

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

**Ministère de l'Enseignement Supérieure et de la Recherche Scientifique**

**Université des Sciences et de la Technologie Houari Boumediene**

**Faculté d'Electronique et Informatique**



Mémoire de Magister en Electronique

**Spécialité** Communication Parlée

Présenté par : Mr. Bouferroum Othmane

THEME

**Etude du locus de consonnes arabes**

Soutenu publiquement, le 09/12/2010 devant le jury composé de :

|                      |                                    |                    |
|----------------------|------------------------------------|--------------------|
| <b>Mr-B. Boudraa</b> | Maître de conférence/A, à L'USTHB. | Président          |
| <b>Mr-A. Djeradi</b> | Professeur, à L'USTHB.             | Directeur de thèse |
| <b>Mme-A. Oukil</b>  | Maître de conférence/A, à L'USTHB. | Examineur          |
| <b>Mme-L. Falek</b>  | Maître de conférence/A, à L'USTHB. | Examineur          |
| <b>Mr-H. Teffahi</b> | Maître de conférence/A, à L'USTHB. | Examineur          |

à la mémoire de mes parents

## **Remerciements**

Le travail présenté dans ce mémoire a été effectué au laboratoire de communication parlée et traitement de signal (LCPTS) de la faculté d'Electronique et d'Informatique de l'USTHB

Je remercie Dr. L. Falek pour m'avoir bien introduit dans le domaine de la coarticulation et du locus et aussi pour m'avoir aidé dans l'enregistrement du corpus audio.

Je remercie Madame Djeradi pour m'avoir fourni de la documentation et les encouragements tout au long de ma recherche.

Je remercie Monsieur le Professeur A. Djeradi pour m'avoir proposé d'étudier le locus par l'analyse discriminante et pour avoir corrigé la rédaction.

# SOMMAIRE

|                                                              |    |
|--------------------------------------------------------------|----|
| Introduction.....                                            | 10 |
| Chapitre I. Etat de l’art                                    |    |
| 1. Introduction .....                                        | 13 |
| 2. Théorie du locus.....                                     | 14 |
| 3. Droite de locus.....                                      | 20 |
| 4. Droite de locus et coarticulation .....                   | 23 |
| 5. Droite de locus et lieu d’articulation .....              | 25 |
| Conclusion .....                                             | 27 |
| Chapitre II. Etude expérimentale                             |    |
| 1. Introduction .....                                        | 28 |
| 2. Matériau linguistique .....                               | 29 |
| 3. Enregistrement des phrases .....                          | 29 |
| 4. Analyse et mesure des caractéristiques formantiques. .... | 30 |
| 5. Résultat.....                                             | 30 |
| - Détermination des droites de locus .....                   | 30 |
| - Détermination du locus .....                               | 35 |
| - Locus, droite de locus et segmentation .....               | 39 |
| Conclusion .....                                             | 45 |
| Chapitre III Analyse discriminante                           |    |
| 1. Introduction .....                                        | 46 |
| 2. Aspect théorique .....                                    | 46 |
| 3. élaboration des vecteurs de représentation.....           | 47 |
| 4. Algorithme de Bayes .....                                 | 49 |
| 5. Résultats et interprétations .....                        | 51 |

|                                                           |    |
|-----------------------------------------------------------|----|
| - Analyse des résultats .....                             | 52 |
| - Variabilité du 3 <sup>e</sup> formant .....             | 57 |
| - Interaction entre phonèmes .....                        | 59 |
| 6. Espace de redescription .....                          | 61 |
| - Introduction .....                                      | 61 |
| - Construction de l'espace RKHS .....                     | 62 |
| - Algorithme de Bayes dans l'espace de redescription..... | 66 |
| Conclusion et perspectives .....                          | 69 |
|                                                           |    |
| Annexe 1 : Outil d'analyse                                |    |
| 1. Introduction .....                                     | 70 |
| 2. L'oscillogramme .....                                  | 70 |
| 3. La fenêtre du signal. ....                             | 71 |
| 4. Le spectrogramme .....                                 | 72 |
| 5. Suivi de formants par prédiction linéaire .....        | 72 |
| 6. Procédure de mesure.....                               | 73 |
|                                                           |    |
| Bibliographie .....                                       | 76 |

## TABLEAUX

|                                                                                               |    |
|-----------------------------------------------------------------------------------------------|----|
| <b>Tableau 1.</b> Locus des fricatives américaines.....                                       | 20 |
| <b>Tableau 2.</b> Paramètres caractéristiques des droites de locus.....                       | 31 |
| <b>Tableau 3.</b> Paramètres caractéristiques des droites de locus par locuteur.....          | 34 |
| <b>Tableau 4.</b> Locus déterminé par convergence de F2, F2+F3, et par $S=1/(1-a)$ .....      | 39 |
| <b>Tableau 5.</b> Intersection des droites de locus de [ʃ], [q] et [h].....                   | 41 |
| <b>Tableau 6.</b> Résultat de classification en utilisant l'algorithme de Bayes .....         | 51 |
| <b>Tableau 7.</b> Résultat de classification binaire en utilisant l'algorithme de Bayes ..... | 51 |
| <b>Tableau 8.</b> Résumé de la contribution du 3 <sup>e</sup> formant.....                    | 57 |
| <b>Tableau 9.</b> Variances des composantes (b2, b3) du vecteur de représentation.....        | 58 |
| <b>Tableau 10.</b> Résultat de classification dans l'espace de redescription.....             | 68 |

## TABLE DES FIGURES

|                                                                                                                                                                                                                |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 1.</b> Transitions entre cibles de phonèmes .....                                                                                                                                                    | 10 |
| <b>Figure 2.</b> Principe de fonctionnement du synthétiseur « Pattern play-back ».....                                                                                                                         | 14 |
| <b>Figure 3.</b> Schémas spectrographiques utilisés pour déterminer l'effet de la position<br>du bruit dans la perception des occlusives sourdes.....                                                          | 15 |
| <b>Figure 4.</b> Schémas spectrographiques de la voyelle [a] indiquant l'effet auditif qui<br>résulte des changements dans les transitions du second formant pour<br>la perception du lieu d'articulation..... | 16 |
| <b>Figure 5.</b> Transitions formantiques de F2 produisant les occlusives [b], [d] et [g].....                                                                                                                 | 17 |
| <b>Figure 6.</b> Frontières entre [b], [d] et [g] sous l'effet de F3.....                                                                                                                                      | 18 |
| <b>Figure 7.</b> Gamme de variation des fréquences des formants droits utilisée pour la<br>détermination du locus .....                                                                                        | 20 |
| <b>Figure 8.</b> Transitions CV du 2 <sup>e</sup> formant et droite de locus.....                                                                                                                              | 21 |
| <b>Figure 9.</b> Droite de locus pour [g].....                                                                                                                                                                 | 22 |
| <b>Figure 10.</b> Variabilité du locus classique pour [b], [d] et [g].....                                                                                                                                     | 23 |
| <b>Figure 11.</b> Trajectoires F2 pour une coarticulation minimale et une coarticulation<br>maximale.....                                                                                                      | 24 |
| <b>Figure 12.</b> Trajectoires F2 de syllabes /dvd/ synchronisées à l'onset et à l'offset.....                                                                                                                 | 25 |
| <b>Figure 13.</b> Configuration du conduit vocal dans la production de [h] et [q].....                                                                                                                         | 28 |
| <b>Figure 14.</b> Position de la langue dans la production des voyelles [i], [a] et [u].....                                                                                                                   | 29 |
| <b>Figure 15.</b> Spectrogrammes des 3 sons [ɿ], [q] et [h].....                                                                                                                                               | 30 |
| <b>Figure 16.</b> Droites de locus de [ɿ], [q] et [h] pour l'ensemble des locuteurs.....                                                                                                                       | 31 |
| <b>Figure 17.</b> Droites de locus par locuteur de [ɿ], [q] et [h].....                                                                                                                                        | 33 |
| <b>Figure 18.</b> Droites de locus dans le plan pente - ordonnée à l'origine.....                                                                                                                              | 35 |
| <b>Figure 19.</b> Détermination graphique du locus à partir de la droite de locus.....                                                                                                                         | 35 |
| <b>Figure 20.</b> Relation locus - pente par locuteur pour les sons [ɿ], [q] et [h].....                                                                                                                       | 35 |
| <b>Figure 21.</b> Erreur projetée sur le locus.....                                                                                                                                                            | 36 |
| <b>Figure 22.</b> Interpolation linéaire d'une transition /CV/ .....                                                                                                                                           | 37 |
| <b>Figure 23.</b> Convergence vers le locus des sons [ɿ], [q] et [h].....                                                                                                                                      | 38 |
| <b>Figure 24.</b> Relation locus - pente par locuteur pour les sons [ɿ], [q] et [h] obtenue<br>en utilisant la méthode de convergence en D2.....                                                               | 38 |

|                                                                                                                                                  |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 25.</b> Relation locus - pente par locuteur pour les sons [ɿ], [q] et [h] obtenue<br>en utilisant la méthode de convergence en D3..... | 38 |
| <b>Figure 26.</b> Transitions F2 dans une syllabe /CV/.....                                                                                      | 40 |
| <b>Figure 27.</b> Droites de locus pour différentes segmentations.....                                                                           | 41 |
| <b>Figure 28.</b> Intersection des droites de locus par locuteur.....                                                                            | 42 |
| <b>Figure 29.</b> Transition F2 vers deux locus différents. ....                                                                                 | 43 |
| <b>Figure 30.</b> La régression linéaire.....                                                                                                    | 46 |
| <b>Figure 31.</b> Surfaces régressives de [ɿ], [q] et [h] .....                                                                                  | 48 |
| <b>Figure 32.</b> Histogrammes des composantes du vecteur de représentation.....                                                                 | 49 |
| <b>Figure 33.</b> Résultat de discrimination entre [ɿ] et [h] en utilisant F2 seul.....                                                          | 52 |
| <b>Figure 34.</b> Résultat de discrimination entre [ɿ] et [q] en utilisant F2 seul.....                                                          | 53 |
| <b>Figure 35.</b> Résultat de discrimination entre [h] et [q] en utilisant F2 seul.....                                                          | 54 |
| <b>Figure 36.</b> Contribution du 3 <sup>e</sup> formant dans la discrimination entre [ɿ] et [h].....                                            | 56 |
| <b>Figure 37.</b> Contribution du 3 <sup>e</sup> formant dans la discrimination entre [ɿ] et [q].....                                            | 56 |
| <b>Figure 38.</b> Contribution du 3 <sup>e</sup> formant dans la discrimination entre [q] et [h] .....                                           | 56 |
| <b>Figure 39.</b> Intersection des droites de locus.....                                                                                         | 57 |
| <b>Figure 40.</b> Variabilité de F3 par rapport à la droite de locus F2.....                                                                     | 59 |
| <b>Figure 41.</b> Droites de locus de [b], [d] et [g].....                                                                                       | 60 |
| <b>Figure 42.</b> Ellipses de dispersion de (F2 <sub>onset</sub> , F3 <sub>onset</sub> ) de [d] et [g] avec voyelles arrondies.....              | 60 |
| <b>Figure 43.</b> Transformation en utilisant le noyau gaussien.....                                                                             | 65 |
| <b>Figure 44.</b> Oscillogramme de « men saqila ».....                                                                                           | 71 |
| <b>Figure 45.</b> Spectrogramme du /VCV/ /aqi/.....                                                                                              | 72 |
| <b>Figure 46.</b> Visualisation de toutes les répétitions dans le fichier son « aqí6 ».....                                                      | 73 |
| <b>Figure 47.</b> Visualisation d'une seule répétition.....                                                                                      | 74 |



## GLOSSAIRE

**Burst** : bruit d'explosion dans la production des occlusives

**CV** : consonne - voyelle

**Formants** : fréquences de résonance qui correspondent aux cavités dans le conduit vocal.

**F1** : 1<sup>er</sup> formant

**F2** : 2<sup>nd</sup> formant

**F3** : 3<sup>ème</sup> formant

**Offset** : fin de la voyelle avant la consonne

**Onset** : début de la voyelle après la consonne..

**RKHS (Reproducing Kernel Hilbert Space)**: Espace de Hilbert à Noyau Reproducteur

**Transitions formantiques** : changement rapide de la fréquence des formants lors de la production des consonnes.

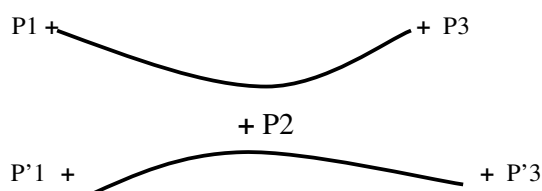
**VCV** : voyelle - consonne - voyelle

**Voisement** : vibration des cordes vocales

## Introduction

Lorsque l'on écoute la parole on a non seulement l'impression qu'elle est composée d'un enchaînement de mots mais que chacun de ces mots est formé d'une suite de sons distincts, que l'on a appelés *phonèmes*, de la même façon que l'écriture l'est de graphèmes successifs. Ceci a amené les chercheurs dans les années 1950s à réaliser une machine qui lit pour les aveugles, un convertisseur « texte to speech » basé sur un alphabet de sons phonémiques. Plusieurs alphabets de sons, de forme acoustique fixe, bien conçus ont été essayés et malgré les durs entraînements pour de longues périodes, on a pas pu dépasser le dixième de l'efficacité (en débit) de la parole. L'être humain peut parler avec un débit maximum de 400 mots par minutes (Friedman & Williams 1965)[1,2], ce qui donne une moyenne, pour la langue anglaise, de 30 phonèmes par seconde. On connaît par la psychophysique de l'audition (Miller & Taylor 1948)[1,3], que ceci dépasse largement la résolution temporelle de l'oreille. Même avec 15 phonèmes par seconde qui est très rare dans une conversation, l'oreille n'arriverait pas à se débrouiller bien si la parole était une chaîne de sons phonémiques discrets.

Au niveau de la production, chaque phonème est considéré comme une cible que les articulateurs essayent d'atteindre. Les organes de la parole ne peuvent pas transiter d'une certaine configuration cible à une autre instantanément. Chaque articulateur évolue de façon continue entre les positions articulatoires. Au lieu de donner à chaque phonème une configuration invariante et prendre tout le temps nécessaire pour transiter d'une cible à l'autre, c'est une course rapide à travers la séquence qui est observée. Dès que la cible soit suffisamment approchée de sorte que ce soit intelligible à l'auditeur, les articulateurs changent de destination vers la cible suivante et ainsi de suite. Donc, le phonème pourrait être approché de différents cotés selon la position des phonèmes adjacents, voir figure 1.



**Figure 1.** Transitions entre cibles de phonèmes

Le phonème P2 est approché selon deux trajectoires différentes en fonction des cibles des phonèmes adjacents P1 P3 ou bien P'1 P'3. Donc une variabilité articulatoire et par conséquent acoustique due à l'influence des segments adjacents ou proches est observée. C'est ce que nous appelons: *coarticulation*. Le fait que l'influence d'un segment souvent s'étend au delà de ses frontières, l'information sur ce segment est disponible au niveau de la perception plus longtemps que ne serait le cas si les indices étaient confinés à l'intérieur de ses frontières. Ceci permet le traitement parallèle de l'information au niveau de la perception. *La fonction de coarticulation permet donc de véhiculer l'information phonémique d'une manière rapide et fiable*. Ceci augmente le débit certes mais au détriment d'une relation complexe entre le phonème et sa représentation par l'information acoustique.

Le problème se pose : comment identifier les phonèmes à partir du signal acoustique dont la forme change avec le contexte. Cette question de l'invariance, issue d'une longue tradition, a donné lieu à de nombreux travaux visant à déterminer les paramètres descripteurs permettant une distinction des unités perçues. Malgré des avancées spectaculaires dans le dégagement de règles fixes et reproductibles concernant le rôle des traits et indices acoustiques dans la discrimination phonétique, il n'existe toujours pas de modèle général du fonctionnement de la perception de la parole qui ferait l'unanimité parmi les chercheurs.

Deux tendances principales existent. A la théorie motrice de Lieberman et Mattingly (1985)[4], selon laquelle l'invariance est de nature articulatoire, s'oppose la théorie acoustico-auditive de Stevens (1996)[5] qui postule que les informations acoustiques spécifient l'identité phonémique. Le modèle du *locus* est la forme la plus élaborée du modèle acoustico-auditive de la perception. Selon ce modèle, d'abord initié par Delattre, Liberman et Cooper (1955), les transitions du second formant dans les différents contextes vocaliques d'une syllabe /CV/ pointent vers un point (une fréquence), appelé *locus*, considéré comme un invariant caractérisant le lieu d'articulation. Par la suite, d'autres travaux ont montré que le locus n'était pas stable. En 1991 Sussman a reformulé le modèle. Au lieu d'un point fixe (locus), c'est une relation linéaire (*droite de locus*) entre le début de transition du deuxième formant ( $F2_{\text{onset}}$ ) des différentes voyelles suivant la consonne et la fin de transition ou leurs états stables correspondant ( $F2_{\text{voyelle}}$ ). Cette droite de locus est supposée être un invariant pour chaque lieu d'articulation. Ce modèle aussi a reçu plusieurs critiques. En plus, ces droites s'entrecoupent dans une certaine mesure, donc elles ne sont que partiellement distinctives. Dans ces régions d'intersection le 3<sup>e</sup> formant ( $F3_{\text{onset}}$ ) semble apporter l'information

distinctive. C'est dans ce contexte que se situe notre travail qui consiste à étudier le locus des consonnes arabes [ʔ] (ع), [q] (ق), et [h] (ه). Dans une première étape, nous avons déterminé les droites de locus ainsi que les caractéristiques pouvant être utilisées pour discriminer ces trois sons. Notre démarche a consisté à faire des enregistrements d'un corpus audio, puis à mettre au point un outil d'analyse de ces caractéristiques. Dans une seconde étape, nous avons réalisé un classifieur en tenant compte de l'apport du troisième formant.

Dans le premier chapitre nous avons présenté quelques travaux entrepris dans ce domaine et nous avons posé notre problématique. Dans le deuxième chapitre, nous avons décrit les procédures utilisées, partant de l'étape d'enregistrement du corpus audio jusqu'au tracé des droites de locus. Une comparaison des résultats avec ceux trouvés dans la littérature a été donnée ainsi qu'une première interprétation de ces résultats. Aussi, nous avons présenté une analyse détaillée du locus et l'importance de la droite de locus. Au troisième chapitre, nous avons présenté la méthode mise au point concernant l'analyse discriminante des consonnes étudiées. Après une description de la régression linéaire, nous avons présenté l'algorithme de Bayes. Nous pensons que notre analyse dans un espace RKHS (Reproducing Kernel Hilbert Space), de dimension plus élevée, fournirait de meilleurs résultats. Des interprétations sont données et enfin nous avons terminé par une conclusion avec des perspectives.

# Chapitre I

## Etat de l'art

### 1. Introduction

La parole se distingue des autres sons par des caractéristiques acoustiques ayant leurs origines dans les mécanismes de production. Le conduit vocal est un tube de forme complexe constitué essentiellement du larynx avec les cordes vocales, du pharynx, des cavités buccales et nasales. Et il peut varier considérablement en fonction des mouvements de la langue, du voile du palais et des lèvres. Un écoulement d'air peut rencontrer à tous les points de son trajet un rétrécissement ou une fermeture momentanés susceptibles de donner naissance à un bruit. Ainsi sont fabriquées les consonnes constrictives (rétrécissement) et les consonnes occlusives (fermetures). Chacune de ces productions pouvant être accompagnée (consonnes sonores) ou non (consonnes sourdes) de la vibration des cordes vocales. Celles-ci, lors du *voisement*, s'ouvrent et se ferment rapidement pour fragmenter la colonne d'air en bouffées successives. Si un son est propulsé de la sorte mais que le passage reste ensuite ouvert, on parlera de voyelles ou de semi-voyelles s'il se produit un léger resserrement.

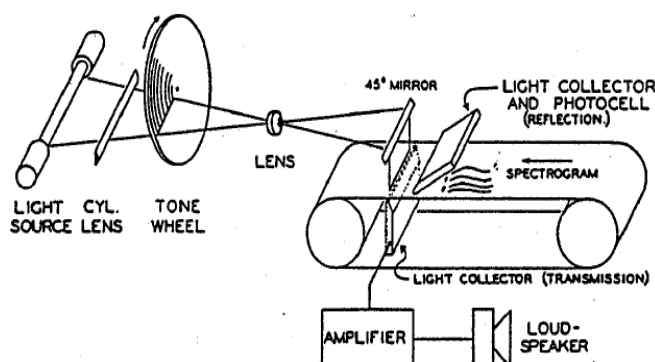
Le signal de parole émis au niveau des cordes vocales est complexe. Il va être modifié en passant par le conduit vocal. Les cavités résonantes de ce dernier renforcent dans le spectre sonore les harmoniques dont la fréquence est proche de la leur. Les zones de fréquence ainsi renforcées sont appelées *formants* et ce sont les éléments qui permettent la reconnaissance et l'identification des sons.

Lors de la production d'une voyelle, le conduit vocal se modifie très peu. La fréquence des formants des voyelles que l'on veut produire est donc relativement stable. Par contre, la forme de ce conduit vocal, lors de l'émission de certaines consonnes, évolue rapidement, ce qui produit un changement de la fréquence des formants propres aux voyelles qui les précèdent et qui les suivent. Ces phases transitionnelles que l'on a appelées *transitions formantiques* sont caractéristiques de la consonne prononcée. Dans le but de comprendre les effets de la coarticulation sur le signal de parole, l'étude des transitions formantiques a ouvert

la voie à une série de recherches acoustiques prometteuses, notamment le concept du *locus* et de la *droite de locus*.

## 2. Théorie du locus

Les transitions formantiques sont des éléments essentiels dans la perception de la parole. La synthèse spectrographique de la parole l'a confirmée. Ceci a été rendu possible par l'emploi de formes spectrographiques peintes à la main et transformées en son par un synthétiseur de parole nommé « *pattern play-back* » [6], voir figure 2.



**Figure 2.** Principe de fonctionnement du synthétiseur « Pattern play-back » [6].

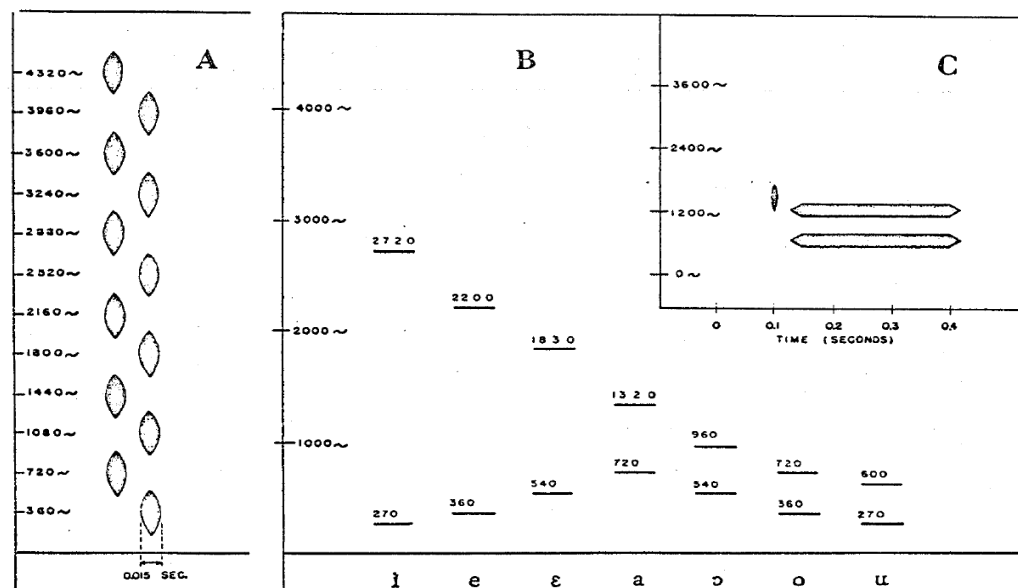
Ce synthétiseur peut reproduire des photocopies de spectrogrammes réels ou bien des modèles artificiels recomposés à la main avec une peinture blanche sur une base d'acétate de cellulose. La production des sons se fait par l'intermédiaire d'un faisceau lumineux modulé par une roue tonale et réfléchi par les parties blanches du spectrogramme. Celui-ci produit de cette manière un son complexe composé d'une fréquence fondamentale de 120 Hz et pouvant comporter les 50 premières harmoniques de ce son.

Maintenant qu'il est devenu possible de transformer en son des spectrogrammes artificiels, une série d'expérimentations a permis l'évaluation par l'oreille des effets d'une grande variété de modifications dans le schéma acoustique. La méthode utilisée consiste, d'abord à examiner des spectrogrammes ordinaires, essayer de reproduire visuellement les traits caractéristiques et en faire des modèles artificiels, et ensuite modifier ces modèles, par

de multiples tâtonnements, afin d'obtenir à l'audition des sons intelligibles, semblables à ceux de la parole réelle. Ceci est dans le but de simplifier encore davantage le schéma acoustique. En d'autres termes, le signal de parole sera réduit juste à ses composantes phonémiques.

L'avantage de la parole artificielle est qu'elle permette d'isoler à tour de rôle chacun des divers facteurs de la perception d'un phonème. Par exemple, pour une consonne donnée, on élimine le bruit et on laisse les transitions formantiques et vice versa. Et de cette façon, on détermine lequel des indices acoustiques a plus de poids dans la perception.

Dans cette perspective, le bruit d'explosion (*burst*) émis dans la production des occlusives [p], [t] et [k] a été isolé; afin de voir si cela pouvait servir pour distinguer entre ces trois occlusives sourdes [7]. Pour cela une série de schémas spectrographiques (bruit + voyelle) a été préparée, voir figure 3.



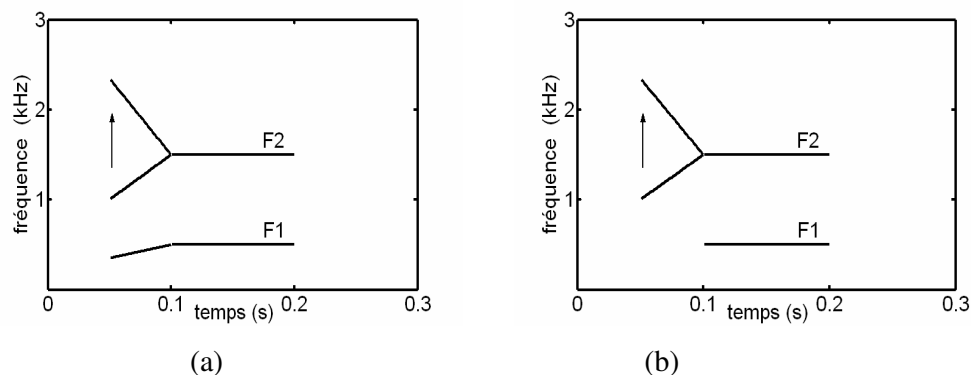
**Figure 3.** Schémas utilisés pour déterminer l'effet de la position du bruit dans la perception des occlusives sourdes. (A) positions des fréquences des bruits utilisés. (B) formants des voyelles (C) une des 84 syllabes utilisées (Cooper 1952)[7].

En passant ces schémas dans le synthétiseur, le [t] a été reconnu lorsque le bruit était à des fréquences élevées supérieures à 3000 Hz, et cela indépendamment de la voyelle qui suit. Avec un bruit de fréquences dans la bande 300-3000 Hz, la perception de [p] et [k] dépendait

non seulement du bruit mais aussi de la position du bruit relativement à la voyelle. En d'autres termes, *la perception nécessite une syllabe CV comme une unité acoustique minimale*.

L'idée que les transitions formantiques fournissent des indices pour le lieu d'articulation des consonnes remonte à Potter, Kopp et Green (1947) [8]. Dans leur discussions au sujet des traits caractéristiques qui peuvent être utilisés pour lire les spectrogrammes, ils ont noté que les transitions entre occlusive et voyelles, spécialement dans le second formant, sont similaires pour chaque paire [b]-[p], [d]-[t] ou bien [g]-[k].

Pour voir si les transitions formantiques, en fait, peuvent être utilisées seules pour distinguer entre les occlusives; la voyelle [a] a été choisie avec une variété de transitions aux deux formants de la voyelle, voir figure 4a. La transition du premier formant (F1) est fixe dans la position qui convient à toutes les occlusives. Celle du deuxième formant (F2) change en pointant vers les différentes fréquences de 0.6 kHz à 3 kHz. Lorsque cette série de schémas spectrographiques passe dans la machine à synthétiser la parole, le son change légèrement à chaque nouveau schéma, commençant par /ba/, passant par /da/ et aboutissant à /ga/.



**Figure 4.** Schémas spectrographiques de la voyelle [a] indiquant l'effet auditif qui résulte des changements dans les transitions de F2 pour la perception du lieu d'articulation. (a): [b], [d] et [g]. (b) : [p], [t], [q].

En éliminant la transition du premier formant (F1), on obtient la figure 4b. En synthétisant ces schémas, le son change aussi à chaque nouveau schéma mais commençant cette fois ci par /pa/, passant par /ta/ et aboutissant à /ka/. Pour l'oreille, seul le lieu d'articulation change. Le lieu d'articulation est donc relié à la direction de la transition du deuxième formant.

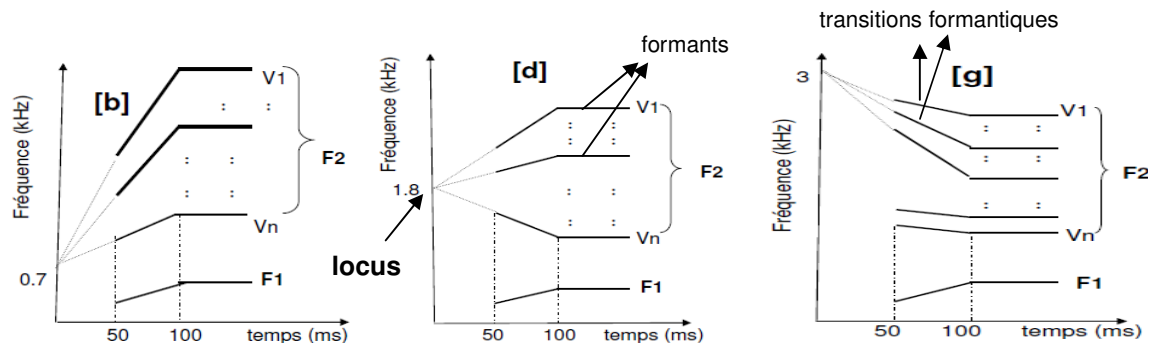


Une autre expérience similaire a été réalisée pour voir si les mêmes transitions du second formant pouvaient servir aussi pour distinguer entre les nasales [m], [n] et [ɱ] qui sont reliées aux occlusives par le lieu d'articulation respectivement. L'occlusion buccale est donc, bilabiale pour [m], [p], [b], alvéolaire pour [n], [t], [d] et vélaire pour [ɱ], [k], [g].

La distinction des sons perçus n'était pas aussi bonne comme dans le cas des occlusives mais en général les résultats sont similaires. Les transitions qui ont servi pour la synthèse de [b] et [p] ont fait entendre le [m], Ceux pour [d] et [t] ont fait entendre le [n] et finalement ceux pour [g] et [k] ont fait entendre le [ɱ].

*Le lieu d'articulation est donc relié à la direction de la transition du second formant*

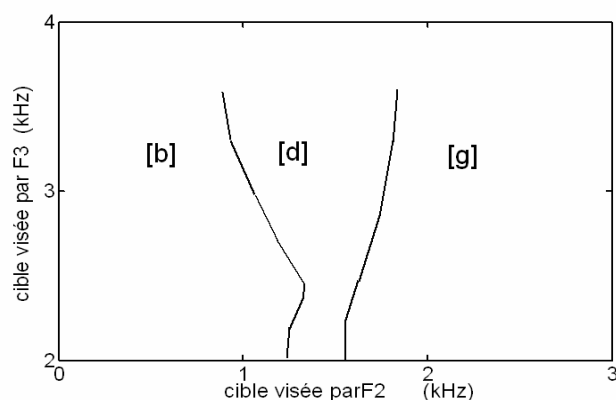
Si maintenant la voyelle change, toutes les transitions qui font entendre un même lieu d'articulation visent un même point [9]. Dans la figure 5, toutes les transitions font entendre un [d] et toutes visent un point dont la fréquence est à quelque 1800 Hz. Bien que très différentes les unes des autres, ces transitions ont donc un invariant en commun, ce point qu'elles visent et auquel on a donné le nom de **locus**. Les transitions du second formant de [b] visent un point de fréquence assez basse (700 Hz). Dans le cas de [g], elles fonctionnent selon qu'elles sont suivies de voyelles écartées ou de voyelles arrondies. Devant les voyelles écartées les transitions F2 visent un point de fréquence élevée égale à 3000 Hz qui est leur locus. Mais devant les voyelles arrondies les transitions F2 ne visent plus ce point.



**Figure 5.** Transitions formantiques du second formant produisant les occlusives [b], [d] et [g]

Il a été observé que la durée du silence nécessaire pour faire entendre la meilleure qualité du son est en moyenne de 50 % de la durée totale entre le locus et l'état stable de la voyelle qui est de 100 ms pour un débit moyen de la parole. Pour voir ce qui pourrait se passer avec les autres débits, l'expérience a été répétée avec des intervalles entre le locus et l'état stable de la voyelle de 40, 60, 80, et 120 ms. Pour les durées de 80 et 120ms, les résultats étaient similaires. Pour le cas des débits rapides dont l'intervalle total est de 40 et 60 ms, la qualité du son obtenu n'était pas bonne pour tout intervalle de silence.

La transition du 3<sup>e</sup> formant a aussi un rôle à jouer dans la perception des occlusives. Ceci a été montré dans les travaux de Harris et ses collaborateurs (1957) [10]. Dans leur expérience pour synthétiser [b], [d] et [g] en utilisant les trois premiers formants, deux problèmes majeurs ont été rencontrés affectant ainsi la procédure utilisée et limitant la nature des conclusions qu'on pouvait espérer. La première difficulté rencontrée est comment fixer la valeur du 3<sup>e</sup> formant de la voyelle qui suit. Ceci ne posait pas problème dans l'étude des transitions du 2<sup>e</sup> formant puisque la valeur du 2<sup>e</sup> formant des voyelles est déterminée avec précision. Le 3<sup>e</sup> formant a peu ou pas d'effet sur l'identité phonémique de la voyelle. Pour cela, la valeur du 3<sup>e</sup> formant qui a été trouvée par Peterson et Barney [10,11] typique pour les voyelles américaines, a été choisie.



**Figure 6.** Frontières entre [b], [d] et [g] sous l'effet de F3.  
Transitions avec voyelle [æ] (Harris 1957) [10].

Un second problème est la difficulté d'isoler la transition du 3<sup>e</sup> formant afin d'obtenir une mesure directe de son effet. La difficulté est que le 3<sup>e</sup> formant ne pourrait être présenté sans les deux premiers formants. Le 1<sup>er</sup> formant ne pose pas problème puisque sa transition

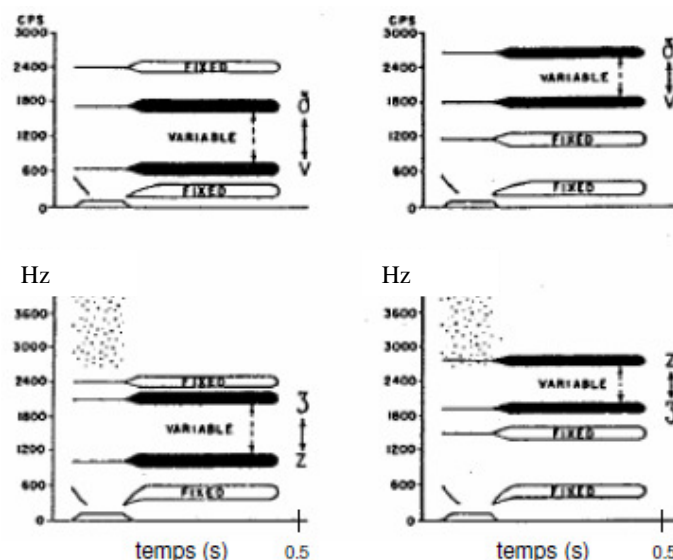
n'affecte pas la distinction entre [b], [d] et [g]. Cependant les transitions du 2<sup>e</sup> formant sont des traits acoustiques potentiels pour distinguer entre ces sons. Même si ces transitions sont droites elles contribuent appréciablement à la perception de [b], [d] et [g]. Donc, il était nécessaire de combiner une transition du 3<sup>e</sup> formant avec les différentes transitions du second formant. Ceci ne permet pas de savoir directement et exactement, comme il aurait été souhaité, combien est la contribution du 3<sup>e</sup> formant relativement au 2<sup>e</sup> formant. Tout d'abord, il est nécessaire de mentionner que d'autres travaux [12] ont montré que la perception des occlusives est catégorielle. Il n'y a pas de sons intermédiaires entre [b] et [d] ou bien entre [d] et [g]. Comme on le voit encore dans la figure 6, qui résume les résultats obtenus. L'effet de F2 et F3 combinés change tout simplement les positions des frontières b-d et d-g.

Ce qui a facilité l'étude des transitions pour les occlusives, c'est le fait que, à l'exception des vélaires devant voyelles arrondies, leurs lieux d'articulation puissent se reconnaître, à la synthèse, sans aucun appui du bruit d'explosion.

Pour les consonnes fricatives américaines [f], [θ], [s], [ʃ], [v], [ð], [z] et [ʒ], l'identification du lieu d'articulation dépend du bruit de friction aussi bien que des transitions des deuxième et troisième formants [13]. Pour isoler ces derniers indices des premiers, il a fallu neutraliser le bruit de friction. Cela s'est fait en ne manipulant que les fricatives sonores en opposant [v] à [ð] sans aucune friction et une friction neutre quand il s'agit d'opposer [z] à [ʒ]. Dans cette étude les quatre fricatives sonores étaient suivies chacune de trois voyelles différentes, une voyelle antérieure, une centrale et une postérieure.

Le premier formant est fixé à une valeur arbitraire qui n'interfère pas avec les variations du second formant et qui a une transition appropriée à la classe des fricatives. Pour chaque paire de fricative, le second formant varie tandis que le troisième formant est fixé à une valeur neutre. Ensuite, c'est le troisième formant qui varie tout en fixant le second formant à une valeur neutre. En passant donc ces schémas spectrographiques, voir figure 7, dans le synthétiseur, les fricatives sont reconnues d'une façon claire seulement lorsque les valeurs des formants sont proches du locus. En admettant que les locus des fricatives sonores sont également valides pour les fricatives sourdes correspondantes, les locus des 2<sup>e</sup> et 3<sup>e</sup> formant des fricatives américaines pour un locuteur masculin ont approximativement les fréquences montées dans le tableau 1.

En résumé, c'est donc de la direction des transitions F2 et F3 par rapport à un locus que dépend la perception du lieu d'articulation consonantique. Durant les années 1960s-1980s différentes études acoustiques (Lehiste & Peterson, 1961 ; Öhman, 1966 ; Fant, 1973 ; Kewley-Port, 1982) ont essayé de vérifier les valeurs du locus dans la parole naturelle. A part le [d] où une grande convergence a été observée, ces études n'ont pas pu confirmer l'idée d'un locus invariant en général.



**Figure 7.** Gamme de variation de fréquences des formants droits utilisée pour la détermination du locus (Delattre et al, 1962) [13].

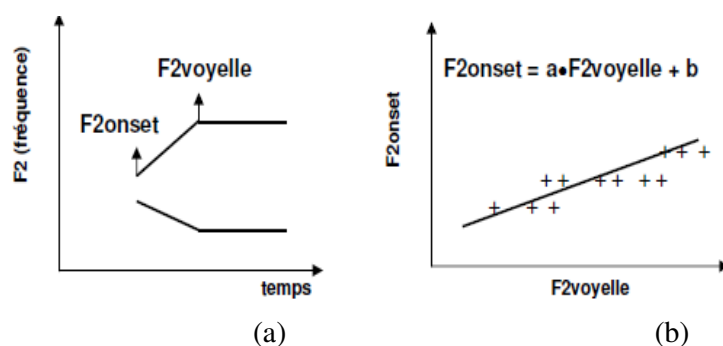
| Fricatives | F2 Locus | F3 Locus |
|------------|----------|----------|
| [f] , [v]  | 700 Hz   | 1900 Hz  |
| [θ] , [ð]  | 1400 Hz  | 2700 Hz  |
| [s] , [z]  | 1600 Hz  | 2700 Hz  |
| [ʃ] , [ʒ]  | 2000 Hz  | 2200 Hz  |

**Tableau 1.** Locus des fricatives américaines.

### 3. Droite de locus

Le concept de *droite de locus* a été introduit par Lindblom en 1963. Ce sont des droites de régression qui passent à travers un ensemble de points dont les coordonnées sont les  $F2_{\text{onset}}$

(début de la transition F2 de voyelle après la consonne) et  $F2_{\text{voyelle}}$  (fin de transition ou état stable de la voyelle), voir figure 8(a). La figure 8(b) montre la droite de régression ou bien droite de locus .représentée par l'équation :  $F2_{\text{onset}} = a \cdot F2_{\text{voyelle}} + b$  où «  $a$  » est la pente et «  $b$  » l'ordonnée à l'origine. Lindblom a testé un seul locuteur suédois produisant des corpus /CVC/ où /C/ est [b], [d] et [g] dans les deux positions initiale et finale et /V/ l'ensemble des huit voyelles centrales de la langue suédoise. Les pentes des droites étaient 0.69, 0.28 et 0.95 pour [b], [d] et [g] et les ordonnées à l'origine 410Hz, 1225Hz et 360Hz respectivement.



**Figure 8.** (a) Transitions CV du 2<sup>e</sup> formant. (b) Droite de locus

La droite de locus de Lindblom diffère dans deux points importants du concept classique du « locus virtuel » des laboratoires de Haskins. Le premier point est que le locus utilisé par Lindblom a une valeur concrète, qui est la fréquence du deuxième formant mesurée à la frontière /CV/. Le deuxième point concerne la dépendance du contexte entre la position du second formant en début de la voyelle et le noyau vocalique. Tandis que le « locus » original des Haskins est un point abstrait dans l'échelle des fréquences, observé comme le point de départ de la transition F2. Donc le « locus » dans le sens classique est un invariant acoustique fixe ne dépendant pas du contexte.

Malgré que le locus de Lindblom diffère de celui des Haskins, il y'a une relation entre les deux. Si le locus des consonnes dans le sens des Haskins existe en fait ; il serait vu comme le point d'intersection de la droite de l'équation de locus et la droite  $F2_{\text{onset}} = F2_{\text{voyelle}}$ . Ce point d'intersection est équivalent à un F2 stable (transition horizontale) qui pointe vers le « locus ». Tant que ces points d'intersection existent; les deux locus, un fixe et absolu, l'autre variable et relationnel peuvent être comparés.

$$F2_{\text{onset}} = a F2_{\text{voyelle}} + b \quad \text{et} \quad F2_{\text{onset}} = F2_{\text{voyelle}} = F2_{\text{locus}} \quad \Rightarrow \quad F2_{\text{locus}} = b / (1 - a).$$

Dans les travaux de Sussman et al (1991) [14] les comparaisons ont été faites. Dans cette expérience, des équations de locus ont été dérivées pour [b], [d] et [g] en prenant dix locuteurs féminins et dix locuteurs masculins. Le problème des transitions du [g] avec voyelles arrondies qui ne pointent pas vers le locus réapparaît ici sous une autre forme. Les points correspondant aux voyelles arrondies semblent déviées par rapport à la linéarité observée pour les points correspondant aux voyelles écartées, voir figure 9. Pour cela, deux droites de locus pour le [g] ont été tracées. Le [g] avec voyelle écartées est un [g] palatal et celui avec voyelles arrondies est un [g] vélaire.

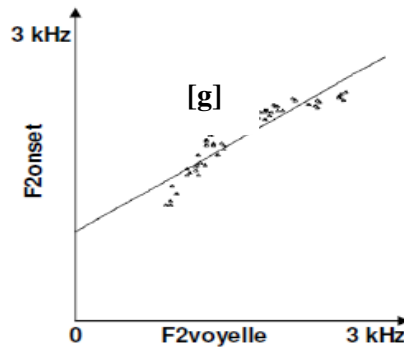
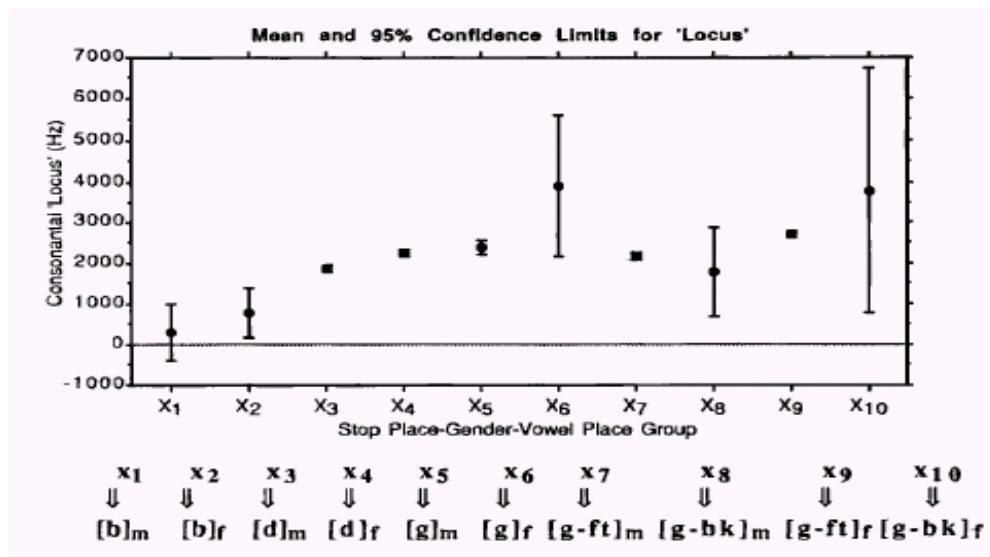


Figure 9. Droite de locus pour [g] (Sussman 1991) [14]

Pour comparer la droite de locus au locus virtuel, des droites de locus ont été dérivées pour chaque locuteur et le locus a été déduit en utilisant la formule citée ci-dessus. En ce qui concerne le [g], d'abord la droite de régression à travers toutes les voyelles a été tracée, ensuite les droites pour le [g] palatal avec voyelle écartées et le [g] vélaire avec voyelles arrondies ont été tracées. L'analyse a révélé que l'existence d'une droite de régression de l'équation de locus n'implique pas nécessairement qu'un locus de consonne unique et absolu existe. La figure 10 présente la variabilité du locus virtuel montrant un intervalle de confiance de 95% de chaque côté de la moyenne pour les 03 occlusives [b], [d] et [g].

Les labiales et le [g] vélaire ne révèlent pas une valeur de locus consistante. Seulement le [d] et le [g] palatal révèlent une valeur de locus stable. Cependant ces valeurs se recouvrent; donc le locus traditionnel ne pourrait être utilisé pour distinguer entre les occlusives. En ce

qui concerne la droite de locus, l'analyse discriminante a révélé qu'en prenant la pente et l'ordonnée à l'origine comme descripteurs une classification de 100% a été obtenue.



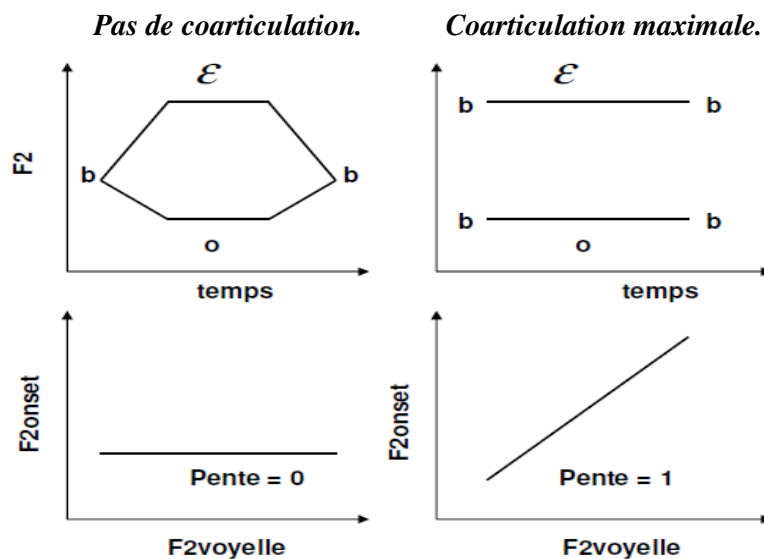
**Figure 10.** Variation du locus classique pour [b], [d] et [g]. L'indice 'm' correspond à masculin et 'f' à féminin. [g-ft] correspond à [g] palatal et [g-bk] correspond à [g] vélaire. (Sussman et al, 1991) [14]

En 1993 Sussman a étendu l'étude des équations de locus à trois autres langues « l'arabe égyptien », « le thaï » et « le urdu ». La linéarité observée était très forte avec un coefficient de corrélation moyen de 0.86 [15]. La linéarité des équations de locus dans d'autres langues a été observée aussi, notamment l'arabe moderne marocain : Yeou, 1997 [16]; l'arabe dialectique jordanien: Tamimi, 2004 [17] ; le vietnamien: Eric Castelli, 2006 [18]; l'Espagnol : Celdran et Villalba, 1995; On peut voir cette forte corrélation des équations de locus dans toutes ces langues comme un fait bien établi et probablement dans tout le langage universel. Ceci est très remarquable comme une statistique robuste permettant de prédire le  $F2_{\text{onset}}$  à partir de  $F2_{\text{voyelle}}$ .

#### 4. Droite de locus et coarticulation

Krull (1987) [19,20] a montré que la pente de la droite de locus pourrait être une mesure du degré de coarticulation de la voyelle V sur la consonne C. Plus une consonne est influencée par la voyelle, moins les transitions formantiques convergent vers un locus

commun et plus raide est la pente de la droite de locus. Ceci est illustré par les transitions F2 de deux syllabes /beb/ et /bob/, voir figure 11.

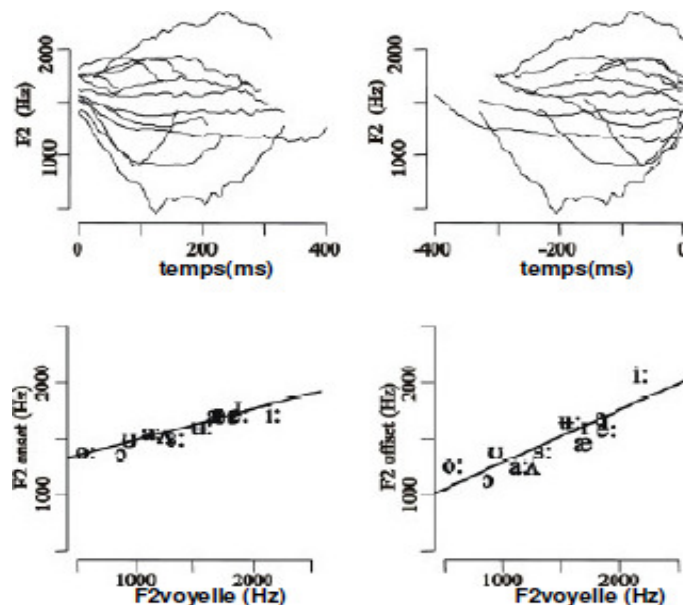


**Figure 11.** En haut : Trajectoires F2 de /beb/ et /bob/ à gauche absence de coarticulation et à droite coarticulation maximale.  
En bas : Droites de locus correspondantes.

A gauche, les transitions F2 convergent vers un locus commun. Dans ce cas,  $F2_{\text{onset}}$  est complètement non affectée par la voyelle qui suit (l'anticipation de la voyelle sur la consonne à l'instant de l'onset est nulle). D'un autre point de vue ; il y'a aucun moyen de prévoir la voyelle sachant seulement le  $F2_{\text{onset}}$ , puisque /be/ et /bo/ ont le même  $F2_{\text{onset}}$ . A droite c'est la coarticulation maximale. La coarticulation de V sur C est tellement forte qu'il n'y a pas convergence vers un locus commun. Dans ce cas, le formant  $F2_{\text{onset}}$  est le même que le formant  $F2_{\text{voyelle}}$  ce qui veut dire que l'on peut prévoir la voyelle sachant seulement le  $F2_{\text{onset}}$ . Et à partir de ces figures, il est évident que les deux cas de coarticulation minimale et maximale correspondent aux pentes des droites de locus qui sont 0 et 1 respectivement. Ce qui est important, c'est que le fait que la pente varie entre 0 et 1 pourrait être utilisée pour mesurer le degré de la coarticulation V sur C. Ce principe est illustré pour quelques syllabes /dVd/ produites par un Australien, voir figure 12. En haut, le suivi de formant F2 synchronisé d'abord à l'onset de la voyelle (à gauche) et ensuite à l'offset de la voyelle (à droite). Nous remarquons qu'il y'a une plus grande convergence vers un F2 commun à l'onset plutôt qu'à l'offset. A partir de cela, il pourrait être inféré que la coarticulation V sur C au début de la



syllabe /dV/ est moins importante qu'à la fin de la syllabe /Vd/. En bas de la figure les droites de locus correspondantes sont tracées montrant une pente plus élevée pour la fin de la syllabe.



**Figure 12.** En haut : Trajectoires F2 de syllables /dvd/ synchronisées à l'onset de la voyelle (à gauche) et à l'offset (à droite). En bas : Droites de locus correspondantes (Harrington, 2006) [20].

## 5. Droite de locus et lieu d'articulation

Afin de vérifier si la droite de locus renseigne sur le lieu d'articulation en général deux expériences ont été faites, l'une en 1994 par Fowler [21] et l'autre en 1996 par Sussman et Shore [22]. L'objectif était donc d'appliquer les droites de locus à des consonnes se partageant le même lieu d'articulation mais variant dans la manière d'articuler. Pour Fowler, la pente de la droite de locus reflète la coarticulation plus ou moins directement. Elle renseigne sur le lieu d'articulation indirectement, seulement parce que la variation du lieu d'articulation est une source importante dans la variation de la coarticulation. Dans son expérience, elle a confirmé ses prédictions. Pour [z] et [d] qui ont le même lieu d'articulation, la pente pour [z] était d'une façon significative moins raide que celle de [d]. Aussi [g] et [v] qui ont des lieux d'articulation différents ont d'une manière significative la même pente. En plus l'analyse discriminante où la pente et l'ordonnée à l'origine étaient utilisées comme descripteurs était infructueuse pour distinguer entre des consonnes représentant six lieux d'articulation. Elle conclut que les droites de locus fournissent de l'information sur le lieu

d'articulation mais pas suffisamment pour spécifier le lieu d'articulation. Sussman et Shore (1996) [22] ont testé de nouveau les droites de locus pour les obstruents se partageant le même lieu d'articulation et variant dans la manière d'articuler. Cette étude était une réponse et extension à celle de Fowler (1994) [21] ; c'était effectuée avec cinq manières d'articulation et en utilisant plus de locuteurs (22). Les pentes moyennes à travers les obstruents [d], [z] et [n] ont été trouvées avec des différences non significatives, ce qui contredit les résultats de Fowler. Les 22 locuteurs ont produits les cinq coronaux [d], [t], [n], [z], et [s]. Les droites de locus déterminées ensemble avec celles de [b] et [g] (20 locuteurs) prises d'une étude précédente (Sussman et al 1991 [14]) ont été parfaitement séparées en trois classes distinctes, les labiaux, les coronaux et les vélaires. Les résultats ont confirmé que les droites de locus peuvent servir comme des descripteurs phonétiques effectifs du lieu d'articulation des consonnes en général.

## Conclusion

Sussman et ses collaborateurs ont proposé les équations de locus comme des invariants relationnels que les auditeurs peuvent utiliser pour identifier le lieu d'articulation de la consonne à partir du signal acoustique de la parole. Bien sûr les droites de locus, même l'information acoustique nécessaire pour les construire ne sont pas présentes dans le signal acoustique CV que l'auditeur pourrait écouter. Lorsqu'un locuteur produit un corpus /CV/, seulement un couple de points  $F2_{\text{onset}}$  et  $F2_{\text{voyelle}}$  sont disponibles et non pas une droite ou bien une équation pour tracer la droite. Le modèle proposé par Sussman est que la variété de couples de points  $F2_{\text{onset}}$ ,  $F2_{\text{voyelle}}$  que l'auditeur pourrait recevoir constitue la droite de locus au niveau de la perception. Par la suite, lorsqu'un nouveau couple de points ( $F2_{\text{onset}}$   $F2_{\text{voyelle}}$ ) sera perçu, il sera proche de l'une des droites de locus et de cette façon le lieu d'articulation sera déterminé. Le fondement de ce concept est que la droite de locus est liée à la consonne. Chaque droite de régression peut définir une classe d'équivalence de transitions CV représentant une consonne particulière. Ces classes s'entrecoupent dans une certaine mesure, donc *les droites de locus ne sont que partiellement distinctives*. Dans ces régions d'intersection autres indices, notamment le « burst » est supposé apporter l'information distinctive [23]. En 1997 Fruchter et Sussman dans leur expérience [24], ont observé que l'effet du 3<sup>e</sup> formant ( $F3$ ) ressemble à l'effet du burst dans l'expérience faite par Dorman en

1977 [23]. Comme avec le burst,  $F3_{\text{onset}}$  apporte l'information distinctive dans ces régions où les droites de locus s'entrecoupent. Cependant il n'y a pas effet de  $F3$  dans ces régions où il n'y a pas de recouvrement. Donc pour bien comprendre les schémas du burst et de  $F3_{\text{onset}}$  dans la perception des occlusives, il faut les observer par rapport aux droites de locus. C'est dans ce contexte que nous allons utiliser le  $F3_{\text{onset}}$  ensemble avec la droite de locus pour discriminer les lieux d'articulation des consonnes que nous allons étudier dans les chapitres suivants.

## Chapitre II

### Etude expérimentale

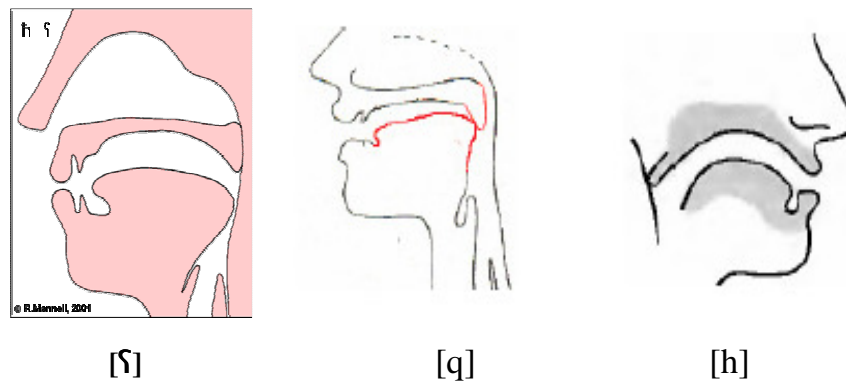
#### 1. Introduction

L'étude porte sur trois consonnes arabes, la glottale [h] (هـ), la pharyngale [ʕ] (ع) et l'uvulaire [q] (ق). Toutes produites par une constriction dans la partie postérieure du conduit vocal.

*La glottale* [h] est produite sans aucune constriction au niveau du conduit vocal. Plutôt, elle est caractérisée par la petitesse de l'ouverture de la glotte produisant ainsi un mouvement d'air turbulent que le conduit vocal, par sa configuration neutre, voir figure 13, laisse passer librement. Les caractéristiques spectrales du son généré dépendent de l'ouverture de la glotte ainsi que la cavité résonante à l'onset de la voyelle [25].

*La pharyngale* [ʕ] est produite avec un rétrécissement du pharynx dû aux rétractions de la racine de la langue et de l'épiglotte, voir figure 13. Aussi, le larynx est relevé considérablement [26,27].

*L'uvulaire* [q] Durant la production de l'occlusive [q], il n'y a pas de vibration des cordes vocales. La constriction est totale par le biais de la partie arrière de la langue en s'élevant contre l'uvula [26,27], voir figure 13.

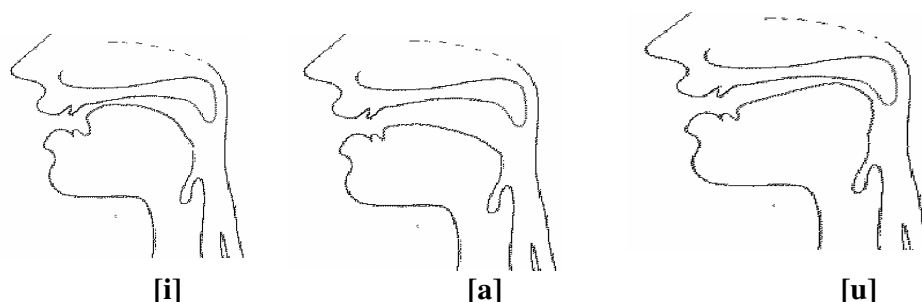


**Figure 13.** Configuration du conduit vocal dans la production de [ʕ], [q] et [h].

## 2. Matériau linguistique

Les 3 consonnes étudiées sont placées dans les contextes vocaliques [a], [i] et [u]. L'articulateur principal dans la production de ces voyelles est la langue dont les différentes positions [25] sont montrées dans la figure 14. Le corpus se compose d'un /CV/ en milieu de mots arabes. La phrase porteuse est de forme interrogative afin de donner un aspect naturel. Six locuteurs masculins et un locuteur féminin adultes Algériens ont enregistré un corpus volumineux de 630 phrases (trois consonnes X trois contextes vocaliques X dix répétitions X sept locuteurs). Les locuteurs ont été invités à produire les phrases suivantes :

|                |                 |                 |
|----------------|-----------------|-----------------|
| /men saʃala ?/ | /men saqala ?/, | /men sahala ?/, |
| /men saʃila ?/ | /men saqila ?/, | /men sahila ?/, |
| /men saʃula ?  | /men saqula ?/, | /men sahula ?/, |



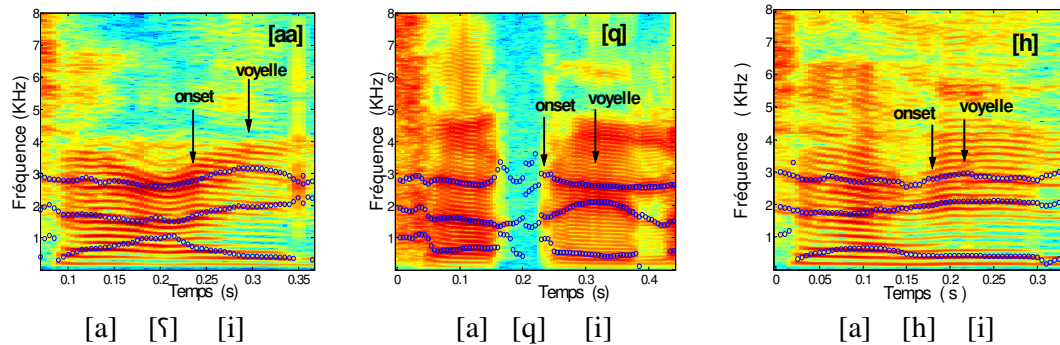
**Figure 14.** Position de la langue dans la production des voyelles [i], [a] et [u].

## 3. Enregistrement des phrases

Les enregistrements ont été faits dans une chambre calme du laboratoire d'acoustique de la faculté d'électronique. D'abord, il a été demandé au locuteur de prononcer correctement et naturellement la phrase une seule fois. Si l'expérimentateur entendait la phrase bien dite, il demande alors au locuteur de répéter dix fois et procède à l'enregistrement. Si la phrase n'a pas été bien prononcée, la prononciation correcte lui serait donnée. Le locuteur la répète jusqu'à ce que cette phrase soit bien prononcée et ensuite répéter correctement une dizaine de fois pour l'enregistrement. L'enregistrement a été fait sur un PC Pentium IV muni d'une carte son de type « Sound Blaster » à l'aide d'un microphone professionnelle. Le signal enregistré a été digitalisé à 16 kHz avec 16 bits.

#### 4. Analyse et mesure des caractéristiques formantiques

L'analyse des segments de parole est faite dans le domaine fréquentielle. Pour localiser le début de transition et le point stable de la voyelle nous avons utilisé des procédures similaires à celles décrites par Sussman et autres (1991) [14] et Fowler (1994) [21]. Des mesures de F2 ont été prises au début de la transition consonne voyelle ( $F2_{\text{onset}}$ ) et sur un point stable de la voyelle ( $F2_{\text{voyelle}}$ ). Pour la plosive [q] (ق), le début de la voyelle a été identifié comme début du voisement de la voyelle suivant le relâchement de l'occlusive. Pour le point stable de la voyelle, quand le suivi de formants a la forme en U ou U inversé, c'est la valeur de la cuvette ou de la crête qui est prise, si la variation de la fréquence est monotone, la valeur du formant au milieu temporel de la voyelle est choisie. Pour les fricatives [h] (ه) et [ʕ] (ع), le  $F2_{\text{onset}}$  est relevé à l'instant où se termine la friction. Pour plus de détail, voir le paragraphe 6 dans l'annexe 1 (Outil d'Analyse). Nous montrons le onset et l'état stable de la voyelle sur le spectrogramme des trois sons [ʕ] (ع), [q] (ق) et [h] (ه), dans la figure 15.

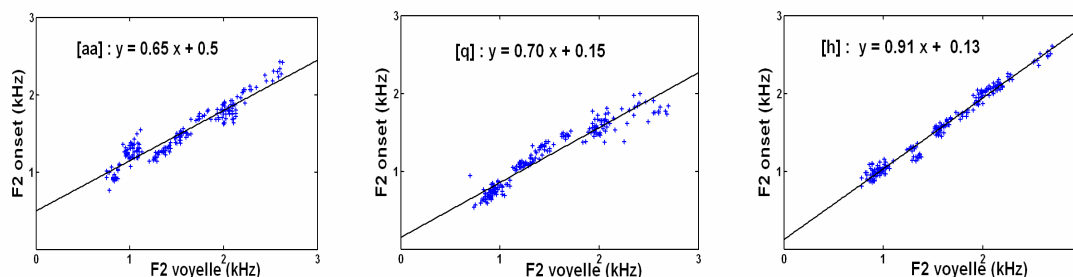


**Figure 15.** Spectrogrammes des 3 sons [ʕ] (ع), [q] (ق) et [h] (ه)

#### 5. Résultats et interprétations

##### *a- Détermination des droites de locus*

La droite de locus est calculée pour chaque phonème, en utilisant la régression linéaire. D'abord en prenant les sept locuteurs ensemble, voir figure 16. Les pentes **a**, les ordonnées à l'origine **b** et les coefficients de corrélation **R<sup>2</sup>** des droites de locus correspondantes sont montrés dans le tableau 2.



**Figure 16.** Droites de locus de [ʕ] (aa), [q] et [h] pour l'ensemble des locuteurs

| Phonèmes | a    | b    | R <sup>2</sup> |
|----------|------|------|----------------|
| [ʕ]      | 0.65 | 0.50 | 0.90           |
| [q]      | 0.70 | 0.15 | 0.92           |
| [h]      | 0.91 | 0.13 | 0.98           |

**Tableau 2.** Paramètres caractéristiques des droites de locus  
a pente b ordonnée à l'origine R<sup>2</sup> coefficient de corrélation.

Toutes les droites de régression sont obtenues avec un coefficient de corrélation supérieur à 0.90. Ce qui assure la fiabilité du processus de mesure utilisé dans notre expérience.

Nous observons que dans le cas de [h], les échantillons sont bien alignés sur leur droite de locus. Dans le cas de [q], Nous observons une plus grande variabilité dans la région de la voyelle [i]. Aussi dans la région de la voyelle [u] les échantillons semblent dévier un peu vers le bas. Ceci ressemble au cas de l'occlusive [g] où dans la région des voyelles arrondies, les échantillons sont déviés vers le bas par rapport à leur droite de locus [14] [21]. Dans le cas de [ʕ], les échantillons ne sont pas bien alignés sur leur droite de locus.

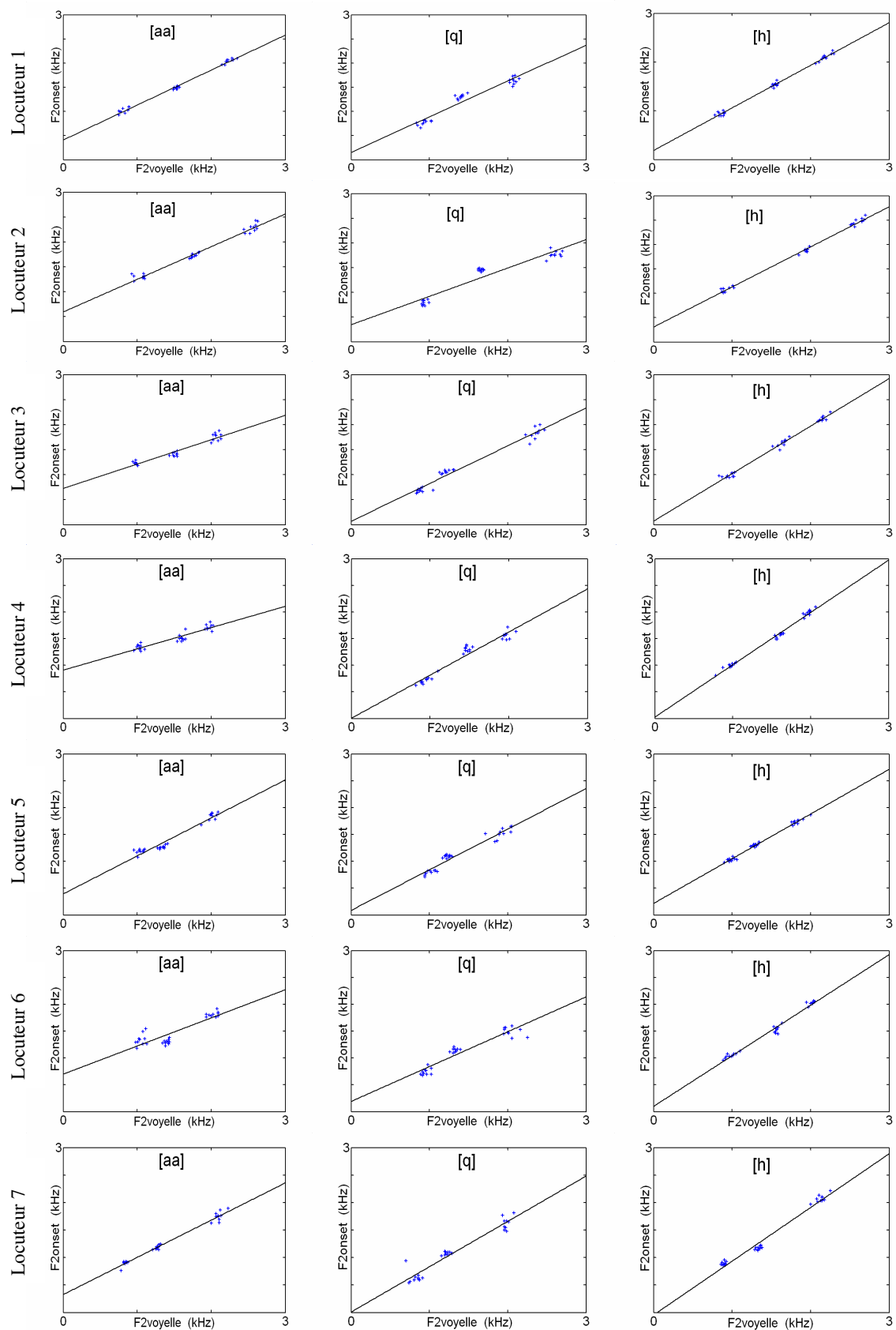
La glottale [h] co-articule mieux avec les voyelles. Ceci est caractérisé par la pente élevée égale à 0.91. La pharyngale [ʕ] et l'uvulaire [q] montrent plus de résistance à la coarticulation avec des pentes respectives de 0.65 pour le [ʕ] et 0.70 pour le [q]. Cependant, on peut observer une nette différence dans l'ordonnée à l'origine, 0.15 kHz pour le [q] et 0.5 kHz pour [ʕ].

On peut expliquer la différence dans la coarticulation par le fait que la racine de la langue soit utilisée dans la production des pharyngales et des uvulaires, mais pas dans la production des laryngales [26] [27]. Donc la voyelle qui suit n'a pas une liberté totale pour coarticuler avec [q] ou [ʕ] comme dans le cas de [d] expliqué par Fowler (1994) [21]. Cependant, dans la production de la glottale [h], la langue n'est pas du tout impliquée, la voyelle qui suit a une liberté considérable pour coarticuler avec cette consonne. Pour comparer nos résultats avec ceux trouvés dans la littérature, il y'a seulement ceux de la pharyngale [ħ](ح) et de l'uvulaire [χ](خ) obtenus dans les travaux de Yeou 1997 [16]. Pour le [ħ] se partageant le même lieu d'articulation avec le [ʕ], la pente moyenne obtenue est égale à 0.68, donc très proche de notre résultat pour la pharyngale [ʕ] avec une pente moyenne égale à 0.65. Pour le [χ] se partageant le même lieu d'articulation avec le [q], la pente moyenne obtenue est égale à 0.90. Ceci contredit notre résultat obtenu pour le [q] où la pente moyenne est égale à 0.70. Une explication possible est que dans la langue arabe, il n'y a pas de fricative vélaire contrastant ainsi le [χ]. Ce dernier a donc plus d'espace pour mieux coarticuler avec les voyelles sans qu'il y'ait de confusion perceptuelle entre le [χ] postérieure et le [χ] antérieure. Cette explication est similaire à celle qui a été donnée par Fowler 1994 [21] pour le vélaire [g] qui a une pente élevée malgré que la langue soit impliquée dans sa production. Ces deux phonèmes [q] et [χ] qui ont le même lieu d'articulation ont des droites de locus distinctes. Donc, *la droite de locus ne reflète pas d'une façon générale le lieu d'articulation*. Ainsi, nous confirmons les résultats de Fowler 1994 [21].

Aussi, nous avons tracé la droite de locus pour chaque locuteur. Nous observons les mêmes phénomènes cités ci haut. Dans le cas de [ʕ] et [q], les échantillons ne sont pas aussi bien alignés sur leur droite régressive correspondante comme dans le cas de [h], voir figure 17. Dans le cas de [q], nous observons une plus grande variabilité dans la région de la voyelle [i] et une déviation des échantillons vers le bas dans la région de la voyelle [u].

Les pentes **a**, les ordonnées à l'origine **b** et les coefficients de corrélation **R<sup>2</sup>** des droites de locus correspondantes sont montrés dans le tableau 3. Nous observons pour chaque consonne une corrélation entre la pente et l'ordonnée à l'origine. Lorsque la pente augmente, l'ordonnée à l'origine diminue. Nous observons aussi que la variabilité de la pente de chaque consonne augmente lorsque la pente diminue. Ceci se voit encore mieux dans la figure 18.



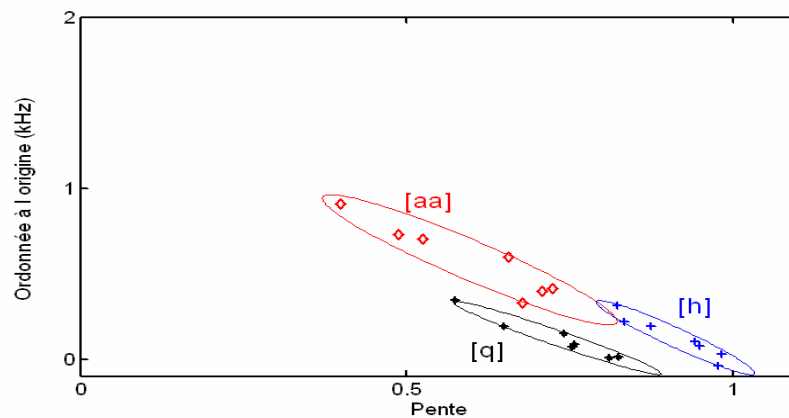


**Figure 17.** Droites de locus par locuteur des 3 sons [ʎ] (aa), [q] et [h].

| locuteur | [ʃ]  |      |                | [q]  |      |                | [h]  |       |                |
|----------|------|------|----------------|------|------|----------------|------|-------|----------------|
|          | a    | b    | R <sup>2</sup> | a    | b    | R <sup>2</sup> | a    | b     | R <sup>2</sup> |
| 1        | 0.72 | 0.41 | 0.99           | 0.74 | 0.15 | 0.93           | 0.87 | 0.19  | 0.99           |
| 2        | 0.66 | 0.60 | 0.98           | 0.57 | 0.35 | 0.91           | 0.82 | 0.31  | 0.99           |
| 3        | 0.49 | 0.73 | 0.92           | 0.75 | 0.07 | 0.97           | 0.95 | 0.08  | 0.99           |
| 4        | 0.40 | 0.91 | 0.88           | 0.81 | 0.01 | 0.94           | 0.98 | 0.03  | 0.99           |
| 5        | 0.71 | 0.40 | 0.95           | 0.76 | 0.09 | 0.94           | 0.83 | 0.22  | 0.99           |
| 6        | 0.52 | 0.70 | 0.77           | 0.65 | 0.19 | 0.90           | 0.94 | 0.10  | 0.98           |
| 7        | 0.68 | 0.33 | 0.98           | 0.82 | 0.01 | 0.93           | 0.98 | -0.04 | 0.98           |

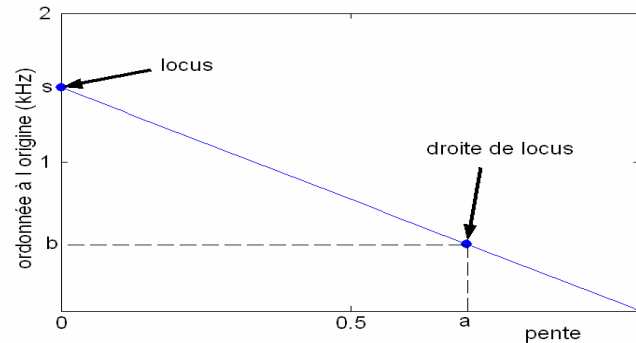
**Tableau 3.** Paramètres caractéristiques des droites de locus par locuteur  
**a** pente ; **b**, ordonnée à l'origine ; **R<sup>2</sup>** coefficient de corrélation.

Traditionnellement, pour mieux montrer la séparabilité des droites de locus des différentes consonnes, on utilise le plan **a**, **b** (pente, ordonnée à l'origine) où les paramètres « **a** » et « **b** » représentent les axes principaux. Chaque droite sur le plan  $F2_{\text{onset}}$ ,  $F2_{\text{voyelle}}$  est donc représentée par un point de coordonnées  $a_0$ ,  $b_0$  (pente, ordonnée à l'origine). Les droites de locus par locuteur des 3 phonèmes sont montrées dans la figure 18. Sur ce plan, on voit bien comment les classes sont séparées. Les ellipses de [ʃ] (aa) et [h] sont orientées vers une même fréquence autour de 1.5 kHz sur l'axe de l'ordonnée à l'origine tandis que celle de [q] vers une fréquence plus basse 1 kHz. Ce sont les locus de ces sons.



**Figure 18.** Droites de locus par locuteur pour les sons [ʃ](aa),[q] et [h]. Les ellipses de dispersion sont déterminées sur un intervalle de confiance de 0.8.

C'est la relation entre le locus et l'ordonnée à l'origine  $s = \frac{b}{1-a}$ , voir paragraphe 3 chapitre I, qui donne cette orientation. Ceci est montré dans la figure 19.

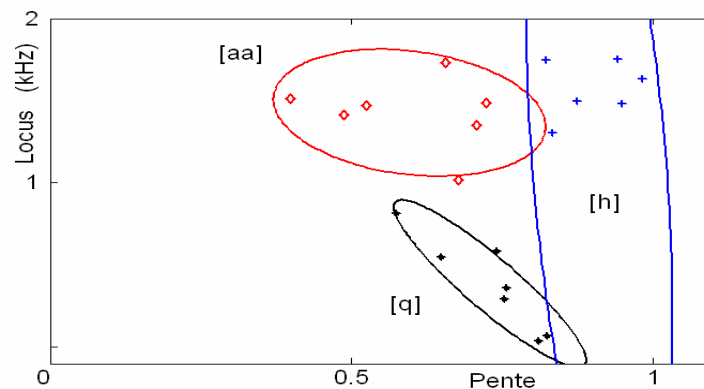


**Figure 19.** Détermination graphique du locus à partir de la droite de locus

Ceci est schématisée par une simple interpolation du point (1,0) avec la droite de locus (a,b), l'intersection avec l'axe **ordonnée à l'origine** donne la valeur du locus.  $\frac{s}{1} = \frac{b}{1-a}$

#### *b- Détermination du locus*

Nous allons utiliser la formule  $s = \frac{b}{1-a}$  pour transformer la figure 18 afin d'obtenir les ellipses positionnées horizontalement au niveau de leur locus. Donc, en divisant b par le facteur (1-a) on obtient la figure 20.



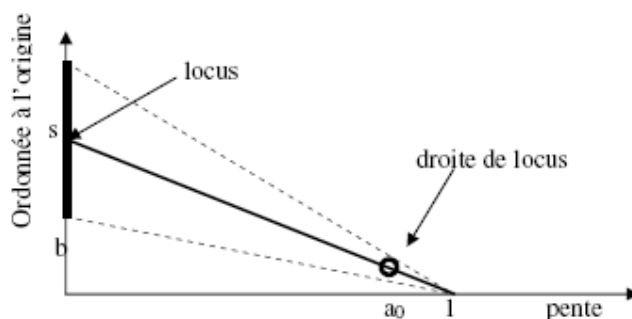
**Figure 20.** Relation locus pente par locuteur pour les trois phonèmes [ʔ](aa), [q] et [h], obtenue en utilisant la formule  $s=b/(1-a)$ .

Pour le [ʃ] c'est acceptable mais pour les autres sons ce n'est pas le résultat escompté. La formule utilisée pour déterminer le locus à partir des paramètres de la droite de locus fournit des résultats instables notamment pour le [q] et surtout le [h] malgré que les droites de locus soient obtenues avec un coefficient de corrélation assez élevé. La question que l'on se pose, est ce que c'est la méthode utilisée pour déterminer le locus qui n'est pas fiable ou bien c'est le locus qui n'est pas stable par rapport à la droite de locus ?

En utilisant la formule  $s = \frac{b}{1-a}$  l'incertitude relative de la valeur du locus « s » est :

$$\frac{\Delta s}{s} = \frac{\Delta b}{b} + \frac{\Delta a}{1-a} = \frac{\Delta b}{b} + \frac{\Delta a}{a} \times \frac{a}{1-a}$$

On voit bien que l'erreur sur la valeur du locus ne dépend pas seulement des erreurs estimées de la pente et de l'ordonnée à l'origine mais aussi du facteur  $a/(1-a)$ . Pour les pentes  $a < 0.5$ , ce facteur tend à réduire l'erreur sur le locus, en d'autres termes la formule nous donne un locus plus stable. Tandis que lorsque la pente est assez élevée par exemple pour  $a = 0.9$ , le facteur  $a/(1-a) = 9$ , ce qui permet d'obtenir un locus moins stable. C'est ce que Sussman a obtenu dans ses travaux en 1991[14] en utilisant cette formule, des locus stables pour [d] et [g] palatal correspondant aux droites de locus avec des pentes faibles tandis que pour le [b] et le [g] vélaire avec des pentes élevées, les locus obtenus étaient instables, voir figure 10 chapitre I.

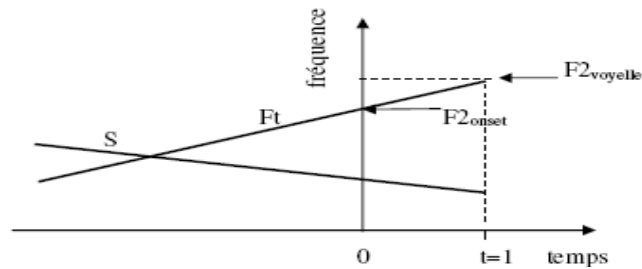


**Figure 21.** Erreur projetée sur le locus

En se référant à la figure 21, il est clair que lorsque la pente est élevée (proche de  $a=1$ ) une petite erreur engendre une grande déviation sur l'axe des **b** et donc sur la valeur du locus. L'erreur est projetée à partir du point (1,0) sur l'axe de l'ordonnée à l'origine. Donc, en

déterminant le locus de cette façon, on ne peut en conclure que la droite de locus est plus stable que le locus classique.

Une autre façon de déterminer le locus est tout simplement de trouver le point de convergence des différentes transitions /CV/. Pour simplifier le problème, nous utilisons une interpolation linéaire reliant les deux points  $F2_{\text{onset}}$ ,  $F2_{\text{voyelle}}$ . Donc, la transition est une droite passant par ces deux points. Choisissons  $t=0$ , où la droite passe par le  $F2_{\text{onset}}$ , et  $t=1$  où la droite passe par le  $F2_{\text{voyelle}}$  comme on le voit dans la figure 22.



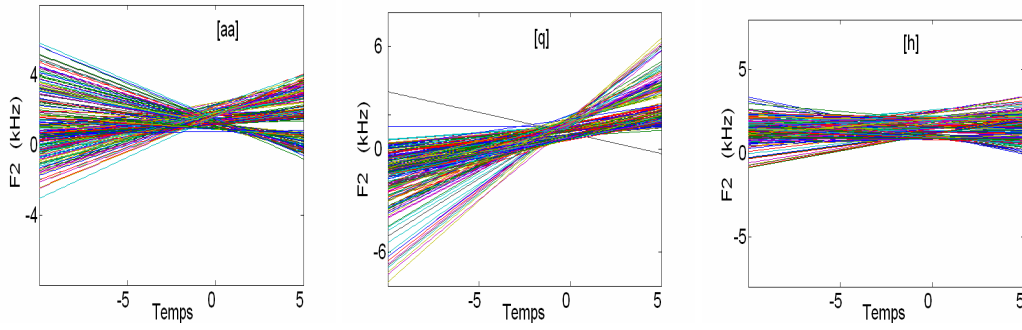
**Figure 22.** Interpolation linéaire d'une transition /CV/

La fréquence de la transition est :  $F_t = (F2_{\text{voyelle}} - F2_{\text{onset}}) * t + F2_{\text{onset}}$

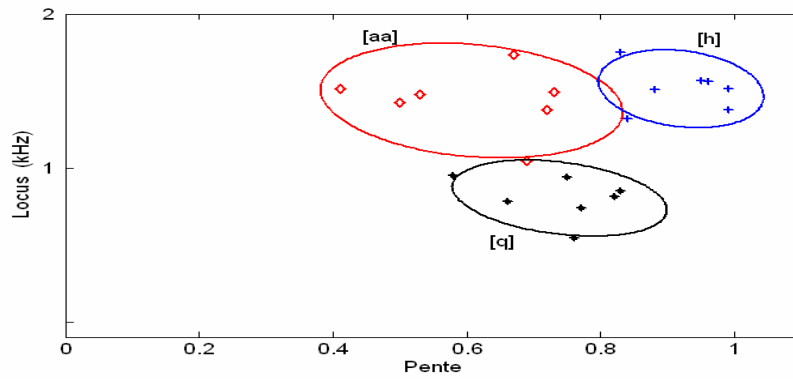
Pour une consonne donnée, en appliquant ce modèle aux différentes transitions /CV/, nous obtenons le point de convergence là où la variance de  $F_t$  est minimale, et la valeur du locus c'est la moyenne de  $F_t$  en ce point. Les convergences des transitions formantiques vers les locus des trois (03) sons étudiés sont montrées dans la figure 23, et le résultat obtenu est montré dans la figure 24. Maintenant, si on utilise une transition dans un espace à 3 dimensions en considérant les deux formants  $F2$  et  $F3$ , la transition sera caractérisée par les deux équations suivantes :

$$F_{t2} = (F2_{\text{voyelle}} - F2_{\text{onset}}) * t + F2_{\text{onset}} \quad \text{et} \quad F_{t3} = (F3_{\text{voyelle}} - F3_{\text{onset}}) * t + F3_{\text{onset}}$$

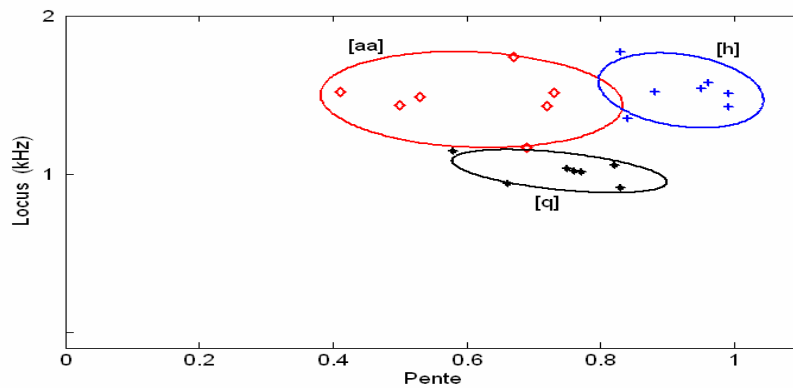
Nous obtenons le point de convergence là où la somme des variances de  $F_{t2}$  et  $F_{t3}$  est minimale. Les locus obtenus sont encore plus stables, voir figure 25.



**Figure 23.** Convergence vers le locus des sons [ʔ](aa), [q] et [h].



**Figure 24.** relation locus pente par locuteur pour les sons [ʔ], [q] et [h].  
obtenue en utilisant la méthode de convergence en 2 dimensions.



**Figure 25.** Relation locus pente par locuteur pour les sons [ʔ], [q] et [h].  
obtenue en utilisant la méthode de convergence en 3 dimensions.

Le tableau 4 résume les résultats obtenus sur la stabilité du locus en utilisant la formule  $s = \frac{b}{1-a}$  et la méthode de convergence.

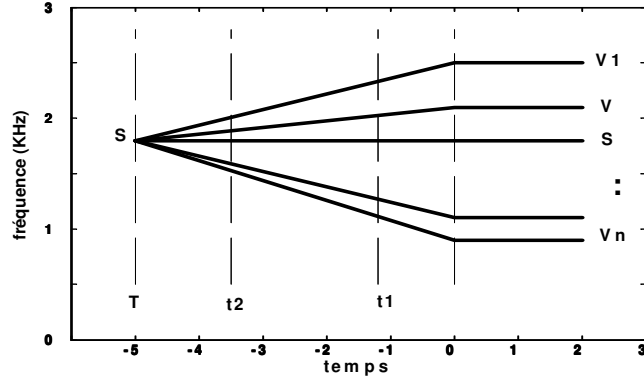
| Locuteur | [ʃ]  |      |      | [q]  |      |      | [h]   |      |      |
|----------|------|------|------|------|------|------|-------|------|------|
|          | Sf   | D2   | D3   | Sf   | D2   | D3   | Sf    | D2   | D3   |
| 1        | 1.49 | 1.49 | 1.51 | 0.59 | 0.94 | 1.03 | 1.50  | 1.51 | 1.52 |
| 2        | 1.73 | 1.74 | 1.74 | 0.81 | 0.95 | 1.15 | 1.75  | 1.75 | 1.77 |
| 3        | 1.42 | 1.42 | 1.43 | 0.29 | 0.55 | 1.02 | 1.48  | 1.56 | 1.57 |
| 4        | 1.51 | 1.51 | 1.52 | 0.04 | 0.81 | 1.06 | 1.63  | 1.51 | 1.51 |
| 5        | 1.35 | 1.38 | 1.43 | 0.36 | 0.74 | 1.01 | 1.30  | 1.32 | 1.35 |
| 6        | 1.47 | 1.48 | 1.49 | 0.55 | 0.78 | 0.94 | 1.75  | 1.57 | 1.54 |
| 7        | 1.02 | 1.04 | 1.17 | 0.07 | 0.85 | 0.92 | -1.83 | 1.38 | 1.42 |

**Tableau 4.** D2 : locus déterminé par convergence de F2, D3 : locus déterminé par convergence de F2+F3, Sf : locus déterminé par la formule  $S=1/(1-a)$ .

Le locus déterminé par la méthode de convergence en D3 paraît le plus stable, notamment pour le [q] où le locus calculé par la formule varie de 0.07 kHz à 0.81 kHz, celui déterminé par la méthode de convergence en D2 varie de 0.55 kHz à 0.95 kHz tandis que celui obtenu par la méthode de convergence en D3 varie de 0.92 kHz à 1.15 kHz. Ce résultat laisse penser que les transitions formantiques sont utilisées comme une simple transition dans un espace de dimension plus élevée, D3 si on utilise F2 et F3 et D4 si on rajoute encore F1, au lieu d'être utilisées séparément comme des projections dans de sous espaces.

#### c- Locus, droite de locus et segmentation

Pour une consonne donnée dans une syllabe /CV/ toutes les transitions de F2 pointent vers un point appelé « locus », voir figure 26. L'axe vertical c'est la fréquence et l'axe horizontal c'est le temps. La cible de la voyelle est atteinte à l'instant  $t = 0$ , et le locus c'est le point de convergence « S ». La droite de locus est obtenue à partir d'une segmentation. Soit  $t_1$  l'instant où les  $F2_{onset}$  sont relevés.



**Figure 26.** Transitions F2 dans une syllabe /CV/

$$\frac{(F2_{\text{voyelle}} - F2_{\text{onset}})}{t_1} = \frac{F2_{\text{voyelle}} - S}{T} \quad \text{ce qui donne :} \quad F2_{\text{onset}} = a F2_{\text{voyelle}} + b$$

$$\text{où} \quad a = 1 - \frac{t_1}{T} \quad \text{et} \quad b = S \frac{t_1}{T} \quad (1)$$

$$\text{Soit deux droites de locus : } F2_{\text{onset1}} = (1 - \frac{t_1}{T})F2_{\text{voyelle}} + S \frac{t_1}{T} \quad \text{et} \quad F2_{\text{onset2}} = (1 - \frac{t_2}{T})F2_{\text{voyelle}} + S \frac{t_2}{T}$$

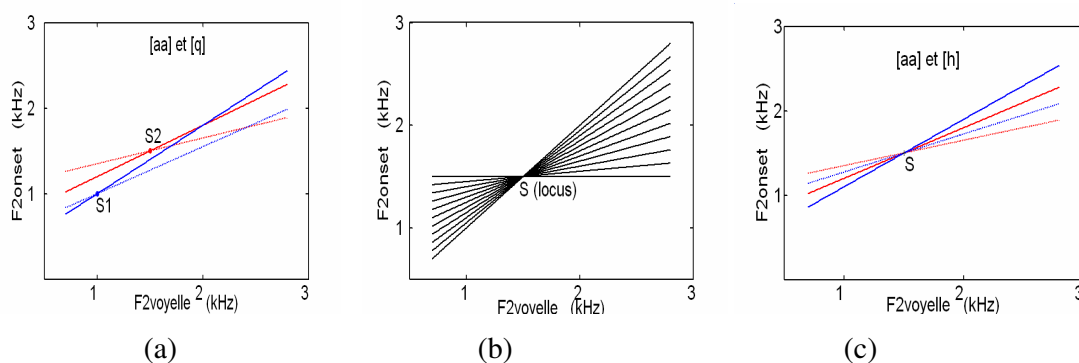
obtenues à partir de deux segmentations différentes  $t_1$  et  $t_2$ . Ces deux droites se coupent au point  $F2_{\text{onset1}} = F2_{\text{onset2}}$ , ce qui donne :

$$F2_{\text{voyelle}} = \frac{S (\frac{t_1}{T} - \frac{t_2}{T})}{(1 - \frac{t_2}{T}) - (1 - \frac{t_1}{T})} = S$$

Lorsque  $t_1$  tend vers 0 (vers la voyelle) la pente « a » tend vers 1 et « b » tend vers zéro. Et lorsque  $t_1$  tend vers T (vers le locus) la pente « a » tend vers 0 et « b » tend vers S. Donc, en faisant varier  $t_1$  de T à zéro (de S vers V), on obtient un ensemble de droite de locus qui se croisent exactement au locus « S ». *La droite de locus tourne autour du locus lorsqu'on fait varier la segmentation*, voir figure 27b. Maintenant, nous allons voir comment change l'intersection de [ɣ] avec [q] et [ɣ] avec [h] en variant la segmentation. Dans le cas de [ɣ] avec [h], l'intersection des droites de locus ne change pas avec la segmentation puisque la segmentation permet aux droites de tourner autour du locus S, voir figure 27c, qui est le même pour ces deux consonnes. Dans le cas de [ɣ] avec [q], l'intersection change avec la



segmentation comme on le voit dans la figure 27a. Ceci parce que la segmentation permet aux droites de tourner autour de deux axes différents S1 et S2.



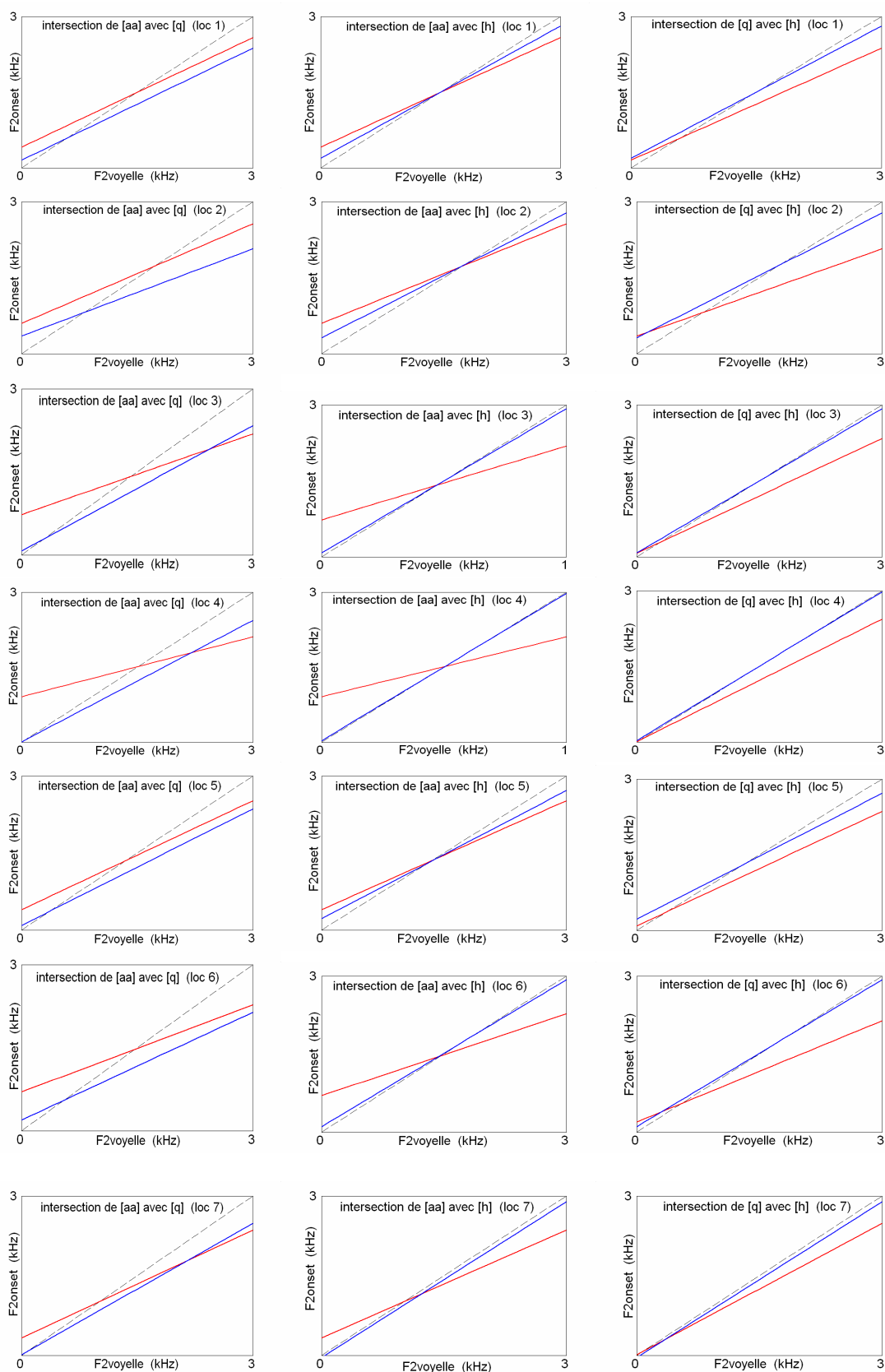
**Figure 27.** (a) Droites de locus pour deux segmentations dans le cas de [ʔ] et [q]  
(b) Droites de locus pour différentes segmentations.  
(c) Droites de locus pour deux segmentations dans le cas de [ʔ] et [h]

Les intersections des droites de locus des différentes paires de sons pour chaque locuteur sont montrées dans la figure 28, et se résument dans le tableau 5.

| Locuteur | [ʔ] [h] |                |                | [ʔ] [q] |                 |                | [q] [h] |                |                |
|----------|---------|----------------|----------------|---------|-----------------|----------------|---------|----------------|----------------|
|          | X       | S <sub>ʔ</sub> | S <sub>h</sub> | X       | S <sub>aa</sub> | S <sub>q</sub> | X       | S <sub>q</sub> | S <sub>h</sub> |
| 1        | 1.48    | 1.49           | 1.51           | 15.63   | 1.49            | 0.90           | -0.28   | 0.94           | 1.51           |
| 2        | 1.72    | 1.74           | 1.75           | -3.06   | 1.74            | 0.95           | 0.15    | 0.95           | 1.75           |
| 3        | 1.41    | 1.42           | 1.56           | 2.45    | 1.42            | 0.55           | -0.02   | 0.55           | 1.56           |
| 4        | 1.51    | 1.51           | 1.51           | 2.20    | 1.51            | 0.81           | -0.12   | 0.81           | 1.51           |
| 5        | 1.42    | 1.38           | 1.32           | 6.22    | 1.38            | 0.74           | -1.73   | 0.74           | 1.32           |
| 6        | 1.43    | 1.48           | 1.57           | 4.05    | 1.48            | 0.78           | 0.31    | 0.78           | 1.57           |
| 7        | 1.23    | 1.04           | 1.38           | 2.15    | 1.04            | 0.85           | 0.35    | 0.85           | 1.38           |

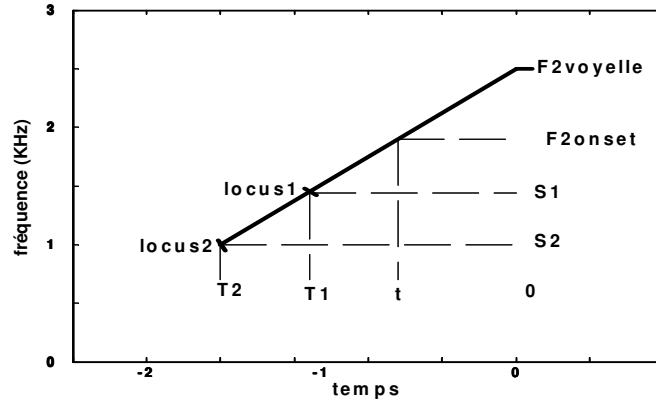
**Tableau 5.** Intersection de droites de locus. X : intersection. S<sub>aa</sub>, S<sub>q</sub> et S<sub>h</sub> : locus de [ʔ], [q] et [h].

On peut remarquer que dans le cas de [ʔ] avec [h], les droites se croisent au locus qui est le même pour les deux consonnes et localisé dans la plus part des cas exactement au voisinage de la voyelle [a] et pas entre deux voyelles (intersection des droites de locus avec la droite en discontinu  $F2_{onset} = F2_{voyelle}$ ). Cependant, dans le cas de [ʔ] avec [q] les droites ne se croisent pas au locus et pas forcément au voisinage des voyelles.



**Figure 28.** Intersection des droites de locus par locuteur.

Maintenant, nous allons montrer l'importance de la droite de locus par rapport au locus. Soit deux consonnes de locus S1 et S2 et positionnées dans le temps T1 et T2 comme on le voit dans la figure 29.



**Figure 29.** Transitions F2 vers deux locus différents

A partir de la figure ci-dessus, on peut écrire l'équation suivante :

$$\frac{(F2_{\text{voyelle}} - S_1)}{T_1} = \frac{(F2_{\text{voyelle}} - S_2)}{T_2} \quad \text{ce qui donne} \quad F2_{\text{voyelle}} = \frac{T_2 S_1 - T_1 S_2}{T_2 - T_1} = V$$

Cette équation représente la région de confusion déterminée par la voyelle V qui pointe vers deux différents locus à la fois. Afin d'obtenir les droites de locus des deux consonnes, on relève les  $F2_{\text{onset}}$  au temps t. Les équations des deux droites sont :

$$F2_{\text{onset1}} = \left(1 - \frac{t}{T_1}\right) F2_{\text{voyelle}} + S_1 \frac{t}{T_1} \quad \text{et} \quad F2_{\text{onset2}} = \left(1 - \frac{t}{T_2}\right) F2_{\text{voyelle}} + S_2 \frac{t}{T_2}$$

Ces deux droites se croisent au point  $F2_{\text{onset1}} = F2_{\text{onset2}} = F2_{\text{onset}}$ . Ceci nous donne

$$\left(1 - \frac{t}{T_1}\right) F2_{\text{voyelle}} + S_1 \frac{t}{T_1} = \left(1 - \frac{t}{T_2}\right) F2_{\text{voyelle}} + S_2 \frac{t}{T_2} \quad (2)$$

ou bien

$$F2_{\text{voyelle}} = \frac{T_2 S_1 - T_1 S_2}{T_2 - T_1} = V$$

Donc, indépendamment de la segmentation, on retrouve la même région de confusion la voyelle V qui pointe vers les deux locus dans la même direction. Dans ce cas, la droite de

locus n'apporte rien de plus par rapport au locus classique. Maintenant, si on segmente les deux consonnes à des moments différents  $t_1$  et  $t_2$  ; l'équation (2) devient alors :

$$(1 - \frac{t_1}{T_1}) F2_{\text{voyelle}} + S_1 \frac{t_1}{T_1} = (1 - \frac{t_2}{T_2}) F2_{\text{voyelle}} + S_2 \frac{t_2}{T_2} \quad (3)$$

posons  $\frac{t_1}{T_1} = q_1$  et  $\frac{t_2}{T_2} = q_2$  ; de l'équation (3) on déduit

$$F2_{\text{voyelle}} = V = \frac{(q_2 S_2 - q_1 S_1)}{(q_2 - q_1)} \quad (4)$$

En variant  $q_1$  ou  $q_2$  (décalage de la segmentation d'une consonne par rapport à l'autre), on fait varier la région de confusion  $V$  (intersection et frontière entre les droites de locus). C'est ce qui marque l'importance de la droite de locus par rapport au locus. *Donc, en contrôlant le « timing » au niveau de la production, on peut faire changer les régions de confusion. Ceci semble nécessaire surtout dans les VCV en faisant des régions de confusion différentes à droite et à gauche.* Le second formant est choisi tout simplement à cause de sa portée large par rapport aux autres formants. Donc, plus d'espace est attribué pour séparer entre les régions de confusion.

En se référant au paragraphe 4 chapitre I, il a été observé que deux droites de locus différentes ont été tracées pour une même consonne. Montrant ainsi une coarticulation (pente) à gauche plus importante que celle à droite. Aussi, il a été observé que la durée du bruit des fricatives est plus longue que la durée du silence des occlusives [12]. Donc pour le cas de [ʃ] avec [q], en considérant deux droites de locus par consonne (à gauche et à droite), il serait fort possible que les échantillons qui seront dans la région de confusion des premières droites de locus seront séparés dans les secondes droites de locus et par conséquent une meilleure séparation pourrait être obtenue. Dans le cas de [ʃ] avec [h] qui ont le même locus, la segmentation ne peut pas changer la région de confusion comme on peut voir ça, une seconde fois, à partir de l'équation (4). Si  $S_1 = S_2 = S$  on obtient  $V = S$  (même région de confusion « V »). Donc, le 3<sup>e</sup> formant, ou bien peut être le premier formant, devrait prendre des valeurs distinctes dans cette région de confusion.

## Conclusion

Dans ce chapitre, nous avons trouvé que les droites de locus obtenues reflétaient bien la coarticulation. Ceci a été expliqué en suivant les interprétations données dans la littérature. Quant au lieu d'articulation, les résultats obtenus pour le [χ] dans les travaux de Yeou (1997) contredisent nos résultats trouvés pour le [q] se partageant le même lieu d'articulation. Donc, la droite de locus ne permet pas de spécifier le lieu d'articulation en général. D'autres formants notamment le troisième doit apporter sa contribution.

En observant les intersections entre droites de locus, nous avons remarqué que ces régions de conflit varient considérablement d'un locuteur à l'autre dans le cas où les droites régressives ne correspondent pas à un locus commun des phonèmes considérés. L'étude de ces régions de conflit entre locus et droite de locus nous a amené à proposer une analyse montrant ainsi l'importance de la droite de locus. Cette dernière permet donc de faire déplacer les intersections en contrôlant le timing au niveau de la production. Ceci semble utile aussi dans le cas des /VCV/ en faisant des régions de conflit différentes à droite et à gauche. Donc, dans le cas d'une consonne en milieu de mot, deux droites de locus seront nécessaires à droite et à gauche, pour déterminer le lieu d'articulation.

Dans le cas où les consonnes ont le même locus, l'intersection ne change pas avec la segmentation. La contribution des autres formants, notamment le 3<sup>e</sup> formant, devient impérative dans la détermination du lieu d'articulation. C'est ce que nous allons voir dans le chapitre suivant.

# Chapitre III

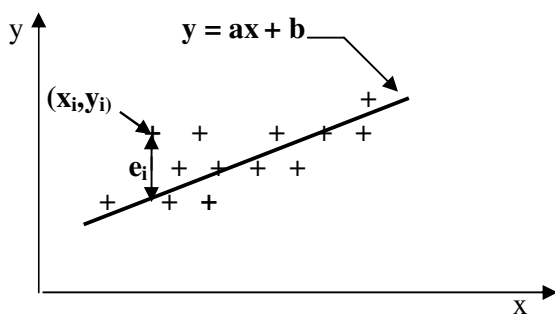
## Analyse discriminante

### 1. Introduction

L'objectif est de réaliser un classifieur pour discriminer les trois phonèmes [ʔ] (ع), [q](ق) et [h](ه). Le nombre total d'échantillons par consonne est de 210. La moitié des échantillons sera utilisée pour l'apprentissage et l'autre moitié pour le test. En plus du second formant, le troisième formant sera aussi utilisé puisqu'il est impliqué directement dans la perception des consonnes. Afin de mesurer l'apport du 3<sup>e</sup> formant dans la discrimination de ces sons, d'abord le 2<sup>e</sup> formant sera utilisé seul ensuite les deux formant F2 et F3 (F3<sub>onset</sub>) seront utilisés ensemble.

### 2. Aspect théorique

Nous allons utiliser la procédure mathématique suivante [28, 29] pour déterminer la régression linéaire de nos échantillons.



**Figure 30.** La régression linéaire

Dans la figure 30 nous considérons la distribution de nos échantillons déterminés par leur coordonnées  $(x_i, y_i)$ , la régression linéaire consiste à trouver la droite du type  $y = ax + b$ , qui passe à travers le plus grand nombre de ces points  $(x_i, y_i)$  de la sorte que la somme des carrés des offsets verticaux  $e_i$  (écarts par rapport à la droite) soit minimum.

$$e^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - (ax_i + b)]^2 \quad \text{doit être minimum.}$$

Ce qui implique

$$\begin{cases} \frac{\delta e^2}{\delta b} = -2 \sum_{i=1}^n [y_i - (ax_i + b)] = 0 \\ \frac{\delta e^2}{\delta a} = -2 \sum_{i=1}^n [y_i - (ax_i + b)] x_i = 0 \end{cases}$$

ce qui mène aux équations

$$\begin{cases} nb + a \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ b \sum_{i=1}^n x_i + a \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

dont les solutions

$$a = \frac{\left( \sum_{i=1}^n x_i y_i \right) - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad \text{et} \quad b = \bar{y} - a \bar{x}$$

où  $\bar{y}$  et  $\bar{x}$  sont respectivement les moyennes  $\frac{\sum_{i=1}^n y_i}{n}$  et  $\frac{\sum_{i=1}^n x_i}{n}$ .

« **a** » et « **b** » sont donc respectivement la pente et l'ordonnée à l'origine de la droite régressive représentant l'ensemble des échantillons, c'est la droite de locus :

$$F2_{\text{onset}} = a \cdot F2_{\text{voyelle}} + b$$

### 3. Elaboration des vecteurs de représentation

On suppose qu'il existe une droite parallèle à la droite de locus pour tout point  $(x_i, y_i)$ , par conséquent ce point peut être caractérisé uniquement par son ordonnée à l'origine  $b_i$ .

$$b_i = -F2_{\text{voyelle } i} \cdot a + F2_{\text{onset } i}, \quad \text{« a » étant la pente de la droite de locus.}$$

Dans le cas où on veut utiliser en même temps, les deux formants F2 et F3, on peut considérer la surface régressive  $F3_{\text{onset}} = \text{fonction}(F2_{\text{onset}}, F2_{\text{voyelle}})$ . Le modèle d'équation utilisé pour déterminer les surfaces régressives est de la forme quadratique :

$$z = a_0 + a_1(y+px) + a_2(y+px)^2 \quad , \quad z = f_{3\text{onset}} \quad , \quad y = f_{2\text{onset}} \quad , \quad x = f_{2\text{voyelle}}$$

et  $p$  étant la pente de la droite de locus de la consonne correspondante. C'est une surface orientée dans la direction de la droite de locus.

Comme dans le cas précédent, pour tout point donné  $(F3_{\text{onset}}, F2_{\text{onset}}, F2_{\text{voyelle}})$ , il existe une surface parallèle à la surface régressive passant par ce point. L'intersection avec l'axe des  $z$  donne l'ordonnée à l'origine  $b_3$  et est déterminée de la même manière :

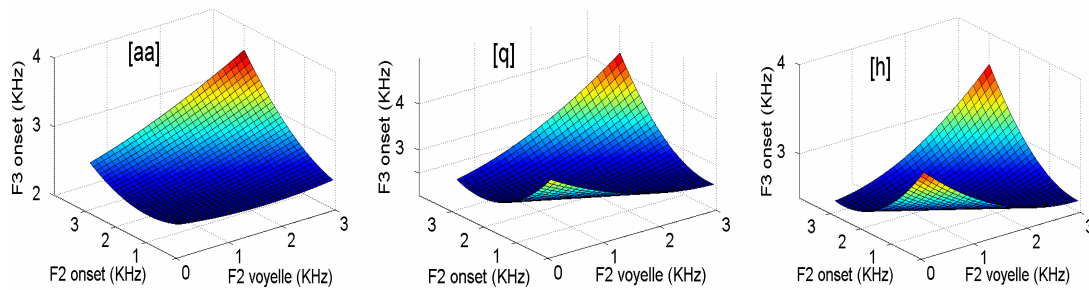
$$F3_{\text{onset}} = a_0 + a_1(y+px) + a_2(y+px)^2$$

Donc, l'ordonnée à l'origine est :

$$b_3 = a_0 = F3_{\text{onset}} - a_1(y+px) - a_2(y+px)^2 \quad \text{où} \quad x = F2_{\text{onset}}, \quad \text{et} \quad y = F2_{\text{voyelle}}.$$

On obtient alors le vecteur descripteur :  $\mathbf{V} = (b_2, b_3)^T$ .

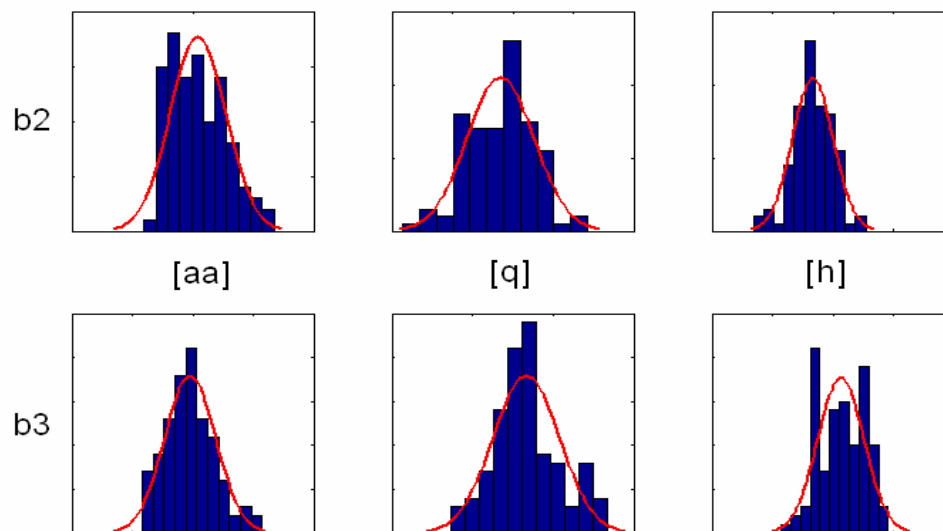
Pour le cas des phonèmes  $[\gamma]$ ,  $[q]$  et  $[h]$ , la figure 31 montre comment le  $F3_{\text{onset}}$  varie à travers la région occupée par les points  $(F2_{\text{onset}}, F2_{\text{voyelle}})$ .



**Figure 31.** Surfaces régressives  $F3_{\text{onset}} = \text{fonction}(F2_{\text{onset}}, F2_{\text{voyelle}})$  de  $[\gamma]$  (aa),  $[q]$  et  $[h]$  .



Afin d'observer la distribution de nos échantillons, les histogrammes superposés à la distribution normale des composantes  $b_2$  et  $b_3$  sont montrés dans la figure 32.



**Figure 32.** Histogrammes des composantes du vecteur de représentation.  $V = (b_2, b_3)^T$ .

#### 4. Classification en utilisant l'algorithme de Bayes

Il est classique en reconnaissance de forme d'utiliser des approches paramétriques qui consistent à faire une hypothèse sur la nature de la distribution des données puis à estimer les paramètres de la distribution en utilisant les données d'apprentissage.

L'algorithme de *Bayes* [30] est basé sur l'hypothèse que les données de chacune des classes suivent une distribution normale multi variable :

$$p(x|\omega_j) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_j|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)\right) \quad (1)$$

où  $\Sigma_j$  et  $\mu_j$  sont la matrice de covariance et le vecteur moyenne des données de la classe  $\omega_j$  et  $d$  la dimension des vecteurs. Maintenant si on observe la distribution de nos échantillons

sur la figure 32, on voit bien que les histogrammes sont proches de la normale. Le modèle de Bayes pourrait donc être appliqué.

La phase d'apprentissage consiste alors à utiliser les données d'apprentissage afin d'estimer pour chacune des classes la moyenne et la matrice de covariance. Etant donnée  $X$  le vecteur paramétrique représentant les échantillons d'apprentissage ;  $\mu$  c'est la moyenne de  $X$  et  $\Sigma$  c'est la covariance de  $X$ . Ces paramètres sont sauvegardés de manière à pouvoir être utilisés par la suite pour classer les exemples de test en utilisant la règle de Bayes afin d'estimer les probabilités à posteriori :

$$P(\omega_j | x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)} \quad (2)$$

où  $P(\omega_j)$  représente la probabilité à priori de la classe  $\omega_j$  et  $p(x)$  l'évidence qui est définie par :

$$p(x) = \sum_{j=1}^c p(x|\omega_j)P(\omega_j) \quad (3)$$

où  $c$  représente le nombre de classes du problème. En pratique il est préférable de calculer les fonctions discriminantes associées à chacune des classes qui sont obtenues en prenant le logarithme naturel de l'équation (2) et en supprimant les termes constants :

$$d_j(x) = \ln P(\omega_j) - \frac{1}{2} \ln |\Sigma_j| - \frac{1}{2} (x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j) \quad (4)$$

Notons que dans notre cas, les trois classes sont équiprobables ; c'est à dire que toutes les classes ont la même probabilité à priori  $P(\omega_j) = 1/c$ . le premier terme de l'équation (4) n'intervient pas dans la décision et peut donc être supprimé. On obtient alors :

$$d_j(x) = -\frac{1}{2} \ln |\Sigma_j| - \frac{1}{2} (x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j)$$

Enfin, la décision est prise en cherchant la valeur maximale parmi les fonctions discriminantes associées à chacune des classes.

## 5. Résultats et interprétations

Afin d'observer la contribution du troisième formant dans la discrimination, on utilise d'abord la transition du second formant ( $F2_{\text{onset}}$ ,  $F2_{\text{voyelle}}$ ) ; ensuite les deux formants F2 et F3 ( $F2_{\text{onset}}$ ,  $F2_{\text{voyelle}}$ ,  $F3_{\text{onset}}$ ) en même temps. Les résultats sont montrés dans le tableau 6.

|       | [ʃ] | [q] | [h] |
|-------|-----|-----|-----|
| F2    | 58  | 9   | 33  |
| [ʃ]   |     |     |     |
| F2+F3 | 71  | 9   | 20  |
| F2    | 10  | 89  | 2   |
| [q]   |     |     |     |
| F2+F3 | 9   | 89  | 3   |
| F2    | 4   | 5   | 91  |
| [h]   |     |     |     |
| F2+F3 | 8   | 5   | 88  |

**Tableau 6.** Résultat de classification en utilisant l'algorithme de Bayes.

Nous remarquons que le 3<sup>e</sup> formant contribue dans l'identification de [ʃ] et non dans l'identification des autres sons [q] et [h]. Aussi, le taux d'identification de ces derniers est plus élevé.

Si nous utilisons notre classifieur pour discriminer deux classes et en considérant toutes les paires possibles nous obtenons les résultats montrés dans le tableau 7.

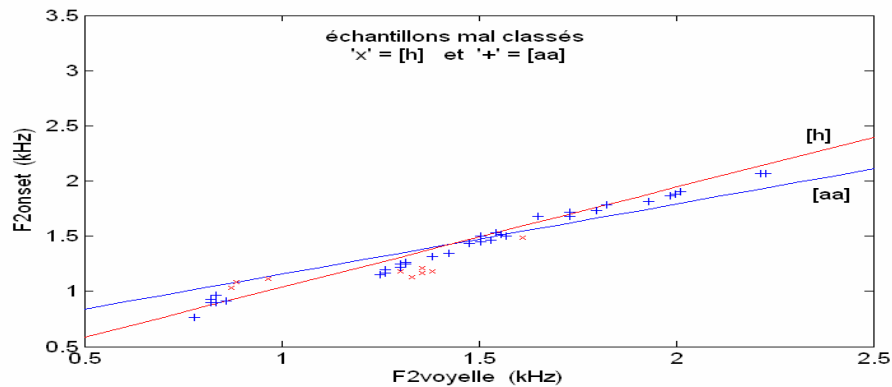
|         | [ʃ] | [q] | [ʃ] | [h] | [q] | [h] |
|---------|-----|-----|-----|-----|-----|-----|
| F2      | 90  | 89  | 65  | 91  | 95  | 95  |
| F2 + F3 | 89  | 90  | 78  | 87  | 95  | 92  |

**Tableau 7.** Résultat de classification binaire en utilisant l'algorithme de Bayes.

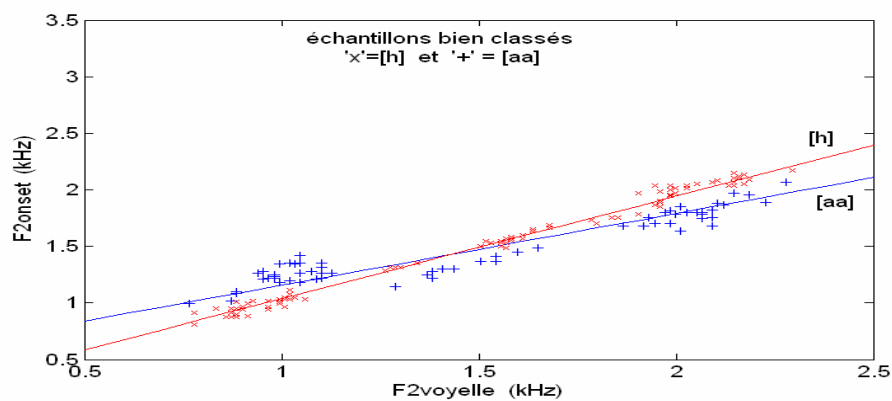
Aussi la même remarque, dans le cas de [ʃ] avec [h], c'est le [ʃ] qui a le taux d'identification le moins élevé et aussi seulement dans le cas de ce dernier que le 3<sup>e</sup> formant contribue dans son identification.

### a- Analyse des résultats

Comme nous avons déterminé les composantes de nos vecteurs de représentation par rapport aux droites de locus pour le second formant et aux surfaces régressives pour le 3<sup>e</sup> formant, nous allons observer les échantillons mal classés par rapport à ces courbes régressives et essayer de trouver les explications.



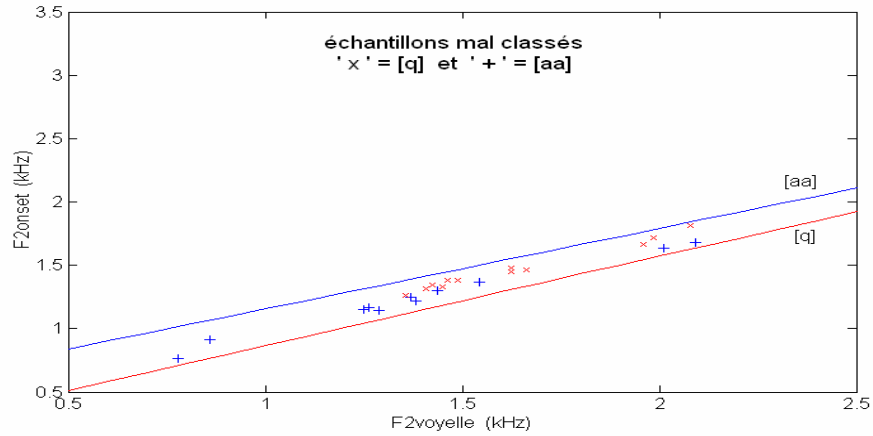
(a)



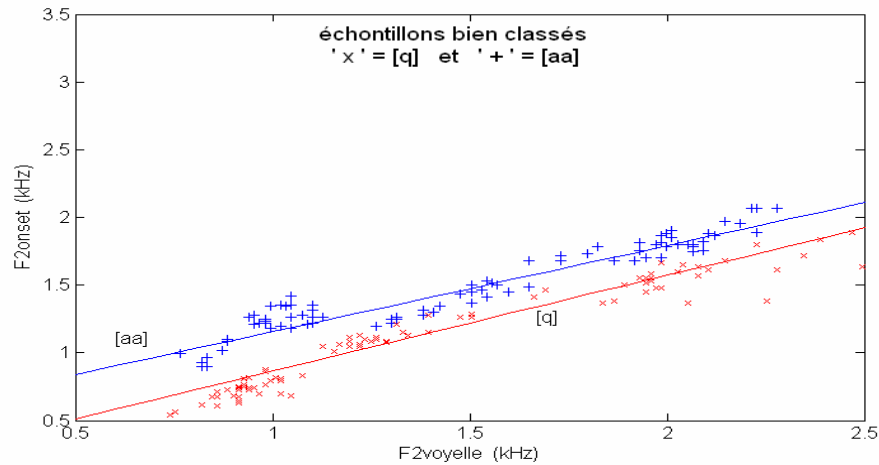
(b)

**Figure 33.** Résultat de discrimination entre [ɣ] et [h] en utilisant F2 seul  
(a) Echantillons mal classés. (b) Echantillons bien classés

En se référant aux figures 33a pour le cas de [ɣ] avec [h] où seul le second formant est utilisé pour discriminer ces sons, on observe bien que les échantillons mal classés se trouvent autour de l'intersection des deux droites de locus ainsi qu'entre ces dernières proches de leur frontière commune. Cependant, les échantillons bien classés, voir figure 33b, sont en général bien alignés sur les droites ou bien à l'extérieure de ces dernières du côté de leur classe correspondantes.



(a)

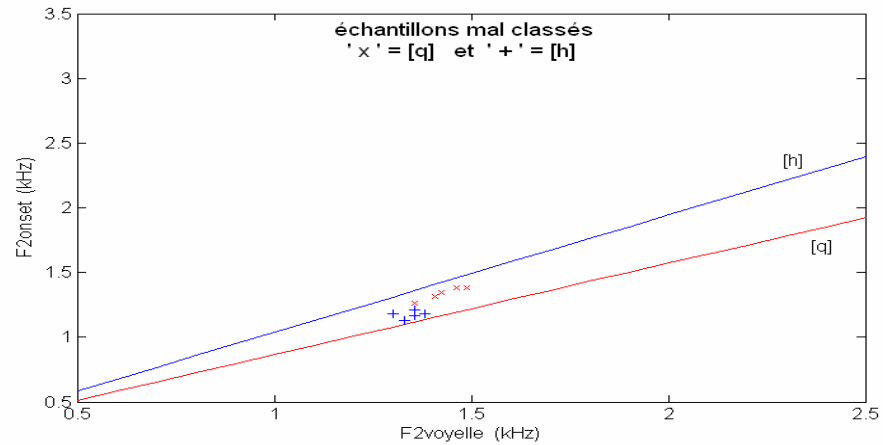


(b)

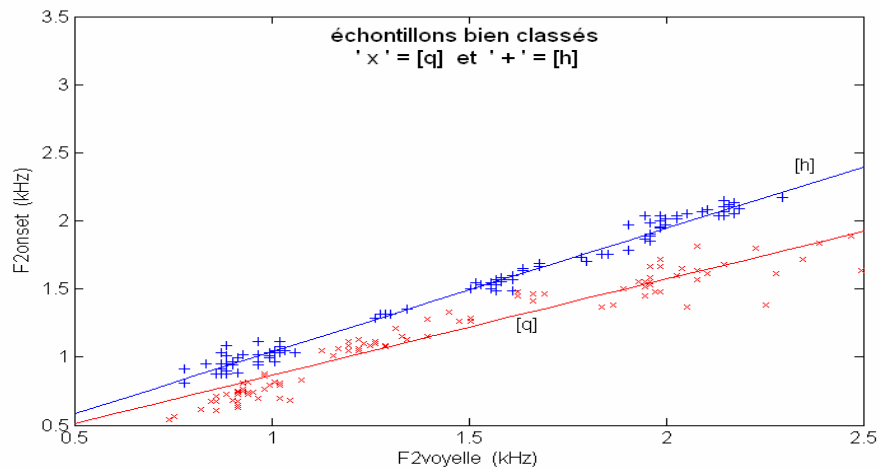
**Figure 34.** Résultat de discrimination entre [ʃ] et [q] en utilisant F2 seul  
(a) Echantillons mal classés. (b) Echantillons bien classés

Dans le cas de [ʃ] avec [q], les droites de locus des deux classes sont presque parallèles. Les échantillons mal classés se trouvent entre les deux droites, le long de leur frontière commune dont la plus part est dans la région de la voyelle [a], voir figure 34a. Ceci est dû au fait que dans cette région les échantillons de [q] sont au dessus de leur droite de locus tandis que ceux de [ʃ] sont un peu au dessous, voir figures 16 et 17. D'autres échantillons mal classés sont dans la région de la voyelle [i]. Ceci est dû au rapprochement de leurs droites de locus dans cette région. Cependant, les échantillons bien classés sont proches de leur droite correspondante, aussi bien de l'intérieur que de l'extérieur, voir figure 34b.

Dans le cas de [q] avec [h], les échantillons mal classés se trouvent dans la région de la voyelle [a], voir figure 35a, pour la même raison que dans cette région les échantillons de [q] sont au dessus de leur droite de locus, donc proches de celle de [h]. Aussi, les échantillons bien classés sont proches de leur droite correspondante, aussi bien de l'intérieur que de l'extérieur, voir figure 35b.



(a)



(b)

**Figure 35.** Résultat de discrimination entre [h] et [q] en utilisant  $F2$  seul  
(a) Echantillons mal classés. (b) Echantillons bien classés

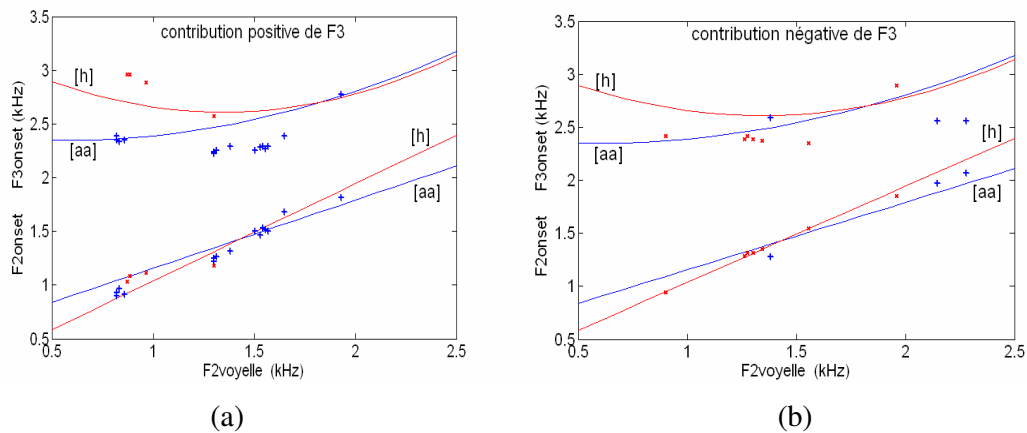
La contribution du 3<sup>e</sup> formant est visualisée dans les figures ci-dessous. Notant que les surfaces régressives montrées dans ces figures sont des intersections avec le plan perpendiculaire au plan  $F2_{onset}$ ,  $F2_{voyelle}$  passant par la frontière (médiane) commune entre les deux droites de locus.

Dans le cas de [ɣ] avec [h], la figure 36a montre la contribution positive, en d'autres termes, les échantillons qui ont été mal classés en utilisant le 2<sup>e</sup> formant seul et qui sont devenus bien classés en utilisant les deux formants F2 et F3. Ces échantillons sont au nombre de dix-neuf (19). Sept (07) de ces échantillons, se trouvent dans la région de la voyelle [u] où les surfaces régressives des deux classes représentant le 3<sup>e</sup> formant sont bien séparées et les échantillons aussi. Onze (11) autres se trouvent dans la région de la voyelle [a], où malgré que les surfaces régressives sont très proches, ces échantillons sont séparés parce qu'ils se trouvent à l'extérieur et du côté de la classe correspondante ([ɣ]).

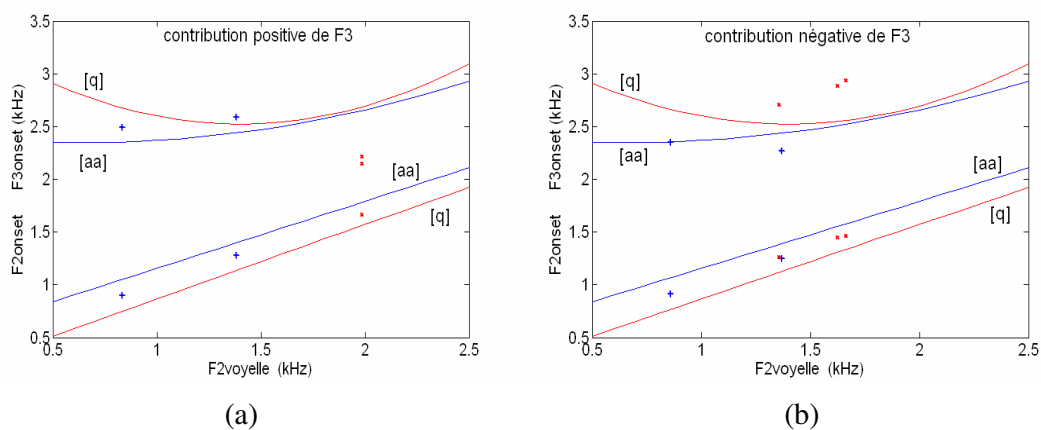
La figure 36b montre la contribution négative. Ce sont les échantillons qui ont été bien classés avec le 2<sup>e</sup> formant seul et devenus mal classés en ajoutant le 3<sup>e</sup> formant. Ces échantillons sont au nombre de onze (11). Sept (07) d'entre eux se trouvent dans la région de la voyelle [a] où les surfaces régressives sont très proches et trois (03) dans la région de [i] où les surfaces régressives sont confondues.

Dans le cas des autres paires [q] avec [ɣ] et [q] avec [h], aucune intersection des droites de locus n'est observée, voir figures 37 et 38. Dans le cas de [ɣ] avec [q], il n'y a pratiquement pas de contribution du 3<sup>e</sup> formant avec l'apport d'un seul échantillon. Les échantillons mal classés sont dans la région des deux voyelles [a] et [i] où les surfaces régressives des deux classes sont confondues. Dans le cas de [q] avec [h], on observe plutôt la perte d'un échantillon. Les surfaces régressives de ces deux sons sont aussi confondues.

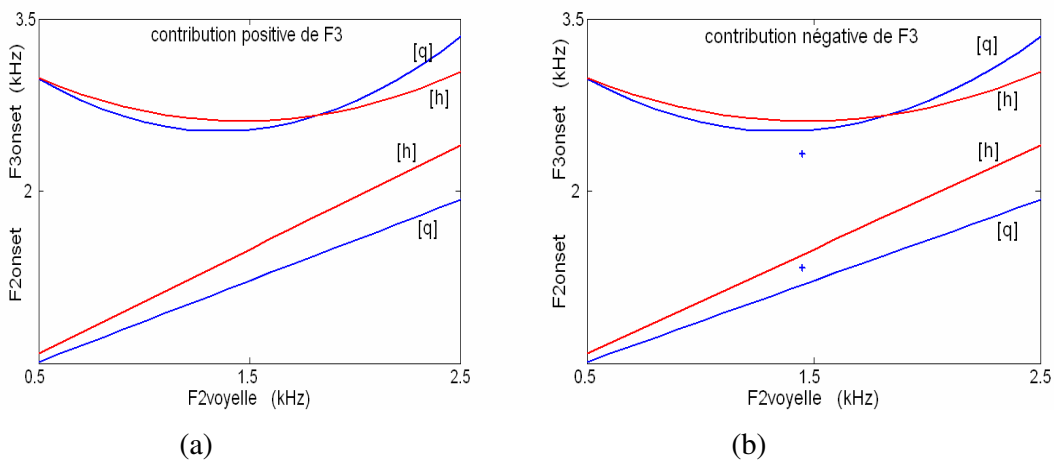
L'ensemble de ces observations se résume dans le tableau 8. Dans le cas de [ɣ] avec [h], autour de la voyelle [a], l'apport du 3<sup>e</sup> formant est de onze (11) moins sept (07), donc seulement quatre (04) échantillons sur trente (30) mal classés avec F2 seul. Dans la région de la voyelle [u], la contribution du 3<sup>e</sup> formant est de sept moins un, donc six (06) échantillons sur neuf (09) mal classés avec F2 seul. Relativement, la contribution de F3 est plus importante dans la région de la voyelle [u] que dans celle de [a] où se croisent les droites de locus, voir figure 39. *Le 3<sup>e</sup> formant n'apporte pas vraiment l'information distinctive là où se croisent les droites de locus comme avancé dans la littérature [24].*



**Figure 36.** Contribution du 3<sup>e</sup> formant dans la discrimination entre [ʃ] et [h]



**Figure 37.** Contribution du 3<sup>e</sup> formant dans la discrimination entre [ʃ] et [q]



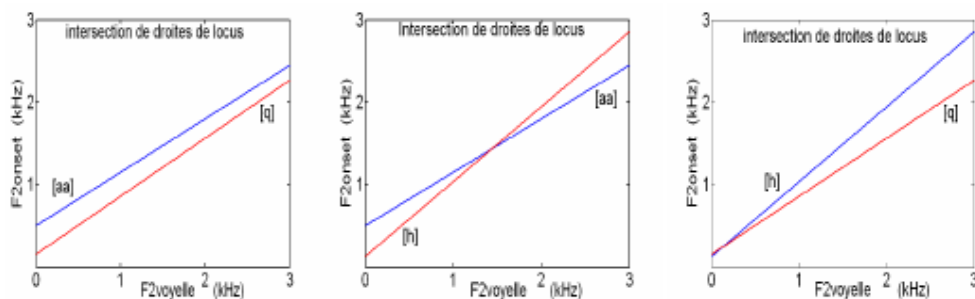
**Figure 38.** Contribution du 3<sup>e</sup> formant dans la discrimination entre [q] et [h]



La contribution relative du 3<sup>e</sup> formant pour les différentes paires est de quatre (04) négatifs pour cinq (05) positifs dans le cas de [ʃ] avec [q], onze (11) négatifs pour dix neuf (19) positifs dans le cas de [ʃ] avec [h] et un (01) négatif pour zéro (0) positif dans le cas de [q] avec [h]. *La contribution négative du 3<sup>e</sup> formant laisse penser que les frontières entre ces sons étudiés sont beaucoup plus complexes que l'algorithme de Bayes ne pourrait générer.*

| Phonèmes                                        | [ʃ] - [q] |     |     | [ʃ] - [h] |     |     | [q] - [h] |     |     |
|-------------------------------------------------|-----------|-----|-----|-----------|-----|-----|-----------|-----|-----|
|                                                 | [u]       | [a] | [i] | [u]       | [a] | [i] | [u]       | [a] | [i] |
| Echantillons mal classés (F2 seul utilisé)      | 2         | 16  | 5   | 9         | 30  | 7   | 0         | 10  | 0   |
| Contribution positive du 3 <sup>e</sup> formant | 1         | 2   | 2   | 7         | 11  | 1   | 0         | 0   | 0   |
| Contribution négative du 3 <sup>e</sup> formant | 1         | 1   | 2   | 1         | 7   | 3   | 0         | 1   | 0   |
| Contribution moyenne du 3 <sup>e</sup> formant  | 0         | 1   | 0   | 6         | 4   | -2  | 0         | -1  | 0   |

**Tableau 8.** Résumé de la contribution du 3<sup>e</sup> formant



**Figure 39.** Intersection des droites de locus

#### *b- Variabilité du 3<sup>e</sup> formant*

Le modèle utilisé par l'algorithme de Bayes est caractérisé par la moyenne représentée par les droites et surfaces régressives que nous avons utilisées et aussi la matrice de

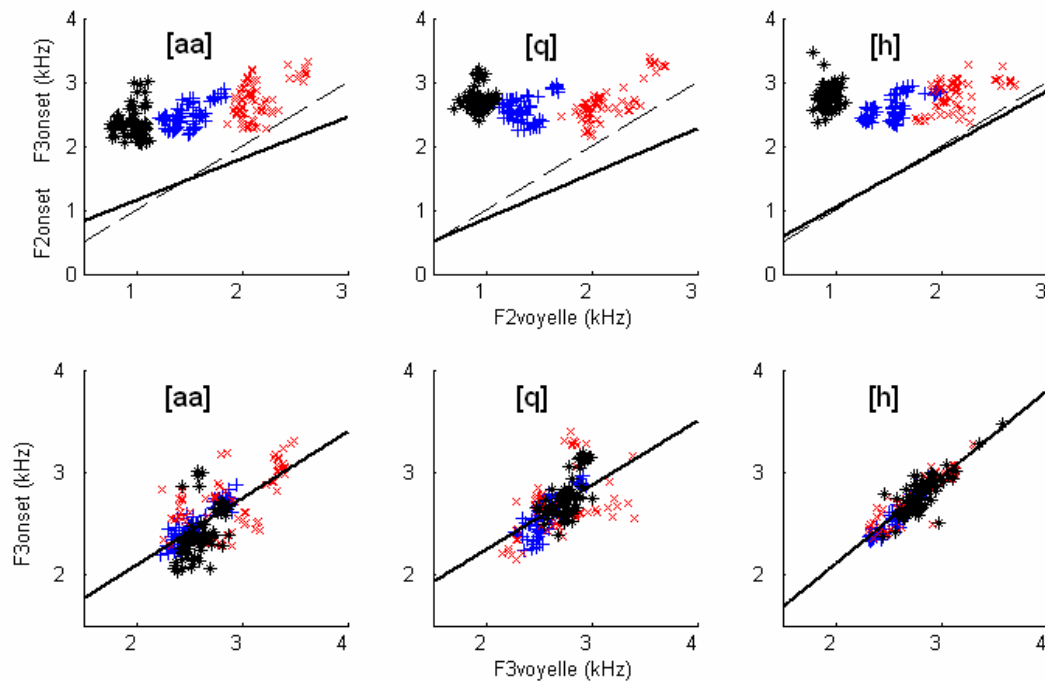
covariance représentant la variabilité des échantillons que nous allons observer par rapport aux droites de locus qui est visualisée dans la figure 40.

En considérant toutes les consonnes, on peut remarquer que le  $F3_{\text{onset}}$  est plus stable dans la région où la coarticulation est maximale (la droite en discontinu  $F2_{\text{onset}} = F2_{\text{voyelle}}$ ). Et plus on s'éloigne de cette droite, plus la variabilité augmente, voir figure 40 (en haut). Donc la variabilité du troisième formant augmente lorsque la coarticulation diminue. Ceci se confirme encore dans la figure 40 (en bas) qui nous montre la distribution des  $(F3_{\text{onset}}, F3_{\text{voyelle}})$ . On observe une régression linéaire pour le cas de [h] qui coarticule bien avec les voyelles et une distribution aléatoire dans le cas de [ʃ] et [q] qui coarticulent moins bien avec les voyelles. Les variances par voyelle des composantes de notre vecteur de représentation sont montrées dans le tableau 9.

|     | [u]    | [a]    | [i]    |
|-----|--------|--------|--------|
| b2  | 0.0138 | 0.0062 | 0.0102 |
| [ʃ] |        |        |        |
| b3  | 0.0507 | 0.0250 | 0.0581 |
| b2  | 0.0048 | 0.0029 | 0.0160 |
| [q] |        |        |        |
| b3  | 0.0386 | 0.0357 | 0.0511 |
| b2  | 0.0037 | 0.0048 | 0.0031 |
| [h] |        |        |        |
| b3  | 0.0375 | 0.0326 | 0.0400 |

**Tableau 9.** Variances des composantes (b2, b3) du vecteur de représentation

On peut remarquer que *la variabilité du troisième formant suit tout simplement celle du second formant mais plus amplifiée*. Dans le cas de [ʃ], la variance est minimale avec la voyelle [a] qui est proche de la droite  $F2_{\text{onset}} = F2_{\text{voyelle}}$  et prend de grandes valeurs avec les voyelles [i] et [u] qui s'écartent de la droite  $F2_{\text{onset}} = F2_{\text{voyelle}}$ . Dans le cas de [q], la variance est maximale avec la voyelle [i] qui est la plus éloignée de la droite  $F2_{\text{onset}} = F2_{\text{voyelle}}$ .

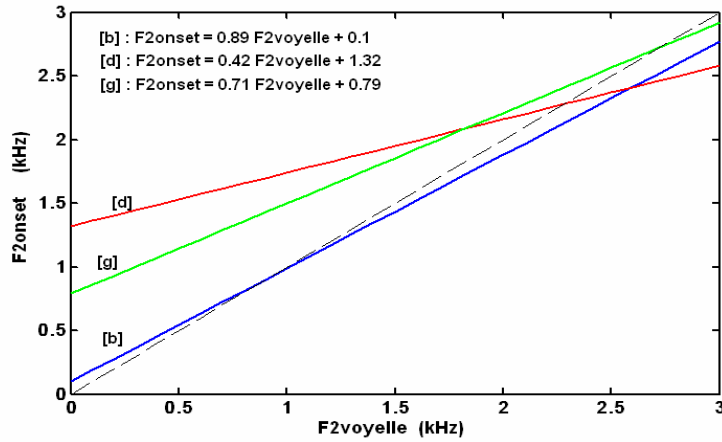


**Figure 40.** Variabilité de F3. [a] = '+'; [i] = 'x'; [u] = '\*'  
 En haut : F3<sub>onset</sub> par rapport à la droite de locus F2.  
 En bas: droite de locus F3.

### c- Interaction entre phonèmes

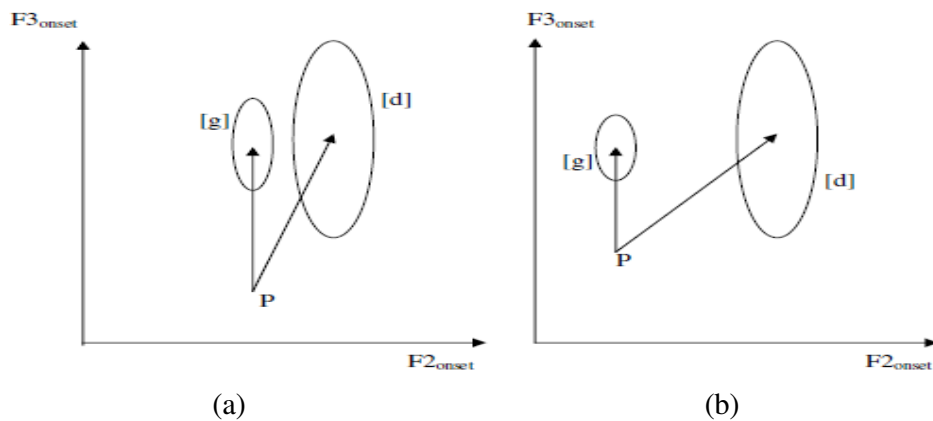
La variabilité du 3<sup>e</sup> formant observée dans le paragraphe précédent pourrait donner une explication au problème du phonème [g] avec les voyelles arrondies lors des transitions /CV/ qui ne pointent pas vers le locus, voir figure 5. En d'autres termes les points (F2<sub>onset</sub>, F2<sub>voyelle</sub>) sont un peu déviés de la droite de locus, voir figure 9. En fait, des travaux de synthèse [9] ont montré qu'en pointant la transition de [g] avec voyelle arrondie vers le locus c'est le [d] qui est entendu au lieu de [g].

Nous considérons les valeurs utilisées dans la littérature [14,24] pour le tracé des droites de locus de [b], [d] et [g], voir figure 41. Dans la région des voyelles arrondies, la droite de locus de [d] est plus éloignée que celle de [g] par rapport à la droite F2<sub>onset</sub> = F2<sub>voyelle</sub> (droite en discontinu). Donc, la variabilité du 3<sup>e</sup> formant de [d] est supérieure à celle de [g] dans cette région des voyelles arrondies. La distance de Mahalanobis qu'on a utilisée dans l'algorithme de Bayes est une distance normalisée par la variance dans la direction de la moyenne.



**Figure 41.** Droites de locus de [b], [d] et [g]

En se référant à la figure 42a le point P est plus proche du centre de [g] que du centre de [d]. Mais si on normalisait par la variance dans la direction du centre, on voit bien que le point P est à moins de trois mesures de [d], tandis que par rapport à [g] est à plus de trois mesures et c'est la raison pour laquelle on entend le [d]. Pour entendre le [g], le point P doit dévier un peu en s'éloignant de [d] (figure 42b) et c'est le décalage des points par rapport à la droite de locus de [g] observé dans la région des voyelles arrondies. Cette interaction entre phonèmes est observée aussi dans le cas de [ɣ] et de [q] où les points ( $F2_{\text{onset}}$ ,  $F2_{\text{voyelle}}$ ) ne sont pas aussi bien alignés sur leurs droites respectives que dans le cas de [h], voir figures 16 et 17.



**Figure 42.** Ellipses de dispersion de ( $F2_{\text{onset}}$ ,  $F3_{\text{onset}}$ ) de [d] et [g] avec voyelle arrondie.

En appliquant l'algorithme de Bayes, nous avons supposé que les composantes de notre vecteur de représentation suivent une distribution normale. Ceci est équivalent à ce que les échantillons soient contenus dans une ellipse. Seulement, dans notre cas cette ellipse n'est pas stable à travers les différentes voyelles. En d'autres termes, cette ellipse se rétrécit dans la région du locus. Par conséquent, les résultats de classification fournis par l'algorithme de Bayes dans cet espace ne pourraient être fiables. Pour cela, nous devons faire subir à nos échantillons une transformation afin d'obtenir une variance stable à travers les voyelles, et par conséquent, l'algorithme de Bayes pourrait être utilisé en donnant certainement de meilleurs résultats.

## 6. Espace de redescription

### *a- Introduction*

En analyse de données, il est classique de procéder à un changement d'espace afin d'accéder à une représentation des données plus informative. Ce changement d'espace implique l'utilisation d'une fonction  $\Phi$  et d'un espace  $F$ .

$$\Phi : E \rightarrow F$$

Les données de l'espace original (ou de départ)  $E$  sont appelées « attributs » et les données transformées « caractéristiques » (feature). Maintenant, nous définissons le produit scalaire dans ce nouvel espace. Un produit scalaire noté  $\langle \cdot, \cdot \rangle$  est une forme bilinéaire (linéaire pour chacune des deux variables) et possède de plus les trois propriétés suivantes :

- Un produit scalaire est **symétrique**, ce qui signifie :  $\forall x, y \in E \quad \langle x, y \rangle = \langle y, x \rangle$
- Un produit scalaire est **positif**, ce qui signifie :  $\forall x \in E \quad \langle x, x \rangle \geq 0$
- Un produit scalaire est **défini**, ce qui signifie :  $\forall x \in E \quad (\langle x, x \rangle = 0 \Rightarrow x = 0)$

Donc un **produit scalaire** est une forme bilinéaire symétrique définie positive.

Une façon efficace de calculer le produit scalaire dans l'espace de redescription est d'utiliser la fonction noyau ou kernel. Une fonction noyau est une fonction à deux variables qui permet de calculer directement le produit scalaire  $\langle \Phi(x_i), \Phi(x_j) \rangle$  comme une fonction de  $x_i$  et  $x_j$ , soit  $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ . Pour que la fonction  $k$  soit un noyau, elle doit répondre aux conditions de Mercer [32,33] suivantes :

- elle doit être symétrique :  $\forall x_1, x_2 \in E^2 \quad k(x_1, x_2) = k(x_2, x_1)$
- et définie positive  $\forall (x_1, x_2, \dots, x_n) \in E^n, \forall (a_1, a_2, \dots, a_n) \in \mathbb{R}^n, \sum_{i,j=1}^n a_i a_j k(x_i, x_j) > 0$ ,

en d'autres termes la matrice Gram, définie comme la matrice dont les éléments  $a_{ij} = k(x_i, x_j)$ , doit être symétrique, définie positive.

Les méthodes à noyau (kernel), après leur succès dans le traitement du cas non linéaire avec les svms [32,33], ont connu ces dernières années une explosion considérable dans tous les domaines de l'apprentissage artificiel. Ces méthodes ont des fondements théoriques solides et offrent des résultats expérimentaux attractifs dans la plupart des problèmes de classification.

Les fonctions d'apprentissage à noyau reposent sur une structure de représentation vectorielle appelé espace de Hilbert à noyau reproduisant (*RKHS : Reproducing Kernel Hilbert Space*). La théorie des espaces hilbertiens à noyau reproduisant est très classique en analyse fonctionnelle et l'article de Aronszajn (1950) [34], constitue une synthèse de référence. Un espace Hilbertien est essentiellement comme un espace euclidien mais de dimension infini. C'est un espace vectoriel où le produit scalaire est défini. Un espace de Hilbert à noyau reproduisant est l'association d'un espace de Hilbert à une fonction noyau et à un ensemble fini de données de départ.

#### b- Construction de l'espace RKHS

Pour construire un espace RKHS associé à une fonction noyau et à un espace de données, nous considérons que la projection  $\Phi$  appartient à un espace vectoriel. Nous

définissons un produit scalaire dans cet espace et nous montrons que le produit scalaire possède la propriété  $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ .

Soit  $\Phi$  une transformation projetant les données d'entrée dans un espace des caractéristiques.

$$\begin{aligned}\Phi : E &\rightarrow F \\ x &\mapsto k(.,x)\end{aligned}$$

donc  $\Phi(x) = k(.,x)$  est la fonction  $k(x',x)$  déterminée pour la valeur  $x'$  de  $E$ . En considérant toutes les combinaisons linéaires de ces fonctions :

$$f(.) = \sum_{i=1}^m \alpha_i k(.,x_i) , \quad x_i \in (x_1, x_2, \dots, x_m)$$

nous obtenons un espace vectoriel dans lequel nous allons définir le produit scalaire.

Soit deux fonctions  $f(.) = \sum_{i=1}^m \alpha_i k(.,x_i)$  et  $h(.) = \sum_{i=1}^n \beta_i k(.,y_i)$

Nous définissons le produit scalaire :  $\langle f, h \rangle = \sum_{i,j=1}^{m,n} \alpha_i \beta_j k(x_i, y_j)$

Maintenant nous avons besoin de vérifier les propriétés du produit scalaire citées ci-haut.

Nous remarquons qu'il est symétrique, aussi il est positif puisque :

$$\langle f, f \rangle = \sum_{i,j=1}^{m,n} \alpha_i \alpha_j k(x_i, x_j) > 0 \quad (1)$$

Pour vérifier la bilinéarité, soit :

$$\langle f, h \rangle = \sum_{i,j=1}^n \alpha_i \beta_j k(x_i, y_j) = \sum_{i=1}^n \alpha_i \sum_{j=1}^n \beta_j k(x_i, y_j) = \sum_{i=1}^n \alpha_i h(x_i)$$

On voit bien que le produit scalaire est une combinaison linéaire de la fonction  $h$ .  
 Similairement pour la fonction  $f$  :

$$\langle f, h \rangle = \sum_{i,j=1}^n \alpha_i \beta_j k(x_i, y_j) = \sum_{j=1}^n \beta_j \sum_{i=1}^n \alpha_i k(x_i, y_j) = \sum_{j=1}^n \beta_j f(y_j)$$

Donc, le produit scalaire que nous avons défini dans l'espace des caractéristiques est aussi une forme bilinéaire. Il reste la dernière propriété à vérifier :

$$\forall x \in E \quad (\langle x, x \rangle = 0 \Rightarrow x = 0).$$

A partir de (1), on peut écrire :

$$\langle k(., x), f \rangle = \sum_{i=1}^m \alpha_i k(x_i, x) = f(x) \quad \text{et en particulier} \quad \langle k(., x), k(., x') \rangle = k(x, x').$$

Par ces deux propriétés  $k$  est appelé un noyau reproduisant. Comme on a vérifié que ce produit scalaire est positif, on peut utiliser l'inégalité de Cauchy Schwarz :

$$|k(x_1, x_2)|^2 \leq k(x_1, x_1) \cdot k(x_2, x_2)$$

*démonstration :*

La matrice de Gram, dont les entrées  $k_{ij} = k(x_i, x_j)$ ,  $K = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}$  est symétrique, définie positive. Donc, ses valeurs propres ne sont pas négatives et ainsi leur produit, le déterminant :

$$k_{11} \cdot k_{22} - k_{21} \cdot k_{12} = k_{11} \cdot k_{22} - k_{21} \cdot k_{12} = k_{11} \cdot k_{22} - |k_{12}|^2 \geq 0 \quad \Rightarrow \quad |k_{12}|^2 \leq k_{11} \cdot k_{22}$$

ou bien 
$$|k(x_1, x_2)|^2 \leq k(x_1, x_1) \cdot k(x_2, x_2)$$

Soit  $f(x) = \langle k(., x), f \rangle \Rightarrow f(x)^2 = \langle k(., x), f \rangle^2 \leq k(x, x) \langle f, f \rangle$  si  $\langle f, f \rangle = 0$  on a  $f = 0$ .

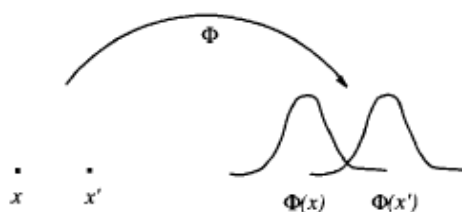


Toutes les propriétés du produit scalaire sont vérifiées. Donc pour tout transformation  $\Phi$  déterminée à partir d'un noyau  $k$ , soit  $\Phi(x) = k(.,x)$ , le produit scalaire possède la propriété :  $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ .

En pratique, on choisit une fonction à deux variables symétrique de type positif et cela garantit l'existence d'un espace hilbertien admettant cette fonction comme noyau reproduisant. L'intérêt principal de l'espace hilbertien à noyau reproduisant ainsi construit est la simplicité des calculs de produits scalaires. En fait, le noyau permet de calculer directement le produit scalaire  $\langle \Phi(x_i), \Phi(x_j) \rangle$  comme une fonction de  $x_i$  et  $x_j$ . Donc, on a pas besoin de connaître la fonction  $\Phi$ , seule la fonction noyau intervient dans les calculs. Ainsi, l'espace de redescription reste virtuel. Pour notre application le noyau Gaussien a été utilisé. Soit :

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$$

Chaque échantillon  $x$  est transformé en une gaussienne centrée en cet échantillon, comme cela est montré dans la figure 43.



**Figure 43.** Transformation en utilisant le noyau gaussien

Le noyau gaussien permet de projeter les échantillons d'entrée dans un espace de dimension infini. Aussi ceci assure la correspondance 'un vers un' entre l'espace d'entrée et l'espace de redescription. En d'autres termes, deux échantillons séparés dans l'espace d'entrée sont aussi séparés dans l'espace de redescription :

$$K(x_1, x_2) = \Phi^T(x_1)\Phi(x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right). \text{ Maintenant, si } \Phi(x_1) = \Phi(x_2),$$

nous avons :  $K(x_1, x_2) = \Phi^T(x_1)\Phi(x_2) = \Phi^T(x_1)\Phi(x_1) = K(x_1, x_1) = 1 \Rightarrow x_1 = x_2$ .

c- Algorithme de Bayes dans l'espace de redescription

La distance Mahalanobis dans l'espace augmenté ou espace de redescription en utilisant le kernel [35] est :

$$d^2(x) = (\Phi(x) - M)^T \Sigma^{-1} (\Phi(x) - M)$$

où  $\Phi(x)$  est une fonction qui transforme les données de l'espace d'entrée vers un espace de dimension plus élevé. La moyenne  $M$  et la matrice de covariance  $\Sigma$  sont :

$$M = \frac{1}{N} \sum_{j=1}^N \Phi(x_j) \quad , \quad \Sigma = \frac{1}{N} \sum_{j=1}^N (\Phi(x_j) - M)(\Phi(x_j) - M)^T = \Phi^T(X) \left( \frac{I_N - 1_N}{N} \right) \Phi(X)$$

$$\Phi(X) = \begin{pmatrix} \Phi^T(x_1) \\ M \\ \Phi^T(x_N) \end{pmatrix} \quad I_N = \begin{pmatrix} 1_{11} & 0 \\ & 1 \\ 0 & & 1_{NN} \end{pmatrix} \quad 1_N = \begin{pmatrix} \frac{1}{N_{11}} & \dots & \frac{1}{N_{1N}} \\ \vdots & \frac{1}{N} & \vdots \\ \frac{1}{N_{N1}} & \dots & \frac{1}{N_{NN}} \end{pmatrix}$$

Soit  $Z = \left( \frac{I_N - 1_N}{N} \right)^{\frac{1}{2}}$  donc  $\Sigma = \Phi^T(X) Z^2 \Phi(X)$  et

$$d^2(x) = (\Phi(x) - M)^T \left( \Phi^T(X) Z^2 \Phi(X) \right)^{-1} (\Phi(x) - M)$$

En utilisant l'égalité :  $t^T (X^T A X)^n u = t^T X^T \left( A^{\frac{1}{2}} \left( A^{\frac{1}{2}} X X^T A^{\frac{1}{2}} \right)^{n-1} A^{\frac{1}{2}} \right) X u$

on obtient :  $d^2(x) = (\Phi(X) \Phi(x) - \Phi(X) M)^T \left( Z \left( Z \Phi(X) \Phi^T(X) Z \right)^{-2} Z \right) (\Phi(X) \Phi(x) - \Phi(X) M)$

Ou bien  $d^2(x) = (T - W)^T \left( Z (Z G Z)^{-2} Z \right) (T - W)$

Où  $G = \Phi(X) \Phi^T(X)$  ,  $W = \Phi(X) M$  et  $T = \Phi(X) \Phi(x)$

sont calculées en utilisant le noyau gaussien :  $k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$

la composante  $i$  de  $T$  est :  $T_i = \exp\left(-\frac{(x_i - x)^T (x_i - x)}{2t^2}\right)$

la composante  $i$  de  $W$  est :  $W_i = \frac{\sum_{j=1}^N \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{2t^2}\right)}{N}$

et  $G(i, j) = \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{2t^2}\right)$  étant la matrice de Gram ou matrice kernel,  $t$  étant le paramètre du noyau gaussien,  $x$  l'échantillon test et  $x_i, x_j$  des échantillons d'apprentissage.  $N$  étant le nombre d'échantillons d'apprentissage.

Le déterminant de la matrice de covariance est calculé en utilisant la propriété :

*déterminant des produits égal au produit des déterminants*

$$|\Sigma| = |\Phi^T(X)Z^2\Phi(X)| = |\Phi(X)\Phi^T(X)Z^2| = |GZ^2|$$

La fonction discriminante de Bayes quadratique dans l'espace de redescription devient :

$$D = -|GZ^2| - (T-W)^T \left( Z (ZGZ)^{-2} Z \right) (T-W)$$

L'inverse de la matrice de covariance est calculé par la méthode de décomposition en valeurs singulières (que l'on désigne souvent par son acronyme anglo-saxon "SVD", pour "Singular Value Decomposition"). Comme la matrice  $ZGZ$  est symétrique définie positive, elle sera décomposée en  $U\Lambda U^T$  où la matrice  $U$  est orthogonale et  $\Lambda$  est une matrice diagonale. En général, si une valeur de la diagonale  $\lambda_i$  est inférieure à une certaine tolérance  $\sigma$ , la valeur correspondante dans la diagonale de l'inverse sera remplacée par zéro. De cette façon, le sous-espace correspondant à ces éléments nuls de la diagonale sera négligé. Comme on ne veut pas réduire la dimension de notre espace augmenté afin qu'on soit assuré que la variance de l'hyper ellipsoïde contenant nos échantillons ne varie pas à travers les voyelles, on met la valeur  $1/\sigma$  à la place de zéro. Donc l'inverse de la matrice sera égal au produit suivant:

$$U \begin{pmatrix} \lambda_1^{-1} & & & \\ & \ddots & & \\ & & \lambda_r^{-1} & \\ & & & \frac{1}{\sigma} & \\ & & & & \ddots & \\ & & & & & \frac{1}{\sigma} \end{pmatrix} U^T$$

Cette méthode résiste bien aux problèmes de mauvais conditionnement numérique et aussi est nécessaire pour ne pas se retrouver dans le sur-apprentissage (overfitting), en utilisant la tolérance  $\sigma$  comme paramètre de régularisation.

En fait, les valeurs de la tolérance  $\sigma$  et du paramètre du noyau gaussien  $t$  sont déterminées par la méthode de la validation croisée. Cette méthode consiste à partitionner l'ensemble des échantillons de l'apprentissage, cinq sous-ensembles dans notre application. Quatre de ces sous-ensembles sont utilisés pour l'apprentissage et un seul pour la validation, c'est à dire le test. Le processus est répété cinq fois avec chaque sous-ensemble utilisé seulement une fois comme données de validation. Ensuite on calcule la moyenne des cinq résultats pour produire une seule estimation. On choisira les valeurs de «  $\sigma$  » et de «  $t$  » celles qui donnent la meilleure estimation. Les résultats sont montrés dans le tableau 10.

|             | [ $\zeta$ ] | [q] | [h] |
|-------------|-------------|-----|-----|
| [ $\zeta$ ] | 89          | 5   | 7   |
| [q]         | 6           | 92  | 2   |
| [h]         | 5           | 9   | 87  |

**Tableau 10.** Résultat de classification en utilisant Bayes quadratique dans l'espace de redescription

Il est clair que de meilleurs résultats sont obtenus en utilisant l'algorithme de Bayes dans l'espace de redescription.

## Conclusion et perspectives

Dans ce 3<sup>e</sup> chapitre, après analyse des premiers résultats obtenus en utilisant l'algorithme de Bayes, on a observé la variabilité du 3<sup>e</sup> formant par rapport aux droites de locus. Le 3<sup>e</sup> formant ne prend pas des valeurs précises là où les droites de locus se croisent comme avancé dans la littérature. Après avoir mesuré la variabilité du 3<sup>e</sup> formant par rapport à chaque voyelle, on a constaté que la variabilité du 3<sup>e</sup> formant dépend plutôt de la coarticulation.

Ceci nous a amené à projeter nos échantillons dans un espace de dimension plus élevée (RKHS) en utilisant le noyau gaussien. Dans ce nouvel espace, l'algorithme de Bayes nous a fourni de meilleurs résultats.

En considérant cette nouvelle donnée sur le 3<sup>e</sup> formant, l'espace occupé par les consonnes au niveau de la perception n'est plus une droite simple mais plutôt un ellipsoïde étranglé au niveau du locus pour former deux lobes. Ainsi, la distance de Mahalanobis que la perception devrait utiliser pour discriminer s'impose. Ceci a permis d'expliquer certaines déformations dues aux interactions entre différentes consonnes au niveau de la perception. Nous observons le phénomène de [g], dans le cas des consonnes [ʕ] et [q]. Donc, chaque consonne occupe un espace de géométrie arbitraire au niveau de la perception. D'autres formants, notamment le premier pourrait aussi y contribuer et améliorer les résultats de classification.

Comme perspectives, il serait intéressant d'étendre cette étude à l'ensemble des sons de la langue arabe et vérifier le comportement de notre classifieur, en considérant les trois premiers formants et en tenant compte de la coarticulation à gauche et à droite dans les VCV.

# ANNEXE 1 OUTIL D'ANALYSE

## 1. Introduction

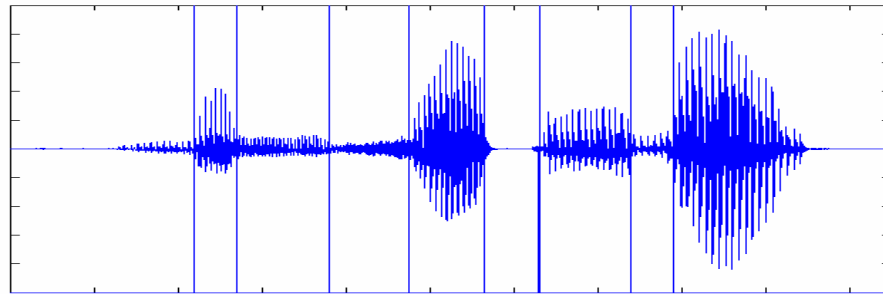
Afin d'étudier les sons produits par la parole; l'analyse instrumentale est indispensable. Comme il y a plusieurs façons de visualiser le signal de parole afin d'essayer de montrer quelques aspects, plusieurs instruments existent notamment, la photographie à rayon X, l'électromiographie (EMG), la spectrographie, le mingographe, le laryngographe etc. Actuellement l'ordinateur est l'instrument le plus disponible et le plus utilisé. Avec l'ordinateur le processus d'analyse est beaucoup plus simple et toujours plus rapide que les autres instruments; cependant; il ne produit pas nécessairement des résultats de meilleure qualité.

Pour ce projet, nous sommes concernés par l'analyse en trajectoires formantiques du signal de parole. Comme il n'existe pas une méthode totalement efficace pour permettre de bonnes estimations de ces trajectoires; deux méthodes différentes sont utilisées; le spectrogramme et la prédiction linéaire (LPC) [25,36]. D'abord nous avons besoin de visualiser le signal de parole; l'oscillogramme est utilisé. Ensuite, une segmentation manuelle va nous permettre de ne prendre qu'une partie de ce signal (/VCV/). Etant donné le caractère non stationnaire du signal de parole, une fenêtre glissante de 30ms pour l'analyse spectrale est un compromis courant. L'ordinateur utilisé est un PC Pentium IV. Les signaux de parole sont enregistrés avec un microphone et échantillonnés à 16 KHz dans l'ordinateur.

## 2. L'oscillogramme

Physiquement, le signal de parole (comme tous les autres sons) est une série de variation de pressions dans le milieu se trouvant entre la source et l'auditeur. La représentation la plus fréquente du signal de parole est l'oscillogramme; l'axe horizontal représente le temps de gauche à droite et l'axe vertical l'intensité qui varie de la même manière que la pression du milieu transmettant le signal à l'auditeur. La phrase en arabe

« man saqila » prononcée par un adulte algérien est visualisée par l'oscillogramme dans la figure 44.



[m] [a] [n] [s] [a] [q] [l] [l] [a]

**Figure 44..** Oscillogramme de « man saqila »

La transcription IPA (International Phonetic Alphabet) a été utilisée. Les traits verticaux ne font pas partie du signal de parole; c'est une segmentation manuelle.

### 3. La fenêtre du signal

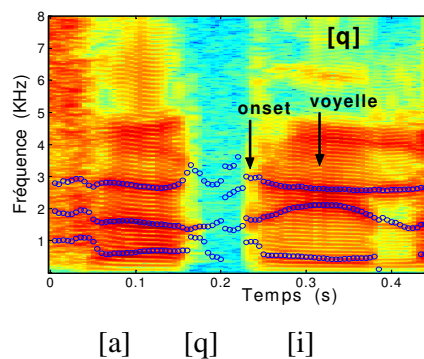
Les trajectoires formantiques sont le résultat des estimations des fréquences des formants sur le spectre d'une fenêtre glissant le long du signal. Cette fenêtre a une durée choisie d'une part assez courte pour que les caractéristiques spectrales puissent être considérées comme étant stables et d'autre part assez longue pour contenir le maximum d'harmoniques possibles pour augmenter le contraste des formants sur le spectre. Une fenêtre de Hanning a été choisie pour atténuer le phénomène de Gibbs et concentrer la répartition de l'énergie sur les basses fréquences. L'équation de cette fenêtre est :

$$W(n) = \frac{1}{2} \left( 1 - \cos\left(\frac{2\pi(n-1)}{N-1}\right) \right)$$

La durée de la fenêtre est de 30 ms avec un glissement de 5 ms

## 4. Le spectrogramme

Selon la théorie en général, chaque signal périodique peut être décrit par la somme de plusieurs sinusoïdes. Il s'agit d'une décomposition de l'onde acoustique en ondes sinusoïdales de différentes fréquences au moyen d'une analyse de Fourier. Le spectre donne une image de la distribution de la fréquence et amplitude à un moment dans le temps. L'axe horizontal représente la fréquence et l'axe vertical l'amplitude. Si on veut visualiser le spectre en fonction du temps, on obtient le spectrogramme. Dans une telle représentation tridimensionnelle, l'axe  $x$  correspond au temps, l'axe  $y$  à la fréquence et le degré de noirceur (couleur) à l'amplitude relative. C'est une représentation importante parce que le niveau de pression (en db) d'un son est proportionnel à l'intensité perçue ; donc ça reflète l'image psycho acoustique du son. La figure 45 montre le spectrogramme du /VCV/ /aqi/.



**Figure 45.** Spectrogramme du /VCV/ /aqi/.

On perçoit les structures dominantes de la parole, les *formants* (les «tâches épaisses» de la figure 45). Leurs positions sont différentes pour différents sons et souvent ils peuvent être prédits pour chaque phonème.

## 5. Suivi de formants par prédiction linéaire

L'analyse par prédiction linéaire est l'une des plus importantes techniques d'analyse du signal de parole. La base de cette technique c'est le modèle source – filtre. C'est une méthode de déconvolution directe qui sépare le signal périodique d'excitation de la fonction de transfert représentant le conduit vocal qui module ce signal. Cette fonction est un modèle tous pôles autorégressif (AR) où chaque pôle correspond aux caractéristiques d'un formant, notamment à

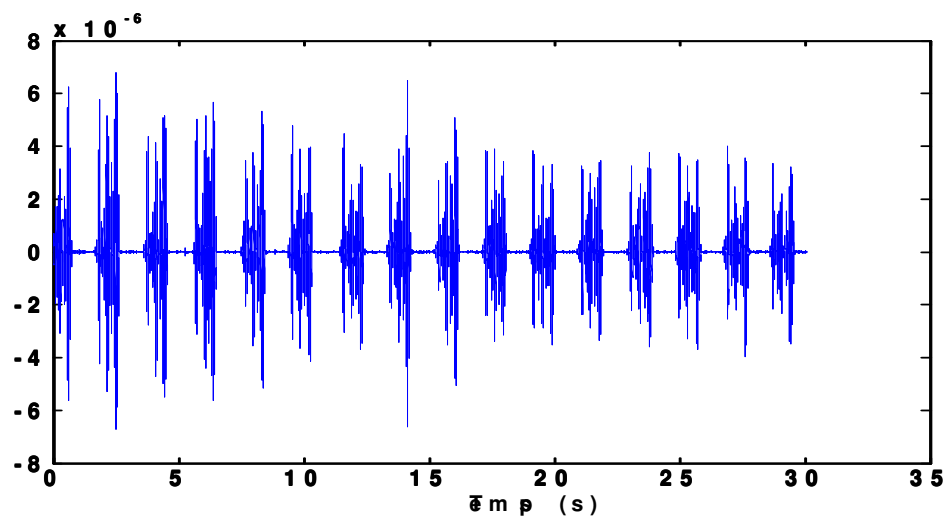


sa fréquence. Le nombre de pôle de cette fonction donne l'ordre du modèle. L'ordre du modèle est ajusté selon le nombre de formants à extraire, dans notre cas, l'ordre du filtre  $p = 2 \cdot N + 1$  (N étant le nombre de formants) Pour  $N=3$ ,  $P= 7$ .

L'algorithme utilisé est très simple. En utilisant Matlab, Il consiste à déterminer d'abord par la commande "LPC" le polynôme du dénominateur de la fonction de transfert; ensuite, en utilisant la commande "ROOTS" on détermine les pôles dans le plan "Z" et n'en prendre que ceux qui appartiennent au demi plan supérieur pour ne pas avoir les formants en double. Et à partir des positions de ces pôles on détermine les formants. Les trois premiers formants sont visualisés en même temps et les points critiques (l'onset et milieu de la voyelle ) sont détectés manuellement.

## 6. Procédure de la mesure

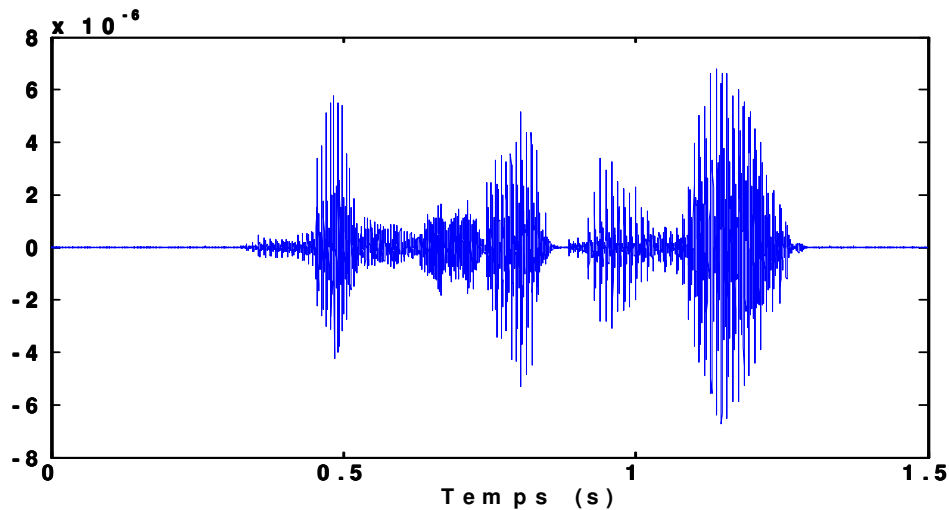
En exécutant la fonction **analyse ('son')**, le fichier « son » à traiter est écouté et visualisé comme on le voit dans la figure 46.



**Figure 46.** Visualisation de toutes les répétitions dans le fichier son 'aqi6'

On voit l'ensemble des répétitions de l'expression enregistrée. Ces répétitions sont traitées dans l'ordre l'une après l'autre. La première est sélectionnée en cliquant sur la souris une seule fois dans le silence qui suit et on valide par la touche 'entrée'.

Maintenant, la première répétition est visualisée dans la figure 47 qui s’affiche automatiquement.



**Figure 47.** Visualisation d’une seule répétition

En utilisant toujours la souris, le /VCV/ est sélectionné en cliquant sur les deux points limitant le /VCV/ et valider par la touche « entrée ». Cette fois-ci, c’est le spectrogramme superposé au suivi de formant qui apparaît (voir figure 45). De la même manière, les quatre points  $F3_{\text{onset}}$ ,  $F3_{\text{voyelle}}$ ,  $F2_{\text{onset}}$  et  $F2_{\text{voyelle}}$  sont relevés en utilisant la souris et validés par la touche « entrée ». Automatiquement, ces points seront emmagasinés dans une matrice appelée « locus » et puis enregistrés dans un fichier de type ‘mat’ portant le même nom que le fichier son. Si le nombre de points sélectionnés est différent de quatre, l’échantillon n’est pas considéré et on passe à l’échantillon suivant. Ensuite, seulement les répétitions non traitées sont visualisées et le cycle recommence et continue. Après la dixième répétition le programme s’arrête.

### Programme matlab

*function analyse(son)*

*% son='aqi6' par exemple : 'men saqila du locuteur 6.'*

```
[x,fe]=wavread(son); %lire le fichier son
soundsc(x,fe); %écouter
x=filter([1 -0.95],1,x); %filtre de préaccentuation
ms30=fe*0.03; %longueur fenêtre d'analyse
ms25=5*ms30/6; %longueur du recouvrement
nfft=1024; %longueur de la fft
n=1;

while n<11;
    plot(x);[t f]=GETPTS; %segmentation manuelle
    t=fix(t);x1=x(1:t);x(1:t)=[];hold off,
    plot(x1);[t f]=GETPTS; %selectionner le /VCV/
    t=fix(t);x1=x1(t(1):t(2));hold off,
    spectrogram(x1,nfft,fe,ms30,ms25);hold on %visualiser le spectrogram
    formants(x1); %superposer le suivi formantique
    [t f]=GETPTS; %selectionner F3onset,F3voyelle,F2onset,F2voyelle
    if length(f)==4
        locus(n,:)=f;n=n+1;save aqil locus; %enregistrer
    end
    hold off;
end
```

=====

*function formants(x)*

```
x=resample(x,1,2); %rééchantillonner(1/2)
fe=8000;
ms30=fe*0.03;ms5=ms30/6;
w=hanning(ms30); %utiliser la fenêtre de hanning
nf=floor((length(x)-ms30+ms5)/ms5); %nombre de fenêtre d'analyse
n1=1;n2=ms30;

for t=1:nf
    xf=x(n1:n2).*w; %signal fenêtré
    a=lpc(xf,7);r=roots(a); %calculer les coefficients du filtre et trouver les racines
    r=r(imag(r)>0.01); %r=r(abs(r)>0.9); %ne prendre que les racines positives
    f=(atan2(imag(r),real(r))*fe/(2*pi)); % calculer les fréquences
    if length(f)==0,f=0;end
    tf=t*5/1000;plot(tf,f,'bo');hold on %visualiser les formants
    n1=n1+ms5;n2=n2+ms5;
end
```

## Bibliographie

- [1] A. M. Liberman, F. S. Cooper, D.P. Shankweiler, M. Studdert-Kennedy. « perception of the speech code ». Psychological review. Vol.74,N°6, 1967
- [2] D. B. Orr, H. L. Friedman, & J. C. C. Williams. « Trainability of listening comprehension of speeded discourse ». Journal of Educational Psychology, 1965, 56, 148-156.
- [3] G. A. Miller & W.G. Taylor. « The perception of repeated bursts of noise ». Journal of the Acoustical Society of America, 1948, 20, 171-182.
- [4] A. M. Liberman, I.G. Mattingly. « The motor theory of speech perception revised ». Cognition. 21, pp 1-36 . 1985
- [5] K. N. Stevens. « Articulatory–Acoustic–Auditory Relationships ». The Handbook of Phonetic Sciences. 1999
- [6] A. M. Liberman, P. C. Delattre et F. S. Cooper. « The role of selected stimulus variables in the perception of the unvoiced stop consonants ». The American Journal of Psychology. Vol. LXV pp 497-516 . Octobre 1952.
- [7] R. K. Potter, C. A. Kopf & H. C. Green. « Visible speech ». New York: van Nostrand. 1947
- [8] F. S. Cooper, P. C. Delattre, A. M. Liberman., J.M. Borst et L. J. Gerstman. « some experiments on the perception of synthetic speech sounds ». JASA. Vol. 24 N° 6. pp 597-606. Novembre 1952.
- [9] P. C. Delattre, A. M. Liberman, and F. S. Cooper. « Acoustic Loci and Transitional Cues for Consonants ». JASA. Vol. 27 N° 4 pp 769-773. Juillet 1955.
- [10] K. S. Harris, H. S. Hoffman, Alvin M. Liberman, P. C. Delattre et F. S. Cooper. « Effect of F3 transitions on the perception of the voiced stop consonants ». JASA. Vol.30 N°2 pp 122-126 Fevrier 1958.
- [11] G.E. Peterson & H.L. Barney. « Control methods used in a study of the vowels. » Journal of the Acoustical Society of America 24: 175–184. 1952.
- [12] A. M. Liberman, K. S. Harris, H. S. Hoffman et B. C. Griffith. « The discrimination of speech sounds within and across phoneme boundaries ». Journal of Experimental Psychology. Vol.54 N°5. Novembre 1957.
- [13] P. C. Delattre, A. M. Liberman et F. S. Cooper « Formant transitions and loci as acoustic correlates place of articulation in american fricatives ». Haskins Laboratories. New York.

- [14] H. M. Sussman, H. A. McCaffrey and S. A. Matthews. « An investigation of locus equations as a source of relational invariance for stop place categorization ». JASA. Vol. 90. N° 3. Septembre 1991.
- [15] H. M. Sussman, K. A. Hoemeke and F. S. Ahmed. « A cross-linguistic investigation of locus equation as a phonetic descriptor for place of articulation ». JASA. Vol. 94. N° 3. Septembre 1993.
- [16] M. Yeou « locus equation and degree of coarticulation of Arabic consonants ». *Phonetica*. Vol. 54. pp. 187-202. Juillet 1997.
- [17] J. A. Tamimi. « Equation de locus comme mesure de la coarticulation VC et CV Etude préliminaire en Arabe Dialectal Jordanien ». Laboratoire Dynamique du Langage. Institut des Sciences de l'Homme. Lyon. France. 2004
- [18] E. Castelli et A. Hierholtz « équation de locus pour les consonnes [b], [d] et [] du vietnamien ». Centre de Recherche International MICA. 2006.
- [19] D. Krull. « Second formant locus patterns and consonant vowel coarticulation in spontaneous speech ». *Phonetic Experimental Research Institute of Linguistics*, University of Stockholm, 1989 X, pp. 87-108.
- [20] J. Harrington « Acoustic phonetics ». J. Laver et W. Hardcastle (Eds). *The Handbook of Phonetic Sciences*. Blackwell 2010.
- [21] C. A. Fowler. « Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation ». *Perception and psychophysics* 1994.
- [22] H. M. Sussman and J. Shore. « Locus equations as phonetic descriptors of consonantal place of articulation ». *Perception and psychophysics* 1996.
- [23] M. F. Dorman. « Stop-consonant recognition : Release bursts and formant transitions as functionally equivalent, context dependent cues ». *Perception and psychophysics* 1977.
- [24] D. Fruchter and H. M. Sussman. « The perceptual relevance of locus equations ». JASA. Vol. 102. N° 5. Novembre 1997.
- [25] Quatieri Th.F. (2002) *Discrete Time Speech Signal Processing* (781s)
- [26] Andries W. Coetzee. « Hamza ». University of Massachusetts, USA, and North West University, South Africa. 2002.
- [27] John J. McCarthy. « The Phonetics and Phonology of Semitic Pharyngeals ». University of Massachusetts, Amherst.
- [28] S. Weisberg. « Applied Linear Regression ». John Wiley & Sons, 2005.
- [29] R. A. Maronna, R. D. Martin and V. J. Yohai « Robust Statistics Theory and Methods » WILEY, 2006.

- [30] R. O. Duda, P. E. Harr et D. G. Stork. « Pattern Classification second edition ». John Wiley & Sons, 2000
- [31] B. Scholkopf & A. J. Smola . « Learning with Kernels ». M.I.T. Press. 2002
- [32] A. R. Webb. « Statistical Pattern Recognition » John Wiley & Sons, 2002
- [33] R. Herbrich. « Learning Kernel Classifiers. Theory and Algorithms ».
- [34] N. Aronszajn. « Theory of Reproducing Kernels ». Transactions of the American Mathematical Society, Vol. 68, No. 3, pp. 337-404. May, 1950.
- [35] S. Abe. « Support Vector Machines for Pattern Classification ». Springer-Verlag 2005.
- [36] L. R. Rabiner & R.W. Schafer. « Digital Processing Of Speech Signals ». Prentice Hall Signal processing Series. A.V. Oppenheim, series editor. 1978.