, coefficient de défiance, ...), puis de compléter l'étude en remplaçant les formants mis à 0 pour cause de non-fiabilité par des valeurs déterminées à partir des spectrogrammes.

## BIBLIOGRAPHIE

[CAE, 88] G. Caelen et G. Pérennou, "Phoneticophonological, Prosodic and Frequential Analyses in a Both Global and Local Approach to Speaker Identification and Verification", Bull. L.C.P., vol. 2, pp 425-455, 1988.

[CAL, 89] Calliope, "La Parole et son Traitement Automatique", Collection CNET-ENST, Masson, pp. 79-120, Paris 1989.

[GON, 92] Y. Gong and J.P. Haton. "Non-linear vectorial interpolation for speaker recognition", Proceedings IEEE-International Conference on ASSP, San Francisco, USA, March 1992.

[LAP, 88] Y. Laprie, "SNORRI : un système d'étude interactif de la parole", Actes des 17<sup>e</sup> journées d'Etude sur la Parole, pp. 71-76, Nancy, septembre 1988.

[MAR, 76] J.D. Markel and A.H. Gray Jr., "Linear Prediction of Speech", Springer Verlag, New-York, USA, 1976.

[MEL, 89] O. Mella et M.C. Haton, "Méthodologie d'étude de la pertinence de paramètres phonétiques et acoustiques pour la reconnaissance du locuteur", Séminaire sur la variabilité et la spécificité des locuteurs, Marseille Luminy, juin 1989.

[MON, 83] R.B. Mosen and A.M. Engebretson, "The accuracy of formant frequency measurements: a comparison of spectrographic analysis and linear prediction", Journal of Speech and Hearing Research, 26(1), pp. 89-96, 1983.

[ROS, 76] A. Rosenberg, "Automatic Speaker Verification: A Review", Proc. IEEE, vol. 64, pp. 475-487, April 1976.