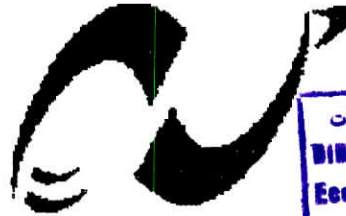


République Algérienne Démocratique et Populaire  
Ministère de l'enseignement supérieur et de la recherche scientifique

## ECOLE NATIONALE POLYTECHNIQUE



المدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

DEPARTEMENT : ELECTRONIQUE

OPTION : TELECOMMUNICATIONS

### THESE DE MAGISTER

Présentée par  
**FARAH Farid**

Ingénieur d'Etat en Electronique (U. S. T. H. B)

#### Thème

CARACTERISATION ACOUSTIQUE DES SONS  
EMPHATIQUES DE L'ARABE STANDARD

**Soutenue publiquement le 08/ 07 / 2003 devant le jury composé de :**

M <sup>r</sup> M. MEHENNI	Maître de Conférence (ENP)	Président
M <sup>me</sup> M. GUERTI	Maître de Conférence (ENP)	Directrice de thèse
M <sup>me</sup> M. BOUDRAA	Maître de Conférence (USTHB)	Examineur
M <sup>me</sup> L. HAMAMI	Chargé de cours (ENP)	Examineur
M <sup>r</sup> B. BOUSSEKSOU	Chargé de cours (ENP)	Examineur

**E.N.P.** 10, Avenue Hassen Badi -El-Harrach- ALGER

## Dédicace

A ma mère, A ma mère, A ma mère

A mon père

A mes frères et sœurs

A mes amis et collègues

Farid

## Remerciements



C'est une grande chance que M. Guerti, ma Directrice de thèse, m'a offerte. Elle m'a ouvert les portes du monde de la parole dont je rêvais mais je pensais alors inaccessible. Elle a été mon initiatrice et mon guide et je la remercie beaucoup. J'associe à ce remerciement deux autres guides, N. Henrich, chercheur au Laboratoire d'Acoustique Musicale (LAM) de Paris, et B. Boudraa, chercheur au laboratoire de la communication parlée à l'USTHB, qui m'ont énormément aidé dans mon travail. J'ai également une pensée aux journalistes de la radio algérienne (chaîne 1) qui n'ont pas hésité à franchir la porte des studios d'enregistrement pour offrir leur talent avec tant de gentillesse. Merci également à tous ceux qui m'ont aidé à avoir accès à ce studio.

Derrière ce travail, il y a une multitude de personnes qui m'ont offert leur temps et leur efficacité : merci à tous, en particulier C. Lounakel, R. Benhmimid, et les amis du ministère des industries. J'ai été touché que les membres qui composent le jury de thèse acceptent spontanément d'y participer, montrant ainsi leur intérêt pour ce travail et je les en remercie chaleureusement.

Je remercie aussi les chercheurs de l'Ecole Nationale Polytechnique (ENP) du département électronique en particulier M. Mehenni, M. Trabelsi, R. Aksas, L. Hamami...

Je remercie également mes parents, mes frères et sœurs, mes proches, et plus particulièrement B. Gacemi, K. Adjiri et L. Martin, pour m'avoir soutenu au quotidien.

Mes remerciements vont aussi à mes amis, ils sont nombreux : Sid Ali, Chafik, Djamel, Hakim, Chaabane, Kamel, Salah...

## SOMMAIRE

<b>Introduction Générale.....</b>	<b>1</b>
<b>Chapitre 1 : Théories articulatoire et acoustique d'un signal vocal.....</b>	<b>6</b>
1. 1. Introduction.....	6
1. 2. Principe de production de la parole humaine.....	7
1. 2. 1. Description physiologique de l'appareil phonatoire humain.....	8
1. 2. 1. 1 Vibration laryngée.....	9
1. 2. 1. 2 Quelques ordres de grandeurs.....	11
1. 2. 2. Caractérisation articulatoire et acoustique du mécanisme de la phonation.....	12
1. 2. 2. 1 L'opposition voisée-non voisée.....	14
1. 2. 2. 2 L'opposition oral-nasal.....	15
1. 2. 2. 3 Les caractéristiques de la voix.....	15
1. 2. 3. Classification des sons d'un langage.....	17
1. 2. 3. 1 Point d'articulation et mode d'articulation.....	18
1. 2. 3. 2 Les voyelles.....	18
1. 2. 3. 3 Les consonnes.....	21
1. 2. 4. Les sons de l'Arabe Standard.....	23
1. 2. 5. Les consonnes emphatiques.....	25
1. 2. 6. Conclusion.....	25



## Chapitre 2 : Etude de la modélisation

<b>du système de production vocale.....</b>	<b>26</b>
2. 1. Introduction.....	26
2. 2. Modélisation acoustique du conduit vocal.....	28
2. 2. 1. La propagation du son.....	29
2. 2. 2. Le tube uniforme sans pertes.....	30
2. 2. 2. 1. L'approche de la propagation de l'onde.....	31
2. 2. 2. 2. L'approche de la ligne de transmission.....	37
2. 2. 2. 3. L'effet des pertes.....	38
2. 2. 2. 3. 1 L'impédance des parois du conduit vocal.....	38
2. 2. 2. 3. 2 Le modèle acoustique du rayonnement aux lèvres.....	40
2. 3. Modélisation de la source glottique.....	41
2. 4. Le modèle de signal source-filtre.....	46
2. 5. Conclusion.....	48

## Chapitre 3 : Analyse des données et simulation..... 49

3. 1. Introduction.....	49
3. 2. Mise en exécution du logiciel.....	49
3. 3. Méthodes d'analyse de la parole.....	51
3. 3. 1. Utilisation de la FFT dans l'analyse spectrale à court-terme.....	52
3. 3. 2. Analyse d'un signal vocal par codage prédictif linéaire.....	52
3. 3. 3. Analyse de la parole par prédiction linéaire perceptuelle.....	60
3. 3. 4. Suivi des Formants.....	63
3. 4. Conclusion.....	66

## Chapitre 4 : Application aux sons emphatiques

<b>de l'Arabe Standard.....</b>	<b>67</b>
4. 1. Introduction.....	67
4. 2. Objectif.....	67
4. 3. Synthèse des travaux sur l'emphase.....	69
4. 3. 1. Etudes articulatoires.....	69
4. 3. 2. Etudes acoustiques.....	70
4. 4. Données acoustiques.....	71
4. 4. 1. Procédures des enregistrements sonores.....	71
4. 4. 2. Corpus.....	72
4. 5. Description acoustique.....	73
4. 5. 1. Description spectrographique.....	73
4. 5. 2. Description formantique.....	80
4. 6. Conclusion.....	86
<b>Conclusions Générales et perspectives.....</b>	<b>88</b>
<b>Références bibliographiques.....</b>	<b>91</b>

Figures et Tableaux	Page
Fig. 1.1 Modèle simple de production vocale humaine	07
Fig. 1.2 Description physiologique de l'appareil phonatoire vue d'ensemble (coupe verticale)	08
Fig. 1.3 Visualisation de l'épaisseur d'une corde vocale (coupe verticale)	09
Fig. 1.4 Vue du larynx	10
Fig. 1.5 L'impulsion du débit glottale	11
Fig. 1.6 Cavités résonantes de l'appareil phonatoire	12
Fig. 1.7 Les résonances formantiques d'un spectre vocalique	13
Fig. 1.8 Production des sons voisés	14
Fig. 1.9 Production des sons non voisés	15
Fig. 1.10 Formants de la voyelle [a] pour deux locuteurs masculin et féminin	19
Tableau 1.1 Les fréquences fondamentales minimale, moyenne et maximale pour les hommes, femmes et enfants	19
Fig. 1.11 Forme d'onde, spectre et spectrogramme de la voyelle [a]	20
Fig. 1.12 Le triangle vocalique pour les locuteurs mâles	21
Tableau 1.2 : Quelques caractéristiques des phonèmes.	22
Tableau 1.3 : Durées et fréquences formantiques relatives aux voyelles arabes [2]	23
Tableau 1.4 : Consonnes et voyelles de l'Arabe Standard suivies de leurs transcriptions en Alphabet Phonétique International (API)	24
Fig. 2.1 : Modèle fonctionnel de production de la parole	27
Fig. 2.2 Conduit vocal idéal à aires de section transversale variable	28

Fig. 2.3 Relations de réflexion à la jonction de deux sections sans pertes	32
Fig. 2.4 Conduit vocal sans pertes discret dans le temps	33
Fig. 2.5 Analogies entre les paramètres acoustiques d'un tube et ceux d'une ligne de transmission électrique	37
Fig. 2.6 Les paramètres de la source glottique	43
Fig. 2.7 Le modèle deux masses des cordes vocales	45
Fig. 2.8 Circuit équivalent à de la glotte	45
Fig. 2.9 Le modèle de signal source-filtre	47
Fig. 3.1 : Entrée au Programme d'analyse	50
Fig. 3.2 : Options disponibles	50
Fig. 3.3 : Signal d'entrée, phonème [s <sup>o</sup> i]	51
Fig. 3.4 : Analyse spectrale par FFT d'un segment du phonème[s <sup>o</sup> i]	52
Fig. 3.5 : Modèle simplifié de la production de la parole par prédiction linéaire	53
Fig. 3.6 : Pôles à l'intérieur du cercle unité d'un segment du phonème [s <sup>o</sup> i]	57
Fig. 3.7 : (a) Analyse spectrale de la voyelle [a] (b) Pôles à l'intérieur du cercle unité	58
Fig. 3.8 : Organigramme de la méthode de l'autocorrélation	58
Fig. 3.9 : Analyse perceptuelle par prédiction linéaire	61
Fig. 3.10 : Analyse par autocorrélation et PLP d'un segment du phonème [s <sup>o</sup> i]	63
Fig. 3.11 : Signal vocal et analyse formantique du mot [ħaya:ta]	65
Fig. 4.1 : Principe articulatoire de l'emphase	68



Fig. 4.2 : Spectrogrammes des consonnes [t <sup>ʔ</sup> ] et [t]	71
Fig. 4.3 : Schéma du dispositif d'enregistrement du corpus	72
Tableau 4.1 : Corpus enregistré	73
Fig. 4.4 : spectrogrammes de [t <sup>ʔ</sup> ]	73
Tableau 4.2 : Caractéristiques de la barre d'explosion	75
Fig. 4.5 : Spectrogrammes de [t <sup>ʔ</sup> ]	76
Fig. 4.6 : Spectrogrammes de [d <sup>ʔ</sup> ]	77
Fig. 4.7 : Spectrogrammes des syllabes [sa] et [s <sup>ʔ</sup> a]	77
Fig. 4.8 : Spectres des phonèmes [sa] et [s <sup>ʔ</sup> a]	78
Tableau 4.3 : Valeurs moyennes du délai d'explosion exprimées en ms	79
Tableau 4.4 : Valeurs moyennes des durées des voyelles avec les occlusives exprimées en ms	80
Tableau 4.5 : Valeurs moyennes des durées des voyelles avec les fricatives exprimées en ms	80
Fig. 4.9 : Contours formantiques des syllabes [ta], [t <sup>ʔ</sup> a], [ti] et [t <sup>ʔ</sup> a]	82
Fig. 4.10 : Contours formantiques des phonèmes [da] et [d <sup>ʔ</sup> a]	83
Fig. 4.11 : Contours formantiques des phonèmes [sa] et [s <sup>ʔ</sup> a]	84
Tableau 4.6 : Les onsets des formants F <sub>1</sub> F <sub>2</sub> F <sub>3</sub> F <sub>4</sub> exprimés en Hz	81
Tableau 4.6 : Les valeurs des formants F <sub>1</sub> F <sub>2</sub> F <sub>3</sub> F <sub>4</sub> après l'explosion exprimés en Hz	85
Fig. 4.12 : Triangle vocalique des voyelles courtes	86



# Introduction Générale

**L**ES dix dernières années témoignent des progrès énormes réalisés dans l'amélioration de la fiabilité des systèmes de Traitement Automatique de la Parole (TAP). En fournissant d'importantes idées sur les propriétés et les caractéristiques du signal vocal, les modèles mathématiques de la production de la parole améliorent sa performance et mènent directement à une analyse et synthèse de la parole de qualité supérieure, à une reconnaissance Automatique de Parole Robuste (RAP), et à la réalisation de codeurs de parole efficaces. En incluant les systèmes de téléphone, du codage et de la compression des données, du courrier vocal (voice mail), des stations de travail (work stations), des ordinateurs individuels et des réseaux. Ces applications montrent le rôle primordial du traitement automatique du langage naturel joué dans le développement du Dialogue Homme/Machine.

La quasi-totalité des travaux de recherche effectués dans le domaine de la production de la parole sont orientés vers une meilleure caractérisation quantitative de son analyse. Cette procédure demande une variété de modèles de production de signaux de parole. Parmi lesquels, on peut noter : les modèles à tuyaux acoustiques, à filtres, pôle-zéro et les modèles de rayonnement de la parole.

Tous les efforts effectués pour atteindre l'objectif principal de l'analyse du signal vocal, à savoir la réduction de sa largeur de bande et sa représentation par un nombre réduit de paramètres pertinents, font face au moins à deux grands défis :

- un taux de données physiologiques insuffisant qui est décisif pour développer des modèles quantitatifs ;
  
- les glissements des caractéristiques acoustiques et articulatoires d'un locuteur vers un autre. C'est pourquoi, une grande attention a été donnée au développement et à l'utilisation de nouvelles techniques d'instrumentation et de mesure dans la collecte des sources de données. Des exemples sur de telles techniques font appel à :
  - l'Imagerie par Résonance Magnétique (RMN) dans l'étude de la géométrie 3D du conduit vocal ;
  - aux technologies de faisceaux à rayons x dans l'étude des mouvements de la langue ;
  - à l'electroglottographie dans l'observation des ouvertures et fermetures glottales ;
  - aux techniques de traitement de signal, tel que le filtrage inverse dans la visualisation et la quantification des données physiques et physiologiques obtenues.

Ainsi, les travaux de recherche sur l'Analyse et la modélisation de la parole naturelle ont un lien acoustico-articulatoire. Par conséquent, et dans le but de développer des modèles acoustico-articulatoire améliorés, la technique de l'Imagerie par résonance a été appliquée à des données différentes, obtenues à partir de plusieurs sujets humains. De bonnes caractérisations qualitatives et quantitatives de la géométrie du conduit vocal, notamment ses dimensions pour les consonnes et les voyelles, ont ainsi été notées.



Cette disponibilité des informations réelles sur la géométrie du conduit vocal, ajoutée à l'utilisation de la simulation numérique dans l'étude de la propagation d'une onde acoustique à travers les modèles 3D du conduit vocal, particulièrement pour les sons de la parole de durée prolongée, font que les modèles acoustiques obtenus montrent une bonne harmonie avec le spectre du signal vocal naturel et assurent ainsi le développement d'une synthèse de haute qualité et d'un codage de parole à bas débit.

Dans le même contexte, un intérêt considérable a été consacré à la solution inverse, c'est-à-dire l'identification des paramètres articulatoires à partir d'un signal de parole. Cette solution possède des applications potentielles dans le codage à bas débit et dans la reconnaissance de la parole, et ce, en dépit du fait que pour lever les contraintes existantes elle fait appel à des algorithmes génétiques et à des techniques des réseaux de neurones dont la viabilité n'est toujours pas montrée.

Notre travail s'applique aux consonnes emphatiques de l'Arabe Standard. A l'aide d'une analyse acoustique, une étude sur certains aspects du processus de production des sons emphatiques de l'AS a été réalisée. Les caractéristiques acoustiques des consonnes de l'Arabe Standard (AS) ont fait l'objet de plusieurs études expérimentales. Des difficultés ont été enregistrées dans la quasi-totalité de ces travaux de recherches notamment l'aspect articulatoire d'un signal vocal et plus précisément, lors de la détermination des caractéristiques acoustiques qui lui sont attribuées.

L'Arabe Standard comprend :

- des fricatives ;
- des Occlusives ;
- des liquides ;
- des nasales ;



- des Coranales pharyngalisées qui en plus de leur articulation antérieure primaire, développent une articulation secondaire causée par la rétraction du dos de la langue vers la paroi postérieure de l'Oro-pharynx (les consonnes emphatiques).

Il est à noter que ce type de consonnes a fait l'objet de plusieurs études acoustique et articulatoire.

Traditionnellement, les quatre articulations définies comme «emphatiques» dans l'AS sont :[s<sup>é</sup>], [t<sup>é</sup>], [d<sup>é</sup>], [z<sup>é</sup>].

Dans la partie de notre travail consacrée aux outils de l'analyse, nous avons introduit des interfaces graphiques GUI lors de l'élaboration du logiciel d'analyse afin de faciliter son utilisation. En effet, il a été conçu sous la version Matlab 5.3 .

Nous avons également étudié les différents algorithmes utilisés dans la programmation notamment celui de l'extraction automatique des formants par le Codage Prédicatif Linéaire (Linear Predictif Coding (LPC)) et celui de l'analyse de la parole par la prédiction linéaire perceptuelle (Perceptual Linear Predictive (PLP)).

En cherchant à examiner certaines propriétés des consonnes emphatiques de l'AS tels que leurs effets sur les voyelles, qui faut-il le souligner se traduit usuellement par un rapprochement des deux premiers formants  $F_1$  et  $F_2$ , ensuite formaliser éventuellement les règles par Son ; il nous a apparu utile d'appliquer ce logiciel à l'analyse à un corpus composé de sons emphatiques de l'AS obtenu à partir de l'enregistrement de six locuteurs.

Le but de l'étude présentée dans ce mémoire consiste en une caractérisation acoustique des sons emphatiques de la langue arabe standard, et ce en utilisant des méthodes d'analyse de la parole dans le domaine fréquentiel, basées essentiellement sur l'algorithme LPC.

Ce mémoire comprend quatre chapitres :

- un premier chapitre où sont exposées des notions fondamentales sur les aspects articulatoires et acoustiques du signal vocal et sur l'aspect acoustique des sons de l'AS.
- un deuxième chapitre synthèse sur la modélisation du mécanisme de production de la parole.
- un chapitre trois où nous exposons les procédures d'analyse de la parole en mettant l'accent sur le domaine fréquentiel et à la méthode de la prédiction linéaire.
- un dernier chapitre où nous élaborons une série d'analyse des consonnes emphatiques du corpus des données. Des résultats seront présentés et commentés.

Une conclusion générale et un compte rendu sur les perspectives de ce travail sont mentionnés.

# **Chapitre 1**

## **Théories articulatoire et acoustique d'un signal vocal**

### **1. 1. Introduction**

Il y a de quelques centaines d'années à peine, les hommes pensaient qu'un corps humain privé de vie ne pouvait plus émettre de son et que la voix était donc l'expression d'une âme. Cependant, quand on est parvenu à produire un son en insufflant de l'air avec un soufflet dans un larynx humain enlevé, on a conclu que l'effort vocal est essentiellement musculaire et respiratoire : Tous les sons de la parole sont formés par un soufflement d'air des poumons vers le conduit vocal.

L'étude du mécanisme de la production de la parole humaine consiste donc à décrire, sur un plan physiologique, les mouvements des articulateurs vocaux qui produisent les mots et les sons et, sur un plan acoustique, l'étude des sons des mots prononcés.

Dans ce chapitre, nous explorerons la physiologie anatomique du mécanisme de production de la parole humaine et les méthodes qui permettent d'étudier sa modélisation numérique. Nous donnerons aussi une brève introduction au système vocal.

## 1. 2. Principe de production de la parole humaine

Au cours d'une phonation, les sons de la parole sont produits soit par les vibrations des cordes vocales (source de la phonation), soit par l'écoulement turbulent de l'air dans le conduit vocal ou à l'issue d'un relâchement d'une occlusion de ce conduit (source de bruit). La parole se distingue donc des autres sons par ses caractéristiques acoustiques ayant leurs origines dans l'interdépendance des paramètres de la source et du conduit vocal.

Dans un processus de communication parlée, pour une langue donnée, l'articulation n'est en fait qu'un processus de production des sons par différentes manipulations des articulateurs, lesquelles changent la configuration du conduit vocal. Ces sons, grâce à leur unité élémentaire appelée phonème, permettent alors la distinction des différents mots [1].

En étudiant ce processus, il est utile d'extraire les caractéristiques importantes du système physique, à l'issue de la réalisation des modèles mathématiques simplifiés [2]. C'est pourquoi, une représentation simple de production de la parole, décrite dans la Figure 1.1, s'avère nécessaire puisqu'elle représente un début de modèle de production selon un aspect physiologique. Le signal acoustique, rayonné à partir des lèvres, est perçu par l'auditeur.

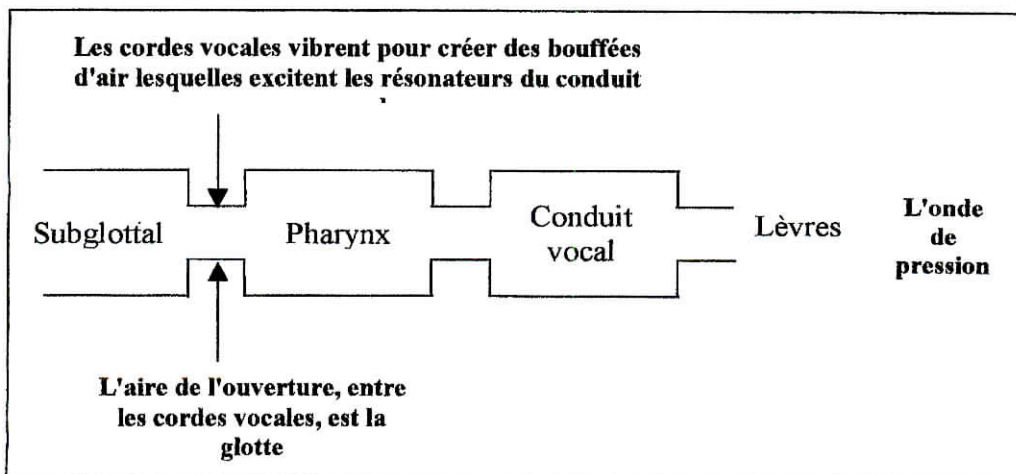


Fig. 1.1 : Modèle simple de production vocale humaine



### 1. 2. 1. Description anatomique de l'appareil phonatoire humain

L'anatomie de l'appareil phonatoire humain montre que les principaux organes phonatoires sont regroupés, selon leur localisation en deçà ou au-delà du larynx, en deux systèmes sous et sus-glottique (Fig. 1.2).

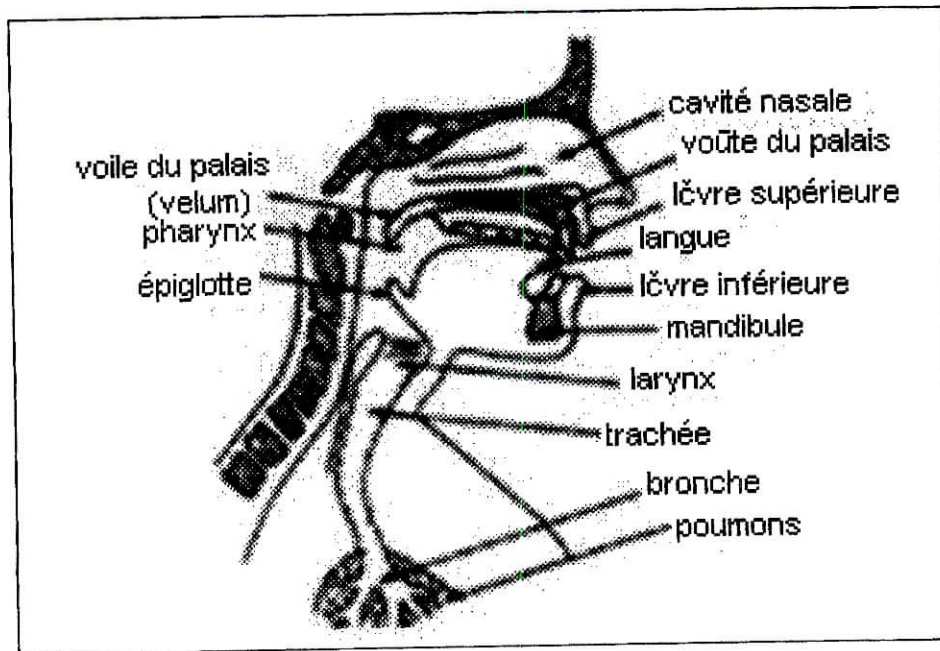


Fig. 1.2 : Description anatomique de l'appareil phonatoire vue d'ensemble (coupe verticale)

Considérés comme de véritables réservoirs d'air, les poumons et la trachée, par laquelle est évacué l'air vers les cordes vocales, forment le système sus-glottique. En jouant donc le rôle d'un générateur de débit d'air, cet ensemble est appelé soufflerie.

Le système sus-glottique est divisé en deux cavités, la cavité orale et la cavité nasale. La première comprend le pharynx et un ensemble d'articulateurs : la langue, les lèvres et la mandibule. La cavité nasale englobe les deux fosses nasales. Le voile du palais qui constitue l'élément mobile de la voûte du palais commande, par ses différentes positions, le degré de couplage nasal dans le processus de phonation.

Relevé, il met complètement les fosses nasales hors circuit et le flux d'air est dirigé alors vers la partie la plus accessible de la cavité orale à savoir les lèvres. Complètement abaissé, il aiguille, vers elles, la totalité de l'onde sonore. Cet ensemble pharyngo-buccalo-nasale (pharyngale, buccale, et nasale) forme le conduit vocal.

En générant l'onde glottale qui véhiculera le signal sonore qui modélisera la forme du conduit vocal, le transformant ainsi en une sorte de voie de passage de la glotte vers les lèvres, le larynx, par le biais des vibrations laryngées des cordes vocales, constitue la partie la plus mobile du système sus-glottale.

### 1. 2. 1. 1. Vibration laryngée

Une corde vocale n'est pas seulement un muscle mais aussi un organe recouvert de plusieurs couches de fibres et d'une couche de muqueuse (Fig. 1.3). Cependant, la phonation d'un son voisé (voyelle), et sous l'effet d'un accroissement de la pression d'air, les cordes vocales sont poussées à part de manière à former une fente appelée glotte. A travers cette dernière la bouffée d'air formée est expulsée.

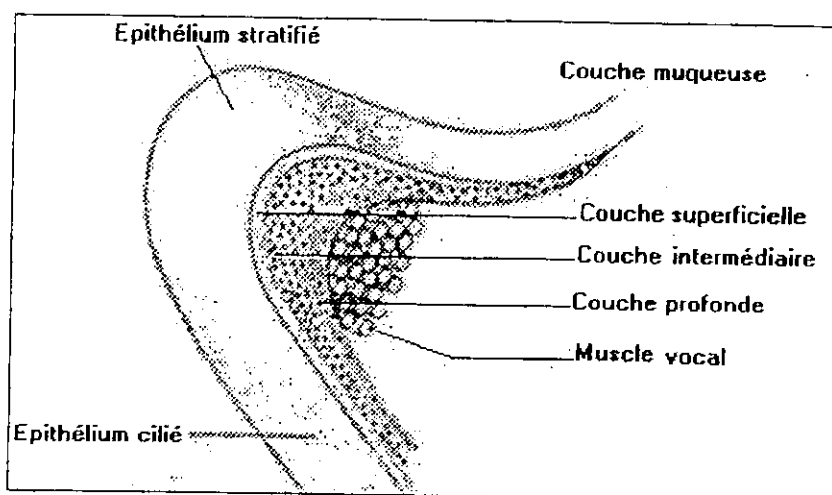


Fig. 1.3 : Visualisation de l'épaisseur d'une corde vocale (coupe verticale)

La vue du larynx dans le plan de la glotte (Fig. 1.4), montre que lors d'un mouvement respiratoire, les aryténoïdes sont écartées laissant le flux d'air circuler sans contrainte. Lorsque le processus de la phonation débute, ces cartilages pivotent pour s'accoler et entraîner une fermeture de la glotte.

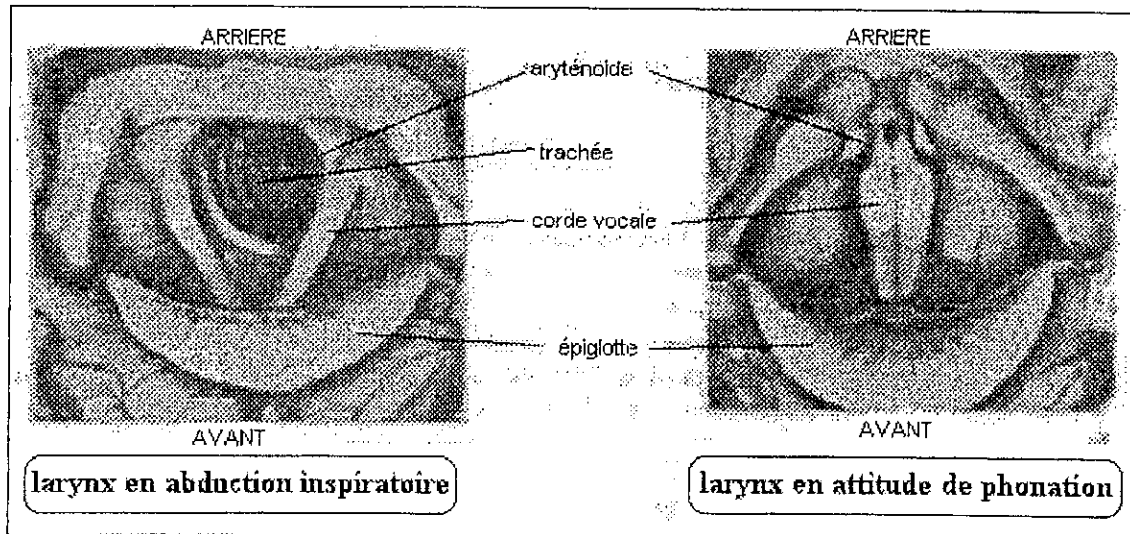


Fig. 1.4 : Vue du larynx

Le principe de vibration laryngée se base sur la théorie aérodynamique-myoélasticité produite par Helmholtz et Muller au le 19<sup>ème</sup> siècle et a été développée, dans les années cinquante, par Van den Berg [3]. Cette théorie définit le mouvement vibratoire comme étant le résultat de l'interaction de deux forces. La pression sous-glottale (en dessous des cordes vocales) s'accroît jusqu'à briser la résistance de la glotte fermée au passage de l'air en provenance des poumons et forcer ainsi l'ouverture glottique.

Dès que la pression sous-glottique décroît, l'effet Bernoulli, l'autre force, fait accoler à nouveau les cordes vocales provoquant ainsi la refermeture de la glotte. En une seconde, les cordes vocales vont donc s'accoler et se décoller quelques centaines de fois (500 fois pour un son de fréquence  $f=500\text{Hz}$ ), générant par ce mouvement vibratoire une série d'ondes acoustiques (Fig. 1.5) qui vont se propager dans le conduit vocal.

## 1. 2. 2. Caractérisation articulatoire et acoustique du mécanisme de la phonation

L'interprétation physique de la production de la parole conclut que l'appareil phonatoire humain correspond à un système acoustique complexe. Il est à rappeler qu'un système acoustique englobe généralement deux composantes : un excitateur et un ou plusieurs résonateurs. L'excitateur délivre un signal-source qui se propage à l'intérieur du résonateur et dont certaines composantes représentant la fréquence de résonance seront affaiblies ou renforcées dans certaines parties du spectre. Ainsi, l'appareil phonatoire humain est composé d'un excitateur : le système complexe glotte-cordes vocales et d'un ensemble de résonateurs (Fig. 1.6).

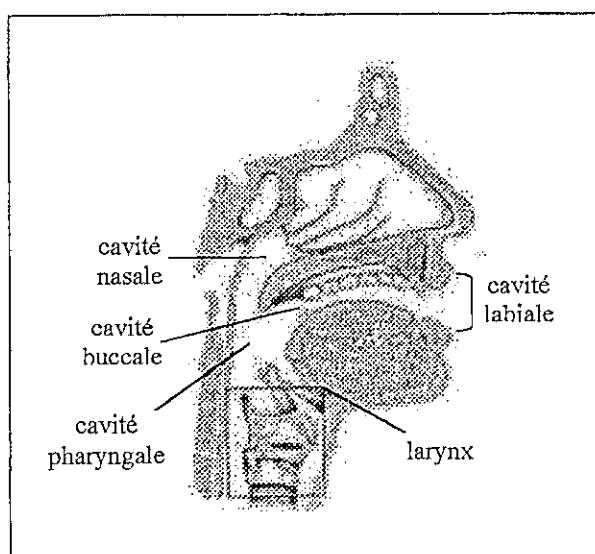


Fig. 1.6 : Cavités résonnantes de l'appareil phonatoire [5]

La majorité des sons du langage produits sont le fait de l'expulsion, des cordes vocales (source d'excitation), d'une colonne d'air, laquelle produit un signal périodique dont la fréquence indique la hauteur de la voix. Ce flux d'air traversera, par la suite, un ou plusieurs résonateurs du conduit vocal. La présence ou l'absence d'obstacles sur son parcours situe la nature du son généré.



Considérée comme l'articulateur principal qui contrôle le conduit vocal, la langue divise ce dernier en deux parties résonantes, le pharynx et la bouche qui, à son tour, englobe trois cavités résonantes : les cavités buccale, labiale et nasale. En tout, le conduit vocal se compose donc de quatre cavités résonantes (Fig. 1.6).

Chacune des cavités résonantes de ce corps sonore de l'appareil phonatoire détermine, par le biais d'une procédure de filtrage, la caractéristique de transmission. Certaines composantes fréquentielles du spectre du signal laryngé seront renforcées afin de permettre un passage avec moins d'atténuation que les autres.

Au moment où la source d'excitation se met en mode vibratoire, un signal laryngé se propagera le long du conduit vocal. Chacune des cavités résonantes de ce corps sonore de l'appareil phonatoire détermine, par le biais d'une procédure de filtrage, la caractéristique de transmission. Certaines composantes fréquentielles du spectre du signal laryngé seront renforcées afin de permettre un passage avec moins d'atténuation que les autres. Des fréquences de résonance ou pics, représentant les formants (Fig. 1.7), seront alors observées dans ce spectre.

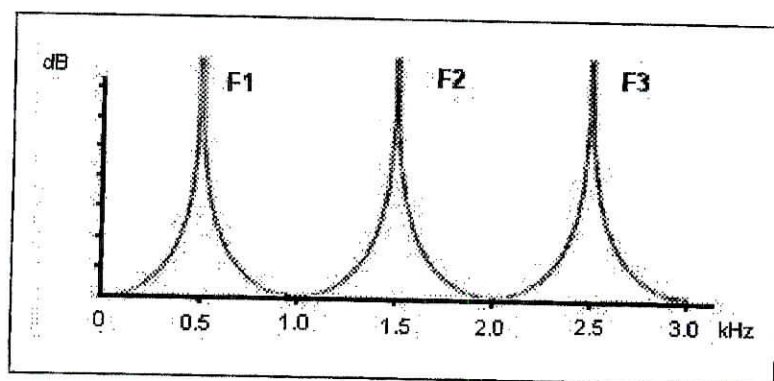


Fig. 1.7 : Les résonances formantiques d'un spectre vocalique

Les formants d'un signal vocal jouent un rôle fondamental dans la caractérisation d'un son voisé. Leur nombre varie selon le volume de la cavité et la surface de l'ouverture du résonateur : d'un seul à (théoriquement) une infinité.

Les deux ou trois premiers formants  $F_1$ ,  $F_2$ ,  $F_3$  suffisent généralement pour la caractérisation perceptuelle des voyelles et des consonnes voisées de la parole. Il est à noter aussi, que le couplage entre les cavités nasale et pharyngobuccale introduit des anti-résonances qui tendent à avoir des formants avec des largeurs de bande plus importants que celles des résonances. Ce sont des "anti-formants".

### 1.2.2.1 L'opposition voisée-non voisée

Tous les sons du langage peuvent être classés, suivant le mode de leur production, en deux catégories. Ils proviennent soit :

- d'une vibration laryngée : "son voisé" ;

ou

- d'un bruit de passage de l'air pulmonaire à travers les organes phonatoires : "son non voisé".

Dans la première catégorie, les mouvements de vibration des cordes vocales sont adductés (fermeture glottique). Le débit glottal émis est alors considéré comme étant la forme de l'onde du signal sonore périodique fourni par la source d'excitation voisée (Fig. 1.8).

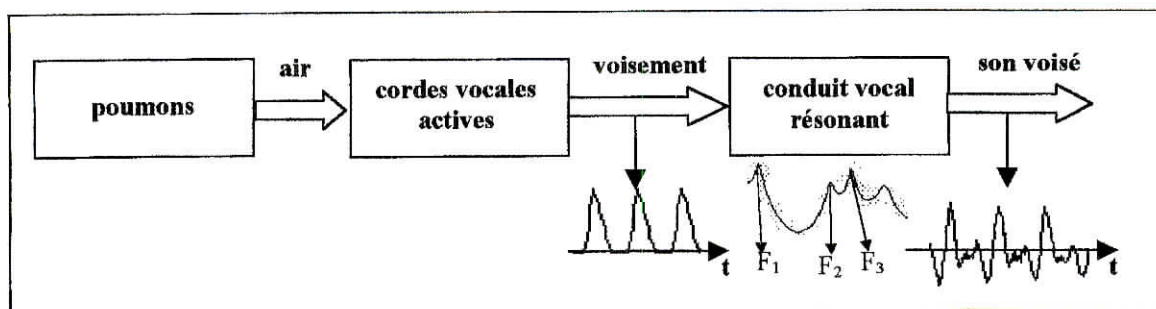


Fig. 1.8 : Production des sons voisés [5]

En revanche, dans la deuxième catégorie, les cordes vocales sont en mouvements d'abduction (ouverture glottique) et sont donc tenues à part de manière à ne pas affecter le passage de l'air pulmonaire par les vibrations glottales. Le signal d'excitation ainsi émis est donc de type apériodique (Fig. 1.9).

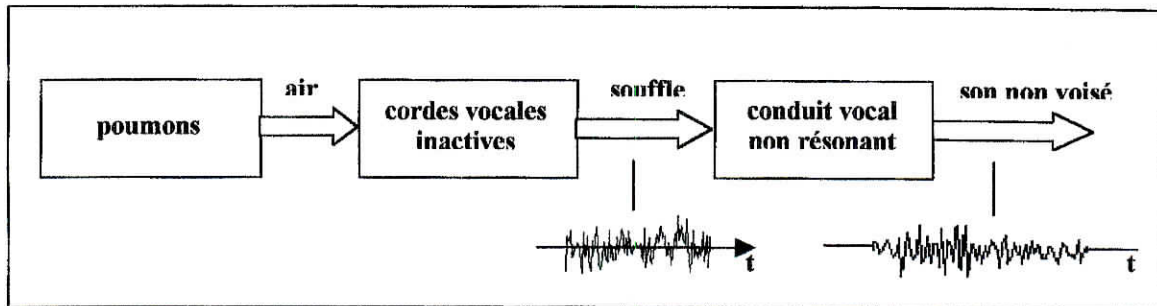


Fig. 1.9 : Production des sons non voisés [5]

### 1.2.2.2 L'opposition oral-nasal

Lors de la production d'un son voisé, la caractéristique de transmission du conduit vocal peut être modifiée par un couplage dans la cavité nasale sous le contrôle du vélum. Si ce dernier est relevé, il interdira alors toute communication entre les conduits nasal et oral. Le son généré sera appelé "son oral". Par contre, si le vélum est complètement abaissé, une connexion entre la cavité nasale, le pharynx et le conduit oral s'effectuera pour laisser place à un libre passage de l'air à travers le conduit nasal. Les "sons nasaux" ainsi produits génèrent des résonances à l'intérieur de la cavité nasale, lesquelles introduisent des zéros dans la fonction de transfert du système vocal.

### 1.2.2.3. Les caractéristiques de la voix

Il est vrai que la forme de l'onde du signal de la source d'excitation influe, d'une façon qualitative, sur le type des caractéristiques vocales d'un signal de parole.



En revanche, il est faux d'affirmer que ceci est toujours le cas. En effet, le type de la voix peut être affecté aussi bien par le mouvement et l'emplacement des articulateurs de l'appareil phonatoire que par d'autres facteurs tels que le débit vocal ou l'accent.

Par exemple, une phrase spécifique peut être parlée en Arabe standard avec un accent espagnol, allemand ou autre. Dans ce cas, les différences perçues sont dues à ceux des mouvements des articulateurs et de la forme de l'onde du signal d'excitation.

Ainsi, la configuration du conduit vocal influera sur les caractéristiques vocales d'un signal de parole. C'est pourquoi, la supposition de l'indépendance entre la source de l'excitation et le conduit vocal, dans la modélisation de la production d'un signal vocal, n'est pas toujours valable. Les paramètres vocaux et ceux de la parole, utilisés pour modéliser un signal vocal, sont donc totalement différents.

Destinés aux différentes applications du traitement automatique de la parole comme la synthèse de la parole et la conversion ou la reproduction des paramètres d'une voix particulière ; plusieurs facteurs deviennent alors importants.

Associés au mode des vibrations des cordes vocales et à leur fréquence fondamentale, les dimensions globales du conduit vocal constituent un facteur primordial puisqu'elles permettent le calcul des fréquences des formants et leurs largeurs de bande.

La fréquence fondamentale est inversement proportionnelle à la longueur de la corde vocale et directement proportionnelle à la racine carrée de la tension de son tissu :

$$F_0 = \frac{1}{2L} \sqrt{\frac{\sigma}{\rho}} \quad (1.1)$$

Avec : L : longueur de la corde  
 $\sigma$  : tension dans le tissu de la corde  
 $\rho$  : densité du tissu de la corde



Un autre facteur peut être inclus. Il s'agit du processus dynamique du mécanisme de la production de la parole qui est affecté par les aptitudes articulatoires et habitudes du parlé d'un locuteur, lesquelles à leur tour influent sur les variations prosodiques telles que l'intonation, le stress et la durée de la parole.

### 1. 2. 3. Classification des sons d'un langage

L'unité sonore la plus petite d'un langage s'appelle "Le phone". Une collection de ces sons articulés forme les éléments de la parole intitulés "les phonèmes". Ces derniers qui peuvent être combinés dans des unités de parole larges, appelées "syllabes", permettent la distinction entre les différents mots d'un langage.

Trois sortes de classification des sons d'un phonème peuvent être citées :

- une classification acoustique qui dépend du type de la source d'excitation : voisée ou non voisée (opposition sourde-sonore),
- une classification articulaire qui est déterminée par le mode d'articulation, le lieu ou le point d'articulation et la nasalité du son produit (opposition nasale-orale),
- une classification par phonème selon que le son émis correspond à une consonne ou une voyelle. Si le passage de l'air s'effectue librement à partir de la glotte, on a affaire à une voyelle. Si, au contraire, il est complètement ou partiellement obstrué en un ou plusieurs endroits du conduit vocal, c'est le son d'une consonne qui sera produit.

### 1. 2. 3. 1. Point d'articulation et mode d'articulation

Le mode d'articulation d'un son vocal obéit à un certain nombre de facteurs. Ces derniers apportent les modifications à la nature du flux d'air expulsé des poumons vers le conduit vocal et parmi lesquelles on distingue :

- la mise en vibration ou le libre passage de la colonne d'air au niveau de la glotte (sonore ou sourde);
- libre passage, ou blocage momentané ou complet, en un point quelconque (point d'articulation) des cavités sus-glottiques (voyelle ou consonne);
- passage par une voie unique ou deux voies différentes (orale ou nasale).

Le point d'articulation constitue un lieu, de la cavité buccale, où le passage du flux d'air est obstrué. D'une manière plus indicative, on assimile le point d'articulation à l'endroit où se place la langue en vue de l'obstruction du passage de la colonne d'air. Afin d'établir une classification articulatoire d'un phonème, nous pouvons utiliser les points d'articulations suivants :

- les lèvres : articulations labiales;
- les dents : articulations dentales;
- les lèvres et les dents : articulations labio-dentales;
- les alvéoles : articulations alvéolaires;
- le palais: articulations pré-palatale, médio-palatale et post-palatale;
- le voile du palais : articulations vélares;
- le pharynx : articulations pharyngales;
- la glotte : articulations glottales.

### 1. 2. 3. 2. Les voyelles

La caractéristique principale d'une voyelle est le libre passage de l'air à travers les cavités sus-glottiques. Le flux d'air ne subira alors que les effets de la résonance à l'origine de

la présence de bandes de fréquences, appelées formants, dont l'énergie est particulièrement élevée (Fig. 1.10).

Ainsi, une voyelle ou tout autre son voisé est considérée comme une réponse du système du conduit vocal à une source de voisement produite par la vibration laryngienne. En d'autres termes, la voyelle possède les propriétés de la source d'excitation modifiées par une action de filtrage du conduit vocal [6]. Le tableau 1.1 indique quelques grandeurs numériques de la fréquence de la vibration laryngienne appelée fréquence fondamentale.

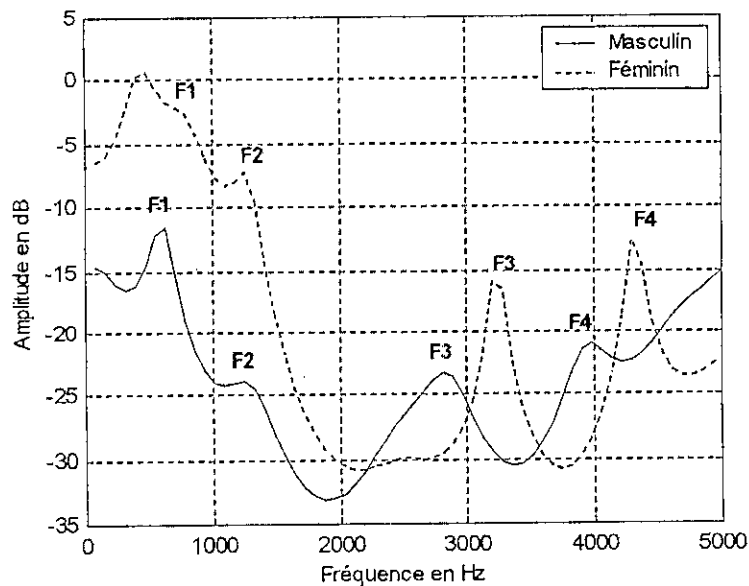


Fig. 1.10 : Formants de la voyelle [a] pour deux locuteurs masculin et féminin

	$F_0$ moyenne (Hz)	$F_0$ minimale (Hz)	$F_0$ maximale (Hz)
<b>Hommes</b>	125	80	200
<b>Femmes</b>	225	150	350
<b>Enfants</b>	300	200	500

Tableau 1.1 : Les fréquences fondamentales minimale, moyenne et maximale pour les hommes, femmes et enfants.

Les voyelles sont rangées selon quatre traits [1] :

- la nasalité : existence, par opposition aux voyelles orales, des voyelles nasales lesquelles sont marquées par l'écoulement d'une partie de l'air pulmonaire à travers les fosses nasales;
- le degré d'ouverture du conduit vocal : ce critère indique le nombre de résonateurs impliqués dans la production de la voyelle prononcée et le volume du résonateur buccal, lequel dépend directement de la distance séparant le point le plus élevé de la langue du palais (degré d'ouverture). On note quatre degrés d'ouverture, du premier, le plus fermé, au quatrième le plus ouvert;
- la position de la constriction du conduit vocal : localisation de l'emplacement de la masse du dos de la langue dans la bouche, ce qui explique l'existence de voyelles antérieure et postérieure;
- l'arrondissement des lèvres : existence de voyelles arrondie et non arrondie.

Les voyelles ont certaines caractéristiques spectrales et sont généralement représentées par la forme d'onde, le spectre et le spectrogramme (Fig. 1.11).

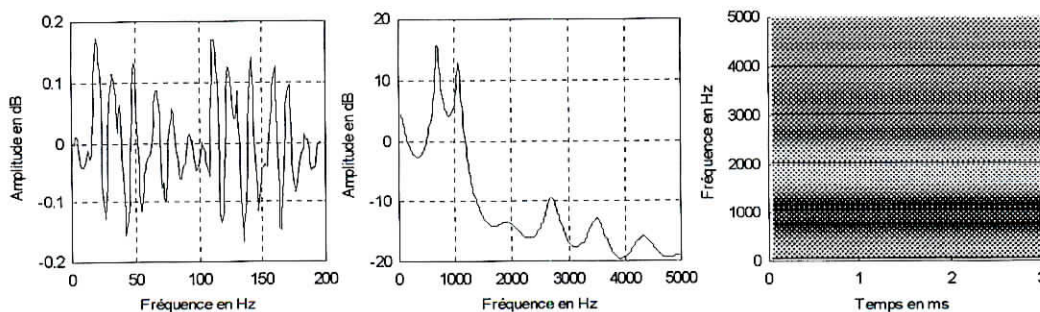


Fig. 1.11 : Forme d'onde, spectre et spectrogramme de la voyelle [a]



Lors d'une réalisation vocalique, une voyelle peut être représentée, suivant les valeurs des fréquences des deux premiers formants  $F_1$  et  $F_2$ , dans un triangle vocalique. (Fig. 1.12).

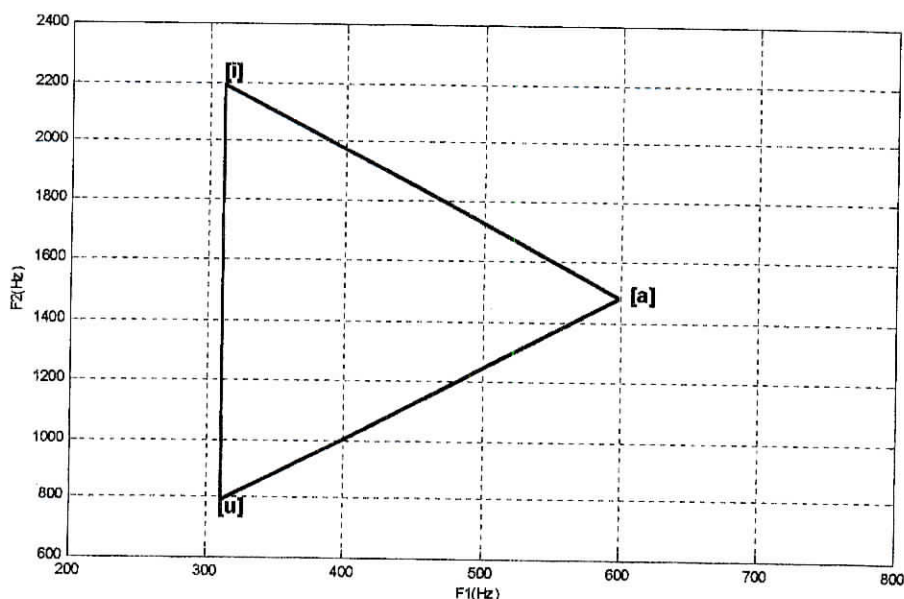


Fig. 1.12 : Le triangle vocalique pour les locuteurs mâles

### 1.2.3.3 Les consonnes

Contrairement aux voyelles, la production d'une consonne est marquée par l'obstruction du passage de l'air pulmonaire. La constriction du flux d'air ainsi produite peut être localisée en un ou plusieurs points du conduit vocal. En tenant compte du voisement, des mode et point d'articulation, les consonnes sont groupées dans plusieurs sous-classes. Par exemple, les consonnes occlusives (plosives) telles que [t], [b] et [k] sont reconnues par le biais du mode de l'articulation. Elles sont générées par une interruption de l'écoulement du flux d'air en un point particulier de la cavité orale, ce qui provoque une accumulation de pression.

Une fois l'occlusion finie, une perturbation acoustique prendra forme et aura l'allure d'une onde de pression, assimilée à une source d'excitation bruitée et due au relâchement brusque de l'air qui était comprimé par la fermeture complète du conduit vocal.

Les consonnes fricatives (constrictives), comme [f] et [s], sont toujours classées selon l'absence ou la présence du voisement dans la source d'excitation. Contrairement aux occlusives, ce genre de consonne est caractérisé par un écoulement turbulent de l'air, à l'origine de la formation d'une constriction dans le conduit vocal et surtout l'apparition d'un bruit de friction. Le type de ce dernier est conditionné par les positions et la tension musculaire des lèvres et de la langue.

Les nasales, un autre genre de consonne, sont produites par la connexion de la cavité nasale au pharynx et au conduit oral en abaissant le vélum.

En présentant, dans certains cas, la particularité de posséder une structure formantique, les liquides et les semi-voyelles peuvent être groupés avec les consonnes. Dans le même contexte, le passage d'une consonne à une voyelle, et inversement, produit des transitions formantiques c'est-à-dire des variations dans la position des formants.

Le tableau 1.2 montre, d'une façon générale, quelques caractéristiques des phonèmes en général et des consonnes en particulier.

NV : Non Voisée M : Mixée				
Son	Energie	Excitation	Passage par zéro	Formants
Voyelles	élevée	V : voisée	bas	oui
Occlusives	basse	NV et M	élevé	non
Fricatives	basse	NV et M	élevé	non
Nasales	moyenne	V	bas	oui

Tableau 1.2 : Quelques caractéristiques des phonèmes.

### 1. 2. 4. Les sons de l'Arabe Standard

La langue Arabe fait partie de la famille des langues du proche-orient intitulée chamito-sémitiques. La combinaison de ces phonèmes en des unités larges de parole, forme, «les syllabes ». (ex. [ma] est la première syllabe du mot bi-syllabique [maça] « Avec »).

Ces phonèmes permettent la distinction entre les différents mots du langage. Acoustiquement, les sons de l'AS sont classés suivant le voisement ou le non voisement de la source d'excitation (opposition sonore-sourde).

Sur le plan articulatoire, ils sont groupés selon le type des mode, lieu et point des articulations et la présence ou l'absence de la nasalité (opposition nasale-orale). Une classification par phonème selon que le son produit correspond à une consonne ou à une voyelle, peut s'ajouter aux deux précédentes.

L'AS comprend vingt huit phonèmes consonantiques, trois voyelles brèves et leurs opposées longues. La voyelle brève que les grammairiens arabes nomment la [ḥaraka] est représentée par un signe diacritique. Des mesures des fréquences des formants et des durées des voyelles de l'AS, ont été établies à partir des spectrogrammes. Les voyelles longues semblent avoir une durée double de celle des voyelles brèves (Tab. 1.3) [7].

Voyelles	Durée (ms)	F <sub>1</sub> (Hz)	F <sub>2</sub> (Hz)	F <sub>3</sub> (Hz)
[i]	300	290	2200	2700
[i:]	600	285	2200	2700
[u]	300	290	800	2150
[u:]	600	285	775	2050
[a]	300	600	1500	2100
[a:]	600	675	1200	2150

Tableau 1.3 : Durées et fréquences formantiques relatives aux voyelles arabes

On distingue trois signes : les [fatḥa], [d<sup>°</sup>ammah] et [kasrah]. Ces derniers représentent respectivement les voyelles brèves [a], [u] et [i]. Quant aux voyelles longues [a:], [u:] et [i:], elles sont respectivement notées par les [ḥuru:f al\_mad] [alif], [wa:w] et [ya:ʔ] (Tab. 1.4).

Voyelles	Courtes	Voisées	[a] [i] [u]	
	Longues		[a:] [i:] [u:]	
Consonnes	Occlusives	Voisées	pharyngalisées	[d <sup>°</sup> ض]
			[ʔ] [ʔ] [ʔ] [b]	
		Non voisées	pharyngalisées	[t <sup>°</sup> ط]
			[k] [q] [t]	
	Fricatives	Voisées	pharyngalisées	[z <sup>°</sup> ظ]
			[h] [ʕ] [ʕ <sup>°</sup> غ] [z] [ð]	
		Non voisées	pharyngalisées	[s <sup>°</sup> ص]
			[ḥ] [x] [ʃ] [s] [ʃ] [θ] [ʔ]	
	Nasales	voisées	[n] [m]	
	Liquides		[l]	
	Semi-consonnes		[r]	
			[w] [y]	

Tableau 1.4 : Consonnes et voyelles de l'Arabe Standard suivies de leurs transcriptions en Alphabet Phonétique International (API)

Les consonnes de l'AS sont groupées suivant le caractère individuel de leur variabilité. Le fait qu'elles soient moins énergétiques que les voyelles, rend leurs caractéristiques instables.

L'originalité du système consonantique de l'AS est indiquée par la présence des consonnes emphatiques pharyngalisées et la prédominance des fricatives par rapport aux occlusives. Nous analysons les consonnes emphatiques sous deux aspects : spectrographique et contours formantiques.



### 1. 2. 5. Les consonnes emphatiques

Les quatre articulations de l'AS, que les anciens Grammairiens Arabes ont appelé [mut<sup>6</sup>baqa], sont traditionnellement définies comme consonnes «emphatiques». Sibawayh, le célèbre grammairien de l'Ecole d'El Basra (Iraq), a décrit dans son livre les trois qualités des consonnes emphatiques : [L'oit<sup>6</sup>ba:q] « fait de couvrir », [l'istiçla] « fait d'être élevé, porté vers le haut » et le [tafxi:m] « expression acoustique grasse et épaisse » [8].

Sur le plan articulatoire, ces consonnes sont définies par une rétraction de la racine de la langue et une association d'une articulation secondaire à une articulation primaire. Malgré cette définition et le fait que l'essentiel de l'articulation emphatique se passe au fond du conduit vocal, le processus articulatoire de production de ce type de consonne demeure quelque peu obscur. Cependant, une analyse acoustique de ces sons s'avère intéressante.

### 1. 3. Conclusion

Les caractéristiques articulatoires et acoustiques de signal vocal ont été décrites dans ce chapitre. La dualité acoustico-articulatoire a été évoquée. Elle souligne la complexité du mécanisme de production vocale, car plusieurs paramètres peuvent agir simultanément dans ce processus, parfois de façon antagoniste. L'approche expérimentale, que nous aborderons dans les chapitres 3 et 4 montrera que toute variation articulatoire des signaux de la parole provoquera certainement des conséquences d'ordre acoustique, et ainsi l'étude de l'aspect acoustique des sons s'avère utile.

# Chapitre 2

## Etude de la modélisation du système de production vocale

### 2.1 Introduction

Après avoir passé en revue la description idéale du mécanisme de production de la parole, nous allons maintenant, dans ce chapitre, approfondir l'étude de ce phénomène physique en présentant ses modèles mathématiques déjà étudiés. Le recours à la modélisation de la production vocale humaine s'est avéré impératif dans la mesure où elle reproduit fidèlement les mécanismes de production de notre parole et mène directement à une synthèse de parole performante et à la réalisation de codeurs de parole à faible débit.

Aujourd'hui, la procédure d'analyse d'un signal vocal fait appel à une variété de modèles de production et de perception de la parole notamment ceux, de type spectral, qui servent à décrire la forme d'onde d'un signal vocal. Nous citons comme exemple, les modèles à tubes acoustiques, les modèles pôle-zéro, les modèles à filtres numériques ...etc.

L'ensemble de ces modèles est englobé dans trois classes :

- les modèles acoustiques;
- les modèles articulatoires;
- les modèles perceptifs.

Les modèles articulatoires transforment les paramètres articulatoires d'un signal de parole en une suite vectorielle d'aires de section transversale du conduit vocal qui, à leurs tour, sont convertis en caractéristiques acoustiques. L'étude des relations existantes entre chacun des formants d'une voyelle et la disposition du conduit vocal facilite la mise en œuvre de ce genre de modèles [9].

La conception des modèles acoustiques est destinée à la simulation de la propagation du son dans le système vocal. On distingue le modèle des conduits vocal et nasal, le modèle des sources d'excitation qui génère la forme de l'onde de l'excitation voisée des voyelles et le flux d'air turbulent de la constriction des fricatives ou des plosives et le modèle de rayonnement de l'énergie acoustique des lèvres ou des narines.

Ces modèles sont décrits par une série d'équations différentielles partielles décrivant les propriétés acoustiques du mouvement d'air dans le système vocal. Leur résolution est d'une extrême difficulté qu'elle nécessite certaines simplifications concernant la forme géométrique du conduit vocal et les pertes d'énergie dans le système vocal (Fig. 2.1) [1, 2].

Par conséquent, la source et le système du conduit vocal seront supposés linéairement séparables, c'est-à-dire que l'effet du couplage entre eux sera négligé.

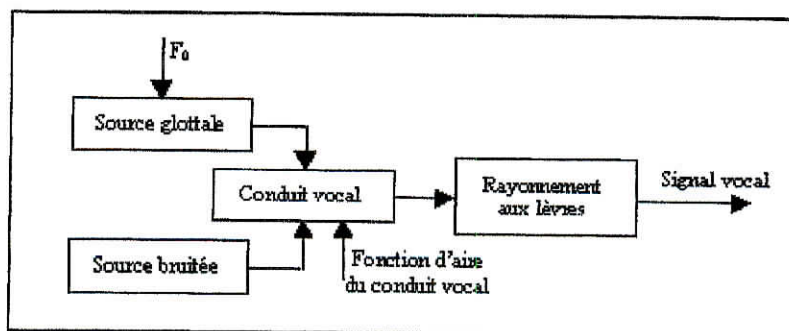


Fig. 2.1 : Modèle fonctionnel de production de la parole

Les modèles de la source d'excitation glottale et des cavités orale et nasale simulent respectivement la génération de l'onde glottale destinée au conduit vocal, et la propagation des ondes sonores le long de ces cavités. Le modèle de la source bruitée génère, pour les consonnes fricatives et occlusives, le signal représentant le flux d'air turbulent, localisé au niveau de la constriction (ou fermeture). Tandis que l'énergie acoustique rayonnée à partir des lèvres et parfois des narines est simulée par le modèle de rayonnement.

## 2. 2 Modélisation acoustique du conduit vocal

Dans la réalité, le conduit vocal est modélisé par un tube acoustique, à pertes, tridimensionnel, légèrement courbé, variant dans le temps et composé de sections transversales non uniformes ainsi que des parois extensibles [5, 7]. Pour lever les différentes limitations de cette modélisation, causées par les valeurs réelles des paramètres acoustiques du conduit vocal tels que les pertes d'énergie dues à la viscosité, la conduction thermique et à la non-rigidité des parois du conduit, la plupart des chercheurs ont évoqué le besoin impératif d'une version simplifiée du modèle acoustique du conduit vocal.

Ainsi, le modèle acoustique élémentaire du conduit vocal est un tube sans pertes droit, formé d'aires de sections transversales circulaires ou elliptiques et variables (Fig. 2.2).

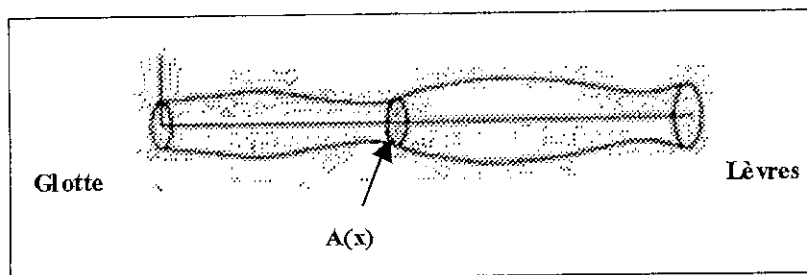


Fig. 2.2 Conduit vocal idéal à aires de section transversale variable



De plus, le fait de considérer 5 khz comme borne supérieure de la largeur de bande du signal vocal et de tenir en compte de la mollesse du tissu le long du conduit vocal, à l'origine de l'empêchement de toute propagation radiale de l'onde sonore, mène à supposer que le type de la propagation du son le long de l'axe du tube n'est autre que plane.

Ce qui signifie que seule l'aire de la section transversale le long du conduit vocal détermine les paramètres acoustiques de ce dernier. Par conséquent, les équations acoustiques seront décrites dans une seule dimension au lieu de trois et la fonction d'aire du conduit vocal sera approximée par un nombre de sections successives suffisamment petit pour que chacune d'elles possédera une aire de section transversale.

### 2. 2. 1. La propagation du son

Après avoir négligé les pertes provoquées par la viscosité de la friction, la conduction thermique et les extensions des parois, et considéré les lois de continuité et celles de Newton; une paire d'équations régissant la propagation d'une onde plane linéaire dans le conduit vocal est exprimée comme suit : [5, 11]

$$\frac{-\partial u(x,t)}{\partial x} = \frac{1}{\rho c^2} \frac{\partial (p(x,t).A(x,t))}{\partial t} + \frac{\partial A(x,t)}{\partial t} \quad (2.1.a)$$

$$\frac{-\partial p(x,t)}{\partial x} = \frac{\rho \partial (u(x,t) / A(x,t))}{\partial t} \quad (2.1.b)$$

Où  $p(x,t)$  et  $u(x,t)$  représentent respectivement, au point  $x$  et à l'instant  $t$ , la variation de la pression sonore dans le tube et le débit volumique,  $\rho$  la densité de l'air dans le tube,  $c$  la vitesse du son et  $A(x,t)$  la fonction d'aire du tube. La résolution de ce système d'équations, qui ne peut être que numérique, n'est effectuée qu'une fois que la fonction d'aire  $A(x,t)$  et les conditions aux limites seront spécifiées.

Le conduit vocal sera alors idéalement modélisé par un tube sans pertes formé d'un accolement de sections uniformes. Les effets de pertes seront incorporés par la suite.

### 2. 2. 2 Le tube uniforme sans pertes

Le tube uniforme sans pertes traduit l'idéalisation de la modélisation acoustique du conduit vocal. Ce tube se compose de  $S_N$  sections élémentaires possédant chacune d'elles une aire de section transversale  $A_k$  et une longueur  $l_k$  où  $1 \leq k \leq S_N$ .

La longueur la glotte à la  $i^{\text{ème}}$  section ( conduit vocal) est indiquée comme suit :

$$x_{i-1} = \sum_{k=1}^{i-1} l_k \quad (2.2)$$

Où  $l_i$  et  $A_i$  représentent respectivement la longueur et l'aire constante de section transversale de la  $i^{\text{ème}}$  section.

Les équations différentielles couplées (2.1) deviennent donc pour cette section élémentaire :

$$\frac{-\partial u_i(x,t)}{\partial x} = \frac{A_i}{\rho c^2} \frac{\partial (p_i(x,t))}{\partial t} \quad (2.3.a)$$

$$\frac{-\partial p_i(x,t)}{\partial x} = \frac{\rho}{A_i} \frac{\partial (u_i(x,t))}{\partial t} \quad (2.3.b)$$

Avec :  $p_i(x,t)$  : la pression le long de l'axe de direction de propagation  $x$ .

$u_i(x,t)$  : le débit volumique le long de l'axe de direction de propagation  $x$ .

$x_{i-1} \leq x \leq x_i$

Les solutions aux équations (2.3) s'écrivent sous la forme suivante [2, 11, 12] :

$$u_i(x, t) = [u_i^+(t - x/c) - u_i^-(t + x/c)] \quad (2.4.a)$$

$$p_i(x, t) = \frac{\rho c}{A_i} [u_i^+(t - x/c) - u_i^-(t + x/c)] \quad (2.4.b)$$

où  $u_i^+(t - x/c)$  et  $u_i^-(t + x/c)$  indiquent respectivement les ondes incidente et réfléchie.

Les conditions aux limites de chaque section déterminent la relation existante entre les ondes mobiles dans les sections adjacentes. La continuité des pressions et du débit volumique, à la fois dans l'espace et dans le temps à un point quelconque de l'intérieur du conduit vocal, permet leur déduction de ces conditions [11].

Ainsi, deux approches de simulation du conduit vocal peuvent être énoncées.

### 2. 2. 2. 1 L'approche de la propagation de l'onde

Cette approche se base sur une résolution analytique des équations (2.4). C'est pourquoi, les solutions, pour une seule section élémentaire d'un tube uniforme sans pertes, seront de la forme :

$$u_i^+(t - x/c) = K_i^+ \exp(j\omega(t - x/c)) \quad (2.5.a)$$

$$u_i^-(t + x/c) = K_i^- \exp(j\omega(t + x/c)) \quad (2.5.b)$$

Avec :  $\omega$  : la pulsation du signal de la source d'excitation.

Dans le même contexte, on remarquera qu'au niveau d'une jonction entre deux sections successives de longueurs et aires de section transversale différentes, chaque onde se caractérisera par une propagation à l'avant et une réflexion à l'arrière (Fig. 2.3).

Un retard de propagation  $\tau$  et un coefficient de réflexion  $r$  seront alors définis.

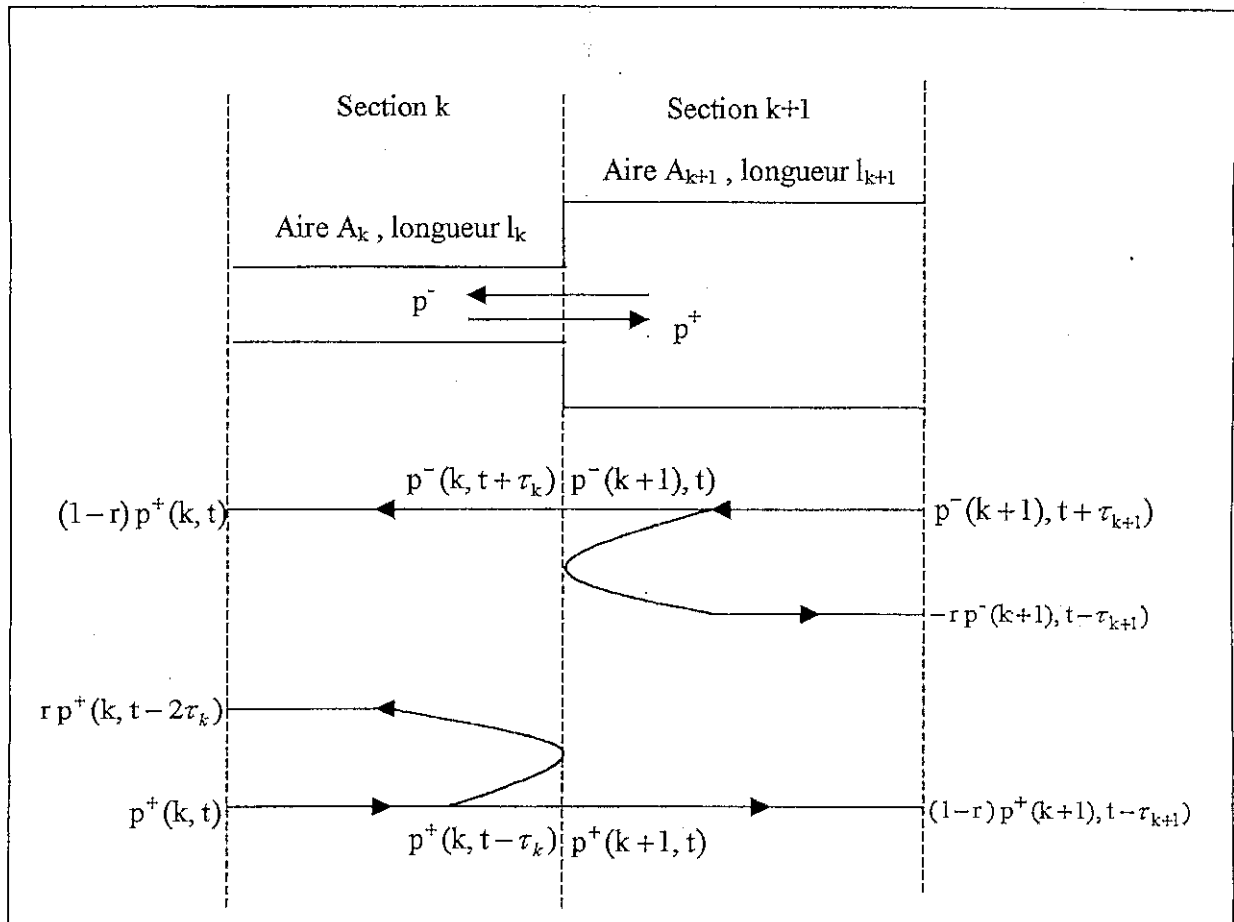


Fig. 2.3 Relations de réflexion à la jonction de deux sections sans pertes

De plus, comme l'indique le graphe de fluence du modèle de tubes sans pertes et son système discret dans le temps équivalent (Fig. 2.4), les conditions aux limites, décrites par les pertes d'énergie aux lèvres et à la glotte, et celles de continuité sont appliquées [2].

L'utilisation du procédé du graphe de fluence pour représenter le modèle acoustique du conduit vocal rends efficace l'extraction de la fonction de transfert d'une concaténation de 'n' tubes.

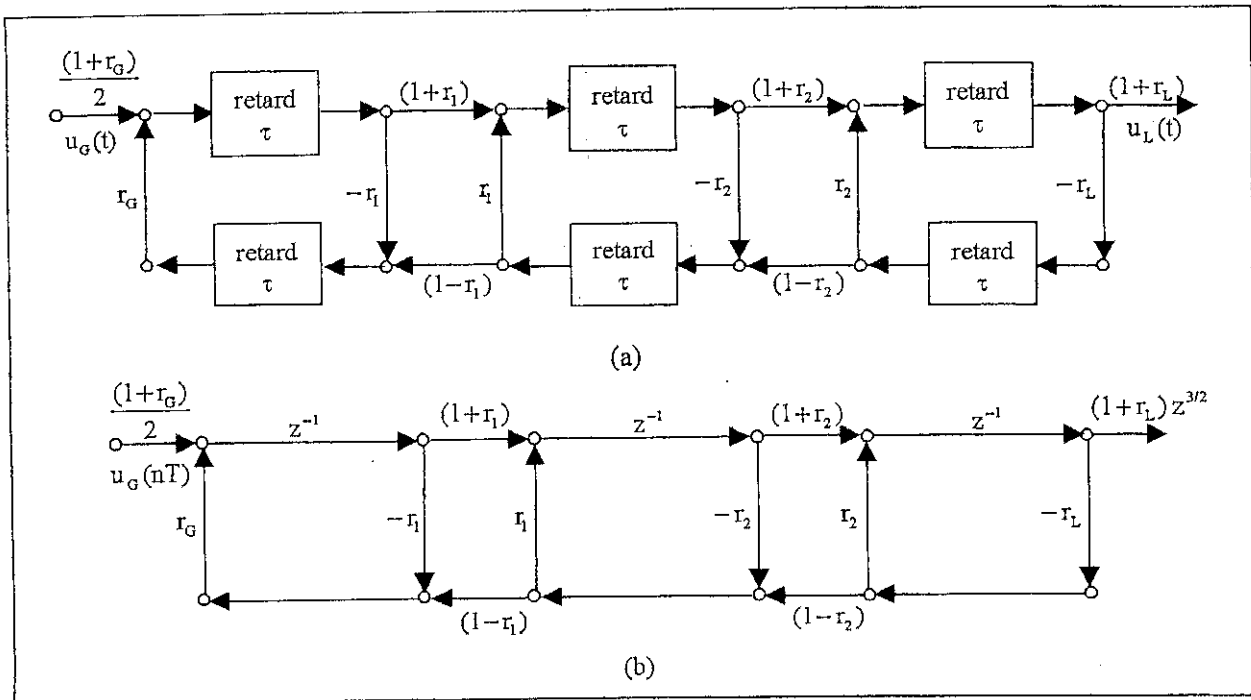


Fig. 2.4 Conduit vocal sans pertes discret dans le temps  
 (a) Graphe de fluence (b) Système équivalent discret dans le temps

Ainsi, nous appliquons, pour le cas de deux tubes mis en bout à bout, les conditions de continuité à la jonction des  $k^{\text{ième}}$  et  $(k + 1)^{\text{ième}}$  sections (Fig. 2.3) comme suit :

$$p_k(l_k, t) = p_{k+1}(0, t) \tag{2.6}$$

$$u_k(l_k, t) = u_{k+1}(0, t) \tag{2.7}$$

En remplaçant les expressions (2.6) et (2.7) dans les équations (2.5), nous obtenons les solutions suivantes :

$$u_{k+1}^+(t) = (1+r_k) u_k^+(t-\tau_k) + r_k u_{k+1}^-(t) \tag{2.8.a}$$

$$u_k^-(t+\tau_k) = -r_k u_k^+(t-\tau_k) + (1-r_k) u_{k+1}^-(t) \tag{2.8.b}$$

Avec :  $r_k$  le coefficient de réflexion pour la  $k^{\text{ième}}$  jonction.



$r_k$  peut être exprimé de la manière suivante :

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \quad \text{Avec } -1 \leq r_k \leq 1 \quad (2.9)$$

Le conduit vocal résonant se modélise, de façon très convenable, par un filtre tout-pôle, auquel on ajoutera des zéros en cas de nasalité. La fonction de transfert  $V(z)$  aura donc la forme suivante :

$$V(z) = \frac{u_L(z)}{u_G(z)} \quad (2.10)$$

Où  $u_L(z)$  et  $u_G(z)$  représentent les transformées en  $z$  respectivement des débits volumiques des lèvres et de la glotte.

Ce qui nous amène à écrire les transformées en  $z$  des équations (2.8) de la manière suivante :

$$u_{k+1}^+(z) = (1+r_k) z^{-1/2} u_k^+(z) + r_k u_{k+1}^-(z) \quad (2.11. a)$$

$$u_k^-(z) = -r_k z^{-1} u_k^+(z) + (1-r_k) z^{-1/2} u_{k+1}^-(z) \quad (2.11. b)$$

Ceci conduit aux solutions suivantes :

$$u_k^+(z) = \frac{z^{1/2}}{(1+r_k)} u_{k+1}^+(z) + \frac{r_k z^{1/2}}{(1+r_k)} u_{k+1}^-(z) \quad (2.12. a)$$

$$u_k^-(z) = \frac{-r_k z^{-1/2}}{(1+r_k)} u_{k+1}^+(z) + \frac{z^{-1/2}}{(1+r_k)} u_{k+1}^-(z) \quad (2.12. b)$$

Les équations (2.12) peuvent se mettre sous la forme matricielle suivante :

$$U_K = Q_K U_{K+1} \quad (2.13)$$

Avec

$$U_K = \begin{bmatrix} u_k^+(z) \\ u_k^-(z) \end{bmatrix} \quad \text{et} \quad Q_K = \begin{bmatrix} \frac{z^{1/2}}{(1+r_k)} & \frac{-r_k z^{1/2}}{(1+r_k)} \\ \frac{-r_k z^{-1/2}}{(1+r_k)} & \frac{z^{-1/2}}{(1+r_k)} \end{bmatrix}$$

En répétant l'application de l'équation (2.13), nous obtenons :

$$U_1 = \prod_{k=1}^N Q_K U_{N+1} \quad (2.14)$$

Puis, la condition aux limites de la glotte s'exprimera comme suit :

$$u_G(z) = \frac{2}{(1+r_G)} u_1^+(z) - \frac{2r_G}{(1+r_G)} u_1^-(z) \quad (2.15)$$

Ou

$$u_G(z) = \begin{bmatrix} \frac{2}{(1+r_G)} & \frac{-2r_G}{(1+r_G)} \end{bmatrix} U_1 \quad (2.16)$$

Et puisque :

$$U_{N+1} = \begin{bmatrix} u_L(z) \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_L(z) \quad (2.17)$$

Alors nous pouvons finalement écrire :

$$\frac{1}{V(z)} = \frac{u_G(z)}{u_L(z)} = \begin{bmatrix} \frac{2}{(1+r_G)} & \frac{-2r_G}{(1+r_G)} \end{bmatrix} \prod_{k=1}^N Q_K \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (2.18)$$

Ainsi, la fonction de transfert  $V(z)$  sera donnée par l'expression suivante :

$$V(z) = \frac{0.5 (1+r_G) \prod_{k=1}^N (1+r_k) z^{-1/2}}{D(z)} \quad (2.19)$$

Où l'expression polynomiale en  $z^{-1}$   $D(z)$  est obtenue par le calcul matriciel suivant :

$$D(z) = [1 \quad r_G] \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \cdots \begin{bmatrix} 1 & -r_N \\ -r_N z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (2.20)$$

En d'autres termes, le retard global et les pôles de la fonction de transfert  $V(z)$  décrivent respectivement le nombre de sections du modèle et les résonances (formants) du conduit vocal.

Dans le cas où  $r_G$  est égale à 1 ( $z_G = \infty$ ), le polynôme  $D(z)$  est calculé grâce à la formule récurrente suivante :

$$D_0(z) = 1 \quad (2.21. a)$$

$$D_k(z) = D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}) \quad k=1, 2, 3, \dots, N \quad (2.21. b)$$

$$D(z) = D_N(z) \quad (2.21. c)$$

Ceci est illustré par l'exemple suivant :

Il est à rappeler qu'en raison de la périodicité de la réponse fréquentielle du modèle à tubes sans pertes, l'approximation du comportement du conduit vocal n'est effectuée que pour la bande fréquentielle suivante :

$$|F| \leq \frac{1}{(2T)} \quad (2.22)$$

Avec :  $T$  la période d'échantillonnage

C'est pourquoi, schafér [10], tenant compte du lien existant entre les choix du nombre de sections et de la fréquence d'échantillonnage du signal de la parole, avait suggéré de prendre :

$$T = 2\tau \quad (2.23)$$

Où  $\tau$ , le temps de propagation dans un sens pour une seule section, peut s'exprimer, pour N sections et une longueur totale l, comme suit :

$$\tau = \frac{l}{Nc} \quad (2.24)$$

Ainsi, en choisissant  $l = 17.5$  cm et  $c = 350$  m/s, nous aurons :

$$\frac{1}{2T} = \frac{Nc}{4l} = \frac{N}{2} (1000) \text{ Hz} \quad (2.25)$$

Pour  $N = 10$  et  $F_c = 10$  KHz, on observera, à chaque 1000 Hz, cinq résonances formantiques dans la bande fréquentielle [0, 5 KHz].

Finalement, afin de calculer la fonction de transfert de 'n' tubes accolés, il suffit simplement de multiplier les fonctions de transfert des différents filtres tout-pôle mis en cascade.

### 2. 2. 2. 2 L'approche de la ligne de transmission

Le modèle du circuit électrique du conduit vocal est axé sur une similitude entre la propagation de l'onde acoustique dans le tube et celle d'une onde électrique le long de la ligne de transmission équivalente. Les analogies permettent de représenter le modèle acoustique du conduit vocal par un quadripôle électrique bien connu (Fig. 2.5) [19].

La résistance série R et la conductance G représentent les pertes d'énergie acoustiques dues respectivement à la viscosité de la friction et la conduction thermique. Composée de trois éléments ( $R_w$ ,  $L_w$ ,  $C_w$ ), l'impédance mécanique acoustique  $Z_w$  décrit les pertes d'énergie des masse-déplacement-viscosité du tissu moelleux des parois du conduit vocal.

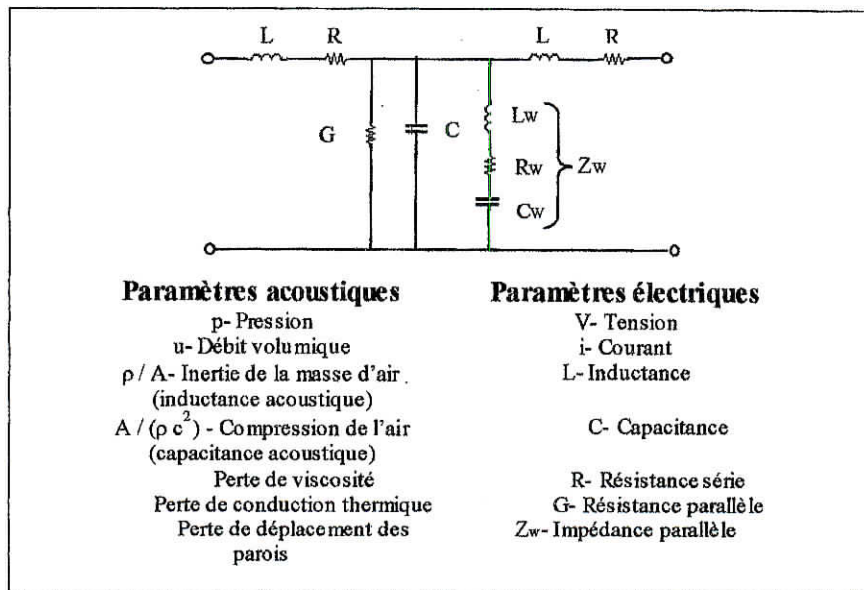


Fig. 2.5 Analogies entre les paramètres acoustiques d'un tube et ceux d'une ligne de transmission électrique

### 2. 2. 2. 3 L'effet des pertes

En réalité, la propagation d'une onde sonore le long du conduit vocal subit l'effet des pertes provoqué par le déplacement de la masse de l'air à l'intérieur du conduit et le rayonnement aux lèvres ou narines.

#### 2. 2. 2. 3. 1 L'impédance des parois du conduit vocal

Les variations de la pression de l'air à l'intérieur du conduit vocal exercent, sur les parois extensibles de ce conduit, une force variable. En supposant que ces parois sont soumises à des réactions locales et que les perturbations de l'aire de la section transversale qui en résultent sont très petites, nous pouvons alors écrire :

$$A(x, t) = A_0(x, t) + \Delta A(x, t) = A_0(x, t) + y(x, t) S_0(x, t) \tag{2.26}$$



Avec :  $A_o(x,t)$  : aire nominale,  
 $\Delta A(x,t)$  : petites perturbations,  
 $y(x,t)$  : déplacement de la section considérée,  
 $S_o(x,t)$  : circonférence du conduit [2].

Les variations de la pression de l'air sont décrites par l'équation différentielle suivante :

$$S_o(x,t) \cdot p(x,t) = m \frac{\partial^2 y(x,t)}{\partial t^2} + b \frac{\partial y(x,t)}{\partial t} + k y(x,t) \quad (2.27)$$

Où  $m$ ,  $b$  et  $k$  indiquent respectivement, par unité de longueur, la masse, la viscosité et l'élasticité de la paroi d'un élément du conduit vocal.

Le débit volumique, généré par la vibration, est défini comme suit :

$$u_w(x,t) = \frac{\partial (y(x,t) S_o(x,t) \cdot l)}{\partial t} \quad (2.28)$$

Avec  $l$  la longueur du conduit vocal.

En combinant les équations (2.26), (2.27) et (2.28) on aura :

$$p(x,t) = \frac{m}{S_o^2(x,t)l} \frac{\partial^2 u_w(x,t)}{\partial t^2} + \frac{b}{S_o^2(x,t)l} \frac{\partial u_w(x,t)}{\partial t} + \frac{k}{S_o^2(x,t)l} u_w(x,t) \quad (2.29)$$

Pour une section élémentaire sans pertes, cette relation devient :

$$p(x,t) = \frac{m}{S_o^2 l} \frac{\partial^2 u_w(x,t)}{\partial t^2} + \frac{b}{S_o^2 l} \frac{\partial u_w(x,t)}{\partial t} + \frac{k}{S_o^2 l} u_w(x,t) \quad (2.30)$$

Le circuit RLC série équivalent (Fig. 2.5) est établi selon l'expression suivante :

$$p(x,t) = L_w \frac{\partial u_w(x,t)}{\partial t} + R_w u_w(x,t) + \frac{1}{C_w} \int u_w(x,t) \quad (2.31)$$

$$\text{Où } L_w = \frac{m}{S_0^2 l} \quad R_w = \frac{b}{S_0^2 l} \quad C_w = \frac{S_0^2 l}{k}$$

Avec :  $L_w$  : Inductance de l'impédance des parois,

$R_w$  : Résistance de l'impédance des parois,

$C_w$  : Capacitance de l'impédance des parois.

Sous la forme d'un élément distribué, l'impédance  $Z_w$  peut être incluse dans chaque section élémentaire du conduit vocal.

L'effet de l'extensibilité des parois le long du conduit vocal, décrit par cette impédance, est assimilé à un amortissement lequel déplacera légèrement les formants à basses fréquences vers d'autres zones de fréquences de telle manière à élargir leur largeur de bande.

Par exemple, le formant à 500 Hz se déplacera à 504 Hz. La largeur de bande passera donc de 50 à 53 Hz soit un effet de l'ordre de 0.8%.

### 2. 2. 2. 3. 2 Le modèle acoustique du rayonnement aux lèvres

L'énergie acoustique s'échappe du conduit vocal via les lèvres ou les narines (les sons nasales). De ce fait, les lèvres seront considérées comme une charge du conduit vocal et seront alors assimilées à une impédance de rayonnement  $Z_r$ .

Cette dernière englobe les pertes de l'énergie acoustique et l'inertie de la masse de l'air lesquelles sont décrites respectivement par ses parties résistive et réactance [13]

$Z_r$  peut s'écrire sous la forme suivante :

$$Z_r = \frac{P_r(\omega)}{U_r(\omega)} \quad (2.32)$$

Avec :  $P_r(\omega)$  : pression rayonnée à une distance particulière des lèvres,  
 $U_r(\omega)$  : débit volumique aux lèvres.

Cette impédance peut être approximée par l'expression suivante :

$$Z_r = R_r + j\omega L_r \quad (2.33)$$

Les résultats empiriques, obtenus par plusieurs chercheurs indiquent qu'approximativement : [10]

$$R_r = \frac{128}{9\pi^2} \quad (2.34.a)$$

$$L_r = \frac{8a}{3c} \quad (2.34.b)$$

Où  $a$  et  $c$  représentent respectivement les rayons de l'ouverture aux lèvres et la célérité du son.

L'effet du rayonnement aux lèvres, décrit par l'impédance  $Z_r$ , tend à « élever » l'amplitude des formants à haute fréquence (zone du troisième formant).

## 2. 3. Modélisation de la source glottique

Le signal glottique fournit une indication importante sur les caractéristiques de la fonction laryngée du signal vocal. Si ce dernier est sourd, alors la forme d'onde prendra l'allure d'une source de bruit aléatoire. En revanche, pour une parole sonore, elle apparaît comme un train d'impulsions quasi-périodique localisé à la glotte. Le déficit enregistré dans la compréhension des caractéristiques complexes du mouvement vibratoire glottique, complique davantage la modélisation de la source d'excitation.

En dépit du fait qu'il s'agit d'une colonne d'air qui traverse l'espace glottique et non pas d'une onde au sens physique du terme (propagation), la fonction qui décrit le débit glottique s'intitule Onde de Débit Glottique (ODG). L'ODG peut être caractérisée par les paramètres suivants :

- l'amplitude de voisement  $A_v$  : ce paramètre est défini comme la dynamique entre les valeurs minimale et maximale du débit glottique;
- la fréquence fondamentale  $f_0$  : appelée aussi fréquence d'oscillations, elle peut être mesurée quantitativement à partir du signal vocal ou du débit volumique glottal;
- le quotient ouvert  $Q_o$  : le rapport entre la durée de la phase ouverte  $T_e$  et la période fondamentale  $T_0$  définit ce paramètre;

$$Q_o = \frac{T_e}{T_0} \quad (2.35)$$

- le quotient de dissymétrie  $Q_D$  : ce paramètre constitue le rapport entre la durée de la phase d'ouverture  $T_p$  et la durée de la phase de fermeture ( $T_e - T_p$ ).

$$Q_D = \frac{T_p}{(T_e - T_p)} \quad (2.36)$$

A titre d'exemple, les valeurs relevées sur la parole humaine naturelle varient dans les limites suivantes : [14]

$$1.7 \leq Q_D \leq 2 \quad (2.37.a)$$

$$0.4 \leq Q_o \leq 0.83 \quad (2.37.b)$$

Il est à noter que ces paramètres, représentés dans la figure 2.6, dépendent grandement de la méthode de calcul choisie pour les déterminer, du couplage entre la source vocale et le conduit vocal et des conditions de phonation [15]. Les différents modèles de la source vocale peuvent être classés en deux catégories : les modèles interactif et non interactif [14].

Fondé sur les aspects physiques du mouvement vibratoire, la modélisation interactive tient en compte les interactions de la source vocale avec le conduit vocal.

Le calcul du débit glottique s'effectue soit à partir de la modélisation de l'air glottal ou de la modélisation physique des cordes vocales à l'aide des systèmes masse-ressort. Son meilleur compromis validité / complexité, et sa simplicité ont poussé de nombreux chercheurs à considérer le modèle à deux masses (Ishizaka et Flanagan, 1972) comme source d'excitation du synthétiseur à formants.

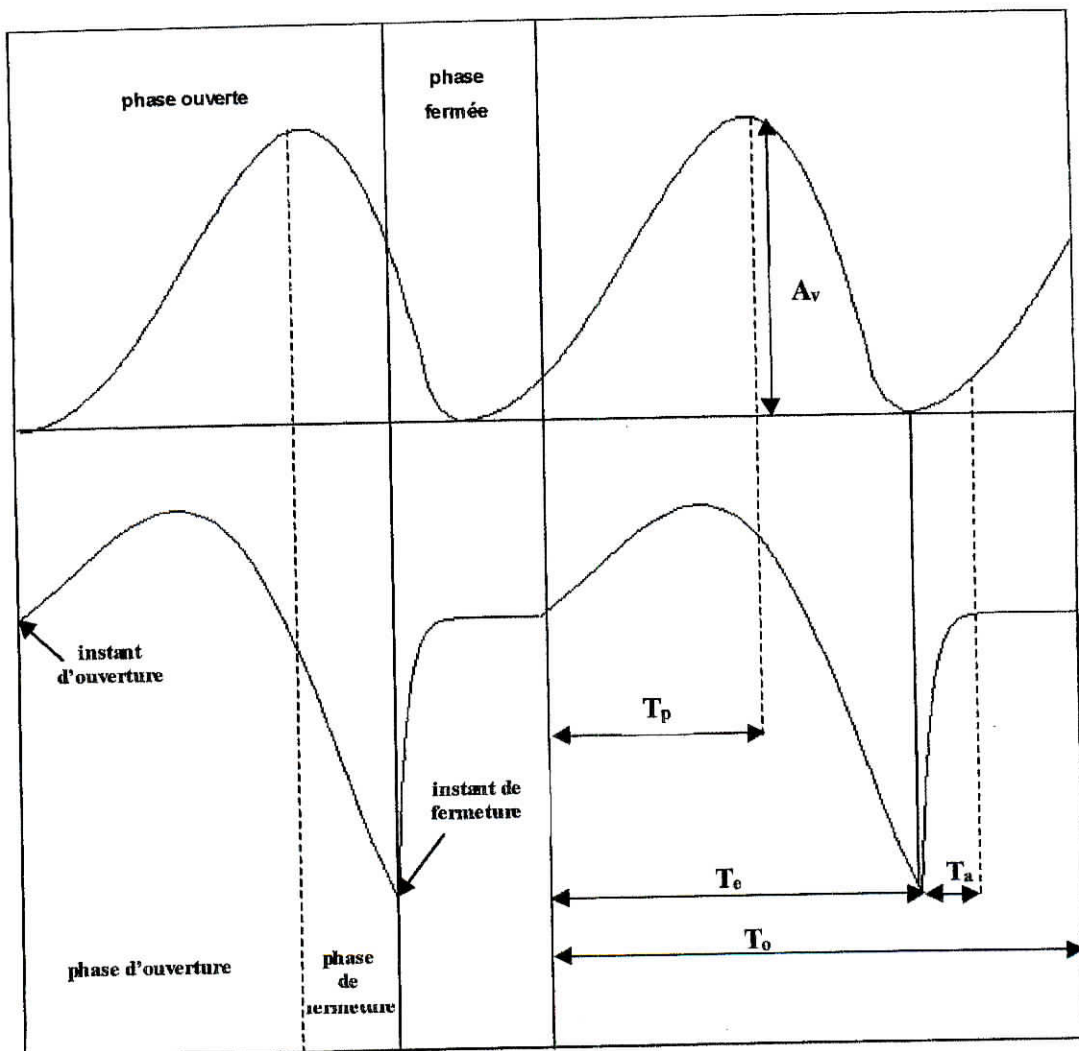


Fig. 2.6 Les paramètres de la source glottique



La modélisation non-interactive se base sur une approximation directe du signal lui-même, et ce, par le biais des fonctions paramétriques empiriques. En estimant directement les paramètres, par filtrage inverse, du signal électroglottographique (EEG), le modèle de Liljencrants-Fant (LF) [14] semble réaliser un large consensus d'utilisation au sein des chercheurs. Cependant, le modèle LF suppose une indépendance entre la source et le filtre (conduit vocal). Néanmoins, le couplage entre les cavités sus et sous-glottiques qui se traduit par une accentuation de l'asymétrie de l'impulsion [14] est pris en compte indirectement.

Les équations du mouvement des masses du modèle à deux masses (Fig. 2.6) sont :

$$m_1 \frac{d^2 x_1}{dt^2} + r_1 \frac{dx_1}{dt} + k_1 x_1 + k_c (x_1 - x_2) + F_1 = 0 \quad (2.38.a)$$

$$m_2 \frac{d^2 x_2}{dt^2} + r_2 \frac{dx_2}{dt} + k_2 x_2 + k_c (x_2 - x_1) + F_2 = 0 \quad (2.38.b)$$

Avec :  $x_i$  : le déplacement latéral des deux masses,

$F_i$  : les forces aérodynamiques exercées sur chaque masse,

$R_i$  : les pertes dues à la viscosité.

Au cours de la fermeture de la glotte, les forces de restauration de raideur sont obtenues comme suit :

$$f_{Si} = k_i x_i (1 + \eta_{ki} x_i), \text{ pour } i=1, 2 \quad (2.39.a)$$

$$f_{hi} = h_i (x_i + x_{oi}) (1 + \eta_{hi} (x_i + x_{oi})), \text{ pour } (x_i + x_{oi}) \leq 0 \quad (2.39.b)$$

Avec :  $i = 1$  : pour la masse inférieure,

$i = 2$  : pour la masse supérieure,

$k$  : paramètre de rigidité linéaire du ressort,

$\eta_k$  : paramètre de rigidité non linéaire du ressort,

$h$  : paramètre de rigidité linéaire du ressort pendant la phase de fermeture glottale,

$\eta_h$  : paramètre de rigidité non linéaire du ressort pendant la phase de la fermeture glottale

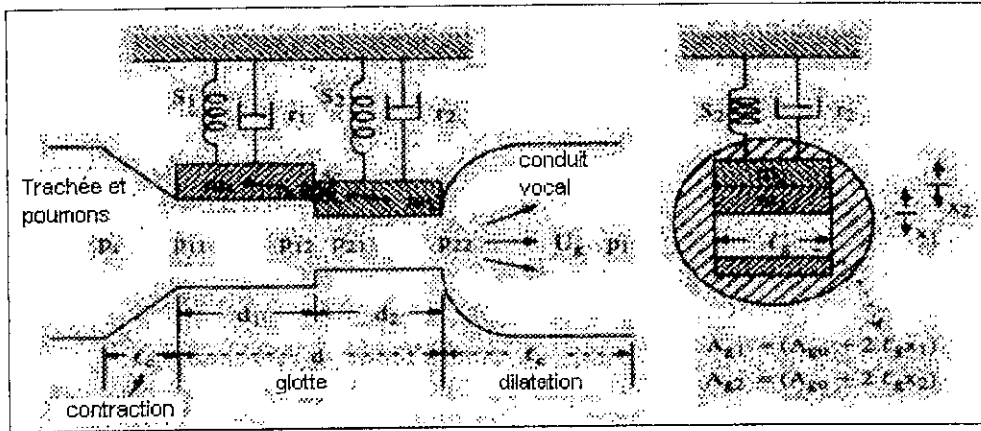


Fig. 2.7 Le modèle à deux masses des cordes vocales

Sur le plan électrique, les éléments de l'impédance acoustique de l'orifice glottale forment le circuit équivalent à la glotte (Fig. 2.8).

$R_{v1}$  et  $R_{v2}$  représentent les pertes de viscosité respectivement des côtés inférieur et supérieur de la corde. La contraction brusque à l'intérieur de la glotte et les perturbations de l'énergie cinétique à la jonction ( $m_1$ ,  $m_2$ ) sont décrits respectivement par  $R_c$  et  $R_{12}$ . Quant à l'effort à l'extérieur de la glotte, il est représenté par la résistance  $R_e$ ; alors que  $L_c$ ,  $L_{g1}$  et  $L_{g2}$  sont les différentes masses de l'air.

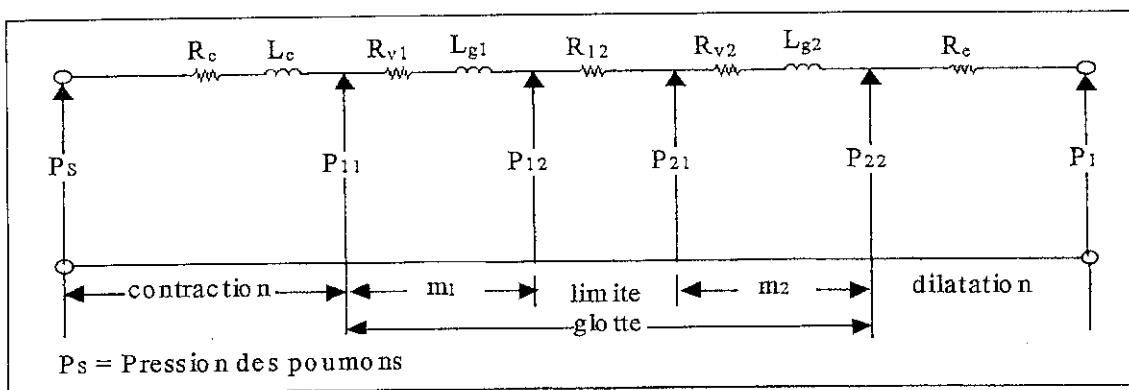


Fig. 2.8 Circuit équivalent à de la glotte

## 2. 4 Le modèle de signal source-filtre

Le modèle de signal source-filtre constitue, malgré ses limites, la base du modèle complet de la production vocale humaine. Elaboré par Fant en 1960, la théorie source filtre, même si elle obéit à une poignée de suppositions, elle permet de modéliser l'acoustique de la production vocale par un système linéaire invariant dans le temps.

Comme l'illustre la figure 2.8, le modèle de Fant englobe trois parties supposées linéairement indépendantes, il s'agit de :

- la source de voisement  $u_g(t)$  ou  $U_g(f)$  : cette partie est marquée par l'émission d'une onde acoustique se propageant le long du conduit vocal;
- la fonction de transfert du conduit vocal  $h(t)$  ou  $H(f)$  : dans cette partie, le conduit vocal est assimilé à un tube acoustique uniforme, de longueur  $L$ , fermé à une extrémité (glotte) et ouvert à l'autre (lèvres). Les ondes stationnaires seront alors identiques à celles dont les multiples impairs de  $\lambda/4$  équivalent à  $L$ . Ainsi, on aura :

$$L = (2n + 1) \frac{\lambda}{4} \quad (2.40)$$

Avec :  $\lambda$  : longueur d'onde de l'onde sonore.

Les fréquences naturelles du conduit vocal, appelées formants, seront alors données par l'expression suivante :

$$F_n = (2n + 1) \frac{c}{4L} \quad (2.41)$$

En remplaçant  $c$  et  $L$  par leurs valeurs respectives, on obtiendra :

$$F_n = (2n + 1) \cdot 500 \text{ Hz} \quad (2.42)$$

Ainsi, on observera un formant tous les 1000 Hz, à la fréquence  $F_n$  (500 Hz, 1500 Hz, 2500 Hz.....)

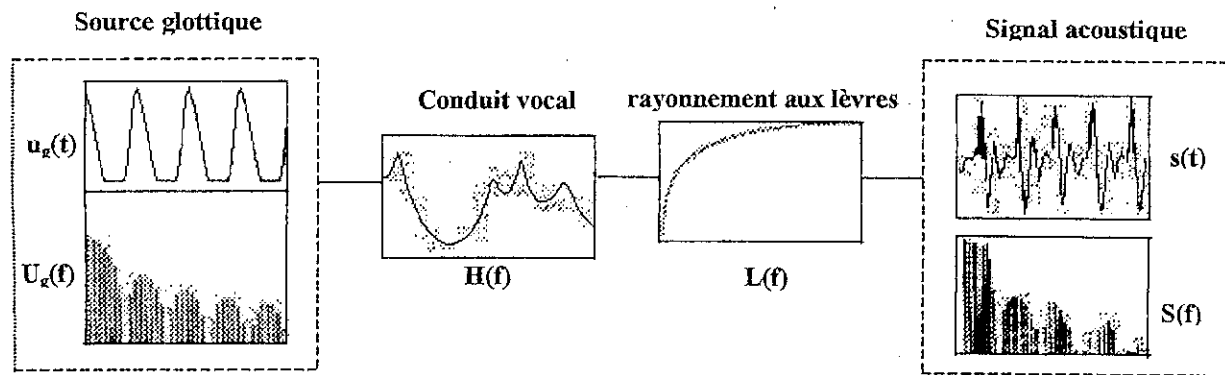


Fig. 2.9 Le modèle de signal source-filtre

Le rayonnement aux lèvres  $l(t)$  ou  $L(F)$  : lorsque les ondes planes arrivent aux lèvres, elles se transforment en ondes sphériques qui se propagent dans le milieu environnant. Le signal acoustique  $S(f)$  ainsi émis correspondra à la convolution temporelle ou multiplication fréquentielle des trois composantes suivantes :

$$S(t) = u_g(t) * h(t) * l(t) \quad (2.43.a)$$

$$S(f) = U_g(f) \cdot H(f) \cdot L(f) \quad (2.43.b)$$

Le modèle source-filtre a eu un impact considérable puisqu'il a été appliqué avec succès dans l'analyse et le codage de la parole par prédiction linéaire [16].

Cependant, en raison de la supposition selon laquelle la propagation des ondes acoustiques dans le conduit vocal est de type plane limite la bande de fréquence utilisée pour ce modèle à 5 KHz. Empêchant ainsi l'élargissement de son utilisation à des voix chantées qui présentent des fréquences très élevées pouvant largement dépasser les 6 KHz [14].

## 2.5 Conclusion

Actuellement, les efforts de modélisation de la production vocale humaine font face, au moins, à deux défis majeurs :

- le manque de valeurs des données physiologiques et physiques justes lesquelles sont décisives dans le développement des modèles quantitatifs;
- la présence des liens intra-locuteurs acoustiques et articulatoires.

Le manque de modèles efficaces, notamment pour la source vocale, est du à la limitation des connaissances dans la compréhension complète du comportement phonatoire des cordes vocales et du conduit vocal. C'est pourquoi, l'introduction d'une troisième dimension dans le modèle du conduit vocal contribuera beaucoup à récupérer l'information perdue et fournira une plate forme adéquate pour les études sur la synthèse articulatoire de la parole et ce, en utilisant les formes géométriques du conduit vocal qui seront déterminées avec un nombre réduit de paramètres articulatoires [17].



# Chapitre 3

## Analyse des données et simulation

### 3. 1 Introduction

Dans ce chapitre, nous exposerons notre démarche expérimentale, en exposant les grandes lignes introduites dans les étapes de l'élaboration de notre outil d'analyse. Nous montrerons l'interactivité de notre logiciel, assurée par l'introduction des fenêtres GUI, dans l'étude des aspects de l'analyse des signaux de la parole.

Nous examinerons aussi les algorithmes des différentes techniques d'analyse introduits dans le logiciel.

### 3. 2 Mise en exécution du logiciel

Après avoir mis en marche le Matlab 5.3, l'utilisateur pourra cliquer sur le bouton «Entrer» pour découvrir les options d'analyse (Fig. 3.1).

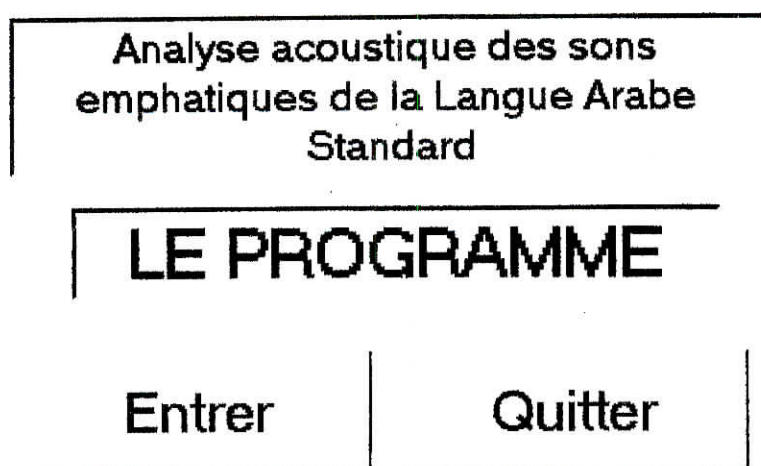


Fig. 3.1 : Entrée au Programme d'analyse

Une fois avoir choisi l'option utilisée (Fig. 3.2), l'utilisateur peut charger un fichier sonore en ASCII qui pourra être visualisé sur une fenêtre appelée signal d'entrée. Par exemple on pourra facilement analyser une syllabe [s<sup>h</sup>i] (Fig. 3.3).

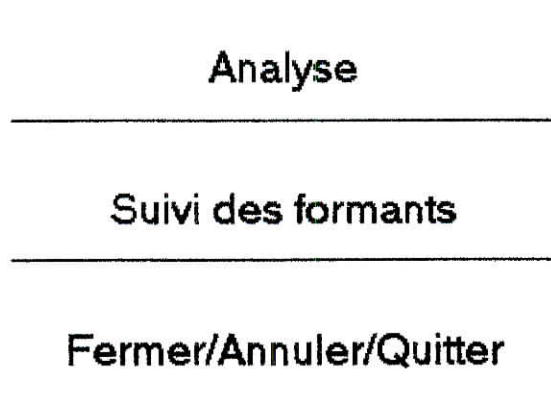
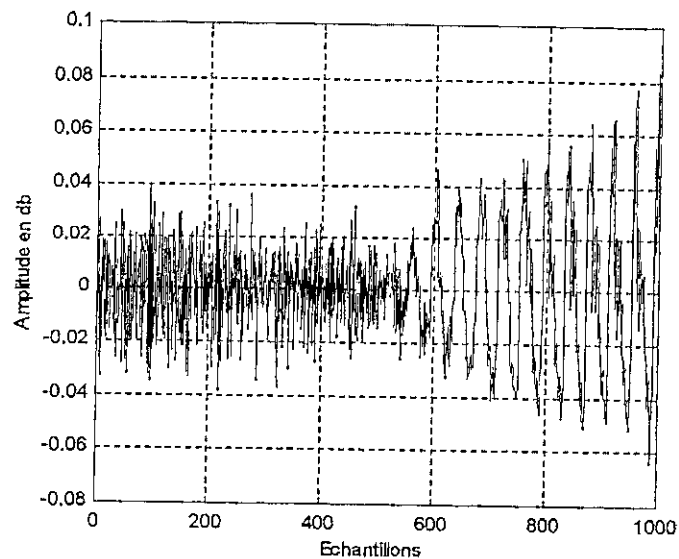


Fig. 3.2 : Options disponibles

Fig. 3.3 : Signal d'entrée, phonème [s<sup>h</sup>]

### 3. 3 Méthodes d'analyse de la parole

Paramétrer un signal vocal en le représentant par un nombre réduit de caractéristiques est l'objectif fondamental de son analyse. Grâce aux deux approches temporelle et fréquentielle, les méthodes d'analyse de la parole, en faisant appel à l'analyse de Fourier à court-terme, fournissent une forme spectrale des données acoustiques d'un signal vocal.

Il convient aussi de préciser que la parole constitue un signal réel, continu, d'énergie finie et non stationnaire, c'est pourquoi la globalisation de son analyse à long terme est impossible. Sa multiplication par une fenêtre devient, donc indispensable.

Ainsi, le choix des méthodes d'analyse, utilisées dans l'estimation des paramètres pertinents du modèle Auto-Régressif (AR) sélectionné pour représenter le signal vocal à analyser, concerne strictement le domaine fréquentiel. Ces méthodes incluent les techniques de l'analyse spectrale par Transformée Rapide de Fourier (Fast Fourier Transform (FFT)), du codage prédictif linéaire (LPC) et de la prédiction linéaire perceptuelle (PLP).

### 3.3.1 Utilisation de la FFT dans l'analyse spectrale à court-terme

En dépit de son incapacité de lever l'intermodulation entre la source de l'excitation et le conduit vocal, à l'origine de la difficulté enregistrée dans la mesure exacte des formants et de  $F_0$  d'un signal de parole, la FFT assure l'obtention, en temps réel, des spectres fréquentiels par l'intermédiaire des processeurs de traitement de signal en simulant la fonction sonographe.

Pour une analyse qualitative, la FFT est nettement suffisante. En revanche, pour une analyse quantitative, d'autres méthodes lui sont préférées [18] (Fig. 3.4).

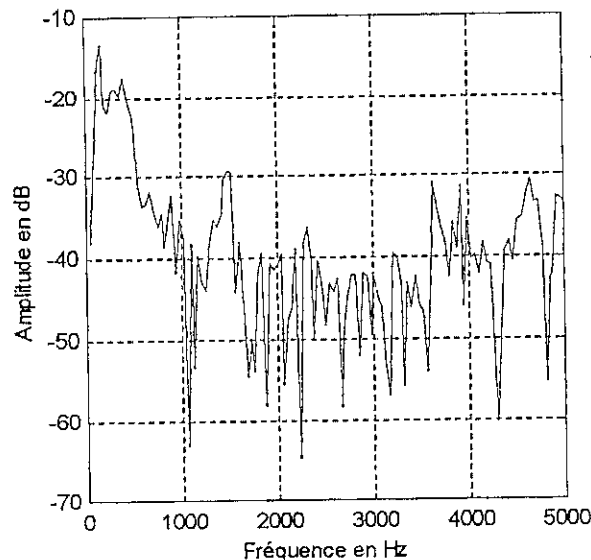


Fig. 3.4 : Analyse spectrale par FFT d'un segment de la syllabe [s'i]

### 3.1.2 Analyse d'un signal vocal par codage prédictif linéaire

Le codage prédictif linéaire (LPC) a été utilisé avec performance dans plusieurs applications en relation avec des systèmes de Traitement Automatique de la Parole [19, 20, 21, 22, 23].

Simple, et pourtant très efficace, cette méthode a suscité un vif intérêt en raison des nombreuses applications en analyse et synthèse de la parole, en particulier pour le codage et la transmission des signaux vocaux [20], ainsi que la synthèse à partir du texte, comme la constitution d'un dictionnaire de diphones (diphonèmes) [5].

Basée sur le modèle de production simple (Fig. 3.5), la technique LPC se fonde sur l'hypothèse qu'un échantillon du signal de la parole  $S(nT)$  ( $T$  étant la période d'échantillonnage), peut être prédit par une somme pondérée linéaire, des  $p$  échantillons qui le précèdent immédiatement.

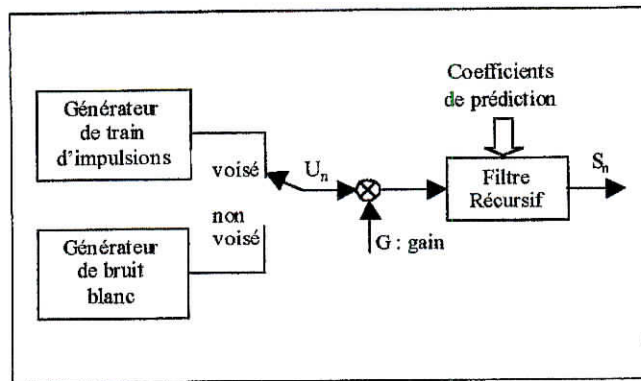


Fig. 3.5 : Modèle simplifié de la production de la parole par prédiction linéaire

Le signal prédit peut se mettre sous la forme suivante [22] :

$$\hat{S}(n) = - \sum_{k=1}^p a_k S(n-k) \tag{3.1}$$

Où  $a_k$  et  $p$  représentent respectivement l'ensemble des coefficients de prédiction et l'ordre du prédicteur.

En considérant que le signal de la source est modélisé par un signal impulsionnel ou un bruit blanc et que le filtre du conduit vocal est approché par un filtre tout-pôle (filtre autorégressif), l'erreur de prédiction entre le signal réel  $S(n)$  et le signal estimé  $\hat{S}(n)$  sera alors donné par :



$$e(n) = S(n) - \hat{S}(n) = G U(n) \quad (3.2)$$

Ainsi, la simplicité de la procédure de l'analyse par la technique LPC se traduit par un calcul des coefficients  $a_k$  par le biais d'une minimisation à court-terme de  $e(n)$ .

Pour cela, le critère retenu est celui d'une minimisation de l'erreur quadratique moyenne par la méthode des moindres carrés.

Les étapes de calcul conduisent à la résolution des  $p$  équations de Yule-Walker (équations normales) suivantes :

$$\sum_{k=1}^p a_k \Phi_n(i, k) = -\Phi_n(i, 0) \quad i=1, 2, \dots, p \quad (3.3)$$

Plusieurs méthodes de résolution de ces équations peuvent être citées. Les méthodes d'autocorrélation et de covariance sont les plus employées.

Le système d'équations exprimé par la méthode d'autocorrélation est le suivant :

$$\sum_{k=1}^p a_k R_{xx}(i-k) = -R_{xx}(i) \quad i=1, 2, \dots, p \quad (3.4)$$

Où  $R_{xx}(i)$  représente la fonction d'autocorrélation de  $S(n)$ .

L'équation (3.4) peut se mettre sous la forme matricielle suivante :

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & R_{xx}(-2) & \dots & R_{xx}(p-1) \\ R_{xx}(1) & R_{xx}(0) & R_{xx}(-1) & \dots & R_{xx}(p-2) \\ R_{xx}(2) & R_{xx}(1) & R_{xx}(0) & \dots & R_{xx}(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{xx}(p-1) & R_{xx}(p-2) & R_{xx}(p-3) & \dots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_{xx}(1) \\ R_{xx}(2) \\ R_{xx}(3) \\ \vdots \\ R_{xx}(p) \end{bmatrix} \quad (3.5)$$

L'inversion matricielle peut résoudre ce système, mais elle n'est pas aussi efficace que l'algorithme de Levinson lequel calcule les valeurs des coefficients  $a_k$  suivant une manière récursive. En nécessitant moins d'opérations de multiplication et d'addition que l'inversion matricielle, cet algorithme devient plus rapide.

Initialiser la récursion

$$a_1^1 = -\frac{R_{xx}(1)}{R_{xx}(0)}$$

$$[E_T^2]^1 = \text{erreur de prédiction du filtre d'ordre 1} = (1 - (a_1^1)^2) R_{xx}(0)$$

Augmenter l'ordre à  $i = 2$ , la récursion de Levinson est comme suit :

$$a_i^i = \frac{R_{xx}(i) - \sum_{j=1}^{i-1} a_j^{i-1} R_{xx}(i-j)}{(E_T^2)^{i-1}} = k_i = \text{coefficient de réflexion}$$

$$a_j^i = a_j^{i-1} - a_i^i a_{i,j}^{i-1}, \text{ pour } j = 1, 2, \dots, i-1$$

$$(E_T^2)^i = (1 - (a_i^i)^2) (E_T^2)^{i-1}$$

Augmenter l'ordre d'une unité

Si l'ordre =  $p+1$  existe

Le coefficient est celui qui apparaît à la jonction de deux tubes sans pertes dans le modèle du conduit vocal. Le  $i^{\text{ème}}$  coefficient de réflexion est donné par :

$$k_i = \frac{A_{i+1} - A_i}{A_{i+1} + A_i} \tag{3.6}$$

Où  $A_i$  est l'aire à section transversale du  $i^{\text{ème}}$  tube uniforme.

La relation (3.6) est une condition nécessaire et suffisante pour garantir la stabilité du filtre. C'est pourquoi  $k_i$  jouit de la propriété suivante :

$$-1 < k_i < 1 \tag{3.7}$$

Le calcul des coefficients d'autocorrelation  $R_{xx}(i)$  exige, au préalable, une pondération du signal  $S(n)$  par la fenêtre de Hamming  $w(n)$  suivante :

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad n = 0, 1, \dots, N-1 \quad (3.8)$$

La durée  $L$  de  $w(n)$  est choisie selon la formule suivante :

$$L = N + 2D \quad (3.9)$$

Où  $D$  représente le débordement remarqué de part et d'autre de la fenêtre d'analyse.

Le nombre des résonances et d'anti-résonances, la nature de la fonction du flux glottal et le rayonnement aux lèvres influent sur un choix juste du nombre des coefficients  $p$  nécessaire à la représentation fréquentielle de chaque segment du signal de la parole. Dans ce contexte, la durée de mémorisation  $pT$  du prédicteur linéaire doit être le double de celle de la propagation des ondes sonores de la glotte vers les lèvres, et ce, pour bien représenter les pôles de la fonction de transfert du conduit vocal [22].

L'ordre du prédicteur  $p$  a été choisi selon la règle de Markel et Gray [14] suivante :

$$p = F_e + N_a \quad \text{Un son voisé ou sonore} \quad (3.10)$$

$$p = F_e \quad \text{Un son non voisé ou sourd} \quad (3.11)$$

Avec :

$F_e$  : fréquence d'échantillonnage en kHz

$N_a$  : nombre des pôles (4 à 5).

Dans nos applications,  $N_a$  représente l'influence de la source de l'excitation et du rayonnement aux lèvres; et  $12 \leq p \leq 17$  ( $F_e = 10$  kHz).

La méthode d'autocorrelation garantie l'obtention d'un filtre prédictif (LPC) stable dont les pôles sont calculés à partir des zéros de  $A(z)$  contenus dans la fonction de transfert  $H(z)$  suivante :

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (3.12)$$

La stabilité de  $H(z)$  est prouvée par la présence de tous ses pôles à l'intérieur du cercle unité dans le plan  $z$  (Fig. 3.6).

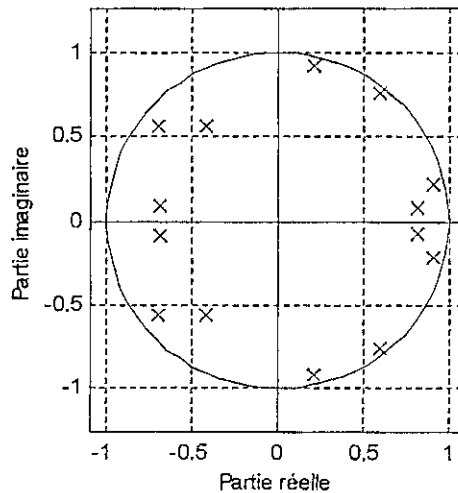


Fig. 3.6 : Pôles à l'intérieur du cercle unité d'un segment du phonème [s°i]

Un exemple de comparaison entre le spectre d'un signal vocal d'une voyelle [a], prononcée par un locuteur masculin, obtenu par FFT avec une estimation spectrale LPC du même signal avec  $p = 14$  sera donné.

Cet exemple montre bien que contrairement aux pics, l'enveloppe de la FFT se superpose avec le spectre LPC dont le lissage (Fig. 3.7 (a)) et la stabilité (Fig. 3.7 (b)) sont parfaits par rapport à d'autres types de spectre.

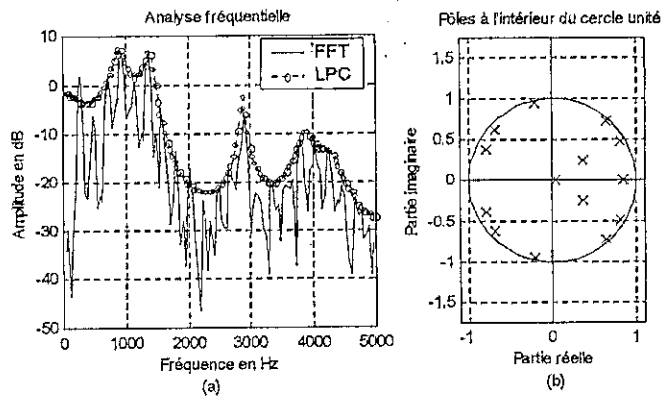


Fig. 3.7 : (a) Analyse spectrale de la voyelle [a]  
(b) Pôles à l'intérieur du cercle unité

L'organigramme de la figure 3.8 donne les détails de calcul des coefficients du prédicteur en utilisant la méthode d'autocorrélation.

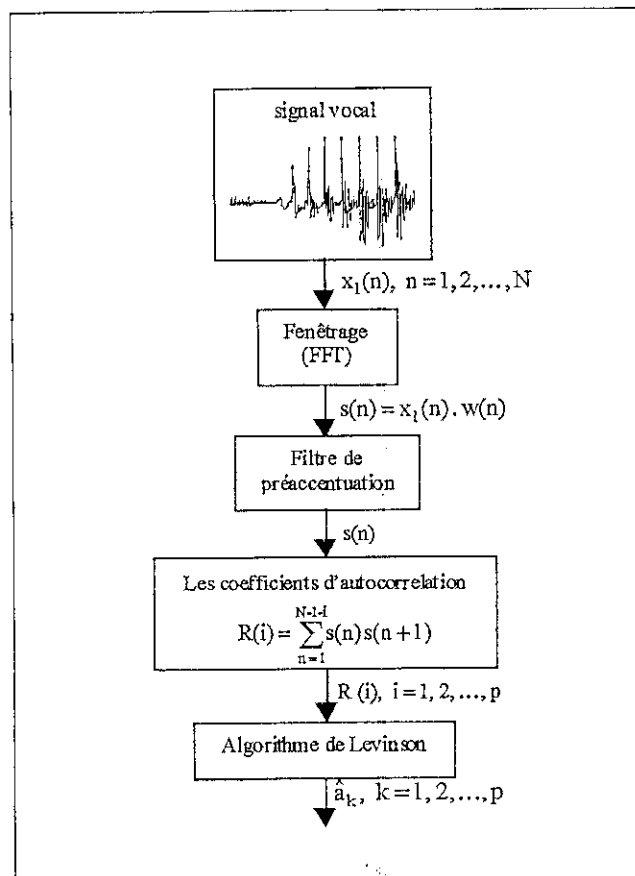


Fig. 3.8 : Organigramme de la méthode de l'autocorrélation

Pour assurer une bonne résolution spectrale, le signal vocal est segmenté en trames d'analyse de 25 ms, il est ensuite préaccentué par un filtre à phase nulle qui élimine les inclinaisons causées par les fréquences basses de la source de l'excitation et du rayonnement aux lèvres pour ne conserver que les pôles du filtre du conduit vocal et d'obtenir finalement un spectre plat.

Destinée au calcul des coefficients du prédicteur linéaire, la méthode de covariance diffère de celle de l'autocorrélation par le fait qu'elle ne suppose aucunement que les données à analyser doivent être fenêtrées puisque les échantillons (N-p) de l'erreur de prédiction sont déjà disponibles.

Par conséquent, les formes des équations normales des deux méthodes sont identiques. Simplement, pour éviter une confusion, entre les deux méthodes, les notations des coefficients d'autocorrélation seront différentes. Ainsi, la forme matricielle sera comme suit :

$$\begin{bmatrix} C_{xx}(1,1) & C_{xx}(1,2) & C_{xx}(1,3) & \dots & C_{xx}(1,p) \\ C_{xx}(2,1) & C_{xx}(2,2) & C_{xx}(2,3) & \dots & C_{xx}(2,p) \\ C_{xx}(3,1) & C_{xx}(3,2) & C_{xx}(3,3) & \dots & C_{xx}(3,p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{xx}(p,1) & C_{xx}(p,2) & C_{xx}(p,3) & \dots & C_{xx}(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} C_{xx}(1,0) \\ C_{xx}(2,0) \\ C_{xx}(3,0) \\ \vdots \\ C_{xx}(p,0) \end{bmatrix} \quad (3.12)$$

En plus de l'absence de la garantie d'obtenir usuellement une stabilité du filtre AR, et même si la matrice  $C_{xx}(i,k)$  demeure symétrique, la résolution du système d'équations normales qui fait généralement appel à l'algorithme de décomposition de Choleski, est rendu difficile puisque les éléments d'une diagonale ne sont plus nécessairement égaux, comme ce fut le cas pour la méthode d'autocorrélation.

Le codage prédictif linéaire par autocorrélation s'adapte mieux à des signaux stationnaires pour lesquels l'analyse s'effectue sur un grand nombre d'échantillons.



En revanche, les signaux non stationnaires sont mieux traités par la méthode de covariance, ce qui explique la précision parfaite de l'analyse des consonnes par covariance et les voyelles par autocorrélation [14].

### 3.3.3 Analyse de la parole par prédiction linéaire perceptuelle

Malgré l'obtention d'une enveloppe spectrale  $P(w)$ , parfaitement lissée, du signal vocal, l'approximation de  $P(w)$  par un modèle tout-pôles autoregressif  $A(w)$  sur la totalité des fréquences de la bande d'analyse constitue tout de même un inconvénient majeur de l'analyse LPC.

En effet, au-delà des 800 Hz, région du spectre où l'oreille humaine est plus sensible, la résolution spectrale diminue avec la fréquence.

Les caractéristiques spectrales de  $P(w)$ , en relation directe avec l'aspect auditif humain, ne sont donc pas toujours préservées par cette méthode [21].

Dans le but de lever la contradiction de l'analyse LPC avec les propriétés acoustiques de l'audition humaine, H. Hermansky a proposé la technique de l'analyse par Prédiction Linéaire Perceptuelle (PLP) et l'a qualifié de la voie la plus efficace utilisée dans l'estimation d'un spectre d'audition en raison de ses liens avec les concepts d'audition psychophysiques (Fig. 3.9).

Une combinaison entre la transformée de Fourier discrète et les techniques LPC donne la Prédiction Linéaire Perceptuelle (PLP).

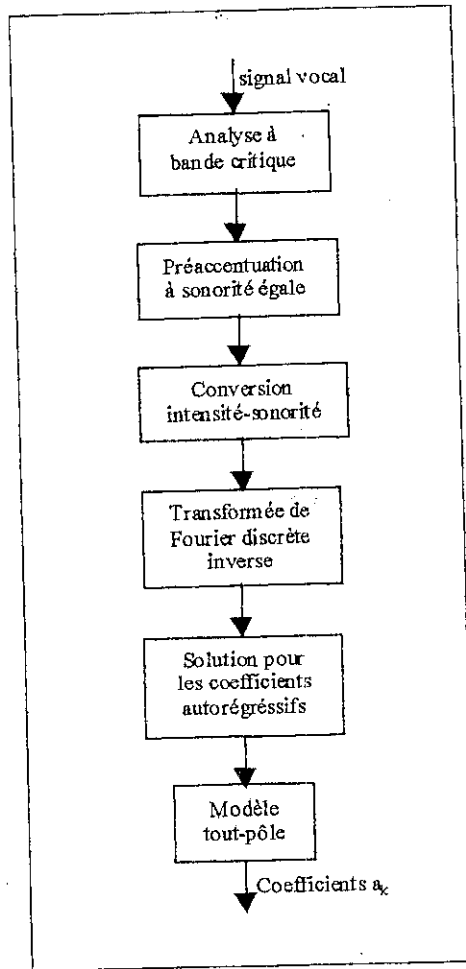


Fig. 3.9 : Analyse perceptuelle par prédiction linéaire

Les échantillons du spectre d'énergie à bande critique  $\theta(\Omega_i)$  du signal de la parole, préalablement filtré par une rangée de filtres à bande critique, sont obtenus à l'issue d'une convolution discrète entre la courbe de recouvrement à bande critique simulée  $\Psi(\Omega)$  et le spectre d'énergie à court-terme  $P(w)$ .

La résolution spectrale de  $\theta(\Omega)$  est alors réduite par rapport à celle de l'originale  $P(w)$ .

$$\theta(\Omega_i(w)) = \sum_{\Omega=-1.31}^{2.5} p(\Omega(w) - \Omega_i(w)) \Psi(\Omega(w)) \tag{3.13}$$

Où  $w$  et  $\Omega(w)$  représentent respectivement la fréquence angulaire en rd/s et la transformation Bark-Hertz donnée par :

$$\Omega(w) = 6 \ln \left\{ \frac{w}{1200\pi} + \left[ \left( \frac{w}{1200\pi} \right)^2 + 1 \right]^{0.5} \right\} \quad (3.14)$$

$\Psi(\Omega(w))$  peut s'écrire comme suit :

$$\Psi(\Omega(w)) = \begin{cases} 0 & \Omega < -1.3 \\ 10^{2.5(\Omega+0.5)} & -1.3 \leq \Omega \leq -0.5 \\ 1 & -0.5 < \Omega < 0.5 \\ 10^{-1.0(\Omega-0.5)} & 0.5 \leq \Omega \leq 2.5 \\ 0 & \Omega > 2.5 \end{cases} \quad (3.15)$$

Le choix de la valeur exacte de l'intervalle d'échantillonnage repose sur l'intégralité des échantillons spectraux qui couvrent entièrement la bande d'analyse.

Typiquement, dix huit échantillons spectraux de  $\theta[\Omega(w)]$  suffisent pour couvrir la bande d'analyse de 0-16.9-Bark (0-5 kHz) suivant des pas de 0.994 Bark.

Cette procédure est suivie par une préaccentuation à sonorité égale effectuée par un filtre dont la fonction de transfert  $E(w)$  peut se mettre sous la forme suivante :

$$E(w) = \frac{(w^2 + 56.8 \times 10^6) w^4}{(w^2 + 6.3 \times 10^6)^2 \cdot (w^2 + 0.38 \times 10^9)} \quad (3.16)$$

En vue d'une réduction des variations spectrales de l'amplitude par l'intermédiaire d'une fonction à racine cubique, un ajustement sonore est exercé sur le spectre à bande critique. La modélisation tout-pôle s'effectue alors avec un modèle d'ordre relativement bas.

L'approximation  $\theta(\Omega)$  par le spectre d'un modèle tout-pôle par le biais de la méthode d'autocorrélation constitue l'étape finale de l'analyse PLP. La fonction d'autocorrélation est obtenue par la TFDI. Les premiers  $(M + 1)$  coefficients d'autocorrélation sont utilisés, par la résolution des équations de Yule-Walker, dans le calcul des coefficients autoregressifs du modèle tout-pôle de  $M^{\text{ième}}$  ordre.

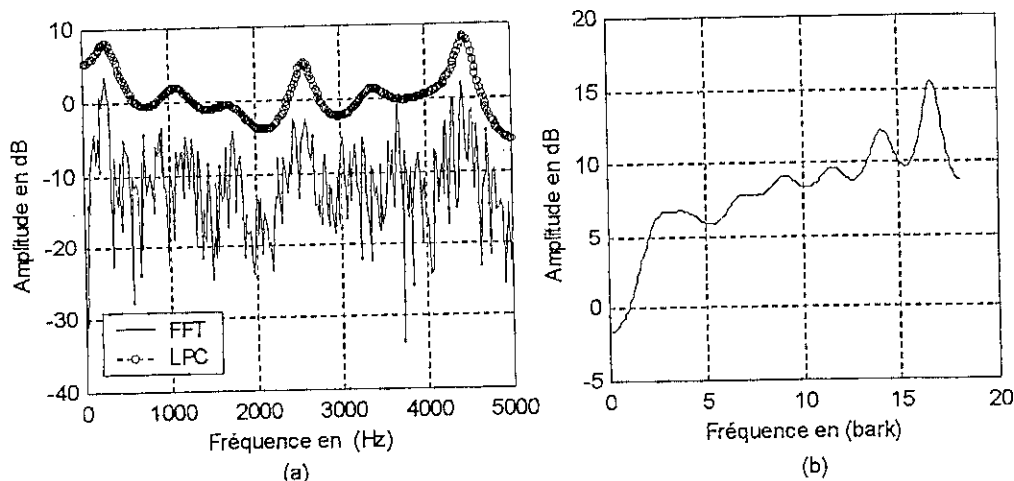


Fig. 3.10 : Analyse par autocorrélation et PLP d'un segment du phonème

### 3.3.4 Suivi des Formants

Les procédures de l'extraction des formants et leurs suivis (ou formant tracking) sont accomplies grâce à un algorithme développé par Mc Candless [19].

A travers la réalisation de l'algorithme automatique de l'extraction des formants et le suivi de leurs contours, Mc Candless tente d'une part de trouver une adéquation parfaite entre les pics de la réponse fréquentielle obtenue du filtre produit à partir des coefficients LPC ( $p = 14$ ) et d'estimer d'autre part, les amplitudes des fréquences des quatre premiers formants durant la totalité des segments voisés du signal vocal [19].

Éliminer les trames à énergie spectrale totale inférieure au seuil du silence, calculer la période du pitch et les instants de fermeture glottale sont les étapes préliminaires à l'application de l'algorithme.

Chaque zone voisée résultante sera alors classée, «voyelle» ou «consonne», suivant le taux de l'énergie fréquentielle totale contenue dans le spectre et celle de la région comprise entre 640 et 2880 Hz.

Il est clair que l'extraction des formants par le calcul des pôles de la fonction de transfert  $H(z)$  du filtre LPC est une procédure idéale (éq. 3.14).

Cependant, le choix des quatre premiers pics formantiques qui représente une technique itérative élaborée par cet algorithme, est beaucoup plus avantageux puisque le taux des opérations de calcul est nettement inférieur.

Initialement, les estimations des quatre premières fréquences formantiques sont une série de valeurs typiques pour la voix d'homme ( $F_1=320$  Hz ;  $F_2=1440$  Hz ;  $F_3=2760$  Hz et  $F_4=3200$  Hz) ou femme ( $F_1=480$  Hz ;  $F_2=1760$  Hz ;  $F_3=3200$  Hz et  $F_4=3520$  Hz).

Pour chaque trame du signal de parole, chacun des pics de la réponse fréquentielle du filtre LPC est comparé avec l'estimation fréquentielle formantique la plus proche.

Les résultats définitifs seront lissés afin de mettre en évidence les suivis formantiques. Ce type d'analyse contribue grandement à l'identification des voyelles dans un contexte CV.

Les trois premières fréquences des formants et leurs amplitudes des formants estimées pour la portion voisée du mot [haya:ta] prononcée par un locuteur masculin sont établis (Fig. 3.11). On pourra facilement reconnaître les structures formantiques de la voyelle [a] dans les trois syllabes de ce mot.

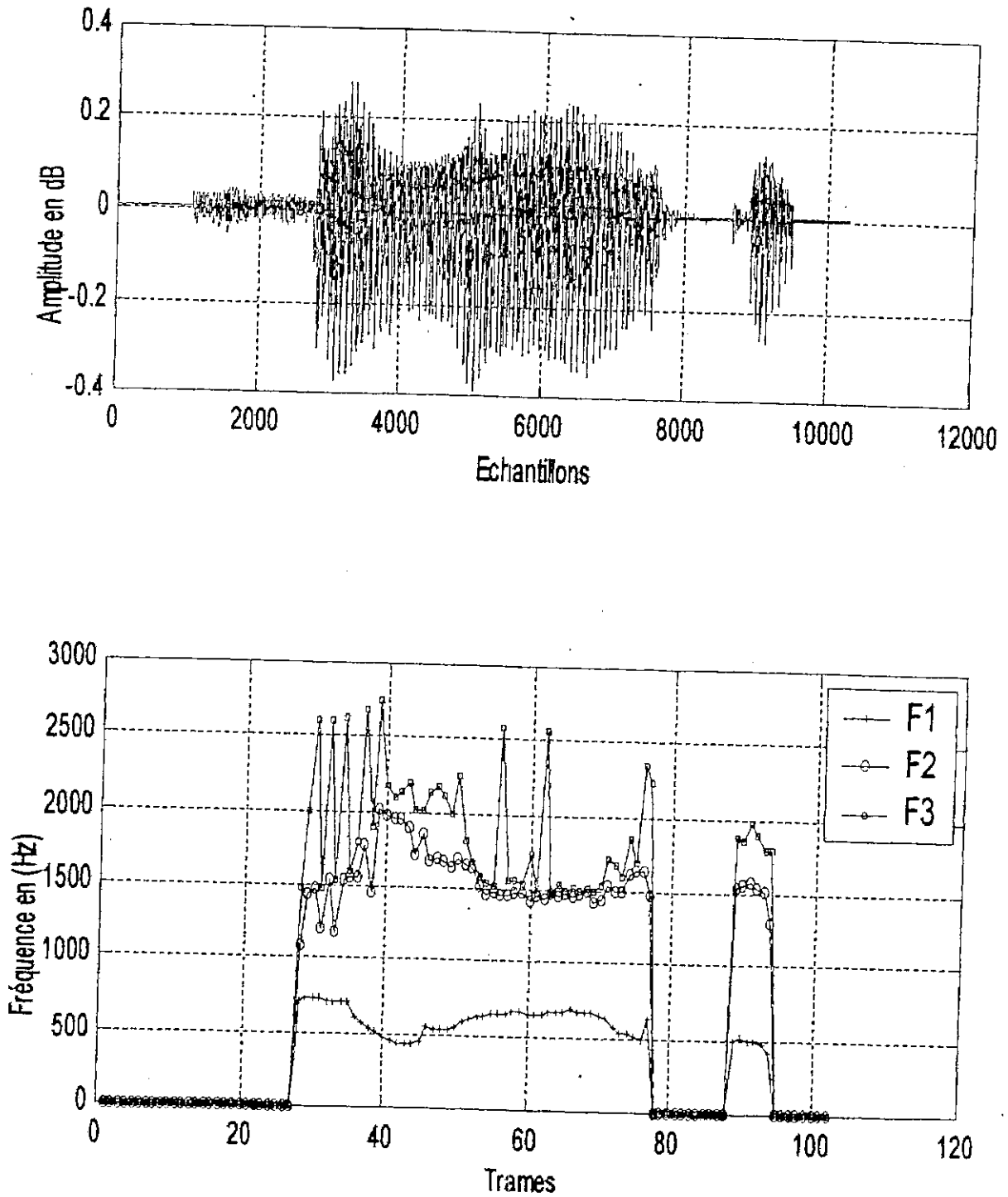


Fig. 3.11 : Signal vocal et analyse formantique du mot [haya:ta]



### 3. 4 Conclusion

La prédiction linéaire classique est une méthode robuste et efficace, qui peut être donner de bons résultats dans le cas où les effets de la source seraient éliminés. C'est en particulier le cas des voyelles de l'AS ayant un premier formant haut en fréquence, dont le [a] est un exemple typique. De plus, étant donné les sons emphatiques de l'Arabe Standard ne contiennent pas de sons nasaux, alors cette technique d'analyse s'adapte bien à notre travail de simulation.

Pour ce qui est de l'analyse par PLP, nous avons jugé qu'elle est destinée exclusivement aux travaux de la perception auditive des sons. En effet, la résolution spectrale est fortement réduite par la convolution entre le spectre dont l'échelle est en Bark avec celui du filtre à bande critique.

Concernant, l'analyse spectrographique, elle demeure la plus efficace. elle permet relever les paramètres acoustiques d'un mot sur tout l'intervalle temporel. Dans le même contexte, l'observation des trajectoires formantiques du même mot peut nous aider à étudier les formants des voyelles ou des sons voisés d'une trame vers une autre.

# **Chapitre 4**

## **Analyse des emphatiques de l'Arabe Standard**

### **4.1 Introduction**

L'emphase est une particularité de l'Arabe Standard (AS). C'est une qualité consonantique qui se traduit au niveau articulaire par une variation dans la cavité pharyngale pour produire deux séries de consonnes : emphatiques (pharynx rétréci) et non-emphatiques (pharynx élargi).

Nous allons nous attacher à décrire dans ce chapitre les résultats d'une analyse acoustique des consonnes emphatiques de l'AS. Nous examinerons, principalement, leur effet sur les voyelles qui les distinguent de leurs correspondantes non-emphatiques.

### **4.2 Objectif**

L'objectif de cette étude expérimentale est de présenter l'intérêt de l'analyse acoustique pour l'étude phonétique de l'emphase dans l'AS. Elle se propose ainsi de faire l'état des lieux de ce type d'approche et de montrer, à travers quelques théories établies, que les résultats de cette analyse peuvent avoir une application aussi bien aux stratégies de synthèse qu'à la théorie de la phonétique générale.

Il est évident que le nombre des consonnes emphatiques peut varier d'un langage à un autre. Pour notre étude, et en accord avec les travaux de la plupart des chercheurs, nous avons inclus quatre sons emphatiques [t<sup>ʕ</sup>], [d<sup>ʕ</sup>], [z<sup>ʕ</sup>] et [s<sup>ʕ</sup>].

La vue générale de la littérature consultée indique que la majorité des chercheurs, même si certains n'ont pas été très explicites, affirment que l'emphatisation de l'AS est associée à plusieurs corrélats articulatoires : une rétraction de la racine de la langue (Fig. 4.1), un volume du pharynx rétréci, une élévation de l'os hyoïde et de l'épiglotte.

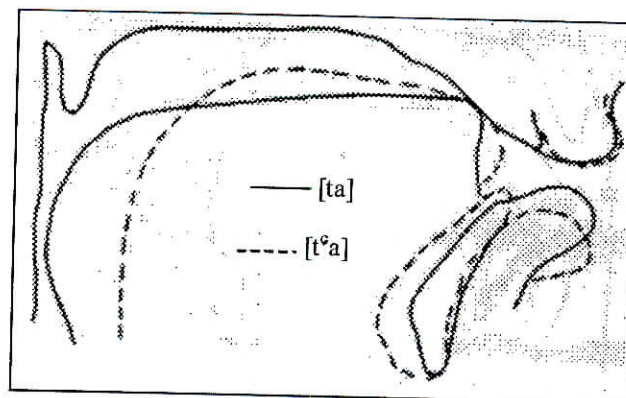


Fig. 4.1 : Principe articulatoire de l'emphase [7]

L'événement le plus important est la constriction de la cavité pharyngale, ce qui autorise à caractériser l'emphase dans l'AS par le trait de pharyngalisation et d'affirmer que la totalité des sons emphatiques et pharyngaux sont produits qualitativement avec la même constriction pharyngale [8, 15, 24, 25, 26, 27, 28].

Ainsi, le son emphatique est distingué de son opposé non-emphatique par le fait que sa production entraîne une articulation secondaire dans un autre lieu avec un degré de constriction plus large que celui d'une «occlusive».

L'utilisation conjointe de la FFT, la LPC et les contours formantiques a élargi notre perspective de l'analyse.

### 4.3 Synthèse des travaux sur l'emphase

L'emphase a été abondamment étudiée pour les différents parlers arabes, tant au niveau articulatoire qu'acoustique. Pour l'AS, nous ne connaissons que deux études acoustiques [7, 8, 23], et plusieurs autres articulatoires.

#### 4.3.1 Etudes articulatoires

Jusqu'au début du XIXème siècle, l'étude de l'emphase est restée marquée par les descriptions traditionnelles des grammairiens arabes. Il a fallu attendre le XXème siècle et les années 40 pour voir les premières analyses instrumentales.

A. Giannini, & M. Pettorino (1982) [23] citent le célèbre grammairien arabe Sibawayhi. Ce dernier explique l'emphatisation d'un son par la présence d'une deuxième articulation réalisée sous forme d'une vélarisation à cause de la rétraction de la langue vers le vélum.

A. Laufer & T. Baer (1988) [27] citent C. Meinhof (1921). Il indique que la caractéristique des articulations emphatiques ne peut pas être déterminée ni par le lieu de l'articulation, puisque la totalité de ces consonnes disposent des lieux d'articulation différents, ni par le mode d'articulation, puisque à l'intérieur de ce groupe il existe aussi bien d'occlusives que de fricatives. Il note ainsi que le processus de la production des consonnes emphatiques de l'AS est soumis à trois phénomènes :

- une élévation de la racine de la langue vers le vélum;
- une contraction de la musculature de l'os hyoïd;
- un abaissement de l'épiglotte.

Ghazali (1987) [27] a basé son travail sur des données cinéroradiographiques d'un corpus de sons du langage tunisien. Il localise la production du segment emphatique dans la partie arrière du conduit vocal : le pharynx. En revanche, Obrecht (1968) [8] emploie le terme vélarisation pour décrire les emphatiques de l'arabe libanais.

A partir de ces travaux, il apparaît que l'emphase se traduit par une combinaison de mouvements articulatoires qui font intervenir la racine de la langue, le pharynx, l'os hyoïde et le larynx. La coexistence des notions de vélarisation et de pharyngalisation pour définir l'emphase sur un plan articulatoire peut s'expliquer par la diversité des locuteurs et des parlers étudiés.

#### 4. 3. 2 Etudes acoustiques

La disposition de la chaîne d'articulateurs (racine de la langue rétractée, pharynx rétréci et larynx élevé) a des répercussions acoustiques.

Depuis l'étude d'Obrecht (1968) [8], l'emphase est associée acoustiquement à l'abaissement du deuxième formant de la voyelle qui suit le segment déclencheur de la pharyngalisation. Il affirme que le locus (valeur à la fermeture) de  $F_2$  d'un son emphatique est localisé dans la région fréquentielle 1000-1400 Hz, alors que celui d'un autre non-emphatique (non-vélarisé selon sa terminologie) se situe à 1800 Hz.

Al-Ani (1970) [7] confirme cet indice acoustique à travers l'analyse spectrographique d'un corpus de sons des parlers irakien et jordanien (Fig. 4.2). A. Giannini, & M. Pettorino (1982) [23], à travers l'analyse acoustique d'un corpus de sons baghdadi, caractérisent par l'abaissement de  $F_2$ , l'élévation non systématique de  $F_1$  et un mouvement inchangé de  $F_3$ . Par contre, Obrecht avait noté un mouvement d'instabilité de  $F_3$ , qu'il qualifie ainsi d'indice irrégulier de l'emphase dans le parler arabe libanais.



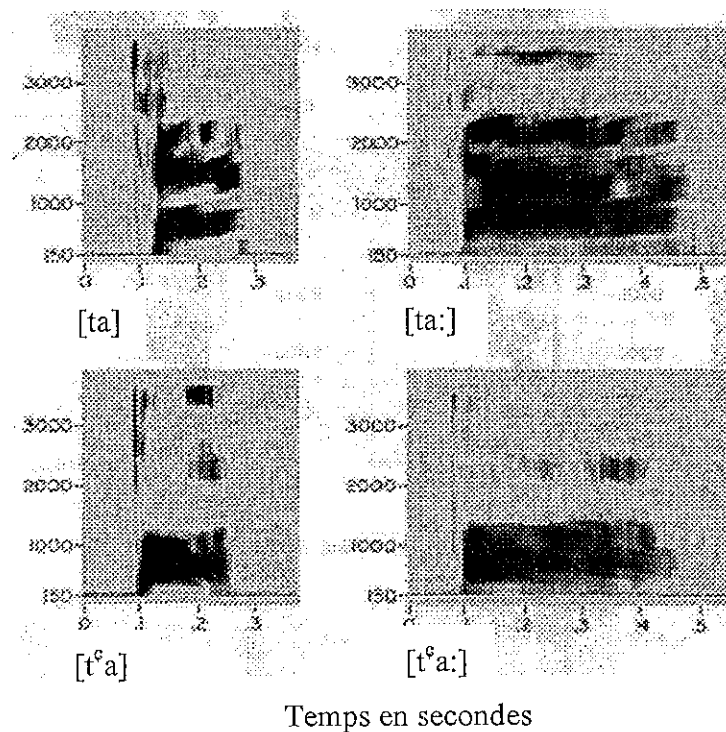


Fig. 4.2 : Spectrogrammes des consonnes [tʰ] et [t] [7]

## 4. 4 Données acoustiques

En chambre sourde, on obtient aisément un enregistrement d'une très grande qualité. Cela encourage à constituer son propre corpus à mesure des besoins, plutôt que de chercher dans les fonds existants.

### 4. 4. 1 Procédure des enregistrements sonores

L'enregistrement sonore des signaux acoustiques a été effectué dans un studio fortement insonorisé de la radio algérienne (chaîne 1).



Le signal sonore est enregistré sur une bande magnétique à l'aide d'un microphone (Brüel & Kjaer) placé à 50 cm du locuteur et un amplificateur (Brüel & Kjaer) (Fig. 4.3).

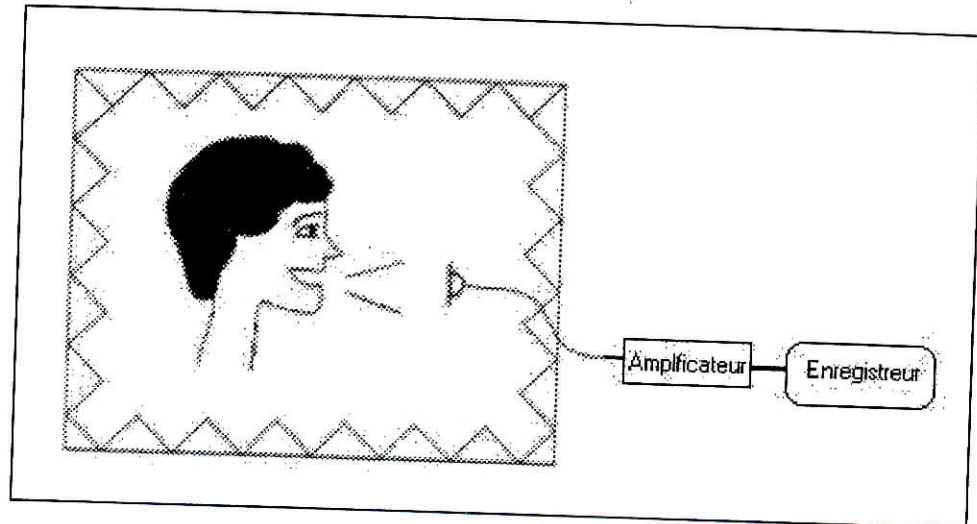


Fig. 4.3 : Schéma du dispositif d'enregistrement du corpus

Les signaux acoustiques enregistrés ont été directement numérisés sur un DAT (PORTADAT PDR1000) à une fréquence d'échantillonnage de 10 kHz sur 16 bits au laboratoire d'Acoustique Musicale (LAM) à Paris (France). Les données numériques obtenues ont été, par la suite, gravées sur un compact disque.

#### 4. 4. 2 Corpus

La nature particulière de notre travail nous a amenés à concevoir un corpus permettant l'établissement d'une opposition à deux termes : consonnes emphatiques / consonnes non-emphatiques. Par conséquent, les paires [t<sup>é</sup>, t], [d<sup>é</sup>, d], [z<sup>é</sup>, ð] et [s<sup>é</sup>, s] ont été combinées avec chacune des six voyelles [a], [u], [i], [a:], [u:] et [i:] dans une position syllabique initiale (Tab. 4.1).

C <sub>E</sub> V	C <sub>E</sub> V:	C V	C V:
[t°a]	[t°a:]	[ta]	[ta:]
[t°i]	[t°i:]	[ti]	[ti:]
[t°u]	[t°u:]	[tu]	[tu:]
[d°a]	[d°a:]	[da]	[da:]
[d°i]	[d°i:]	[di]	[di:]
[d°u]	[d°u:]	[du]	[du:]
[s°a]	[s°a:]	[sa]	[sa:]
[s°i]	[s°i:]	[si]	[si:]
[s°u]	[s°u:]	[su]	[su:]
[z°a]	[z°a:]	[ða]	[ða:]
[z°i]	[z°i:]	[ði]	[ði:]
[z°u]	[z°u:]	[ðu]	[ðu:]

Tableau 4.1 : Corpus enregistré

Le choix judicieux de locuteurs arabes revêt une importance considérable pour notre recherche, dans la mesure où il fallait absolument respecter les règles très strictes de la prononciation classique. Les six locuteurs sélectionnés (03 hommes et 03 femmes) travaillent comme journalistes Arabophones à la chaîne 1 de la radio algérienne. Ils présentent régulièrement des reportages sonores en AS sans aucun substrat dialectal. Les syllabes [CV] sont prononcées en isolation aux intervalles régulières d'environ deux secondes. Une série des spectrogrammes de la consonne emphatique [t°] précédant les voyelles brèves de l'AS a été réalisée (Fig. 4.4).

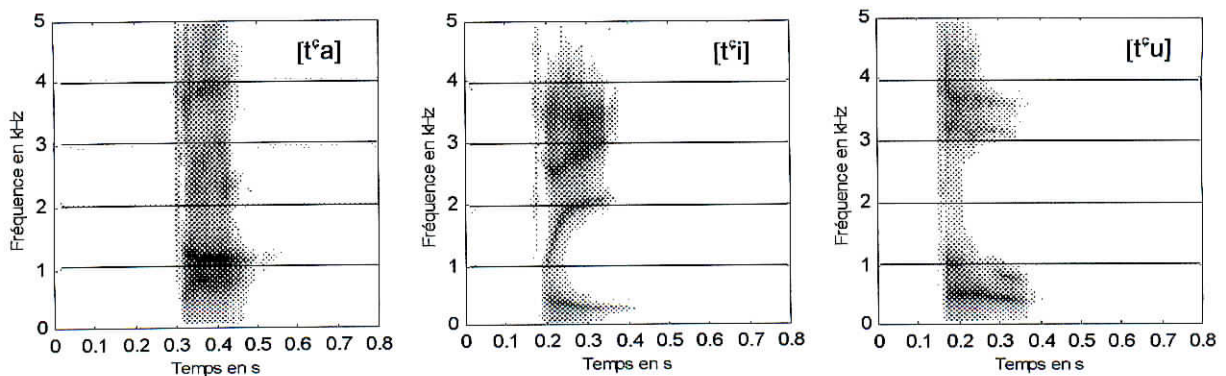


Fig. 4.4 : spectrogrammes de [t°]

## 4. 5 Description acoustique

### 4. 5. 1 Description spectrographique

Pour l'occlusive dentale sourde [t], placé dans une position initiale, la durée de l'ouverture avant la rupture varie de 160 à 200 ms. La concentration intense de l'explosion avec les voyelles [i] et [i:] est localisée dans la région des formants  $F_2$ - $F_3$  (2030-2800 Hz).

Cette barre d'explosion est caractérisée par un bruit apparaissant très haut vers la région de  $F_4$  (04 kHz) (Tab. 4.2).

Pour ce qui est des voyelles [u] et [u:], la barre de l'explosion est située vers la région fréquentielle 1515-2250 Hz, alors qu'avec les voyelles [a] et [a:], elle apparaît vers 1650-2700 Hz.

Pour l'occlusive dentale sourde emphatique [t<sup>é</sup>], placée dans une position initiale, la durée de l'ouverture avant la rupture varie de 96 à 170 ms.

La concentration de l'explosion est plus basse en fréquence que celle de [t]. Elle est située dans la région des  $F_2$  et  $F_3$ , entre les zones fréquentielles 1916-2837 Hz, 1105-2294 Hz et 1467-2535 Hz respectivement pour les voyelles [i] et [i:], [u] et [u:] et [a] et [a:] (Tab. 4.2) (Fig. 4.5).

Concernant l'occlusive dentale sourde emphatique [d], placée dans une position initiale, la durée varie de 64 à 114 ms. Cette consonne apparaît sur les spectrogrammes avec la même allure que [t] à l'exception de la sonorité qui est indiquée par une barre de voisement.

Notons enfin que la durée de l'occlusive dentale sonore emphatique [d<sup>é</sup>], placée dans une position initiale, est comprise entre 70 et 120 ms (Fig. 4.6).

Phonème	Barre d'explosion
[d <sup>°</sup> a]	- vers F <sub>2</sub> (1049 Hz), F <sub>3</sub> (2641 Hz) ou entre F <sub>2</sub> -F <sub>3</sub> - faible vers 5 kHz - moins intense
[d <sup>°</sup> i]	- vers F <sub>3</sub> (1681 Hz) - vers F <sub>2</sub> ou un peu plus bas - plus faible vers 4 kHz - moins intense
[d <sup>°</sup> u]	- entre F <sub>3</sub> -F <sub>4</sub> et plus haut (2.5-4.5 kHz) - rarement entre F <sub>1</sub> -F <sub>2</sub> - moins intense
[da]	- vers F <sub>2</sub> (1476 Hz), F <sub>3</sub> (2764 Hz) ou entre F <sub>2</sub> -F <sub>3</sub> - faible vers 5 kHz
[di]	- vers F <sub>3</sub> (1850 Hz) - plus faible vers 4 kHz
[du]	- entre F <sub>3</sub> -F <sub>4</sub> et plus haut (2.5-4 kHz) - rarement entre F <sub>1</sub> -F <sub>2</sub>
[t <sup>°</sup> a]	- continue entre 1.5 et 5 kHz - maxima vers 5 kHz - entre 1467-2535 Hz - moins intense
[t <sup>°</sup> i]	- continue entre 2 et 4 kHz maxima vers 5 kHz - entre 1916-2837 Hz - moins intense
[t <sup>°</sup> u]	- continue entre 2-3 kHz - entre 1105-2294 Hz - maxima vers 4 kHz - moins intense
[ta]	- continue entre 1.5 et 5 kHz - maxima vers 2 et 5 kHz
[ti]	- continue entre 4.5 et 5 kHz maxima vers 5 kHz - intense
[tu]	- continue entre 1.5-3 kHz - maxima vers 1.7 et 3 kHz

Tableau 4.2 : Caractéristiques de la barre d'explosion

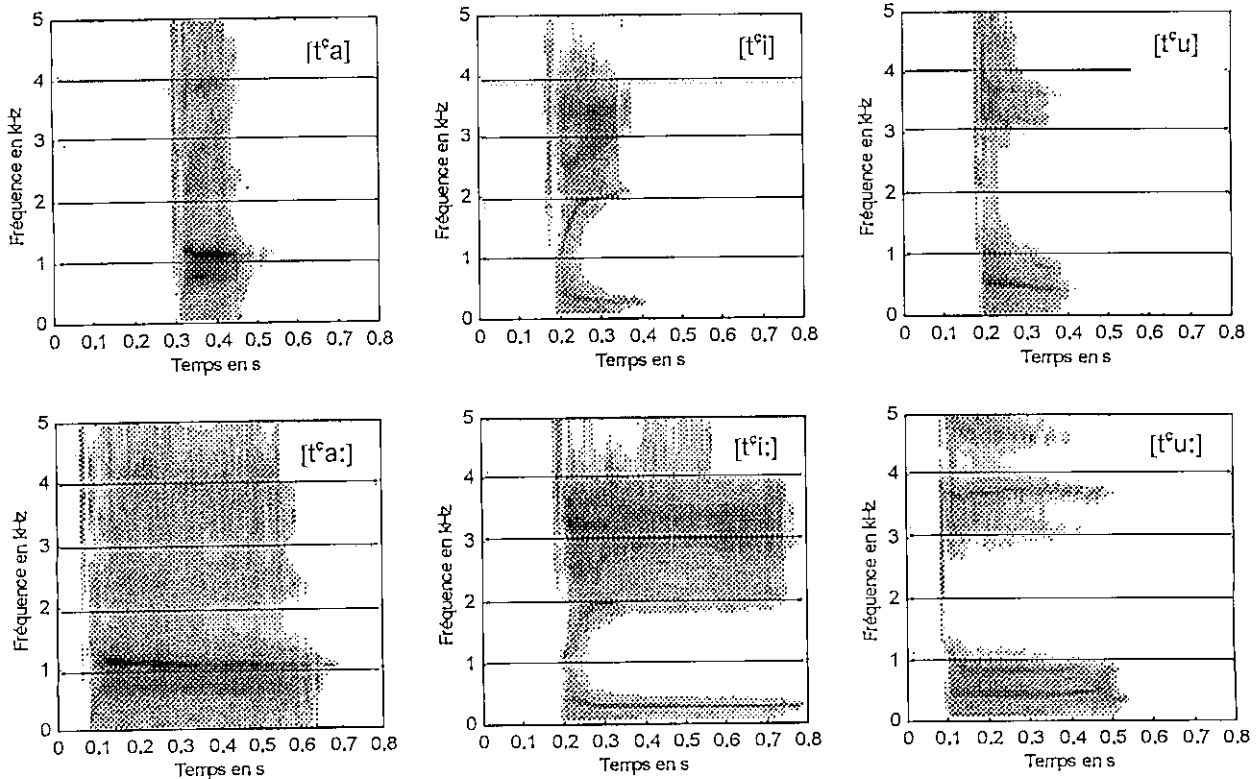


Fig. 4.5 : Spectrogrammes de [t<sup>ʕ</sup>]

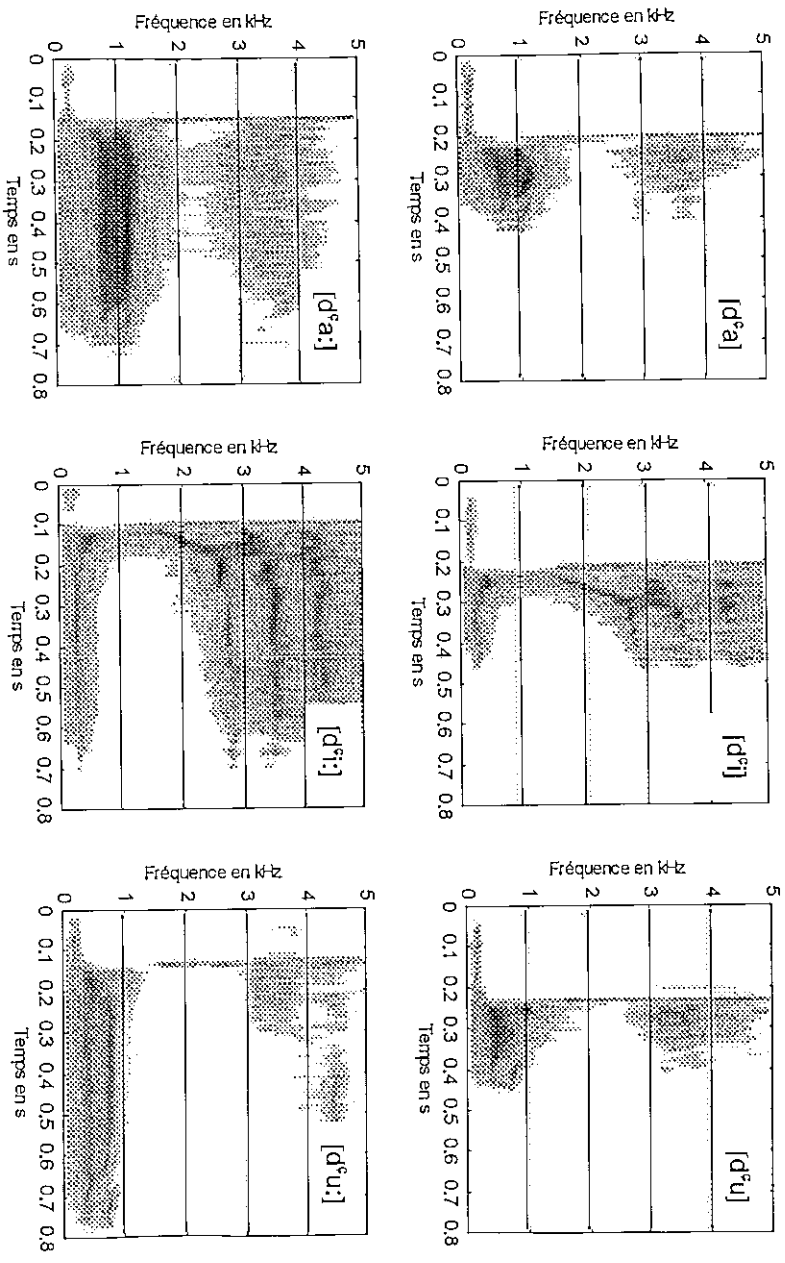
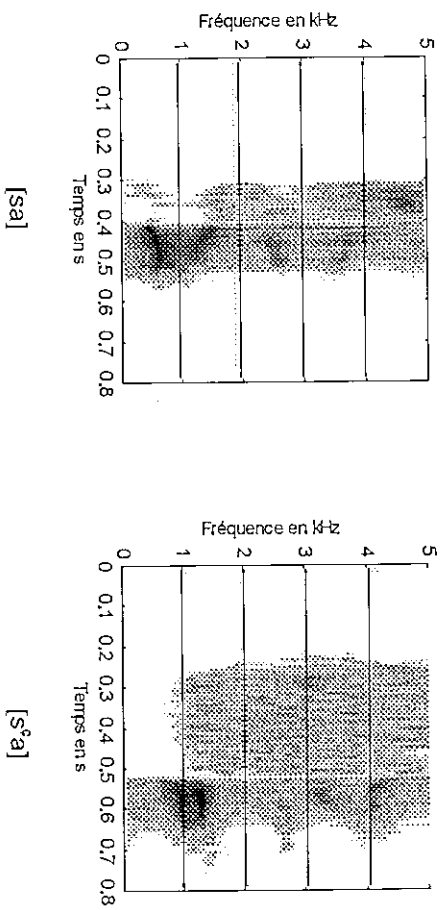
La fricative sifflante sourde apparaît sur un spectrogramme comme un bruit de turbulence de durée moyenne de 250 ms.

Ce dernier est visible aux fréquences supérieures débutant approximativement à 2.5 kHz (Fig. 4.7).

La fricative sifflante sourde emphatique [s<sup>ʕ</sup>] apparaît également sur le spectrogramme comme un bruit de turbulence de durée 300 ms (Fig. 4.7).

Située approximativement à 2.1 kHz, la concentration du bruit est moins intense par rapport à celle de [s].



Fig. 4.6 : Spectrogrammes de [d<sup>s</sup>]Fig. 4.7 : Spectrogrammes des phonèmes [sa] et [s<sup>s</sup>a]



Pour la fricative interdentale sonore [ð], placée dans une position initiale, la durée moyenne est 200 ms. La durée de la fricative interdentale sonore emphatique [z<sup>ʕ</sup>], placée dans une position initiale, est identique à celle de [ð].

Sur un plan spectral, la concentration de l'énergie causée par le bruit de friction de la sifflante pharyngale [s<sup>ʕ</sup>] est beaucoup plus importante que celle de la sifflante [s]. En effet, les spectres des phonèmes [s<sup>ʕ</sup>a] et [sa] montrent bien l'émergence rapide d'un bruit en basse fréquence avec une intensité plus grande dans le contexte emphatique (Fig. 4.8).

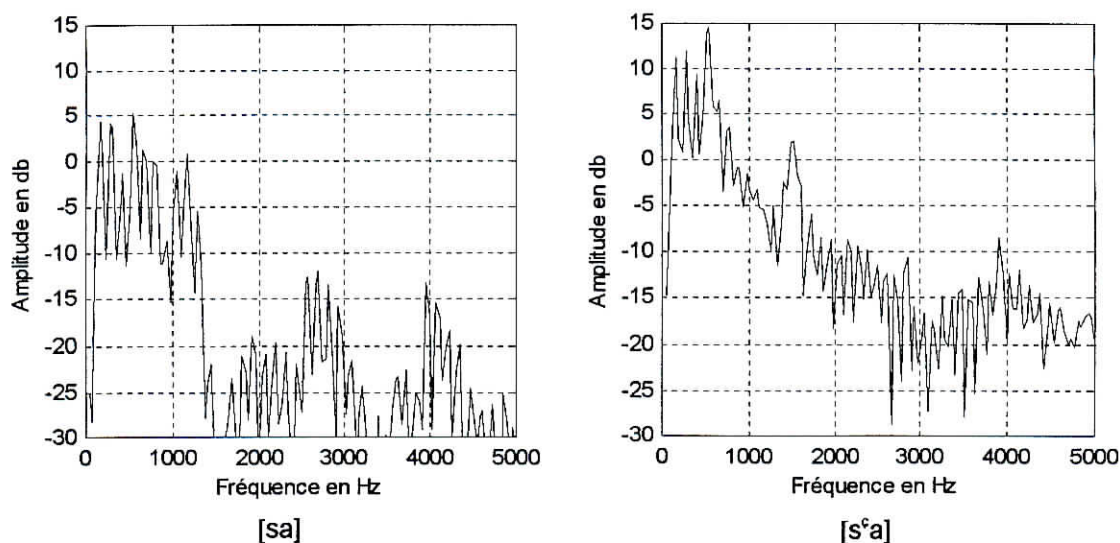


Fig. 4.8 : Spectres des phonèmes [sa] et [s<sup>ʕ</sup>a]

Au cours de cette analyse acoustique, nous avons noté que le délai de l'explosion (burst) des occlusives non-emphatiques [t] et [d] est plus court devant les voyelles [i] et [u]. Dans le même contexte, nous avons relevé que le burst de [t] est 2.7 fois plus grand est que celui de [t<sup>ʕ</sup>].

Nous rejoignons alors l'avis de Panconcelli (1924), cité par A. Giannini, & M. Pettorino (1982) [23], qui indique que le burst de [t] est trois fois plus grand que celui de [t<sup>é</sup>] (Tab. 4.3).

Phonème	[t]	[i:]	[u]	[u:]	[a]	[a:]
[t]	120	120	100	125	120	130
[t <sup>é</sup> ]	050	050	050	050	075	050
[d]	130	135	125	135	160	150
[d <sup>é</sup> ]	114	150	145	175	180	175

Tableau 4.3 : Valeurs moyennes du délai d'explosion exprimées en ms

Dans notre travail, nous avons aussi étudié les durées des voyelles de l'AS dans l'environnement de toutes les paires des consonnes qui composent notre corpus.

Ainsi, la durée de la partie voisée de la voyelle est approximativement identique pour la totalité des fricatives et occlusives emphatiques et non-emphatiques.

Cependant, un léger décroissement des valeurs des durées des voyelles courtes emphatisées par rapport à celles de leurs opposées non emphatisées est noté (Tabs. 4.4, 4.5).

Nous pouvons aussi constater que d'une part, les durées des voyelles longues emphatisées sont supérieures à celles de leurs contreparties non emphatisées.

D'autre part, les durées des voyelles longues sont approximativement 70% plus longues que celles des voyelles courtes dans un contexte non-emphatique.

Par contre nous avons noté une différence de 38% dans le contexte emphatique.

Phonème	[i]	[i:]	[u]	[u:]	[a]	[a:]
[t]	240	400	200	275	250	400
[t <sup>ç</sup> ]	180	607	175	585	220	557
[d]	350	360	225	255	325	425
[d <sup>ç</sup> ]	237	429	162	420	235	543

Tableau 4.4 : Valeurs moyennes des durées des voyelles avec les occlusives exprimées en ms

Phonème	[i]	[i:]	[u]	[u:]	[a]	[a:]
[s]	210	340	206	280	225	380
[s <sup>ç</sup> ]	189	461	173	560	212	480
[ð]	248	320	239	260	250	400
[z <sup>ç</sup> ]	246	524	219	523	207	475

Tableau 4.5 : Valeurs moyennes des durées des voyelles avec les fricatives exprimées en ms

#### 4. 5. 2 Description Formantique

Les débuts (Onsets) de F<sub>2</sub> des voyelles [i] et [i:], au voisinage de la consonne [t], débutent aux fréquences 1916 et 1983 Hz et sont donc légèrement abaissés par rapport à leur état stable habituel estimé à 2200 Hz (Tab. 4.6).

Phonème	Formants	[i]	[i:]	[u]	[u:]	[a]	[a:]
[t]	F <sub>1</sub>	0313	0320	0430	0450	0515	0490
	F <sub>2</sub>	1916	1983	1740	1526	1650	1570
	F <sub>3</sub>	2837	2600	2630	2216	3400	3360
	F <sub>4</sub>	3890	3250	3500	34	3681	3620
[t <sup>ç</sup> ]	F <sub>1</sub>	0430	0400	0460	0530	0688	0658
	F <sub>2</sub>	1250	1600	1023	1105	1077	1044
	F <sub>3</sub>	2600	2500	2653	2294	2264	2386
	F <sub>4</sub>	3341	3548	3140	3147	3890	3458
[d]	F <sub>1</sub>	0340	0330	0335	0328	0482	0480
	F <sub>2</sub>	1850	1750	1720	1700	1476	1657
	F <sub>3</sub>	2600	2500	2462	2500	2764	2840
	F <sub>4</sub>	3400	3200	3321	3400	3227	3201
[d <sup>ç</sup> ]	F <sub>1</sub>	0467	0453	0427	0426	0709	0679
	F <sub>2</sub>	1681	1801	946	1040	1049	1029
	F <sub>3</sub>	2207	2288	2241	2552	2641	2092
	F <sub>4</sub>	3052	3085	3192	3296	3683	3705
[ð]	F <sub>1</sub>	0320	0325	0340	0336	0520	0614
	F <sub>2</sub>	2080	2081	1500	1500	1727	1323
	F <sub>3</sub>	3033	3031	2586	2645	2640	2516
	F <sub>4</sub>	3400	3328	3027	3060	3334	3897
[z <sup>ç</sup> ]	F <sub>1</sub>	0356	0339	0470	0376	0784	0682
	F <sub>2</sub>	1785	1792	1045	1158	1071	997
	F <sub>3</sub>	2793	2854	2386	2545	2795	2056
	F <sub>4</sub>	3759	3908	3327	3360	3981	2705
[s]	F <sub>1</sub>	0300	0301	0320	0317	0575	0609
	F <sub>2</sub>	2068	2000	1350	1400	1427	1300
	F <sub>3</sub>	3223	3100	2720	2819	2506	2424
	F <sub>4</sub>	3766	3712	3480	3451	3800	3160
[s <sup>ç</sup> ]	F <sub>1</sub>	0329	0315	0359	0330	0869	0850
	F <sub>2</sub>	1929	1866	926	821	1138	1225
	F <sub>3</sub>	2697	2100	2620	2736	2870	2500
	F <sub>4</sub>	3650	2780	3488	3461	3842	2789

Tableau 4.6 : Les onsets des formants F<sub>1</sub> F<sub>2</sub> F<sub>3</sub> F<sub>4</sub> exprimés en Hz



En revanche, les onsets de F<sub>2</sub> des [u] et [u:] s'élèvent subitement à 1740 Hz dépassant largement les états stables usuels estimé à 775-800 Hz. Pour la voyelle [a], les onsets et les états stables de F<sub>2</sub> sont égales respectivement à 1570 et 1650 Hz alors que ceux de la voyelle longue [a:], ils sont généralement plus bas, de l'ordre de 1370 et 1450 Hz. A coté de [t<sup>é</sup>], les voyelles [i] et [i:] sont marquées par un F<sub>2</sub> affecté. En effet, les F<sub>1</sub> et F<sub>2</sub> sont influencés par l'abaissement des onsets de F<sub>2</sub> respectivement à 1250-1600 Hz, avant de revenir à leur état stable de 2200Hz (Tab. 4.6).

Ces voyelles sont également influencées par une élévation des onsets de F<sub>1</sub> de 300 à 430 Hz. Les onsets de F<sub>2</sub> de la paire [u] / [u:] s'élèvent, par rapport aux états stables 775-800 Hz, vers 1023-1105 Hz. L'élévation avec [t] est donc plus importante et beaucoup plus aiguë. Pour la voyelle [a], les onsets et les états stables de F<sub>2</sub> varient de 1044 à 1077 Hz alors que ceux de [a:], sont situés entre 1044 et 1116 Hz montrant ainsi que ces deux voyelles ne sont plus différenciées par les onsets de F<sub>2</sub> mais par leurs états stables (Fig. 4.9).

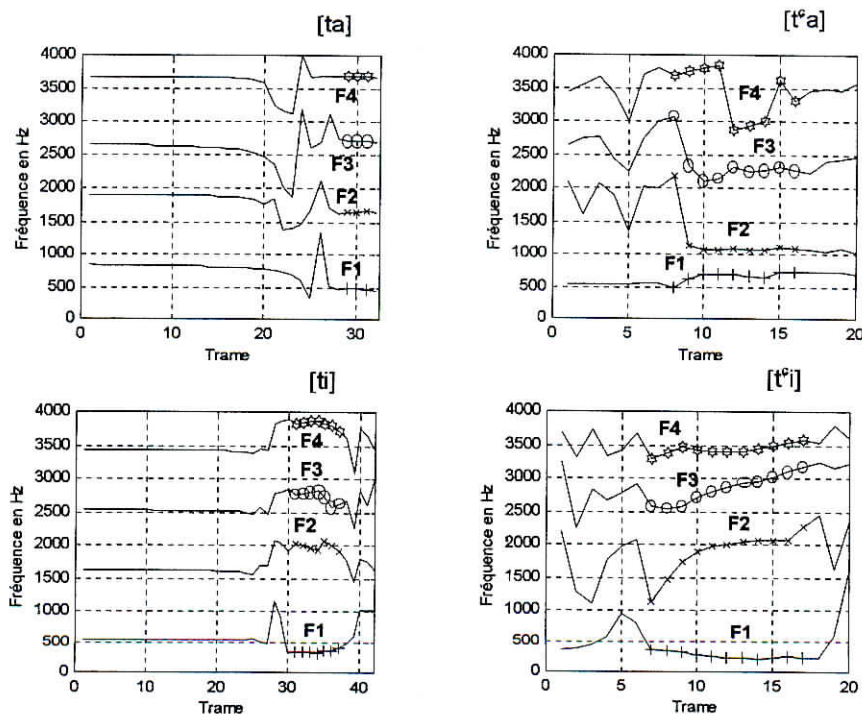


Fig. 4.9 : Contours formantiques des syllabes [ta], [t<sup>a</sup>], [ti] et [t<sup>i</sup>]

De plus, la consonne [d] produit, sur toutes les voyelles, le même effet que celui de [t] et Les voyelles, au voisinage des consonnes [d<sup>◌</sup>] et [t<sup>◌</sup>] sont emphatisées de la même manière c'est-à-dire, un abaissement de F<sub>2</sub> et une élévation de F<sub>1</sub> (Fig. 4.10).

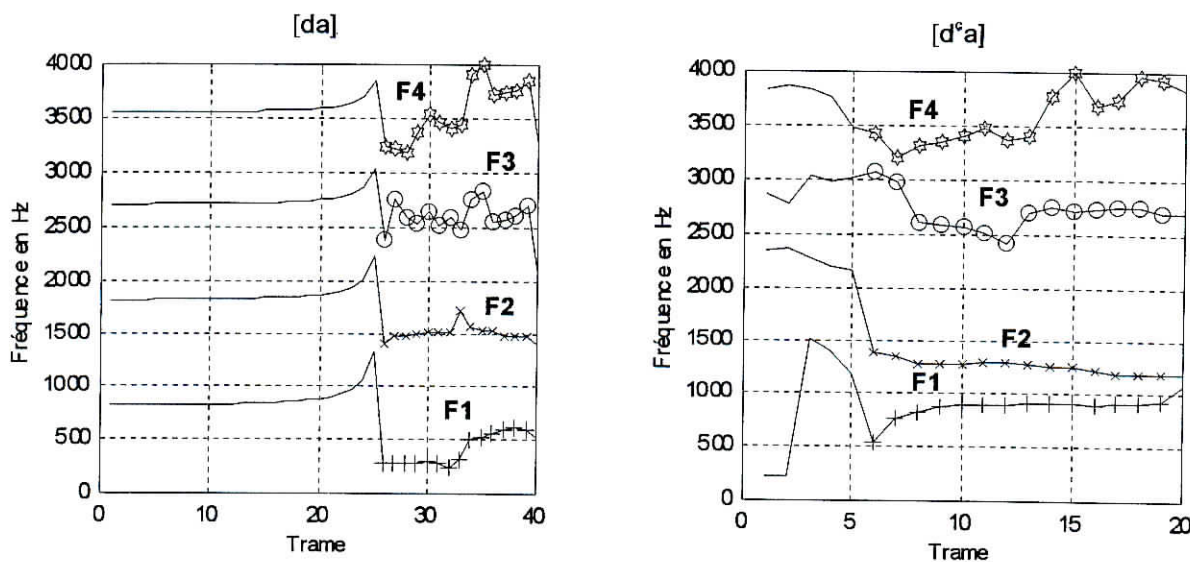


Fig. 4.10 : Contours formantiques des phonèmes [da] et [d<sup>◌</sup>a]

De leurs états stables habituels de 1500-1200 Hz, les onsets de F<sub>2</sub> des voyelles [a] et [a:], au voisinage de [s<sup>◌</sup>], sont abaissés respectivement à 1138 et 1225 Hz (Tab. 4.6).

Beaucoup plus important que celui de [s], l'abaissement de F<sub>2</sub> de [a] est bien indiqué par des transitions qui débutent à 1138 Hz et s'élèvent graduellement pour rejoindre l'état stable.

La consonne [s<sup>◌</sup>] affecte aussi les onsets de F<sub>1</sub> de la paire [i] / [i:] en les élevant de 300Hz à 315-329 Hz (Fig. 4.11). L'effet de [s<sup>◌</sup>] sur F<sub>1</sub> et F<sub>2</sub> est plus proéminent avec [i] qu'avec [i:]. Les onsets de F<sub>2</sub> des voyelles [u] et [u:] s'élèvent graduellement de 821 à 926 Hz mais demeure moins abaissés que ceux notés des mêmes voyelles au voisinage de [s].



Les onsets de F<sub>2</sub> des voyelles [i] et [i:] sont localisés respectivement à 1866 et 1929 Hz.

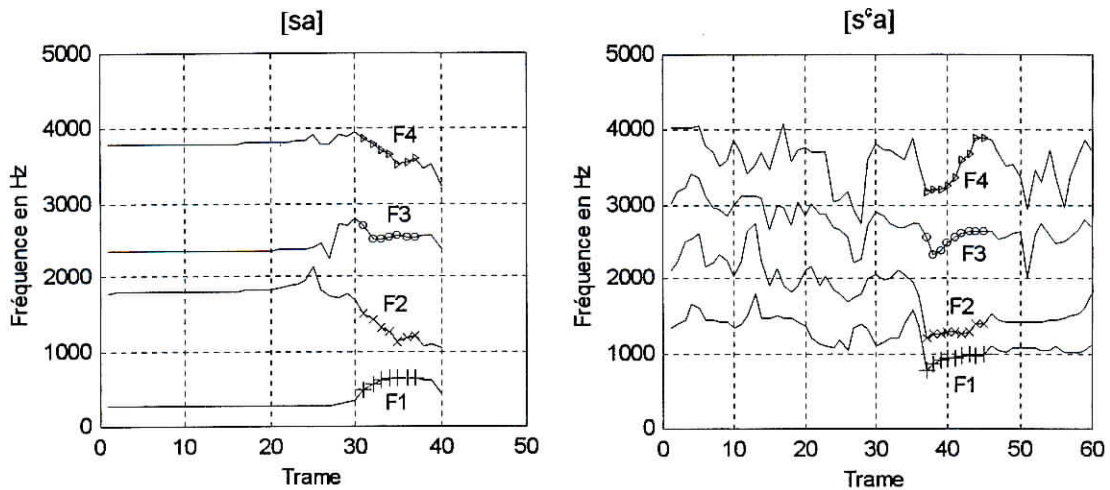


Fig. 4.11 : Contours formantiques des phonèmes [sa] et [s°a]

L'effet de la fricative interdentale sonore [ð], placée dans une position initiale, sur les onsets de F<sub>2</sub> des voyelles [i] et [i:] est indiqué par leurs abaissement à 2081 Hz par rapport à leurs états stables de 2200 Hz.

Les onsets de F<sub>2</sub> de la paire [u] / [u:] sont élevés par rapport à leurs états stables 775-800 Hz pour atteindre 1500 Hz ; Les valeurs respectives de ceux des voyelles [a] et [a:] sont 1727 Hz et 1323 Hz (Tab. 4.6).

Les onsets de F<sub>2</sub> des voyelles [i] et [i:], au voisinage de [z°], sont abaissés par rapport à leurs états stables de 2200 Hz vers les fréquences 1785 et 1792 Hz.

Ceux de la paire [u] / [u:], sont élevés par rapport à leurs états stables usuels de 775-800 Hz pour atteindre 1045 Hz (Tab. 4.5). Et ceux des voyelles [a] et [a:], sont égaux respectivement à 1071Hz et 997Hz.

Dans notre analyse, nous avons aussi cherché à connaître les valeurs des quatre premiers formants des paires d'occlusives de notre corpus juste après l'explosion. Nous avons examiné la variation des formants  $F_1$  et  $F_3$  par rapport à la voyelle associée (Tab. 4.7).

Il est clair que la valeur la plus basse de  $F_2$  de [t], égale à 1526 Hz, se produit avec [u:] alors que son maximum, égale à 1916 Hz, est obtenu avec [i]. Cependant, la valeur minimale de  $F_2$  de [t<sup>ç</sup>] est enregistrée avec [u] et la maximale avec [i:].

Phonème	Formants	[i]	[i:]	[u]	[u:]	[a]	[a:]
[t]	$F_1$	0313	0320	0430	0450	0515	0500
	$F_2$	1916	1850	1740	1650	1700	1600
	$F_3$	2837	2600	2630	2350	3100	2250
	$F_4$	3890	3250	3500	3480	3681	3200
[t <sup>ç</sup> ]	$F_1$	0430	0400	0460	0530	0597	0676
	$F_2$	1255	1608	1023	1105	1197	1467
	$F_3$	2604	2500	2653	2294	2304	2535
	$F_4$	3341	3548	3140	3147	3711	3269
[d]	$F_1$	0410	0415	0420	0410	0500	0500
	$F_2$	1850	1790	1600	1730	1575	1625
	$F_3$	2600	2540	2400	2520	2517	2614
	$F_4$	3400	3260	3200	3410	3365	4000
[d <sup>ç</sup> ]	$F_1$	0463	0426	0465	0413	0619	0610
	$F_2$	1820	1417	1428	1529	1665	2150
	$F_3$	3041	2282	2364	2530	2277	2820
	$F_4$	3686	3110	3257	3237	3121	3668

Tableau 4.7 : Les valeurs des formants  $F_1$   $F_2$   $F_3$   $F_4$  après l'explosion exprimés en Hz

Nous remarquons aussi, à travers les valeurs formantiques contenues dans le tableau 4.6 que le rapprochement des deux premiers formants commence bien avant que le conduit vocal prend la configuration d'une voyelle.

Dans le but d'apporter un intérêt important à la description acoustique des consonnes emphatiques, nous avons calculé les valeurs moyennes des fréquences des deux premiers formants aux deux points de la voyelle courte qui suit. Ces deux points, représentant la partie initiale et la partie stable de la voyelle. Nous avons alors établi un triangle vocalique. (Fig. 4.12)

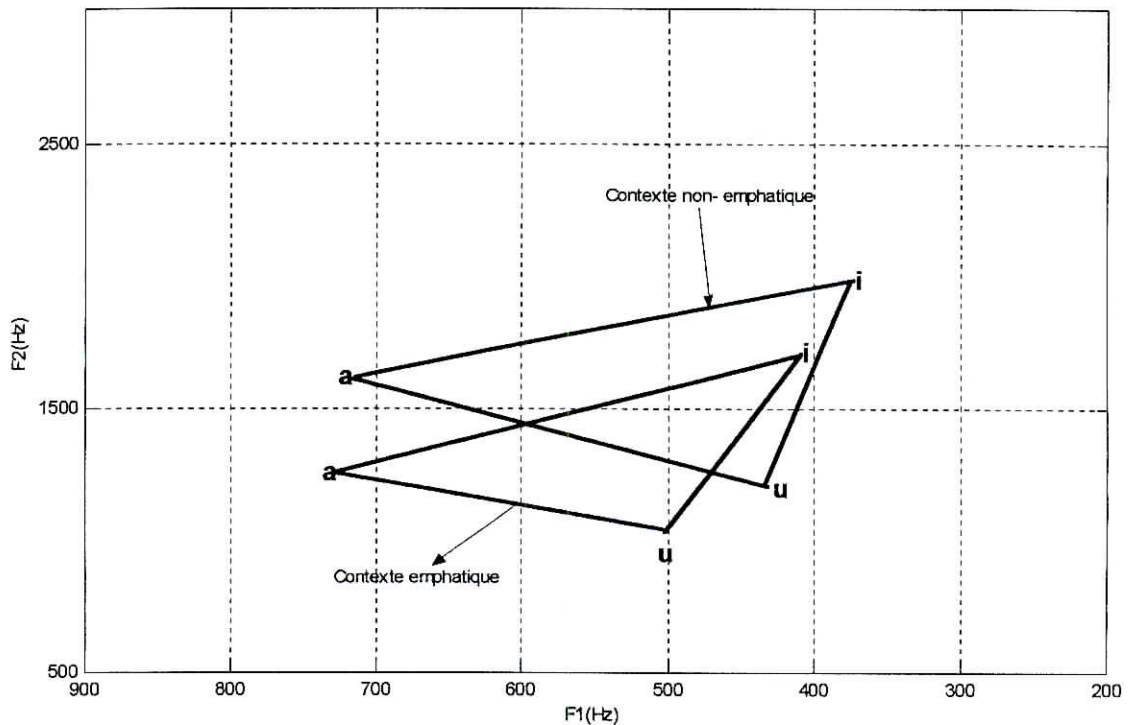


Fig. 4.12 : Triangle vocalique des voyelles courtes

## 4.6 Conclusion

Sur la base de cette analyse acoustique et comme le montre, d'une manière claire, les différents tracés des contours formantiques et les spectrogrammes, la constriction du pharynx des consonnes emphatiques, citée dans la littérature, était toujours accompagnée par un abaissement de  $F_2$  et une élévation de  $F_1$ .

L'abaissement de  $F_2$  se prolonge dans la région stable des voyelles. Des attributs acoustiques similaires ont été rapportés ailleurs pour les sons emphatiques. [7, 8, 23, 24, 25, 27, 28]

Ce rapprochement entre les deux premiers formants est plus apparent pour les occlusives que les fricatives et plus extrême au voisinage de [a] que [i].

Finalement, cette caractérisation acoustique nous mène à affirmer qu'un pharynx est plus «constricte» pour les sons emphatiques que leurs opposés non-emphatiques. Cette étude suggère aussi un degré d'ouverture plus large et une rétraction du lieu d'articulation des voyelles au voisinage des consonnes emphatiques.

# **Conclusions Générales et Perspectives**

Dans cette étude, nous nous sommes attachés à caractériser et à décrire acoustiquement les sons emphatiques de l'Arabe Standard (AS) en voix parlée. Pour aborder cette étude, des techniques d'analyse des signaux de la parole ont été utilisées.

Nous avons englobé les différentes techniques usuelles d'analyse spectrale dans un logiciel. Elles sont considérées comme étant des appuis utiles aux investigations acoustiques, non seulement pour garantir une compréhension plus approfondie des faits et de dépasser ainsi le stade des études impressionnistes, mais aussi pour franchir le pas et proposer des interprétations et des règles nécessaires aux travaux de recherche sur la synthèse de la parole par règles réalisés dans le cadre de l'AS. Cette dernière a bénéficiée d'importantes études tant au niveau quantitatif que qualitatif.

L'effet de l'emphase sur les systèmes vocaliques et consonantiques de l'AS a été le thème moteur de notre caractérisation acoustique des quatre consonnes emphatiques étudiées. Nous nous sommes basés sur une analyse acoustique pour identifier les paramètres affectés par la pharyngalisation.

Pour cela, nous avons utilisé les approches spectrographique et formantique. Les résultats obtenus sont encourageants. En effet, ils sont en accord avec certains résultats obtenus dans d'autres études citées dans la littérature.



L'étude formantique montre que l'intérêt d'une telle approche qui associe suivi articulatoire et analyse acoustique est capitale pour le Traitement Automatique de la Parole (TAP) et l'étude de l'Arabe Standard (AS).

Au cours de notre application nous avons relevé la durée des sons emphatiques étudiés, les valeurs des quatre premiers formants, aux deux extrémités et au milieu des six voyelles de l'AS, ont été notées. De plus, chaque fois que les données nous le permettaient, nous avons tenté de confronter nos résultats à ceux des autres chercheurs cités dans les références bibliographiques.

Il est important de signaler, qu'à travers les conclusions de la littérature consultée, que l'emphase se définit, au niveau articulatoire, comme une pharyngalisation.

Nos résultats acoustiques montrent que l'abaissement des fréquences du deuxième formant, l'élévation de celles du premier et les transitions fortement descendantes constituent les corrélats de l'emphase.

L'élévation du premier formant qui correspond principalement à la partie supérieure du conduit vocal décrit le raccourcissement de cette cavité. Elle prouve alors que le rétrécissement est localisé bas dans le pharynx, appuyant ainsi les conclusions de plusieurs travaux sur ce sujet. Quant à l'abaissement du second formant, il est lié à l'augmentation du volume de la cavité antérieure du conduit vocal.

Les autres conclusions principales dégagées à partir des expériences réalisées pour étudier l'effet du contexte emphatique sur le système vocalique de l'Arabe Standard sont :

- Le délai de l'explosion de l'occlusive non-emphatique [t] est approximativement trois fois plus grand que celui de son opposée emphatique [t<sup>é</sup>].



- La pharyngalisation n'affecte pas la durée vocalique. Néanmoins, nous pouvons conclure que d'une part, les durées des voyelles longues emphatisées sont supérieures à celles de leurs contreparties non emphatisées. Et d'autre part, les durées des voyelles longues sont approximativement 70% plus longues que celles des voyelles courtes dans un contexte non-emphatique. Nous avons aussi noté une différence de 38% entre les mêmes voyelles dans le contexte emphatique.

- Le rapprochement entre  $F_1$  et  $F_2$  est plus marqué pour les voyelles courtes que leurs opposées longues. Une telle différence est justifiée par le rôle spécifique que jouent les voyelles longues dans le système phonologique de l'AS, mais aussi par l'instabilité des caractéristiques articulatoires des voyelles courtes qui les rendent facilement influencés par l'environnement emphatique.

La recherche d'autres indices acoustiques qui contribuent à la distinction d'une articulation emphatique d'une non-emphatique dans le même entourage vocalique est impérative.

Ainsi, en vue de l'extraction des règles qui seront utilisées dans l'élaboration d'un modèle prédictif de la synthèse de la parole par règles de l'AS, ce travail peut être complété par la détermination des limites de l'extension de l'emphase. Il est judicieux d'examiner tous les environnements possibles dans la phonologie de l'AS :  $C_EVC_E$ ,  $VC_EC_EV$  et  $C_EVC_EV$  ( $C_E$  : consonne emphatique). Le but essentiel de cette perspective est de vérifier si «l'emphase» est la caractéristique du segment simple ou elle s'étend de la voyelle qui précède la consonne emphatique à celle qui suit.

## Références bibliographiques

- [1] Calliope, 1989. *La parole et son traitement automatique*. Edition Masson, Paris, France.
- [2] L. R. Rabiner & R. W. Schafer, 1978. *Digital Processing of speech Signals*. Prentice-Hall, Englewood cliffs, N. J.
- [3] J. Van Den Berg, 1958. *Myoelastic-aerodynamic theory of voice production*. J. Speech Hear. Res. 1, pp. 227-44.
- [4] G. Jan. Svec & J. Versely, 2000. *Resonance properties of the vocal folds : In vivo laryngoscopic investigation of the externally excited laryngeal vibrations*. J. Acoust. Soc. Am, vol. 108(4), pp. 1397-1420.
- [5] M. Guerti, 1983. *Contribution à la synthèse de la parole en Arabe standard, synthèse par diphones et technique de prédiction linéaire*. Thèse de Magister, ILP, Alger, Algérie.
- [6] G. Fant, 1991. *On the predictability of Formant levels and spectrum envelopes from Formant frequencies*. Royal Institute of Technology, Stockholm.
- [7] S. H. Al-Ani, 1970. *Arabic Phonolgy : an acoustical and physiological investigation*. Mouton, The Hague, Paris.
- [8] D. H. Obrecht , 1968. *Effets of the Second Formant on the perception of Velarized Consonants in Arabic*. Mouton, The Hague, Paris, France.
- [9] M. Mrayati & R. Carré, 1976. *Relations entre la forme du conduit vocal et les caractéristiques acoustiques des voyelles françaises*. *Phonetica* 33. pp. 285-306.

- [10] M. M. Sondhi, 1979. *Estimation of vocal-tract areas : The need of acoustical measurements*. IEEE Trans. Acoust. , Speech, and Signal Processing, vol. 27(3), pp. 268-273.
- [11] M. M. Sondhi & J. R. Resnick, 1983. *The Inverse problem for vocal-tract : Numerical methods, acoustical experiments, and speech synthesis*. J. Acoust. Soc. Am, vol. 73(3), pp. 985-1022.
- [12] P. Mermelstein, 1967. *Determination of the vocal-tract shape from measured formant frequencies*. J. Acoust. Soc. Am, vol. 41(5), pp. 1283-1294.
- [13] G. Fant & al, 1972. *Speech analysis and speech production*. Q. Prog. Status Rep1. Speech Transm. Lab. (STL-QPSR1). Royal Institute of Technology, Stockholm, Sweden.
- [14] N. Henrich, 2001. *Etude de la source glottique en voix parlée et chantée*. Thèse de doctorat en Acoustique, Université Paris 6.
- [15] N. Q. Tuan, 1984. *Réalisation d'un synthétiseur à Formants numériques en temps réel. Caractérisation de la source d'excitation et des transitions d'amplitude*. Thèse de Doctorat, Institut National Polytechnique de Grenoble.
- [16] H. Wakita, 1979. *Estimation of vocal-tract shapes from acoustical analysis of the speech wave : The state of the art*. IEEE Trans. Acoust. , Speech, and Signal Processing, vol. 27(3), pp. 281-285.
- [17] O, 1999. *Vocal tract modeling in 3D*. TMH-QPSR1-2, pp. 31-38.
- [18] Mc Candless, & S. Stephanie, 1974. *An algorithm for automatic formant extraction using linear prediction spectra*. IEEE Trans. Acoust. Speech, and Signal Processing, vol. 22(2), pp. 135-141.



- [19] B. S. Atal, & S. L. Hanauer, 1971. *Speech analysis and synthesis by linear prediction of speech wave*. J. Acoust. Soc. Am, vol. 50(2), pp. 637-655.
- [20] H. Hermansky, 1990. *Perceptuel Linear Predictive (PLP) analysis of speech*. J. Acoust. Soc. Am, vol. 87(4), pp. 1738-1752.
- [21] J. Makhoul, 1975. *Linear Prediction : A tutorial Review*. Proc. IEEE 63, 4, pp. 561-580.
- [22] A. El-jaroudi, & J. Makhoul, 1991. *Estimation Discrete All-Pole Modeling*. IEEE Transactions. On signal Processing 39, 2, pp. 411-423.
- [23] A. Giannini, & M. Pettorino, 1982. *The emphatic consonants in Arabic*. Speech Laboratory Report, 4, (Istituto Universitario Orientale, Napoli).
- [24] S. Ghazali, 1987. *Etude EMG préliminaire sur les consonnes arrières de l'Arabe*. 16<sup>ème</sup> Journée d'Etude sur la Parole (JEP). Société Française d'Acoustique, Hammamet, Tunisie.
- [25] J. F. Bonnot, 1979. *Recherche expérimentale sur la nature des consonnes emphatiques de l'Arabe classique*. Travaux de l'Institut de phonétique de Strasbourg, France, vol. 9. pp. 47-88.
- [26] A. Alwan, 1989. *On Perceptuel cues for place of articulation for the voiced pharyngeal and uvular consonants*. J. Acoust. Soc. Am, vol. 86(2), pp. 549-556.
- [27] A. Laufer & T. Baer, 1988. *The emphatic and Pharyngeal Sounds in Hebrew and in Arabic*. Status Report on Speech Research, Haskins Laboratories, New Heaven Connecticut, USA.
- [28] F. Farah & M. Guerti, 2003. *Caractérisation des emphatiques sur le plan acoustique*. AJOT, 03 (publication soumise).