

**ECOLE NATIONALE POLYTECHNIQUE**  
DEPARTEMENT D'ELECTRONIQUE

**THESE**

المدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

présentée pour l'obtention du diplôme de  
**MAGISTER EN ELECTRONIQUE APPLIQUEE**

**OPTION : Acquisition et traitement de l'information**

PAR :

**ABINA NADIA**

**INGENIEUR D'ETAT EN ELECTRONIQUE**

**DETECTION DE LA FATIGUE  
A L'AIDE DES  
PARAMETRES PROSODIQUES ET  
FORMANTIQUES DE LA PAROLE**

Soutenue devant le jury composé de :

**Président : M. A. ZERGUERRAS** Maître de conférences (ENP)

**Rapporteur : Melle M. GUERTI** Maître de conférences (ENP)

**Examineurs : Mme M. BEDDEK** Chargée de cours (ENP)

**M. B. BOUSSEKSOU** Chargé de cours (ENP)

**M. B. DERRAS** Maître de conférences (ENP)

**JUIN 1995**

المدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

A ma mère, à mon père.

## Remerciements

Je saisis l'occasion qui m'est donnée pour exprimer ma profonde reconnaissance à Mademoiselle M. Guerti et saluer son sens de la responsabilité. Toujours encourageante, elle n'a cessé de prodiguer son temps et ses conseils pour faire aboutir ce travail. Je la remercie profondément d'avoir accepté de diriger mon travail.

Je remercie Monsieur A. Zerguerras d'avoir accepté de présider le jury chargé d'évaluer mon travail et Madame M. Beddek, Messieurs B. Bousseksou et B. Derras d'avoir bien voulu en être membres.

J'ai une profonde gratitude pour Monsieur B. Derras qui m'a toujours conseillée et fait profiter de ses connaissances de bon coeur. J'aimerais qu'il sache combien j'ai apprécié et la profonde estime que j'en ai conçue.

Je ne saurai oublier Monsieur S. Ait-Chikh dont j'ai apprécié les qualités humaines. J'aimerais ici lui exprimer ma profonde reconnaissance pour son soutien moral et ses encouragements.

J'adresse mes remerciements à Monsieur A. Hadj salah, directeur de l'unité de recherche en science du langage et de la technologie linguistique de l'Université de Bouzaréah, pour avoir accepté de me laisser disposer du matériel de l'unité.

Je remercie aussi vivement Monsieur L. Bechichi, directeur de l'Entreprise Nationale de la Radio Sonore, qui m'a permis de disposer des moyens humains et matériels de l'entreprise.

J'ai une grande reconnaissance pour le Professeur H. Abdelouahab, qui a bien voulu m'ouvrir les portes de son service ORL à l'hôpital Mustapha. Je remercie aussi ma soeur le Docteur Abina Mouni pour son aide bibliographique concernant les pathologies de l'ORL.

Ma sincère gratitude va aussi à Madame N. Tiziri et Mademoiselle D. Remaki, attachées de recherche au laboratoire d'Electronique Acoustique à l'Université de Bouzaréah, qui m'ont très amicalement fait profiter de leurs connaissances et de leurs expériences.

J'adresse mes vifs remerciements à Madame J. Nacer, médecin spécialiste au Centre National de la Médecine du Sport de Ben-Aknoun, qui m'a spontanément et très gentiment offert son aide. J'ai été très touchée de voir qu'elle ne ménageait ni son temps ni ses efforts pour m'aider à comprendre et réaliser la partie sur le processus d'apparition de la fatigue dans le corps humain.

J'aimerais aussi que Mesdemoiselles Rabéa Dahmani, Nadjia Kara et Nassira Labchri sachent combien j'ai été sensible à leurs marques d'amitié. Je les remercie de leurs soutiens et leurs encouragements et leur souhaite de finir leurs thèses dans les meilleurs délais.

Je remercie Mademoiselle Nassira Cherid pour son aide et sa confiance en moi.

Je remercie aussi toutes les personnes qui, ne serait ce que d'une parole d'encouragement, m'ont aidé dans la réalisation de cette thèse.

## TABLE DES MATIERES

INTRODUCTION.....	1
-------------------	---

### CHAPITRE 1 - NOTIONS FONDAMENTALES SUR LA PAROLE ET LA FATIGUE

Introduction.....	4
1.1 Production du signal de parole.....	4
1.1.1 Eléments constitutifs de l'appareil phonatoire.....	4
1.2 Concepts fondamentaux de physico-acoustique et de phonétique.....	5
1.2.1 La fréquence fondamentale.....	5
1.2.2 La mélodie et la hauteur.....	5
1.2.3 Le timbre.....	5
1.2.4 La prosodie.....	5
1.2.5 Le phonème.....	6
1.2.6 La coarticulation.....	6
1.2.7 Représentation mentale des sons.....	6
1.3 Génération d'un signal de parole.....	7
1.4 Fonctionnement acoustique de l'appareil vocal.....	9
1.4.1 Les voyelles.....	9
1.4.1.1 Les voyelles du français.....	9
1.4.1.2 Les voyelles de l'arabe.....	11
1.4.2 Les consonnes.....	12
1.4.2.1 Mode d'articulation.....	12
1.4.2.1 Les consonnes du français.....	13
1.4.2.2 Les consonnes de l'arabe standard.....	14
1.5 Définitions générales et effets de la fatigue.....	17
1.6 Energie de la contraction musculaire.....	17
1.6.1 Processus anaérobie.....	17
1.6.2 Processus aérobie.....	18
1.7 Principaux facteurs de fatigue.....	19
1.7.1 Facteurs externes.....	20
1.7.2 Déplétion en métabolites.....	20
1.7.3 Accumulation de métabolites.....	22
1.8 Fatigue vocale.....	23
Conclusion.....	25

### CHAPITRE 2 - TECHNIQUES DE TRAITEMENT DE LA PAROLE

Introduction.....	26
2.1 Méthodes d'analyse en traitement de la parole.....	26
2.1.1 Le spectrographe.....	26
2.1.2 Le périodogramme.....	28
2.1.3 L'analyse cepstrale.....	29
2.1.4. La prédiction linéaire et le modèle autorégressif.....	30

2.1.4.1. Modélisation du système phonatoire choisie.....	31
2.1.5 Méthodes Auto-Régressives.....	31
2.1.5.1 Méthode de covariance.....	32
2.1.5.2 Méthode d'autocorrélation.....	35
2.1.5.3 Méthode de Burg.....	36
2.1.5.4 Comparaison des méthodes autorégressives.....	38
2.1.5.5 Nombre optimal des coefficients de prédiction.....	39
2.2. Détection du pitch.....	40
2.2.1 Méthode du cepstre.....	41
2.2.2 Codage prédictif.....	42
2.2.3 Méthodes basées sur la fonction d'autocorrélation.....	42
2.2.3.1 Méthode de Sondhi.....	43
2.2.3.2 Méthode de Dubnowski.....	44
2.2.4 Méthode de la Fonction de la Différence d'Amplitude Moyenne (AMDF).....	45
2.2.5 Méthode SIFT (Simplified Inverse Filtering Tracking).....	45
2.2.6 Méthode de Rabiner.....	46
2.2.7 Méthode de l'électroglottographe.....	46
2.2.8 Méthode de passage par zéro du signal.....	46
2.3 Détection des formants.....	46
2.3.1 Méthode du cepstre.....	46
2.3.2 Codage prédictif.....	47
2.3.3 Méthode du périodogramme.....	47
Conclusion.....	48

## CHAPITRE 3 - VARIABILITE ET TRAITS PERTINENTS DANS LA PAROLE

Introduction.....	49
3.1 Variabilité du signal de parole.....	49
3.1.1 Variabilité inter-locuteur.....	49
3.1.1.1 Influence de la taille du conduit vocal.....	50
3.1.2 Variabilité intra-locuteur.....	52
3.1.2.1 Importance de l'articulation.....	52
3.1.2.2 Rôle de la vitesse d'élocution.....	52
3.1.2.3 Rôle de l'intensité de la parole.....	54
3.1.2.4 variation du pitch.....	55
a - Microméodie.....	55
b - Ligne de déclinaison et macroméodie.....	56
3.1.2.5 Importance d'une utilisation adéquate des paramètres prosodiques.....	56
3.1.3 Influence de l'émotion.....	57
3.1.4 Mauvaise hygiène vocale.....	59
3.1.5 Variations pathologiques de la parole.....	60
3.1.6 Effet de la coarticulation.....	61
3.2 Recherche des invariants et identification du locuteur.....	63
3.2.1 Qualité des traits distinctifs.....	63
3.2.2 Le timbre.....	63
3.2.3 L'articulation.....	64
3.2.3.1 La coarticulation.....	64
3.2.3.2 Durée d'établissement des plosives.....	64

3.2.3.3 Enveloppe énergétique.....	64
3.2.4 Les variations prosodiques .....	65
Conclusion.....	66

## CHAPITRE 4 - PROCEDURES D'ACQUISITION ET DE TRAITEMENT UTILISEES

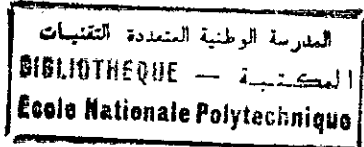
Introduction .....	67
4.1 Génération du corpus.....	67
4.1.1 Population.....	67
4.1.2 Choix du corpus.....	69
4.1.3 Conditions d'enregistrement.....	70
4.2 Chaînes d'analyse utilisées.....	71
4.2.1 Première chaîne d'analyse.....	72
4.2.2 Seconde chaîne d'analyse.....	76
4.2.2.2 Le support logiciel Autostart (Autost).....	77
4.2.2.3 Méthodes d'analyses retenues.....	79
4.2.2.4 Choix des conditions d'analyse .....	79
4.3 Méthode de travail .....	80
4.4 Hypothèses formulées.....	81
4.4.1 La fréquence fondamentale.....	81
4.4.2 Les formants .....	81
4.4.3 Les énergies.....	81
4.4.4 Durée des phonèmes .....	82
Conclusion.....	82

## CHAPITRE 5 -RESULTATS OBTENUS ET INTERPRETATIONS

Introduction.....	83
5.1 Analyse des variations de la fréquence fondamentale sous l'effet de la fatigue.....	83
5.1.1 Evolution du fondamental sous l'influence d'une fatigue non vocale .....	83
5.1.1.1 Variations du fondamental d'une locutrice sur dix jours d'enregistrement.....	84
5.1.1.2 Variations du fondamental pour des enfants.....	84
5.1.1.3 Variations du fondamental de locuteurs adultes.....	85
5.1.1.4 Interprétation.....	88
5.1.2 Evolution du fondamental sous l'effet d'un effort vocal.....	88
5.1.2.1 Pitch des voyelles dans la phrase / lə ʃa sɔt /.....	88
5.1.2.2 Pitch des voyelles de la phrase /Saama ramaDaan/ .....	89
5.1.2.3 Pitch du phonème /m/ en contexte /Saama/ .....	90
5.1.2.4 Interprétation .....	90
5.2 Influence de l'effort vocal sur les durées des phonèmes .....	90
5.2.1 Durées des phonèmes dans la phrase / lə ʃa sɔt / .....	91
5.2.1.2 Interprétation .....	91
5.2.2 Durées des phonèmes dans la phrase /Saama ramaDaan/ .....	91
5.2.3 Interprétation .....	92
5.3 Evolution des énergies sous l'influence de l'effort vocal .....	93
5.3.1 Énergies des voyelles dans la phrase / lə ʃa sɔt / .....	93
5.3.2 Énergies des voyelles dans la phrase / Saama / .....	94

5.4 Détermination et sélection des formants .....	95
5.4.1 Formants de la première catégorie de locuteurs .....	95
5.4.1.1 Comparaison entre les résultats obtenus par les différents modèles AR .....	95
5.4.1.2 Analyse des variations de la fréquence des formants sous l'influence d'une fatigue physique .....	97
5.4.2 Evolution des fréquences des formants sous l'influence de l'effort vocal .....	99
5.4.2.1 Evolution des formants des voyelles dans la phrase « le chat saute » .....	100
5.4.2.2 Evolution des formants des voyelles dans /Saama / .....	104
5.4.2.3 Interprétation .....	106
Conclusion sur l'étude entreprise .....	107
 <b>Conclusion générale</b> .....	 109
 <b>BIBLIOGRAPHIE</b> .....	 111
<b>ANNEXE 1</b> .....	115
<b>ANNEXE 2</b> .....	117

# INTRODUCTION



La voix humaine véhicule, outre son aspect sémantique, des informations sur l'état psychologique et physiologique du locuteur. L'action sur l'être humain d'agents agressifs, organiques ou psychiques, est nommée "charge de travail" [1], [2]. Cette dernière est susceptible de causer des désordres physiologiques et psychologiques tels que la fatigue physique et le stress qui en sont les éléments pathologiques.

Les recherches pour détecter et mesurer les modifications des caractéristiques acoustiques de la voix dues à l'influence d'une charge de travail ont été financées par les différents centres spatiaux existant dans le monde. En effet, les résultats de ces études profitent essentiellement aux astronautes et aux pilotes car ils doivent effectuer des tâches spécifiques très stressantes.

Les charges de travail étaient d'abord des émotions, réelles ou simulées, telles que la surprise, la colère, l'ennui, la tristesse, la peur, le dégoût, le plaisir ou la joie. Le stress émotionnel des cosmonautes dans l'espace par l'analyse des communications vocales a aussi été l'objet de travaux d'études.

Ces derniers ont commencé dès les années quarante [1], [2]. Au début, ils consistaient essentiellement à simuler des états émotionnels. De 1965 à 1975, les études ont porté sur l'influence d'autres charges de travail telles que des messages radio air-sol, des tâches arithmétiques ou des stimulations à la lumière électrique ou ambiante. On a pu développer un évaluateur de stress psychologique nommé PSE (Psychological Stress Evaluator).

De 1975 à 1980 les chercheurs ont utilisé le PSE pour étudier les effets de différents types de charges de travail et pour tester les performances du détecteur de mensonges. Depuis les années quatre-vingt, et bien que l'on ait fait varier les charges de travail psychomotrices, les recherches n'ont pas mis en évidence de nouvelles variations de paramètres acoustiques.



Les techniques médicales existant aujourd'hui et les travaux faits sur la fatigue musculaire permettent de prouver l'existence biologique de la fatigue d'un muscle. Mais, même si l'on a remarqué que les performances des systèmes de reconnaissance de la parole baissent lors d'une période étendue de parole continue [3], nous n'avons pas trouvé de travaux publiés sur la fatigue vocale ou sur l'identification des indicateurs acoustiques de la voix sensibles à une telle charge de travail.

Pourtant la fatigue est invalidante pour toute personne. Elle entraîne une baisse dans l'efficacité des tâches effectuées, un relâchement de l'attention ou de la concentration, une mauvaise appréciation du danger de la portée de certains actes et une baisse de la rapidité des réflexes pour sortir d'une situation dangereuse. Ce qui peut entraîner de graves accidents pour certains métiers où l'on met des vies humaines entre les mains d'un ou plusieurs opérateurs humains.

Notre travail se veut une contribution à la recherche dans le domaine de façon à détecter la fatigue sans analyses médicales. En effet, la détection de la fatigue par l'analyse vocale a l'avantage d'être plus rapide et de pouvoir se faire à distance. De plus, l'analyse se fait sans aucun désagrément pour le locuteur.

Nous analysons la voix pour détecter la fatigue qu'elle soit vocale, c'est-à-dire due au travail de muscles entrant en jeu lors de la phonation, ou qu'elle résulte du travail d'autres muscles. Pour cela, nous travaillons sur des locuteurs susceptibles de présenter une fatigue physique pour certains, et ce après une activité normale d'une journée, et vocale pour les autres. Parmi les premiers, que nous avons enregistrés trois fois tout au long d'une journée, certains avaient participé à une rencontre sportive. Pour la seconde catégorie, nous avons travaillé sur les voix d'animateurs radio et d'acteurs. Ces derniers ont été enregistrés avant et après une émission radiophonique. L'étude s'est faite sans variations émotionnelles réelles ou simulées de la part des locuteurs.

Cet enregistrement nous a permis de créer un corpus assez important de façon à générer une base de données. Ce dernier est composé de phrases en arabe et en français. Nous avons ensuite analysé ces données par différentes méthodes existant en traitement de la

parole [7-10]. Puis, nous avons comparé les résultats obtenus, sans et sous l'influence présumée de la fatigue, de façon à pouvoir détecter les éventuelles modifications des caractéristiques acoustiques de la voix pour de tels locuteurs.

Le chapitre premier de ce travail introduit des notions fondamentales sur la parole et la fatigue. Il présente les mécanismes de la phonation. Les différents phonèmes des langues arabe et française y sont définis et classés. Puis, nous abordons brièvement les causes de fatigue musculaire et les techniques classiques de détection de celle-ci. Nous terminons en exposant les causes de fatigue purement vocale.

Le second chapitre est consacré aux différentes méthodes utilisées pour l'analyse de la parole. Nous y étudions quelques unes des nombreuses méthodes de détection de la fréquence fondamentale.

Le chapitre trois traite de la délicatesse du traitement de la parole. Des variations prosodiques à la coarticulation en passant par l'influence des émotions et des désordres pathologiques, les principales causes de variabilité de celle-ci sont passées en revue. Il présente aussi les critères retenus pour déterminer les traits invariants chez un locuteur.

En chapitre quatre nous présentons les moyens matériels et humains utilisés pour mener à bien notre travail. Le corpus et les différents types de locuteurs retenus y sont exposés.

Puis nous terminons en chapitre cinq en donnant les résultats obtenus avec les interprétations, les conclusions auxquelles nous aboutissons et enfin les perspectives d'amélioration dans le traitement du sujet.

# CHAPITRE 1

## NOTIONS FONDAMENTALES SUR LA PAROLE ET LA FATIGUE

### Introduction

Nous commençons, en première partie de ce chapitre, par présenter les processus de production de la parole. En seconde partie, nous exposons les diverses causes de la fatigue et les transformations métaboliques apparaissant dans le corps humain lors de l'exercice musculaire. Nous donnons ensuite les principales causes de fatigue vocale.

### 1.1 Production du signal de parole

#### 1.1.1 Eléments constitutifs de l'appareil phonatoire

L'appareil phonatoire (Fig. 1.1) comprend l'appareil respiratoire qui est la source d'énergie, le vibrateur laryngé portant les cordes vocales qui sont des muscles élastiques et le pavillon pharyngo-bucco-nasal faisant office de résonateur [4-12].

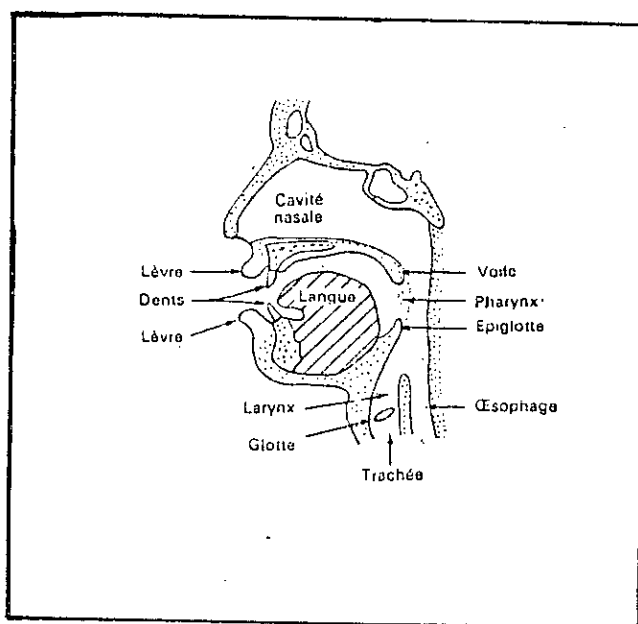


Fig. 1.1- Appareil phonatoire humain (Guibert [10])

## 1.2 Concepts fondamentaux de physico-acoustique et de phonétique

Nous donnons ici la définition des principaux termes utilisés et les notions de base en parole qui sont nécessaires à la compréhension de la suite du travail.

### 1.2.1 La fréquence fondamentale

La vibration, qui est en fait l'accolement puis la séparation, des cordes vocales portées par le larynx détermine la fréquence fondamentale appelée encore pitch ou F0. Elle est comprise entre 75 et 150 Hz chez les hommes, 150 et 300 Hz chez les femmes, et est supérieure ou égale à 300 Hz chez les enfants [4-12].

### 1.2.2 La mélodie et la hauteur

La mélodie de la voix est caractérisée par les fluctuations de F0 en fonction du temps [4-12]. La hauteur est généralement caractérisée par une fréquence donnée de la fréquence fondamentale [7]. Cette dernière varie du grave à l'aigu.

### 1.2.3 Le timbre

Le timbre est caractérisé par la richesse en amplitudes relatives des harmoniques du pitch. En effet, les cavités du résonateur de l'appareil phonatoire ont pour propriété de renforcer certaines harmoniques du son fondamental laryngé, en atténuant les autres [4-12].

### 1.2.4 La prosodie

La prosodie introduit dans la prononciation d'une phrase des nuances qui, dans la langue écrite, demanderaient des ponctuations ou des énoncés différents. Ce sont les caractéristiques prosodiques qui permettent à un auditeur de suivre une conversation même en milieu défavorable. Les principaux paramètres prosodiques sont l'intonation, l'intensité et la durée [7-12].

- L'intonation est un paramètre très important. Elle correspond à une hauteur donnée de  $F_0$  et l'oreille est très sensible à ses variations. Ainsi la phrase "tu viens" a des sens très différents selon l'intonation du locuteur.
- L'intensité donne des informations sur l'amplitude de la voix. Celle-ci pourra être normale, chuchotée ou criée.
- La durée fixe le rythme de la phrase.

### 1.2.5 Le phonème

La succession dans le temps d'unités individualisables constitue la parole. Le phonème est défini comme le plus petit élément acoustique ayant une pertinence de sens dans une langue donnée [6-12].

Les phonèmes emphatiques imposent une forte tension aux différents organes articulaires. Ils résultent d'un resserrement au niveau pharyngal et d'une fermeture des maxillaires. Cela entraîne un report en arrière de la racine de la langue avec un abaissement et un creusement [13], [14].

### 1.2.6 La coarticulation

La coarticulation est définie par l'influence qu'exerce un son sur celui contigu [7-10], [17]. Cela entraîne une modification contextuelle des points d'articulation. Celle-ci résulte d'un effet d'inertie mécanique et d'une réorganisation du geste articulaire en vue d'une minimisation de l'effort articulaire. Son effet est d'autant plus marqué que le débit est rapide [7-10], [17].

### 1.2.7 Représentation mentale des sons

La formation volontaire d'une image auditive [12] est nécessaire au début de toute production phonique. Cette image comprend non seulement la hauteur, mais encore le timbre du son à émettre. La représentation mentale de la hauteur d'un son suffit à déclencher des actes moteurs tels que la fermeture de la glotte et la descente du larynx. Celle du timbre déclenche

une série de mouvements d'accommodation du résonateur pharyngo-buccal utilisant des coordinations acquises au cours de l'enfance lors de l'apprentissage vocal.

### 1.3 Génération d'un signal de parole

La parole résulte de l'action volontaire et coordonnée des appareils respiratoires et masticatoires [4-12]. L'appareil respiratoire fournit l'énergie nécessaire lorsque l'air des poumons est expiré par la trachée artère. Au sommet de celle-ci les cordes vocales modulent la pression de l'air. Ce signal issu de la source vocale est modifié lors de la traversée du conduit vocal. Ce dernier est un ensemble de cavités situées entre la glotte et les lèvres. Il contient la cavité pharyngienne, la cavité buccale et, en dérivation, la cavité nasale.

C'est la soufflerie pulmonaire qui fixe la force, l'intensité et la durée de l'émission sonore. Le vibreur laryngé fixe l'intonation. Les cavités de résonance, quant à elles, fixent la qualité et le timbre du son [4-12].

La position des mâchoires, des lèvres, de la luette et de la langue va déterminer les résonances caractéristiques du son prononcé (Fig 1.2 et Fig 1.3) qui se traduisent par une concentration de la puissance sonore dans des zones de fréquences appelées formants. C'est la position des trois premiers formants qui caractérise le timbre vocalique [6-12].

- F1 naît dans la cavité résonante comprise entre le larynx et le dos de la langue.
- F2 naît dans la cavité résonante située entre le dos de la langue et les lèvres.
- F3 dépend de l'aperture des lèvres.

Les formants d'ordre supérieur donnent des indications sur l'état du locuteur [10].

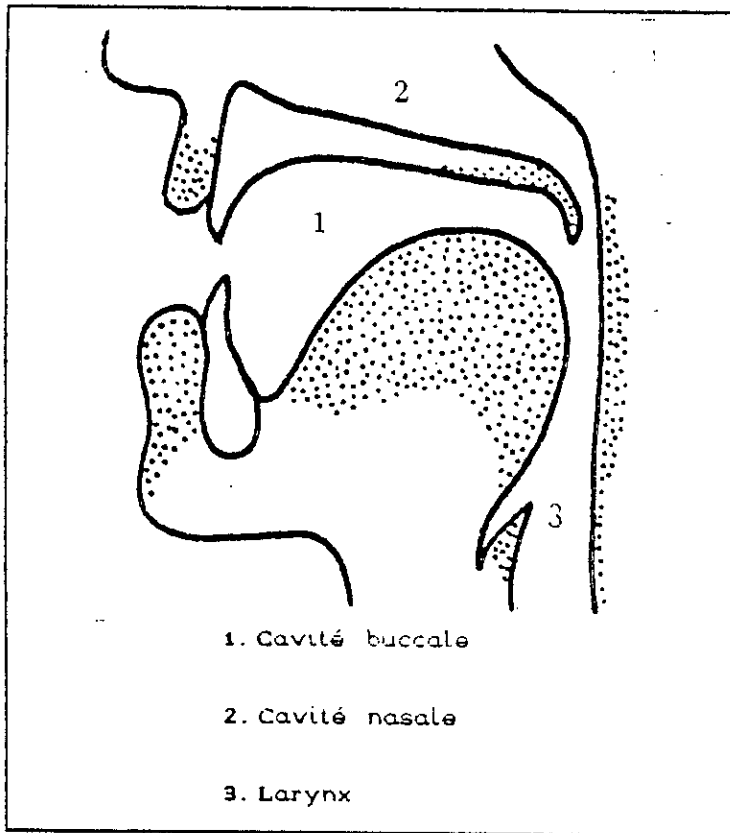


Fig.1.2- Parties mobiles de la bouche et du pharynx ( en pointillé ) permettant la structuration résonnante de ces cavités ( Dejonckere [5] )

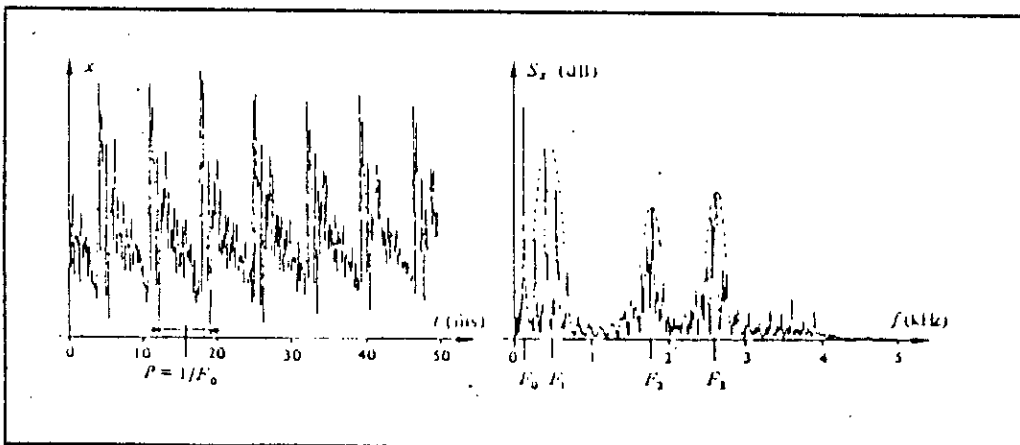


Fig.1.3- Un signal vocal voisé et son spectre ( Kunt [9] )

## 1.4 Fonctionnement acoustique de l'appareil vocal

La parole résulte de l'excitation du conduit vocal par deux types de sources [6-12] :

- la vibration des cordes vocales qui produit les sons voisés ou sonores. Dans ce cas le son émis est périodique. C'est le cas pour toutes les voyelles ;
- une source de bruit qui se crée en un point de resserrement du conduit vocal. Le son émis sera apériodique. Les sons non voisés ou sourds se rencontrent uniquement parmi les consonnes.

La combinaison des deux types de sources donne les consonnes voisées.

Les différents phonèmes que nous exposons dans la suite du travail sont transcrits dans le code de l'Alphabet Phonétique International (API).

### 1.3.1 Les voyelles

Les voyelles sont caractérisées par le degré d'aperture (espace compris entre le haut de la langue et le palais), le lieu d'articulation, la forme des lèvres et la nasalité. Elles sont aussi caractérisées par un passage libre de l'air et par des formants intenses.

#### 1.3.1.1 Les voyelles du français

Il existe deux classes de voyelles : orales et nasales (tableau 1.1). Ces dernières sont souvent représentées dans le plan  $F_1, F_2$ . En effet, les deux premiers formants suffisent à déterminer le timbre et à localiser la majorité des voyelles [7-10]. Mais pour certaines, les antérieures comme /i/ et /e/, le recouvrement entre zones formantiques impose le recours au troisième formant [7].



Voyelles	Code	Exemple
Orales	/i/	lit
	/e/	les
	/ɛ/	lait
	/a/	là
	/u/	loup
	/y/	lu
	/o/	lot
	/ɔ/	lotte
	/ø/	leur
	/ø̃/	leu
	/ɔ̃/	le
Nasales	~ a	lent
	~ o	long
	~ e	lin

Tableau 1.1 - Voyelles du français (Calliope [7]).

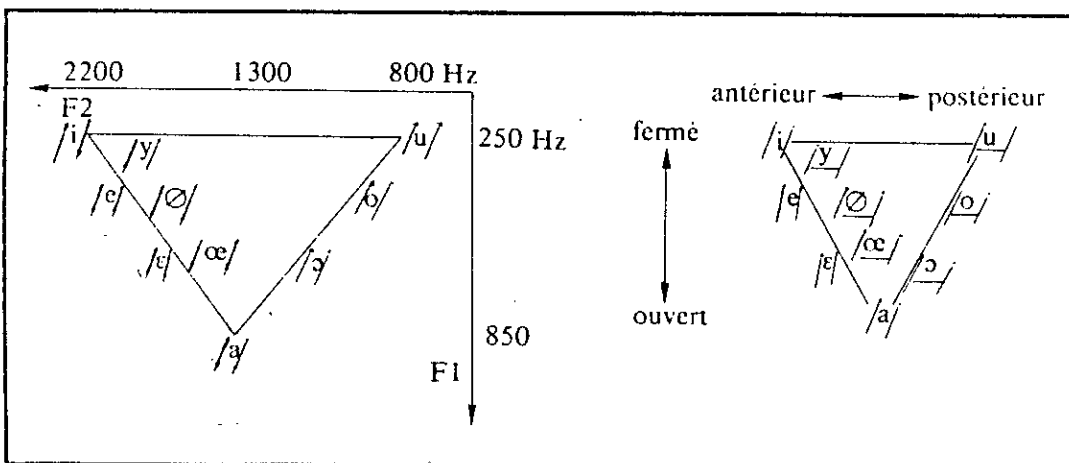


Fig.1.4- Triangle vocalique et articulatoire des voyelles orales du français ( Calliope [7]).

Le triangle vocalique et articulatoire (fig.1.4) représente la position de la langue dans la cavité buccale. Une voyelle est antérieure quand la langue est massée près de la zone dentale et

postérieure vers le voile du palais. De même, la voyelle est ouverte quand la langue est basse et étalée loin du palais [7].

Une augmentation de  $F_1$  est donc le résultat d'une ouverture articuloire et une augmentation de  $F_2$ , une antériorisation de l'articulation [7].

### 1.3.1.2 Les voyelles de l'arabe

L'arabe standard se compose de trois voyelles brèves /a/, /u/ et /i/ et de leurs correspondantes longues /aa/, /uu/ et /ii/. On note généralement V la voyelle brève et VV la longue. La figure 1.5 représente ces voyelles dans le plan  $F_1, F_2$ . Il faut noter qu'au contact d'une consonne emphatique, la disposition spectrale des voyelles change [13].

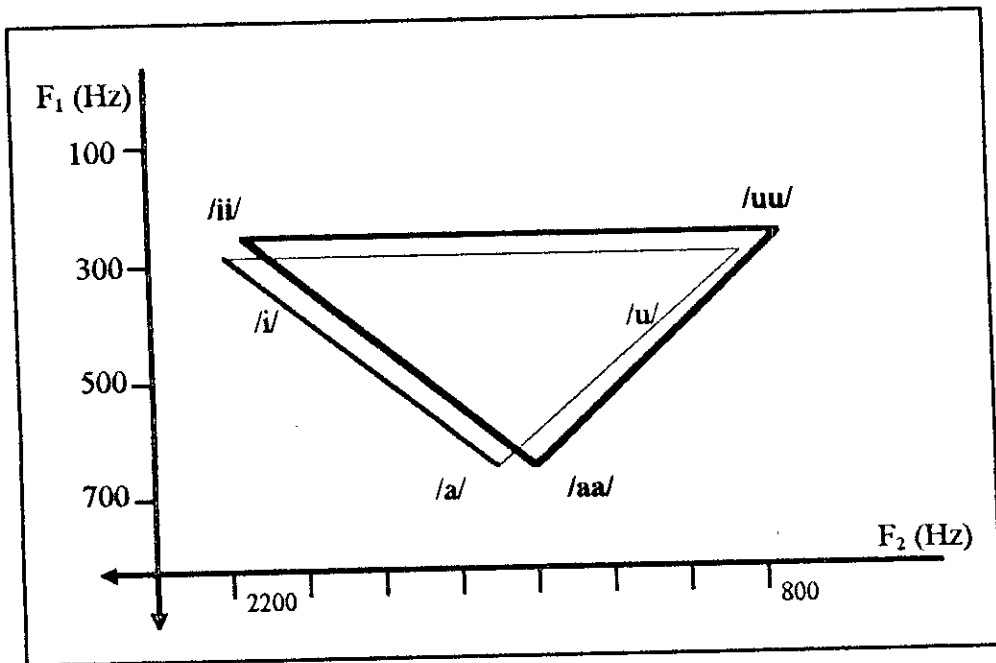


Fig.1.5 - Triangle vocaliques des voyelles arabes (Al ani [13]).

## 1.3.2 Les consonnes

Les consonnes sont classées selon le voisement, l'organe permettant l'articulation, le lieu d'articulation qui désigne la zone du conduit vocal participant à la formation du son et le mode d'articulation [6-10].

### 1.3.2.1 Mode d'articulation

Il existe différents modes articulatoires [6-10] pour la production des phonèmes :

- les occlusives ou plosives résultent d'une obstruction totale, en un point du conduit vocal, et de brève durée, suivie d'une ouverture brutale. Elles peuvent être sonores comme /b/, /d/ ou /g/ et sourdes comme /p/, /t/ ou /k/ ;

- les constrictives ou fricatives sont produites lors du passage de l'air pulmonaire par un rétrécissement en un point du conduit vocal. Elles peuvent être sonores comme /v/, /z/ et sourdes comme /f/, /s/ ;

- les liquides appelées aussi latérales, telle /l/, sont le résultat d'une obstruction partielle du conduit vocal qui fait que l'air peut passer de part et d'autre de celui-ci ;

- les vibrantes, telle /r/, naissent d'une vibration de la langue ;

- les nasales, comme /m/, /n/, résultent de l'action de la cavité nasale pendant la production du son ;

- Il existe aussi un mode particulier produisant des semi-voyelles. Celles-ci sont voisées et présentent une affinité avec les voyelles. Ce sont /w/, /y/ ou /j/ ;

Les vibrantes, nasales, latérales et semi-voyelles sont aussi appelées consonnes sonnantes. Car, comme les voyelles, elles possèdent une structure formantique. Mais, au contact des consonnes sourdes, elles perdent leur voisement.

### 1.3.2.1 Les consonnes du français

Le tableau 1.2 donne la classification, le mode et le lieu d'articulation des différents phonèmes. Ces derniers sont aussi classés selon leur voisement. Ils sont transcrits en alphabet API.

Modes d'articulation	Lieu d'articulation		
	Labiales	Dentales	Vélo-palatales
Occlusives NV* V*	/p/ /b/	/t/ /d/	/k/ /g/ <sup>1</sup>
Fricatives NV V	/f/ /v/	/s/ /z/	/ʃ/ <sup>2</sup> /ʒ/ <sup>3</sup>
Nasales	/m/	/n/	/ŋ/ <sup>4</sup>
Semi-voyelles	/w/ <sup>5</sup>	/y/ <sup>6</sup>	/j/ <sup>7</sup>
Liquide		/l/	
Vibrante			/R/

NV\* : non voisée

V\* : voisée

Tableau 1.2 - Classification des consonnes du français (Calliope [7]).

1 : gai

2 : chez

3 : geai

4 : gagner

5 : ouais

6 : huer

7 : yéyé

### 1.3.2.2 Les consonnes de l'arabe standard

L'arabe standard est la langue des médias et de l'enseignement. Elle a la particularité de posséder des consonnes emphatiques, pharyngales et laryngales en plus des modes articulatoires communs à d'autres langues (Fig 1.6). Le tableau 1.3 donne la transcription de l'alphabet arabe en son équivalent API. Les phonèmes propres à la langue arabe sont :

- la plosive /q/. Cette dernière est une uvulaire sourde ;
- les fricatives /r/ qui est une alvéolopalatale sonore, /q/ une interdentale sourde, /k/ une vélaire sourde et /g/ une uvulaire sonore ;
- des glottales sourdes comme /h/ fricative et /ʔ/ plosive ;
- des pharyngales fricatives comme la sourde /H/ et la voisée /ε/.
- Il existe quatre phonèmes définis emphatiques, deux fricatives comme la sourde /S/ et la sonore /Z/ et deux plosives comme la sourde /T/ et la sonore /D/ [13]. Mais d'autres phonèmes comme /l/ et /r/ s'emphatisent dans un contexte particulier [13].

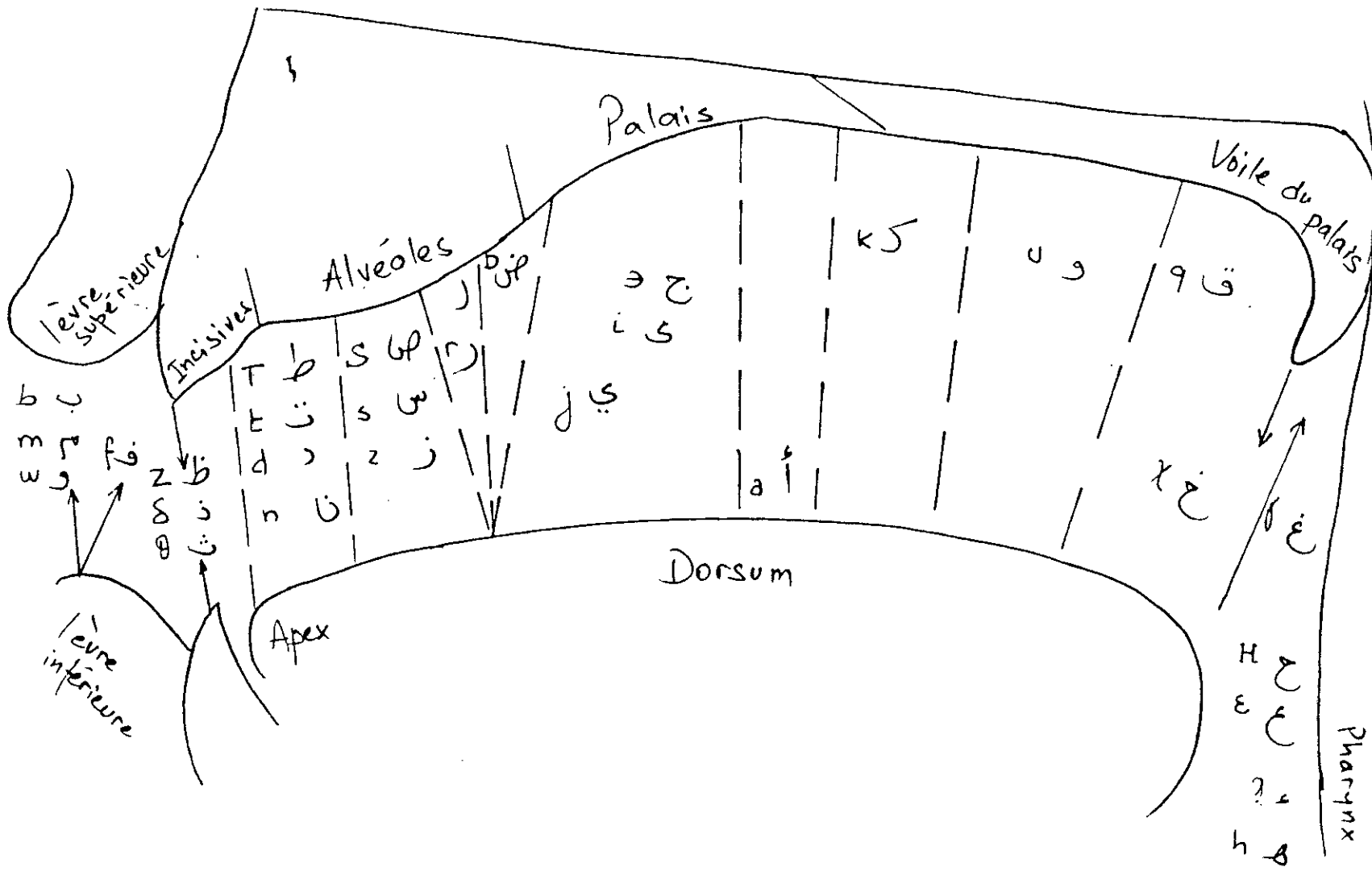


Fig. 1.6 - Lieu et le mode d'articulation des phonèmes arabes [14].

Alphabet Arabe	Alphabet Phonétique International
ء	ʔ
ب	b
ت	t
ث	θ
ج	ʒ
ح	h
خ	χ
د	d
ذ	ð
ر	r
ز	z
س	s
ش	ʃ
ص	s
ض	d
ط	t
ظ	z
ع	e
غ	ɣ
ف	f
ق	q
ك	k
ل	l
م	m
ن	n
هـ	h
و	w
ي	j

Tableau 1.3 - Transcription phonétique de l'arabe standard (Moussa [15]).

### 1.7.1 Processus anaérobie

La seule source d'énergie immédiatement utilisable par le muscle vient d'une molécule particulière : l'Adénosine Tri-Phosphate (ATP). Sans ATP la contraction musculaire - et donc le mouvement - ne peut se faire.

Dans le muscle la concentration d'ATP est de 5 mMol/Kg. Cette faible réserve d'énergie est rapidement consommée. Elle est reconstituée au fur et à mesure grâce à la dégradation réversible d'un autre constituant chimique contenu lui aussi dans le muscle : la créatine-phosphate (CP).

Cependant la concentration du tissu musculaire en CP est elle aussi assez faible (15-20 mMol/Kg) et s'épuise rapidement. Le réapprovisionnement en ATP se fait en le synthétisant grâce à la glycogénolyse [20]. Celle-ci représente la première chaîne de réaction utilisant le glucose et en particulier le glycogène qui est sa forme de réserve. On trouve le glycogène dans le foie (200 g environ), dans le muscle (14 g/Kg) et dans le sang. La dégradation du glycogène en ATP produit aussi de l'acide pyruvique. Celui-ci se dégrade en acide lactique en l'absence d'oxygène, mais est générateur d'énergie en sa présence.

### 1.7.2 Processus aérobie

Si la fourniture en oxygène aux fibres musculaires est adéquate, des processus aérobies reconstituent l'ATP nécessaire au muscle. Ceux-ci se développent à l'intérieur de petites cellules musculaires - les mitochondries- par une chaîne de réaction appelée cycle de Krebs [20], et ce à partir des protéines et des acides gras. Ce processus de formation d'énergie à partir de substrats autres que le glycogène s'appelle la néoglucogénèse.

Ces différents processus de libération de l'énergie sont résumés en figure 1.7.



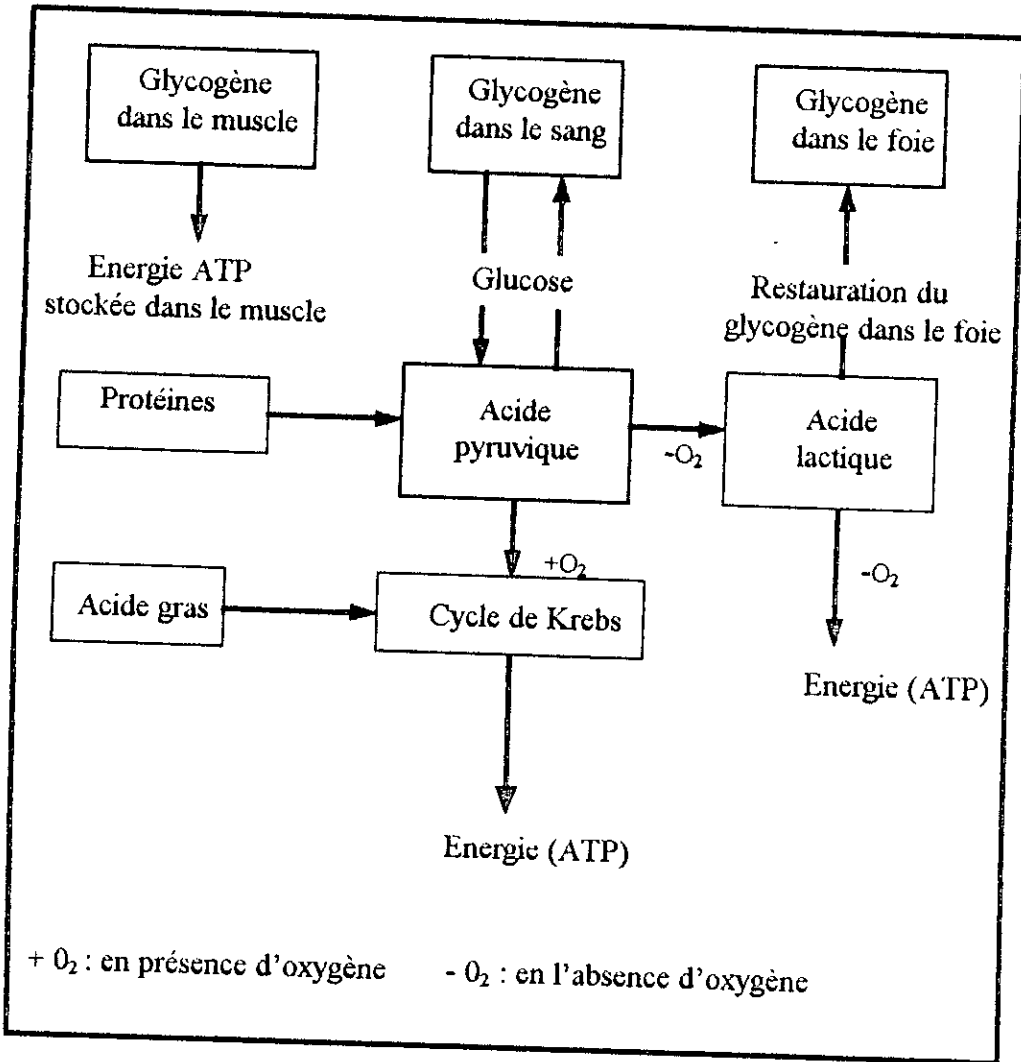


Fig.1.7 - Libération et transport de l'énergie dans la cellule musculaire

## 1.8 Principaux facteurs de fatigue

La cause de la fatigue varie avec la nature de l'activité, l'entraînement, l'état physiologique de l'individu et les conditions d'environnement [19], [20]. Elle peut être due spécifiquement, comme nous le verrons dans ce qui suit, à la déplétion en métabolites-clés dans le muscle, ou à l'accumulation d'autres métabolites qui vont être déversés dans la circulation sanguine et affecter l'ensemble du corps.

## **1.8.1 Facteurs externes**

### **1.8.1.1 Influence de l'entraînement**

La fatigue apparaît plus tôt chez une personne non entraînée à l'exercice que pour un individu entraîné en endurance [19], [20].

### **1.8.1.2 Influence de l'environnement**

Des facteurs environnementaux tels que des conditions thermiques et hygrométriques élevées peuvent entraîner une série de réactions physiologiques qui vont affecter l'endurance et altérer la performance.

Pendant l'exercice en environnement chaud, les pertes sudorales cutanées augmentent et la température corporelle est accrue. Les diverses réactions que cela entraîne représentent des irritants directs du système nerveux central ce qui peut affecter en plus la perception subjective du sujet en ce qui concerne l'exercice [20].

## **1.8.2 Déplétion en métabolites**

### **1.8.2.1 Déplétion en sources d'énergie musculaire**

Des études détaillées sur des muscles isolés stimulés pour se contracter montrent que le développement de la tension est lié au taux de CP [20]. Ainsi, la déplétion en CP amène à la fatigue musculaire. Mais au point de fatigue, à la fois la CP et l'ATP sont déplétés.

### **1.8.2.2 Déplétion en glycogène musculaire**

La déplétion en glycogène dans le muscle squelettique est associée à la fatigue pendant l'exercice prolongé. Quand l'exercice est modéré, le glycogène est dépleté pour tous les types de fibres. Mais quand l'exercice est plus difficile, il y a un recrutement de fibres sélectif. Il est possible de travailler jusqu'à la fatigue à cause de la déplétion en glycogène des fibres musculaires spécifiques, tandis que du glycogène demeure dans les fibres adjacentes [19], [20].

### **1.8.2.3 Déplétion en glucose sanguin ou glycémie**

Pendant des exercices courts et intenses, la glycémie s'élève au dessus du niveau pré-exercice du fait de la stimulation de la glycogénolyse par le système nerveux. L'aptitude du foie à maintenir une libération de glucose soutenue (élevée et durable) est limitée par la quantité de glycogène stockée et par les activités des enzymes hépatiques de la glycogénolyse (dégradation du glycogène en glucose) et de la néoglucogenèse (formation de glucose à partir d'autres substrats).

Quand l'exercice se prolonge, la production de glucose peut se limiter à la néoglucogenèse du fait de l'épuisement des réserves hépatiques de glycogène. La production de glucose peut alors tomber au-dessous de ce qui est nécessaire pour le muscle au travail et les besoins des tissus essentiels comme le cerveau. Lorsque la glycémie s'effondre, l'exercice devient subjectivement plus difficile à cause du "jeûne" du cerveau et de la difficulté d'oxyder les graisses dans le muscle en l'absence de substrats [19], [20].

### **1.8.2.4 Déplétion en oxygène**

Comme la majeure partie de l'ATP nécessaire pour réaliser une activité quelconque de 90 s ou plus doit provenir de la respiration, la fourniture adaptée d'oxygène apparaît essentielle pour assurer le travail [19], [20].

Les effets de l'apport ou de l'utilisation inadaptée d'oxygène peuvent être représentés par une production accrue de lactate ou une baisse du taux de CP : tous deux sont susceptibles de causer la fatigue [20].

### **1.8.3 Accumulation de métabolites**

#### **1.8.3.1 L'acide lactique**

Durant l'exercice court d'intensité élevée, le lactate s'accumule du fait d'une production plus élevée que sa disparition [19], [20]. Il produit par dissociation un proton H qui abaisse le PH.

Les effets négatifs de la production de H sont multiples. Dans le muscle, un PH abaissé ralentit la glycolyse. De plus cet ion H agit sur l'ion calcium Ca interférant ainsi avec la contraction musculaire. Enfin, il stimule les récepteurs de la douleur. L'ion H libéré dans le sang et entrant en contact avec le cerveau peut donner des effets secondaires sévères, incluant la douleur, les nausées et la désorientation. Dans le sang même, il peut inhiber la combinaison d'oxygène avec l'hémoglobine au niveau des poumons [20].

#### **1.8.3.2 L'ion calcium**

L'accumulation de l'ion calcium pendant l'exercice prolongé semble plus débilant que la baisse du PH . Il entraîne une plus grande consommation d'oxygène et il diminue l'énergie potentielle qui permet la formation de l'ATP [20].

## 1.9 Fatigue vocale

Une soixantaine de muscles participe à la phonation de manière coordonnée [5], [8], [12]. La fatigue résulte de la contraction musculaire répétée [5], [19] et [20]. Tous les muscles entrant en jeu lors de la phonation (nous voyons les plus importants en Fig 1.8 à 1.11) subiront donc des variations métaboliques lorsqu'une fatigue vocale existera.

Il existe plusieurs facteurs pouvant induire une fatigue vocale ; les plus importants sont:

- un discours prolongé ou un cours didactique ou même une prestation vocale chantée entraîne un effort vocal important. Une mauvaise acoustique l'augmente [5], [12] ;
- la prononciation mentale d'un texte, nommée parole intérieure par Garde [12], entraîne une fatigue vocale. Cette dernière, ressentie par certains mimes, est plus nerveuse que musculaire puisqu'il n'y a pas de parole réellement prononcée ;
- les conversations pour soutenir un bruit de fond de plus ou moins 80 dB provoquent une augmentation importante du niveau sonore de la voix dont les locuteurs n'ont pas conscience. Cela entraîne un effort vocal important; c'est l'effet de Lombard-Tarneau [5], [12].

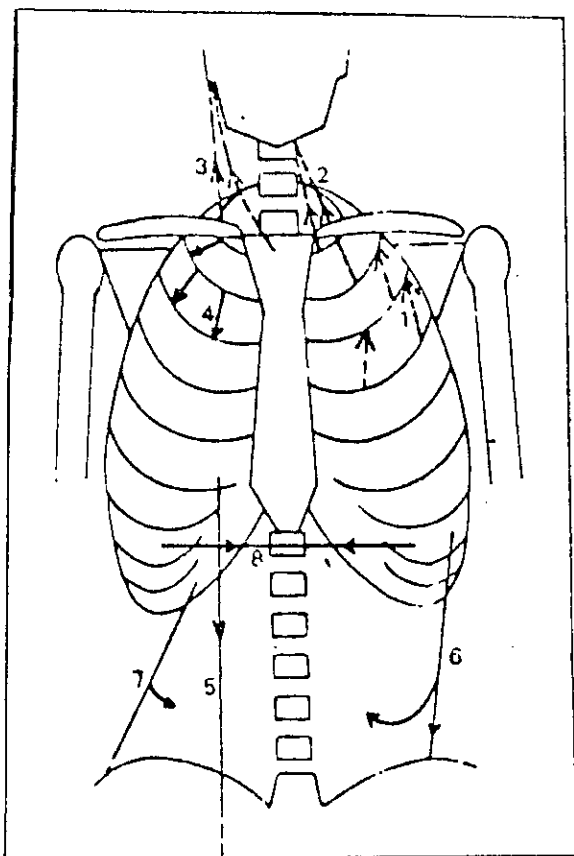


Fig.1.8 - Muscles respiratoires  
 inspiration : 1, 2, 3 ; phonation : 1, 4 à 8

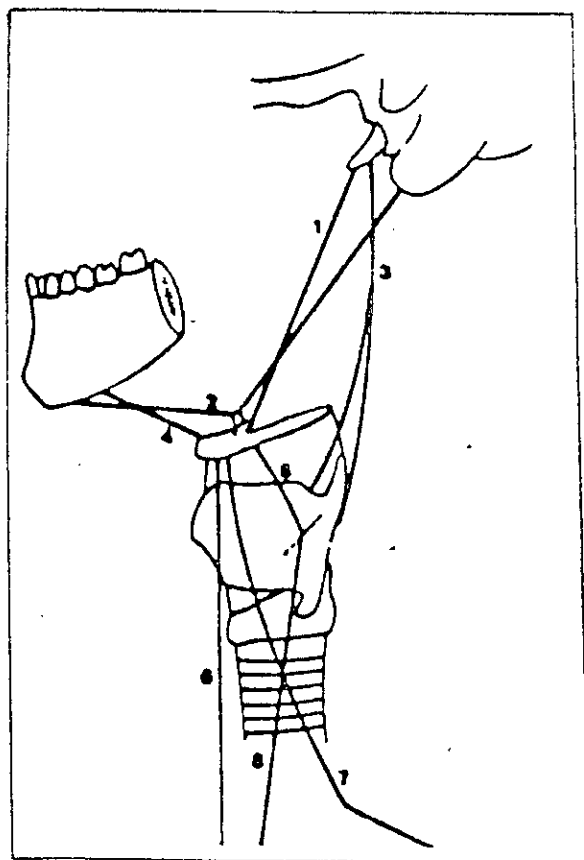


Fig.1.9 - Système musculaire du larynx  
 Élévateurs : 1 à 4 ; Abaisseurs : 5 à 8

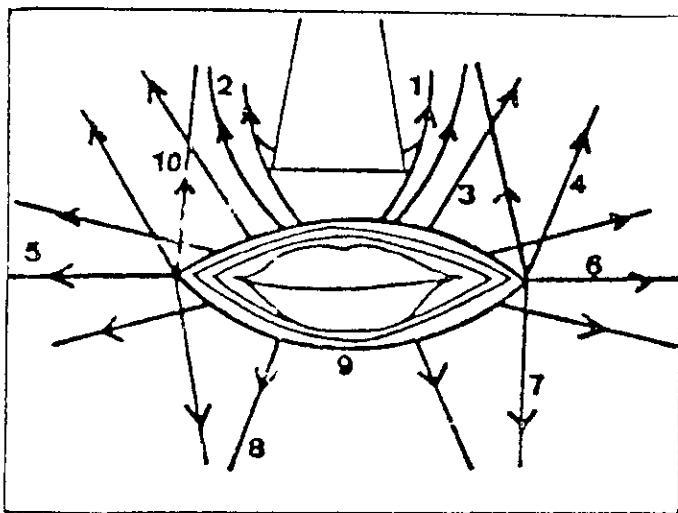


Fig.1.10 - Muscles des lèvres  
 Releveurs : 1 à 8 et 10 ; Constricteur : 9

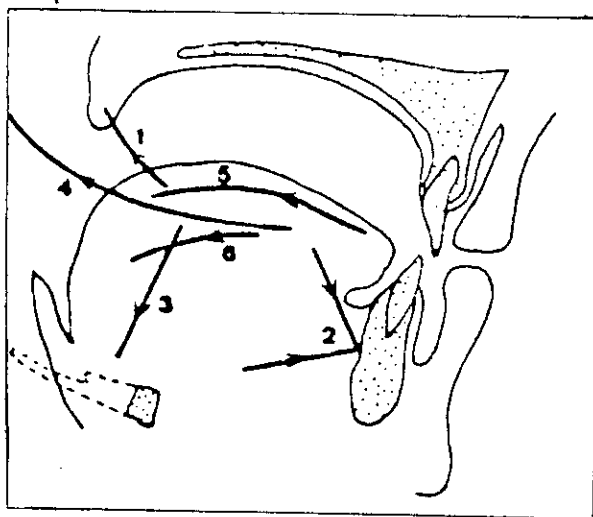


Fig.1.11 - Muscles de la langue

## Conclusion

Ainsi que nous le voyons dans ce chapitre, l'apport de l'énergie nécessaire au mouvement se fait par le repos ou par la consommation d'aliments. Leur absence entraîne la fatigue. Les effets de cette dernière sont assez sévères. En effet, elle inhibe la formation d'énergie nécessaire à la contraction musculaire qui se fait mal jusqu'à disparaître. De plus, elle stimule les récepteurs de la douleur entraînant des nausées et une désorientation.

Nous voyons aussi que la fatigue vocale peut être due au travail des muscles entrant en jeu lors de la phonation et à une mauvaise hygiène vocale ou encore, par induction, au travail d'autres muscles du corps. En effet, les transformations métaboliques conduisant à la fatigue qui apparaissent au niveau d'un muscle affectent finalement l'ensemble du corps par le biais de la circulation sanguine.

Les transformations, qu'entraîne la fatigue sur le métabolisme, sont cernées grâce aux analyses médicales. Nous allons voir - en analysant la parole - si au niveau de la voix des transformations apparaissent aussi. La recherche sera plus ardue car le signal de parole est par nature extrêmement variable.

Pour mener à bien notre travail, nous utiliserons certaines méthodes de traitement de signal appliquées à la parole que nous développons au chapitre suivant.

# CHAPITRE 2

## TECHNIQUES DE TRAITEMENT DE LA PAROLE

### Introduction

Dans ce chapitre, nous décrivons brièvement les différentes méthodes d'analyse existant en traitement de la parole, et nous insistons sur celles que nous avons finalement retenues pour notre étude.

### 2.1 Méthodes d'analyse en traitement de la parole

Avant de pouvoir analyser numériquement un signal de parole, il faut d'abord passer par une étape de prétraitement [9], [21-24] (annexe 1). Les méthodes d'analyse que nous allons voir sont spectrales. Deux de celles-ci se basent sur une connaissance des mécanismes de production. Il s'agit du cepstre et du codage prédictif linéaire ou LPC qui permettent de séparer la source d'excitation du conduit vocal. Ce qui facilite l'extraction de  $F_0$ . Il faut toutefois préciser que la méthode LPC est à la fois temporelle et spectrale.

#### 2.1.1 Le spectrographe

Le spectrographe ou sonagraphe permet d'analyser un signal de parole en faisant sa décomposition spectrale. Il donne une représentation acoustique, en trois dimensions, des sons [6-11] : dans le plan fréquence-temps et la troisième dimension, qui est l'amplitude ou l'énergie du signal vocal, est donnée par des traces plus ou moins foncées.

Selon les mesures que l'on désire effectuer, on variera la bande passante du filtre utilisé pour l'analyse.



- Avec un filtre à bandes étroites (45 Hz), les harmoniques du pitch apparaissent sous forme de raies horizontales espacées de  $F_0$  (fig.2.1.a) mais la résolution temporelle est faible. La valeur du fondamental est donnée par une lecture des intervalles des harmoniques. Ce type d'analyse est utilisé si l'on s'intéresse à l'intonation.

- Avec un filtre à bandes larges (300 Hz), la résolution temporelle est bonne mais on ne distingue plus les harmoniques du pitch. Les formants, par contre, sont bien apparents sous la forme de bandes noires horizontales (fig.2.1.b). Une lecture directe de leurs positions sur l'axe des fréquences donne leurs valeurs [7-10].

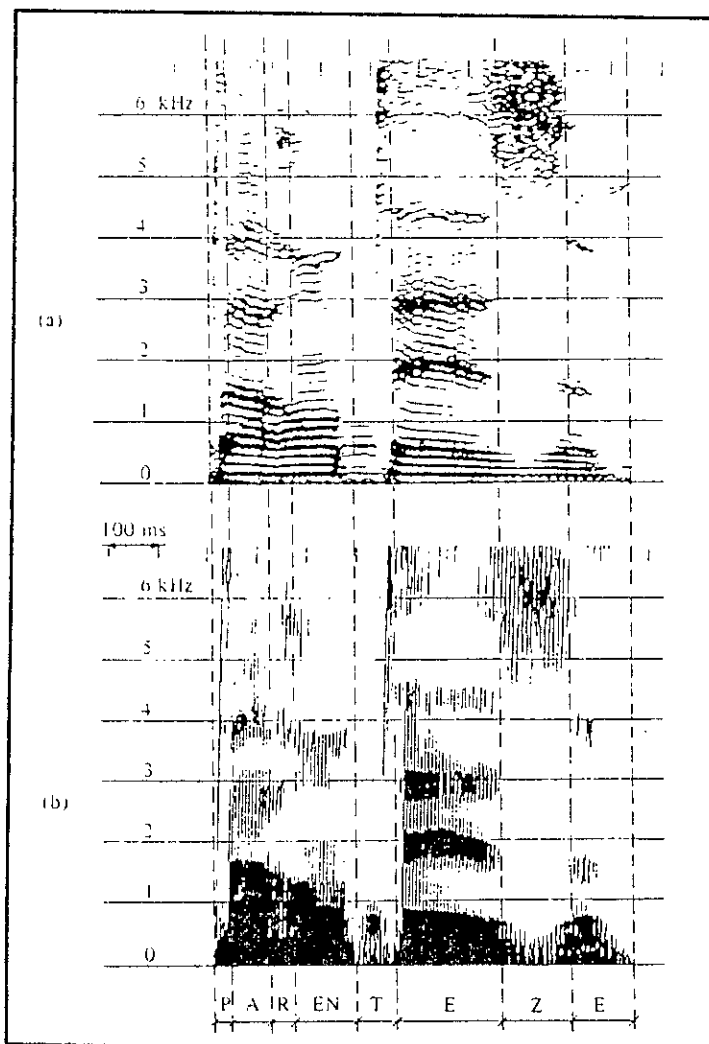


Fig.2.1- Sonagramme du mot parenthèse. (a) bande étroite. (b) large bande. (Kunt [9])

Les paramètres prosodiques sont déterminés par une lecture directe du sonagramme. Ainsi, l'intensité ou l'énergie est donnée par le degré de noirceur des taches donnant les formants. Elle se mesure en décibels.

Une lecture directe, sur l'axe des temps, donne la durée de prononciation d'un phonème, un mot ou une phrase. L'axe des temps nous permet aussi de voir l'évolution du signal spectral et de mesurer les durées et les phases d'existence des différents phonèmes. La hauteur et l'intonation de la fréquence fondamentale sont données par le spectrogramme en bandes étroites.

### 2.1.2 Le périodogramme

La méthode du périodogramme repose sur l'hypothèse de base que les données sont périodiques à l'extérieur de l'intervalle d'observation "N" [25-29]. L'estimée de la densité spectrale de puissance  $P_N(f)$ , qui est le périodogramme, est par définition la transformée de Fourier de l'estimateur biaisé de la fonction d'autocorrélation  $C_{xx}(m)$  :

$$P_N(f) = \sum_{m=-(N-1)}^{N-1} C_{xx}(m) \cdot e^{-j2\pi f m} \quad (2.1)$$

avec :

$$C_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x(n) \cdot x(n+m) \quad (2.2)$$

$|m| < N - 1$ ; n, m sont des variables entières.

Comme la transformée de Fourier d'une séquence réelle finie discrète  $x(n)$  est donnée par :

$$X_N(f) = \sum_{n=0}^{N-1} x(n) \cdot e^{-jn2\pi f} \quad (2.3)$$

Le périodogramme sera donc donné par le module au carré de la transformée de Fourier du signal :

$$P_N(f) = \frac{1}{N} \left| X_N(f) \right|^2 \quad (2.4)$$

### 2.1.3 L'analyse cepstrale

Le défaut majeur de la transformée de Fourier pour le calcul du spectre réside dans l'intermodulation entre la source et le conduit. Ce qui rend difficile la mesure des formants et du fondamental caractéristiques respectivement du conduit et de la source.

Le principe de l'analyse cepstrale est de réaliser leur séparation par déconvolution [7-10], [30]. L'hypothèse de base est que le spectre de la parole  $S(z)$  est le produit du spectre de la source  $G(z)$  par celui du conduit vocal  $B(z)$  :

$$S(z) = G(z).B(z) \quad (2.5)$$

On transforme, par homomorphie, l'expression du spectre en somme de deux termes en prenant son logarithme :

$$\text{Log} |S(z)| = \text{Log} |G(z)| + \text{Log} |B(z)| \quad (2.6)$$

Le cepstre -anagramme de spectre- est la transformée de Fourier inverse du logarithme du spectre du signal ainsi obtenu.

## 2.1.4 La prédiction linéaire et le modèle autoregressif

La méthode d'analyse par prédiction linéaire ou LPC repose sur l'hypothèse fondamentale selon laquelle un échantillon "S(n)" du signal de sortie, est une fonction linéaire des "p" échantillons qui le précèdent, et des "q+1" échantillons de l'excitation [7-10], [28-32].

Nous aurons donc :

$$S(n) = \sum_{i=1}^p a_i \cdot S(n-i) + G \cdot \sum_{k=0}^q b_k \cdot U_{n-k} \quad (2.7)$$

où  $a_i$ ,  $b_k$  et  $G$  sont les paramètres du système prédicteur. La transformée en  $Z$  de l'équation nous permet d'écrire :

$$S(Z) - S(Z) \sum_{i=1}^p a_i \cdot Z^{-i} = G \cdot U(Z) \sum_{k=0}^q b_k \cdot Z^{-k} \quad (2.8)$$

avec  $b_0 = 1$

D'où

$$H(Z) = S(Z) / U(Z) = G \cdot \sum_{k=0}^q b_k \cdot Z^{-k} \bigg/ 1 - \sum_{i=1}^p a_i \cdot Z^{-i} \quad (2.9)$$

$H(Z)$  est la fonction de transfert du système qui est assimilé à un filtre. C'est le modèle général appelé Auto-Regression-Moving-Average (ARMA).

- Si les  $b_k$  ( $k > 1$ ) sont nuls, la fonction de transfert ne possède que des pôles : c'est le modèle dit Auto-Régressif (AR).

Les algorithmes AR sont performants et donnent de hautes résolutions spectrales à l'application. La modélisation ARMA, bien que plus générale, exige la résolution d'un système non linéaire pour la détermination des paramètres. C'est le modèle AR - de par la simplicité de sa mise en oeuvre - qui est le plus utilisé pour l'analyse de la parole.

De plus, même si le signal de parole est du type ARMA (pour les sons nasalisés), on préfère souvent utiliser un modèle AR dont l'ordre est un peu surestimé [9].

#### 2.1.4.1. Modélisation du système phonatoire choisie

Une des principales caractéristiques du signal de parole est sa non-stationnarité. Lors de l'analyse par prédiction linéaire, le signal de parole est considéré comme stationnaire sous une fenêtre de courte durée. La fenêtre est généralement choisie entre 10 et 30 ms, car on tient compte de l'inertie du mécanisme physiologique de la phonation. Le modèle estimé est indépendant du temps à l'intérieur de cette fenêtre.

La prédiction linéaire permet d'extraire les paramètres les plus importants du signal de parole comme le voisement, la valeur du fondamental, les formants.

### 2.1.5 Méthodes Auto-Regressives

Pour un modèle AR, si les coefficients de prédiction "  $a_i$  " sont déterminés et les  $s(n-i)$  mesurés, l'équation (2.7) permet de prédire l'échantillon suivant par :

$$\hat{S}(n) = \sum_{i=1}^p a_i S_{n-i} \quad (2.10)$$

L'erreur commise par la prédiction est :

$$e(n) = S(n) - \hat{S}(n) \quad (2.11)$$

De l'équation (2.7) et (2.10) nous constatons que l'erreur de prédiction est une estimation de  $G \cdot U(n)$  :

$$G \cdot U(n) = S(n) - \sum_{i=1}^p a_i \cdot S(n-i) \quad (2.12)$$

Et donc :

$$H(Z) = G \left/ 1 - \sum_{i=1}^p a_i \cdot Z^{-i} \right. \quad (2.13)$$

Le choix du critère d'identification est un élément important pour la mesure de la ressemblance entre le système et le modèle. Pour déterminer les coefficients "  $a_i$  " nous utilisons le critère des moindres carrés. Cela revient à minimiser l'erreur quadratique totale "  $E$  " pour l'ensemble des échantillons de la séquence du signal de parole à analyser.

$$E = \sum_n e^2(n) \quad (2.14)$$

Le choix des limites de la sommation conduit à différentes méthodes AR [7-10], [28-32] .

### 2.1.5.1 Méthode de covariance

La méthode de covariance suppose que [7-9], [28-33]:

- aucune assertion n'est faite sur les données non disponibles (en dehors de la fenêtre d'observation «  $N$  ») ;

- l'erreur quadratique totale «  $E$  » entre le signal et le modèle est minimisée exclusivement sur les «  $N-p$  » derniers échantillons ( $p \leq n \leq N-1$ ).

- Calcul des coefficients du filtre prédicteur

$$e(n) = S(n) - \hat{S}(n) \quad (2.15)$$

$e(n)$  : erreur de prédiction

$$E = \sum_n (S(n) - \hat{S}(n))^2 \quad \text{avec } p \leq n \leq N - 1 \quad (2.16)$$

Déterminons les coefficients " $a_i$ " pour lesquels " $E$ " est minimale.

$$\delta E / \delta a_j = 0 \quad 1 \leq j \leq p$$

$$2 \cdot \left( \sum_n (S(n) - \sum_{i=1}^p a_i \cdot S(n-i)) \cdot (-S(n-j)) \right) = 0 \quad (2.17)$$

$$\sum_n S(n) \cdot S(n-j) = \sum_n \sum_{i=1}^p a_i \cdot S(n-i) \cdot S(n-j) \quad (2.18)$$

$$\sum_n S(n) \cdot S(n-j) = \sum_{i=1}^p a_i \cdot \sum_n S(n-i) \cdot S(n-j) \quad (2.19)$$

On pose :

$$C(j,i) = \sum_n S(n-i) \cdot S(n-j) \quad (2.20)$$

D'après (2.13), nous avons :

$$C(j,0) = \sum_{i=1}^p a_i \cdot C(j,i) \quad 1 \leq j \leq p \quad (2.21)$$

Sous forme matricielle nous avons :

$$\begin{bmatrix} C(1,1) & C(1,2) & \dots & C(1,p) \\ C(2,1) & C(2,2) & \dots & C(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ C(p,1) & C(p,2) & \dots & C(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} C(1,0) \\ C(2,0) \\ \vdots \\ C(p,0) \end{bmatrix}$$

Pour trouver les « ai », on inverse la matrice de covariance  $C(j,i)$ . Elle est seulement symétrique. La résolution du système d'équations se fait au moyen de l'algorithme de Cholesky [9], [28], [29]. Mais en exploitant le fait que la matrice est proche de celle de Toeplitz [33], nous aboutissons à un algorithme rapide et efficace développé par Morf [33]. Cet algorithme est très élaboré [33]; le tableau 2.1 en donne les grandes lignes.

- Calcul des erreurs quadratiques totales moyennes minimisées progressive et rétrograde
- Calcul de deux vecteurs composé des coefficients de prédiction
- Calcul d'une matrice de Toeplitz Y composée des observations  $S(n)$  dans le domaine de covariance
- Calcul de la matrice R qui est le produit de la matrice Y et sa transposée
- Calcul de deux vecteurs de variables auxiliaires pour annuler les effets de bord de la matrice R qui est non-Toeplitz

Tableau 2.1 - Algorithme de Morf [33]



### 2.1.5.2 Méthode d'autocorrélation

Les hypothèses de la méthode d'autocorrélation impliquent que [7-9], [28-32]:

- le signal est nul à l'extérieur de l'intervalle d'observation « N » considéré, cela est possible en multipliant le signal de parole par une fenêtre temporelle de durée « N » ;

- l'erreur quadratique totale « E » entre le signal et le modèle est minimisée pour l'ensemble des échantillons.

- Calcul des coefficients du filtre prédicteur

En posant  $m = n-j$ , l'expression (2.19) donne :

$$\sum_{m=0}^{N-1} S(m).S(m+j) = \sum_{i=1}^p a_i \cdot \sum_{m=0}^{N-1} S(m+j-i).S(m) \quad (2.22)$$

L'estimateur biaisé de la fonction d'autocorrélation à un facteur  $1/N$  près est donné par :

$$R(j) = \sum_{m=0}^{N-1-j} S(m).S(m+j) \quad 0 \leq j \leq p \quad (2.23)$$

Cet estimateur est généralement préféré car il permet d'avoir une matrice d'autocorrélation définie positive [9]. Il en résultera donc un modèle stable.

$R(j) = R(-j)$ , et donc :

$$R(j) = \sum_{i=1}^p a_i R(j-i) \quad 1 \leq j \leq p \quad (2.24)$$

Sous sa forme matricielle, l'expression (2.19) devient :

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

$R(j-i)$  est la matrice d'autocorrélation qui est carrée d'ordre "p" dite matrice de Toeplitz (les éléments de chaque diagonale sont égaux) symétrique [ ]. La résolution des équations se fait au moyen de l'algorithme de Levinson-Durbin (tableau 2.2) [7-9], [28-32].

<p><u>Initialisation</u></p> <p><math>E(0) = R(0)</math></p> <p><math>a_0(0) = 1</math></p> <p><u>Récursion</u></p> <p><math>k(j) = a_j(j) \quad 0 &lt; j \leq p</math></p> <p><math>k(j) = (R(j) - \sum_{i=1}^j a_{j-1}(i) \cdot R(j-1)) / E(j-1)</math></p> <p><math>a_j(i) = a_{j-1}(i) - k(j) \cdot a_{j-1}(j-i)</math></p> <p><math>E(j) = (1 - k^2(j)) \cdot E(j-1)</math></p> <p><math>a_j</math> : coefficient du filtre</p> <p><math>k_j</math> : coefficient de corrélation partielle ou de réflexion</p>
---

Tableau 2.2 - Algorithme de Levinson-Durbin (Kunt [9]).

### 2.1.5.3 Méthode de Burg

L'hypothèse de base dans la méthode de Burg est que les données à l'extérieur de l'intervalle d'observation sont supposées avoir le maximum de variation ou de désordre, d'où le nom de maximum d'entropie [7-9], [27-35].

Par la méthode de Burg, il est inutile de connaître ou d'estimer la fonction d'autocorrélation. Les coefficients de réflexion sont estimés directement à partir des données. Puis en utilisant la forme récursive de Levinson, nous aboutissons aux coefficients de prédiction.

La méthode de Burg identifie la fonction d'autocorrélation du modèle à celle du signal jusqu'à "k = p" et la prolonge implicitement comme si le signal était effectivement autorégressif d'ordre "p".

Cette technique introduit les erreurs quadratiques moyennes progressive " U<sub>j</sub> (n) " et rétrograde " V<sub>j</sub> (n) " dont la minimisation conduit à un modèle stable. Le tableau 2.3 donne l'algorithme de Burg [7-9], [27-35].

<p><u>Initialisation</u></p> <p>pour <math>0 \leq n \leq N-1</math></p> <p><math>U_0(n) = S(n)</math></p> <p><math>V_0(n) = S(n-1)</math></p> <p><u>Récursion</u></p> <p>pour <math>1 \leq j \leq p</math></p> $a_j(j) = \frac{\sum_{n=0}^{N-2+m} U_{j-1}(n) \cdot V_{j-1}(n-1)}{\sum_{n=0}^{N-2+m} U_{j-1}^2(n) + \sum_{n=0}^{N-2+m} V_{j-1}^2(n-1)}$ <p>pour <math>0 \leq n \leq N-2+m</math></p> <p><math>U_j(n) = U_{j-1}(n) + a_j(j) \cdot V_{j-1}(n)</math></p> <p><math>V_j(n) = V_{j-1}(n-1) + a_j(j) \cdot U_{j-1}(n-1)</math></p>
---

Tableau 2.3 - Algorithme de Burg (Kunt [9]).

### 2.1.5.4 Comparaison des méthodes autorégressives

Le tableau 2.4 montre les avantages et les inconvénients de chacune des méthodes AR. L'algorithme de Levinson-Durbin est très utilisé car le plus rapide et il conduit à une solution toujours stable [30].

<b>Méthodes</b> <b>Critères</b>	<b>Auto- corrélation</b>	<b>Covariance</b>	<b>Burg</b>
<b>Stabilité</b>	Assurée	Non assurée	Assurée
<b>Fenêtre</b>	Nécessaire	Inutile	Inutile
<b>Optimal</b>	Optimal	Optimal	Sous Optimal
<b>Volume de calcul</b>	en $pN + O(p^2)$	en $pN + O(p^3)$	en $pN + O(p^3)$
<b>Estimation</b>	Biaisée	Non biaisée	Biaisée

Tableau 2.4 - Comparaison des performances des 3 variantes LPC (Haton [30]).

La méthode de covariance est une méthode générale qui peut être utilisée sans restriction puisque l'on ne fait aucune assertion sur les données non disponibles. Mais cette méthode pose le problème de la stabilité du filtre résultant et celui de la rapidité de résolution du système. La méthode d'autocorrélation, par contre, donne un filtre stable grâce à l'estimateur biaisé de la fonction d'autocorrélation; mais elle pose un problème de précision des résultats du fait de la troncature du signal original. Pour la méthode de Burg, le coefficient de corrélation partielle assure la stabilité du filtre prédicteur [9], [30] et [32].

Pour réduire le volume de calcul, l'idéal serait d'avoir le nombre d'observation  $N$  le plus petit possible. Mais comme une fenêtre est nécessaire pour l'autocorrélation,  $N$  doit être de l'ordre de plusieurs périodes du fondamental : de 100 à 400 échantillons. De plus, plus  $N$  est grand, dans les limites imposées par la stationnarité, plus les résultats sont fiables [9], [30] et [32].

### 2.1.5.7 Nombre optimal des coefficients de prédiction

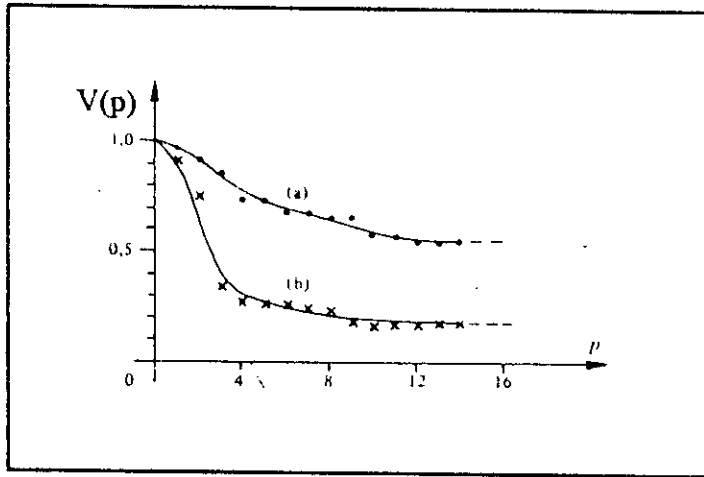
L'une des principales difficultés des méthodes basées sur la modélisation réside dans le choix de l'ordre du modèle. Le nombre "  $p$  " optimal est pris lorsque la variance de l'erreur de prédiction devient minimale.

Certains chercheurs, tel Makhoul [36], pensent que l'estimation s'améliore avec l'augmentation de l'ordre. Mais pour d'autres [37] l'ordre ne doit pas dépasser la moitié du nombre d'échantillons disponibles.

Pour les signaux de parole, on estime [9],[29] en général que l'ordre de prédiction doit avoir la même valeur que la fréquence d'échantillonnage prise en KHz additionné de plusieurs termes (allant de 2 à 5) suivant le résultat désiré.

Une base objective de choix de l'ordre est donnée par l'évolution de l'énergie résiduelle  $V(p) = E(p)/R(0)$  en fonction de l'ordre (fig.2.4). La figure montre qu'au delà de  $p=14$ , théoriquement, l'énergie résiduelle diminue très peu. L'ordre optimal de prédiction est donc choisi très proche de cette valeur.

Cependant, il faudra surestimer l'ordre si on travaille sur un signal nasalisé ou si le signal est bruité [9],[33]. Pour les signaux de parole, le choix de l'ordre optimal est laissé au chercheur jusqu'à l'obtention du signal désiré [9], [29], [33].



a - son non voisé b - son voisé

Fig.2.4- Evolution de l'énergie résiduelle en fonction de l'ordre ( Kunt [9]).

## 2.2. Détection du pitch

L'importance et le rôle du fondamental dans le signal de parole justifient les nombreuses études et méthodes élaborées pour l'extraire. La plupart des problèmes que l'on rencontre dans l'extraction du pitch ont pour origine la difficulté que l'on éprouve à séparer les fréquences des formants, qui constituent l'enveloppe du spectre de parole, du pitch qui en constitue la structure fine. La non-stationarité du signal de parole et certaines irrégularités dans l'excitation glottique compliquent la recherche.

La mesure de  $F_0$  peut se faire à partir du signal de parole dans le domaine temporel, après filtrage du signal, ou dans le domaine spectral, à partir de  $F_0$  du spectre d'un son voisé. L'analyse spectrale demande un fenêtrage du signal, et la valeur de  $F_0$  est donnée sur toute cette fenêtre. Les variations rapides de  $F_0$  ne pourront donc pas être mesurées correctement. Pour cela, les méthodes spectrales résistantes au bruit (provenant des consonnes sourdes) conviennent pour connaître l'évolution de la courbe mélodique ( macro-variations de  $F_0$ ). Nous les avons déjà vues plus haut. Les méthodes temporelles sont souhaitables pour l'analyse de la micro-mélodie ( variation des différentes phases du cycle de vibration glottique) [7]. En plus de l'utilisation du sonographe pour l'obtention du fondamental (§ 2.1.1), nous présentons dans ce qui suit les méthodes d'extraction, de ce dernier, les plus courantes.

## 2.2.1 Méthode du cepstre

Le graphe donnant le cepstre de la figure 2.2 montre un pic étroit, nettement isolé, centré sur la valeur  $T_0$  de la période (anagramme de fréquence). La partie basse fréquence est associée à la structure formantique. En l'éliminant et en faisant la transformée de Fourier du signal obtenu nous retrouvons  $F_0$  [7-10], [30].

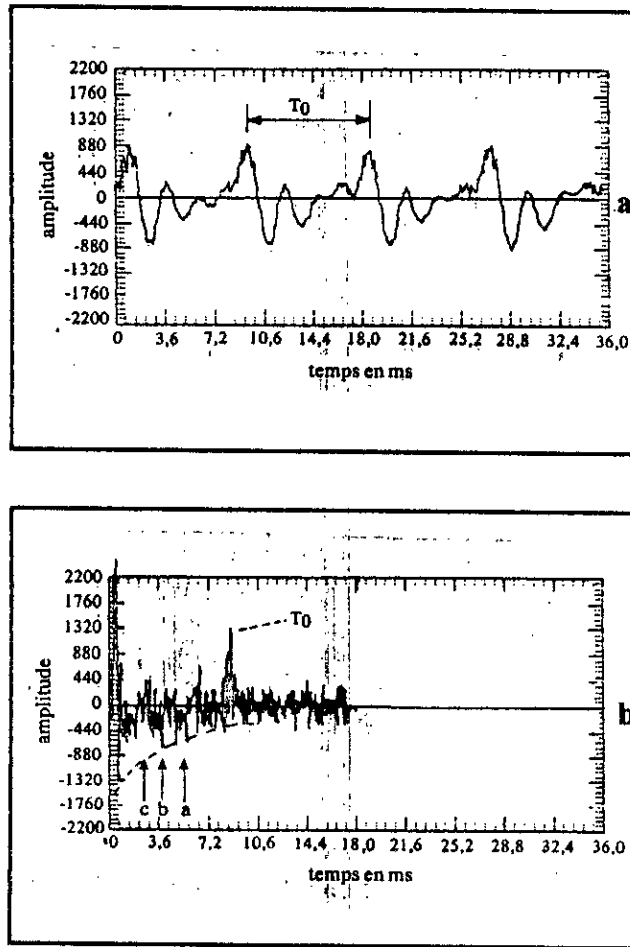


Fig.2.2 - Signal temporel d'un signal voisé(a), le phonème /e/, et son cepstre (b) (Calliope[7]).

## 2.2.2 Codage prédictif

L'erreur de prédiction est donnée par :

$$e(n) = G.U(n) \quad (2.25)$$

où  $U(n)$  est une suite d'impulsions dans le cas des sons voisés. La relation montre que l'erreur est maximale lors des impulsions glottales. Chaque maximum de l'erreur nous donne donc les impulsions du pitch [9], [10].

## 2.2.3 Méthodes basées sur la fonction d'autocorrélation

La majeure partie des méthodes utilisées pour l'analyse dans le domaine temporel se réduit au calcul de la fonction d'autocorrélation. Celle-ci est définie par :

$$r(k) = \sum_{n=1}^{N-1-k} x(n) x(n+k) \quad (2.26)$$

La méthode d'autocorrélation est une méthode de décalage basée sur la détection des maxima de la fonction d'autocorrélation du signal [7-10]. On calcule cette dernière sur une tranche de  $N$  échantillons qui recouvre plusieurs périodes du fondamental. Si le signal de parole est périodique, la fonction d'auto-corrélation est périodique. Son énergie est maximale pour un décalage nul ou multiple de la période du fondamental. Si le signal de parole n'est pas périodique, la fonction ne possède un pic qu'à la position où le décalage est nul (Fig.2.5.a).

Cependant, cette méthode est difficile à mettre en oeuvre à cause des pics plus intenses dus aux formants. Pour cela, on aplatit le spectre par un prétraitement non linéaire, comme l'élimination des crêtes du signal, l'élévation au carré ou au cube, le simple ou le double redressement [7], [10]. Cela a pour effet de renforcer l'amplitude relative de la fondamentale. La méthode a été modifiée en filtrant d'abord le signal de parole à travers un filtre passe-bas.



Mais le coût du calcul reste important. Pour cela, d'autres méthodes ont été élaborées qui permettent d'arranger et de simplifier le signal filtré avant de calculer sa fonction d'autocorrélation.

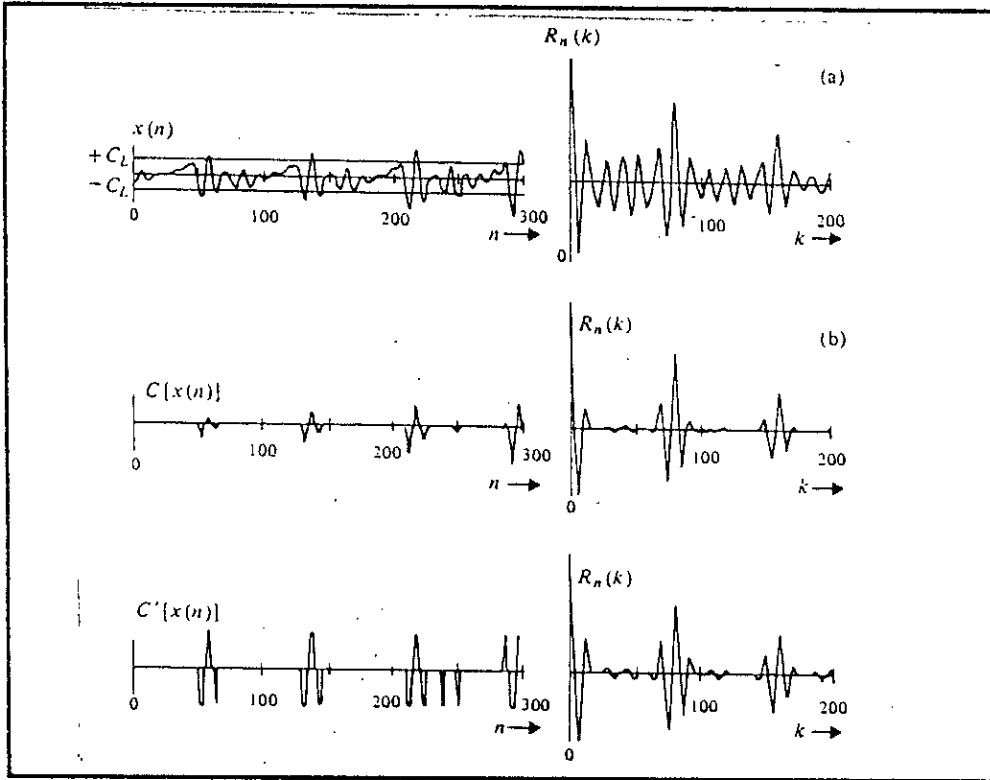


Fig.2.5 - Méthode de l'autocorrélation. (a) signal original. (b) signal sans le noyau  $|x| < CL$ . (c) comme (b) avec codage. ( Kunt [7] )

### 2.2.3.1 Méthode de Sondhi

La méthode de Sondhi est une autocorrélation modifiée avec écrêtage et compression [9], [30]. On ne retient de chaque échantillon  $x(n)$  que la partie qui excède un certain seuil  $CL$  (Fig.2.5.b). Ce dernier est défini par rapport à la plus grande valeur, en module, de  $x(n)$  dans la tranche analysée (30% du signal). On obtient alors un signal dont on a supprimé le noyau  $|x| < CL$  [9], [30]. Le tableau 2.5 donne l'algorithme de la méthode.

- . filtrage,
- . choix d'une fenêtre sur N échantillons,
- . détection du maximum des  $x(n)$  pour le calcul du seuil :  $CL = 0.3 \cdot \text{maximum}$ ,
- . calcul du signal sans noyau  $C [x(n)]$ ,
- . calcul de l'autocorrélation de ce signal,
- . recherche du maximum de l'autocorrélation  $R(\text{max})$  différent de  $R(0)$ ,
- . si  $R(\text{max}) > 0.25 R(0)$  alors  $F_0 = F_e / (\text{abscisse du max})$ ,  $F_e$  étant la fréquence d'échantillonnage

Tableau 2.5 - Algorithme de Sondhi [38].

### 2.2.3.2 Méthode de Dubnowski

Une autre possibilité consiste à substituer une valeur fixe, avec signe, à tout échantillon dont l'amplitude dépasse le seuil de la méthode de Sondhi ( Fig.2.5.c ). Généralement le codage se fait sur trois niveaux , 1 pour les valeurs supérieures au noyau ( $|x| < CL$ ), -1 pour celles qui sont inférieures et 0 pour un signal compris dans le noyau. Le calcul de la fonction d'autocorrélation est donc grandement simplifié [9], [30]. Le tableau 2.6 donne l'algorithme de la méthode.

- . filtrage
- . choix d'une fenêtre sur N échantillons
- . recherche du maximum dans le premier et le dernier tiers de la fenêtre.
- Puis, recherche du plus petit de ces max.  $CL = 0.4 \cdot \text{Minimum}$
- . calcul du signal sans noyau  $C [x(n)]$  et codage de ce signal
- . calcul de l'autocorrélation de ce signal
- . recherche du maximum de l'autocorrélation  $R(\text{max})$  différent de zéro
- . si  $R(\text{max}) > 0.25 R(0)$  alors  $F_0 = F_e / (\text{abscisse du max})$   $F_e$  étant la fréquence d'échantillonnage

Tableau 2.6 - Algorithme de Dubnowski [38].

## 2.2.4 Méthode de la Fonction de la Différence d'Amplitude Moyenne (AMDF)

L'expression de la fonction AMDF est donnée par :

$$\text{AMDF}(r) = \sum_{n=0}^N |s(n) - s(n+r)| \quad (2.27)$$

Il est clair, d'après l'expression, que si l'onde est parfaitement périodique, on doit voir, pour chaque retard « r » égal ou multiple de la période du fondamental, une différence nulle [9], [30].

## 2.2.5 Méthode SIFT (Simplified Inverse Filtering Tracking)

Le conduit vocal est considéré comme un filtre complexe, l'excitation peut donc être récupérée par un filtrage inverse prédictif [9], [10], [30]. Le tableau 2.7 donne l'algorithme de la méthode.

- . Filtrage,
- . Traitement par LPC pour l'obtention des coefficients de prédiction
- . Filtrage prédictif inverse pour l'obtention de l'estimée de la source
- . Calcul de l'autocorrélation de ce signal
- . Recherche du maximum de l'autocorrélation  $R(\max) > \text{seuil donné}$
- .  $F_0 = F_e / (\text{abscisse du max})$ ,  $F_e$  étant la fréquence d'échantillonnage

Tableau 2.7 - Algorithme du SIFT (Haton [30]).

## 2.2.5 Méthode de Rabiner

C'est une méthode de détection hardware basée sur une architecture parallèle. Elle consiste à filtrer le signal par un filtre passe-bas puis à générer six trains d'impulsions caractérisant les pics et les vallées du signal vocal. Chaque train est analysé séparément en vue de l'extraction de  $F_0$ . La plus fiable des six est conservée comme valeur finale [22].

## 2.2.6 Méthode de l'électro-glottographe

Un capteur de pression est mis sur la base du cou pour recueillir les vibrations dues aux cordes vocales en vue de leur traitement. Le signal est insensible à l'effet des formants et des antiformants dus à la physiologie du conduit vocal [7].

## 2.2.7 Méthode de passage par zéro du signal

Le principe consiste à faire passer le signal de parole par un passe-bas pour éliminer les harmoniques. Puis de compter les passages par zéro du signal filtré en un temps déterminé. Le pitch est donné alors par le rapport du nombre trouvé sur deux fois le temps de mesure [7-10].

## 2.3 Détection des formants

En plus de la possibilité d'extraction des formants d'un signal par l'utilisation du sonographe (§2.1.1) différentes autres méthodes existent.

### 2.3.1 Méthode du cepstre

Si nous mettons à zéro toute la partie haute qu'éfrence du cepstre nous faisons disparaître le pic correspondant au pitch ( Fig.2.2 du § 2.2.1). Puis en prenant la transformée

de Fourier du signal obtenu, nous retrouvons le spectre de puissance privé de l'information sur le pitch. Nous déterminerons plus facilement les formants.

### 2.3.2 Codage prédictif

L'estimée de la densité de puissance spectrale (DSP) est donnée par la relation :

$$P(f) = \frac{C}{A(f).A^*(f)} \quad (2.28)$$

où « C » est une constante représentant la variance de l'erreur de prédiction et

$$A(f) = \sum_{i=0}^p a_i . e^{-j2 \pi f i} \quad (2.29)$$

« p » étant l'ordre du filtre prédictif, et « ai » les coefficients de prédiction calculés par les différentes méthodes AR citées plus haut. Les maxima de l'enveloppe du spectre de l'estimée de la DSP calculée par l'équation (2.28) nous donnent les formants.

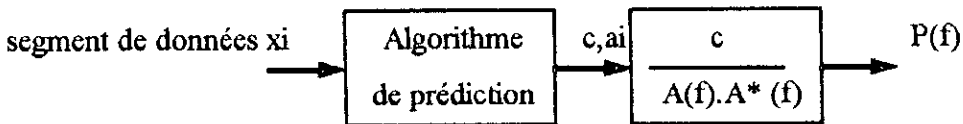


Fig.2.6 - Technique Autoregressive

### 3.2.3 Méthode du périodogramme

Les maxima de l'enveloppe du spectre de l'estimée de la DSP calculée par l'équation (2.4) du § 2.1.2 nous donnent les formants.

## Conclusion

Nous avons présenté différentes méthodes d'analyse. Celle du cepstre est coûteuse en temps de calcul car elle impose le double calcul de la transformée de Fourier. Nous ne l'avons pas retenue pour nos analyses.

Le spectrographe est pratique par la possibilité d'écoute et de visualisation du signal qu'il offre. Il permet aussi l'analyse de tous les paramètres prosodiques. C'est l'outil que nous avons privilégié pour l'extraction des fréquences fondamentales et des paramètres prosodiques.

La méthode du périodogramme appliquée à un signal riche en harmoniques (tel que le signal de parole) donne une résolution spectrale pauvre. Cette dernière est sensiblement améliorée par l'utilisation de méthodes Auto-Régressives ou AR.

En pratique, ces dernières sont très utilisées. Mais même si elles sont fondées sur une connaissance articulatoire, les hypothèses choisies ne permettent pas la modélisation des zéros dus au couplage conduit buccal/conduit nasal et la modélisation du couplage source/conduit.

De plus, les méthodes AR demandent un choix prudent de l'ordre du filtre de prédiction. En effet, un ordre trop petit donne un spectre plat; s'il est trop grand, il y a une apparition de pics parasites dans le spectre estimé de la puissance spectrale. L'utilisateur doit ajuster l'ordre en fonction des résultats.

# CHAPITRE 3

## VARIABILITE ET TRAITS PERTINENTS

### DANS LA PAROLE

## Introduction

Dans ce chapitre, nous insistons sur la variabilité et sur l'importance de la prosodie dans la parole pour bien faire sentir la délicatesse du travail entrepris. Puis nous exposons les paramètres ou traits invariants permettant de reconnaître et d'analyser un son vocal.

### 3.1 Variabilité du signal de parole

Les problèmes rencontrés, lors de toute étude d'un signal de parole, sont liés à la grande richesse et à l'extrême variabilité de celui-ci. Le signal de parole véhicule en effet des informations sur les sons prononcés et leurs structures, sur la personnalité du locuteur, son humeur et son état physique et enfin sur l'interlocuteur.

Les variations du signal de parole proviennent essentiellement des variations inter et intra-locuteurs et de la coarticulation. L'émotion [7], [10] et [11], la fatigue [7-10], une mauvaise hygiène vocale [5], [12] ou la maladie [7], [8] et [10] contribuent aussi à augmenter la variabilité. Le signal de parole recouvre donc des variations qui peuvent être normales mais aussi pathologiques.

#### 3.1.1 Variabilité inter-locuteur :

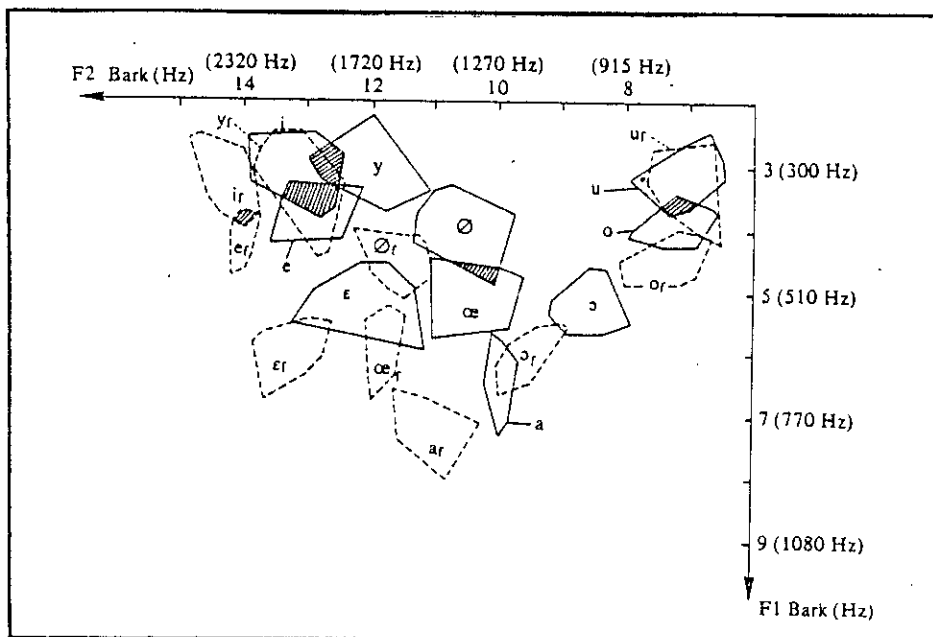
Deux grandes causes de variabilité peuvent être mises en évidence : d'une part les dimensions et proportions de l'appareil vocal, variables selon l'âge, le sexe, l'hérédité et la taille du locuteur; d'autre part les positions et mouvements de l'appareil vocal, les manières de

réaliser les sons, liées à l'anatomie, la physiologie, les habitudes articulatoires de chacun qui sont liées à des facteurs géographiques, sociaux et culturels [5-12].

### 3.1.1.1 Influence de la taille du conduit vocal

La figure 3.1. montre les zones de dispersion des voyelles orales, selon  $F_1$  et  $F_2$ , en échelle de Bark. Cette dernière permet la représentation des fréquences de façon logarithmique. Les hachures délimitent les zones de recouvrement pour un même sexe. Nous voyons, et ceci même à l'intérieur d'une même population de locuteurs hommes par exemple, que les fréquences des formants pour une voyelle donnée peuvent, d'un individu à l'autre, présenter des variations notables.

Les zones formantiques occupées par les différentes voyelles présentent aussi des recouvrements importants lorsque le nombre de locuteurs est élevé. Les formants d'ordre supérieur ne réduisent que légèrement les zones de recouvrement [7].



Hommes : —————

Femmes : .....- - - - -

Fig.3.1- Zones de dispersion des voyelles orales du français sur le plan  $F_1, F_2$  (Calliope [7]).



La figure 3.2 montre les spectrogrammes d'un homme, une femme et un enfant prononçant la même phrase. Nous voyons facilement les différences notables existant dans la valeur de  $F_0$ , des formants et de la durée de prononciation des différents phonèmes de la phrase [8].

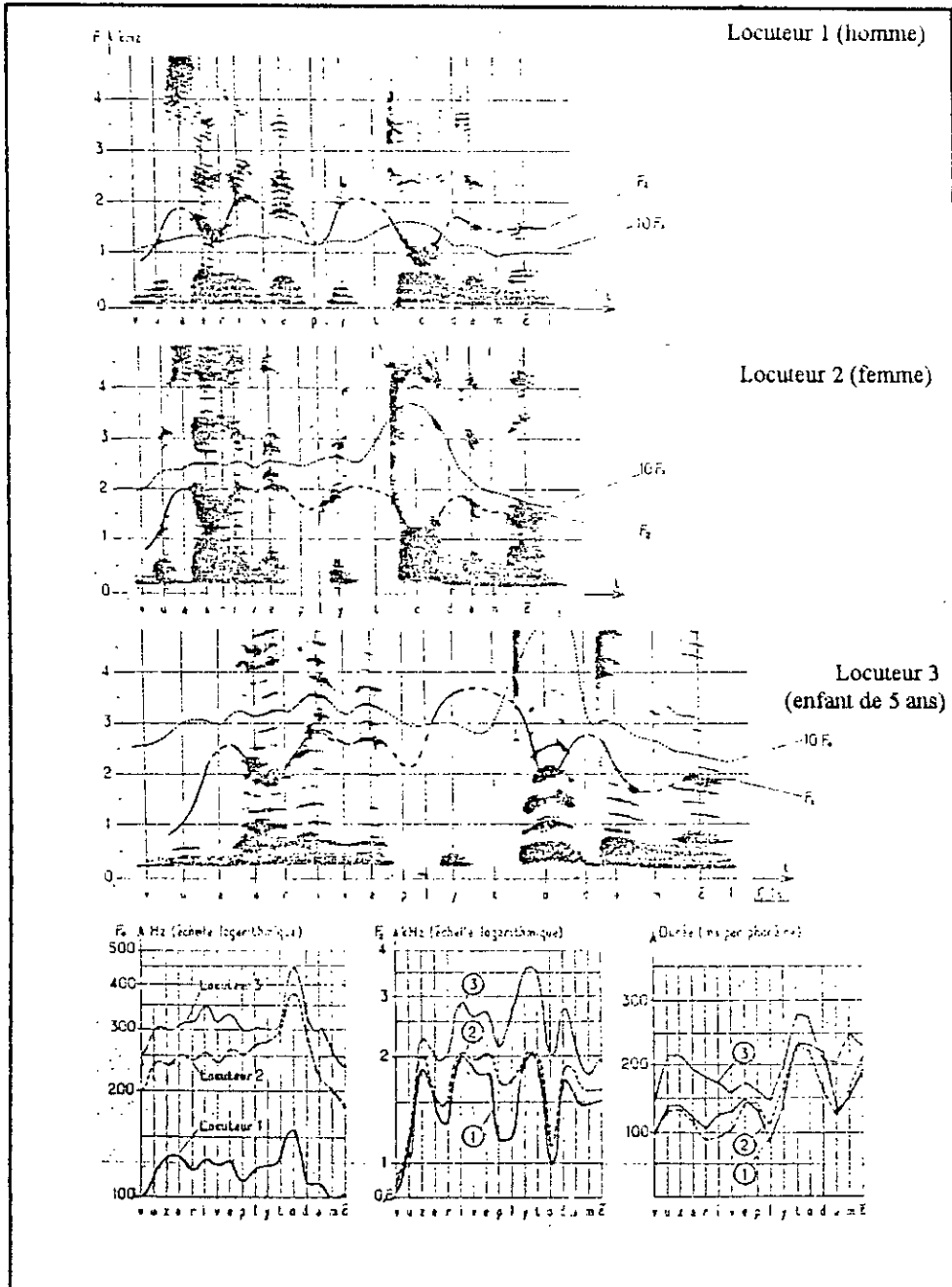


Fig.3.2 - Variabilité du signal « vous arrivez plus tôt demain » suivant la taille du conduit vocal des locuteurs (Liénard [8]).

### 3.1.2 Variabilité intra-locuteur

Le conduit vocal est de forme et de volume extrêmement variables. Sa partie la plus mobile, la langue, peut modeler sa forme à volonté, tout en évoluant rapidement.

D'un jour à l'autre, la voix d'un individu peut se modifier même dans des conditions normales. Ceci est dû soit au manque de précision dans l'articulation ou dans le contrôle des cordes vocales, soit à des variations prosodiques [7-10].

#### 3.1.2.1 Importance de l'articulation

La parole est le résultat d'un ensemble de mouvements synchronisés de l'appareil vocal. Ces mouvements peuvent être plus ou moins précis selon le soin que l'on met pour articuler. La figure 3.3 montre une même phrase prononcée d'abord en prenant bien soin aux mouvements articulatoires (a) puis de façon relâchée, en ouvrant à peine la bouche (b).

Le spectrogramme est plus flou et moins contrasté dans le second cas. En fait, nous constatons que les différences sont nombreuses entre les deux prononciations d'un même locuteur. Elles touchent les fréquences et les largeurs de bande des formants, leur évolution, la richesse du spectre dans les hautes fréquences, les explosions et le voisement des occlusives [8].

#### 3.1.2.2 Rôle de la vitesse d'élocution

La figure 3.3 (c) montre la même phrase qu'en (a) prononcée plus rapidement. Toutes les durées élémentaires des phonèmes subissent une réduction. Mais cette dernière est plus marquée pour les voyelles et les fricatives que pour les occlusives. De plus, en prononciation rapide, les fréquences formantiques de certaines voyelles ne sont pas atteintes [8].

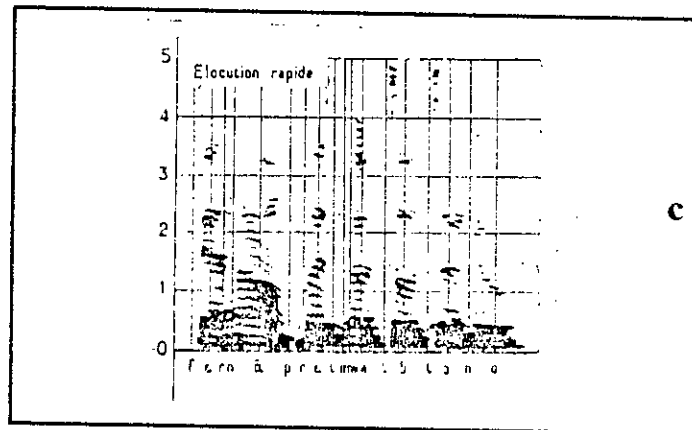
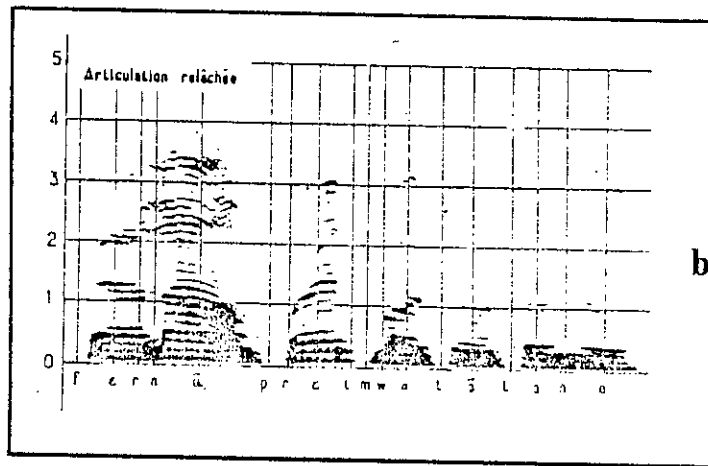
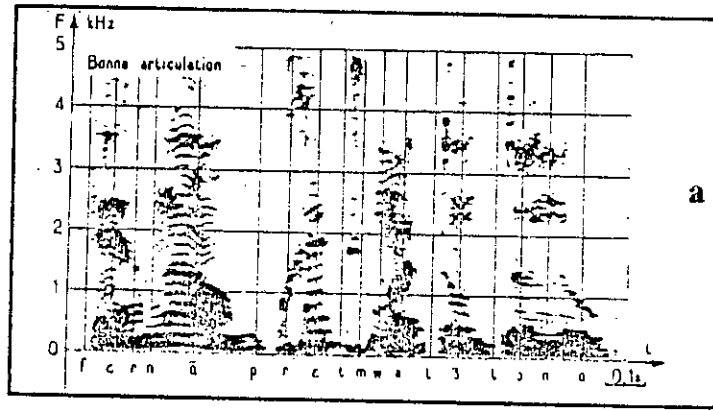


Fig.3.3 - Rôle de l'articulation dans la phrase « Fernand prête moi ton tonneau » (Liénard [8])

### 3.1.2.3 Rôle de l'intensité de la parole

Tout locuteur peut ajuster son niveau sonore de façon qu'il soit audible et qu'il n'ait pas d'effort inutile à faire. Il peut faire varier ce niveau en jouant sur la pression sous-glottique et la tension des cordes vocales.

En voix chuchotée (fig.3.4 a), le locuteur est bien adapté à la communication à très courte distance mais les cordes vocales ne peuvent plus vibrer. Nous avons alors une richesse moindre dans le timbre et l'intonation [8].

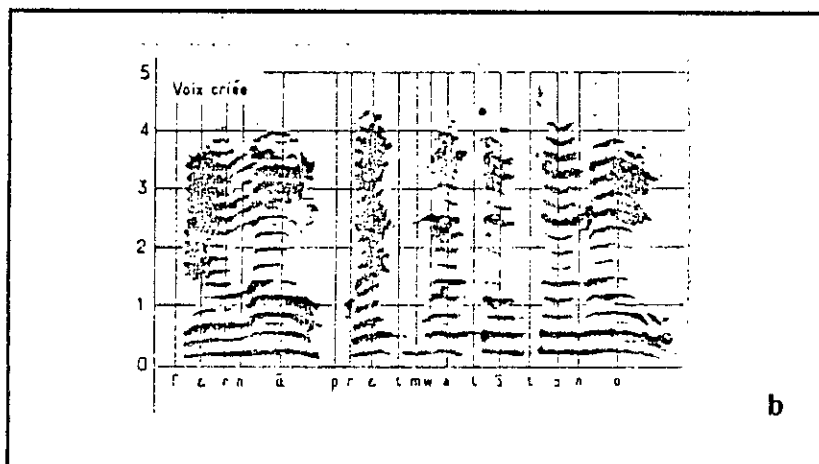
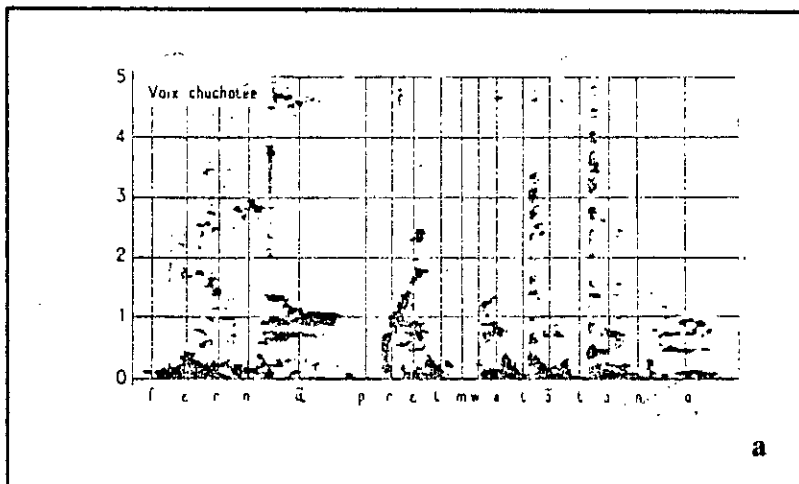


Fig.3.4 - Rôle de l'intensité du signal dans la phrase « Fernand prête moi ton tonneau » (Liénard [8])

En voix criée, le locuteur augmente la pression sous-glottique et tend les cordes vocales tout en les accolant . Cela entraîne une augmentation de l'amplitude et de la fréquence du signal de parole. La figure 3.4 (b) montre un déplacement des formants vers l'aigu et une disparition des phases d'explosion, de silence puis d'écoulement turbulent pour les occlusives.

La tension générale sur les muscles rend les mouvements articulatoires et leur contrôle plus difficile. De plus, cela entraîne une diminution de l'intelligibilité du signal de parole [8].

### 3.1.2.4 variation du pitch

La fréquence fondamentale peut présenter des variations considérables chez un même locuteur selon le type de phrase prononcée et son attitude lors de la parole.

#### a - Micromélorie

La fréquence fondamentale existe lors de l'émission des voyelles et des consonnes voisées. Sa variation au niveau de chaque phonème est la micromélorie. Les consonnes sourdes introduisent une interruption dans la ligne mélodique décrivant l'évolution de  $F_0$  de la même façon que les pauses et les silences.

Parmi les consonnes voisées, certaines font apparaître des chutes de  $F_0$  caractéristiques, de l'ordre de 10 à 20 Hz, par rapport aux voyelles adjacentes et par rapport à l'évolution générale de la courbe mélodique (fig.3.5). Ces chutes sont dues essentiellement à une diminution de la pression sous-glottique pendant la réalisation de ces consonnes. C'est le cas en français pour /b/, /d/, /g/, /z/, /v/ et /r/. Les consonnes /l/, /m/, /n/ et /r/, en position finale, n'entraînent pas de perturbation locale de la ligne mélodique [7], [8].

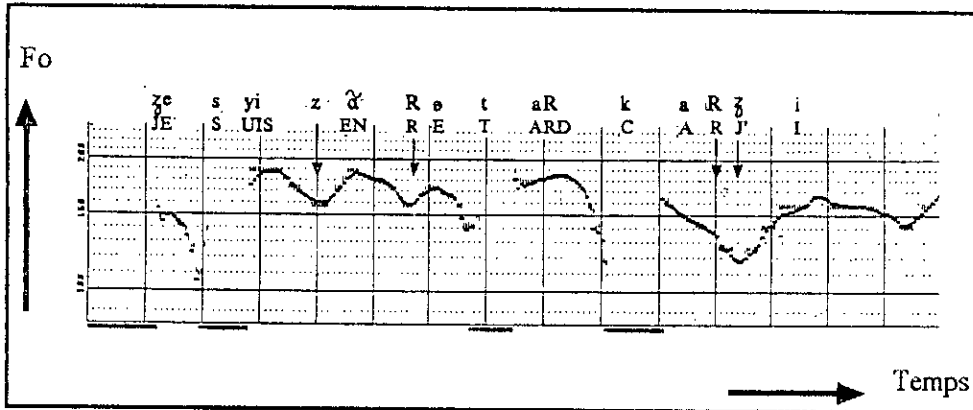


Fig.3.5- Micromélodie des consonnes dans la phrase "je suis en retard car j'ig.." (Calliope [7]).

### b - Ligne de déclinaison et macrométrie

La variation du pitch au niveau de la phrase et de ses groupes syntaxiques est la macrométrie. La ligne mélodique a tendance à décroître du début à la fin de la phrase. Cette ligne de déclinaison [7], [8] qui existe pour de nombreuses langues est due à un relâchement normal des organes de la phonation. Elle sert d'indicateur des limites de la phrase ou de groupes syntaxiques importants. Elle est remise à zéro à chaque pause pour reprendre une valeur élevée en début de phrase.

La ligne de déclinaison d'une phrase existe indépendamment des variations locales existant pour les syllabes. La dénivellation de  $F_0$  au cours de la déclinaison dépend de facteurs linguistiques ou extra-linguistiques tels que l'émotion.

### 3.1.2.5 Importance d'une utilisation adéquate des paramètres prosodiques.

De nombreuses expériences [7] ont montré que tant une déformation de l'organisation temporelle d'un énoncé (durées de segments successifs incorrectes ou introduction de pauses à des emplacements erronés) que l'application d'évolutions de  $F_0$  inappropriées perturbent notablement la compréhension. L'identité de certaines voyelles peut même être modifiée sous

l'influence de la fréquence fondamentale. Car l'identification phonétique et le traitement prosodique sont intimement liés.

La parole doit donc respecter des contraintes de continuité intonative, spectrale et temporelle [7]. Si ces dernières ne sont pas respectées, une même séquence de sons peut être perçue comme provenant de deux locuteurs différents.

Ainsi pour montrer la nécessité d'une continuité intonative, si l'on compose une séquence de 6 voyelles stationnaires alternant d'un  $F_0$  grave à un aigu et séparées par des silences de 150 ms environ (la durée d'une consonne sourde), des écarts de  $F_0$  de plus de 5 demi-tons\* donneront à l'auditeur l'impression que deux locuteurs différents ont prononcé chacun 3 voyelles qui s'intercalent. Alors qu'avec des écarts inférieurs à 2 demi-tons on entend un même locuteur qui modifie la hauteur de sa voix.

La continuité spectrale est concrétisée par les transitions de formants. Ces derniers permettent de percevoir de façon continue des segments que leurs écarts de  $F_0$  auraient pu laisser supposer appartenir à deux locuteurs différents.

La continuité temporelle est régie par des contraintes portant sur les durées relatives des syllabes et des phonèmes en fonction de leur nature phonétique et de leur position dans le mot ou la phrase ainsi que par la durée et la localisation des pauses.

### 3.1.3 Influence de l'émotion

L'émotion du locuteur est un grand facteur de variabilité. Elle est susceptible de faire varier grandement la valeur de la fréquence fondamentale et sa ligne de déclinaison [7]. Elle modifie aussi la valeur des formants, le timbre (Fig.3.6), la coarticulation et même les paramètres prosodiques [1], [2] et [7-11].

\*demi-ton : intervalle entre deux notes consécutives de la gamme musicale.

Sous l'influence de diverses charges de travail, la variation de  $F_0$  augmente avec le niveau du stress émotionnel du locuteur ou la complexité des tâches ou tests à accomplir par ce dernier [1], [2].

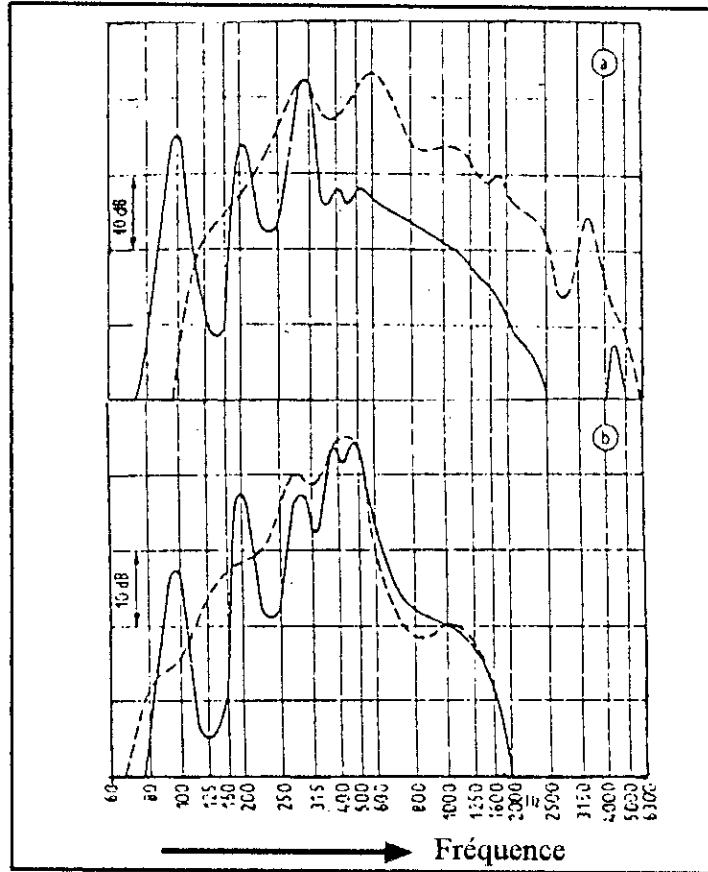


Fig.3.6- Influence de l'émotion sur le timbre des locuteurs (a) et (b) lisant une suite de nombres ( — ) puis un texte (---). (Bordone- Sacerdote [39])

La fréquence fondamentale de la voix dans une situation à haut stress est plus élevée que dans une situation de repos. Les recherches ont mis en évidence un microtremblement des muscles près des cordes vocales à l'origine de la modulation de la fréquence fondamentale [1], [2].

Les fréquences des trois premiers formants augmentent quand le locuteur doit effectuer des tests psychomoteurs ou lorsqu'il simule des émotions telles que la colère, la peur ou la tristesse [1], [2]. Au fur et à mesure de la complexité de tests psychomoteurs imposés au locuteur, le niveau d'énergie augmente [1], [2]. Il augmente aussi d'un dB lorsque le locuteur reçoit des décharges électriques sur la peau puis baisse de 2 dB à l'arrêt de celles-ci [1], [2].



La vitesse d'élocution (en mots par minute) décroît lorsque le locuteur simule une émotion ; le temps mis pour parler augmente par contre. La durée du mot augmente quand il subit un test psychomoteur puis diminue quand celui-ci devient plus complexe [1], [2].

### 3.1.4 Mauvaise hygiène vocale

Certaines habitudes alimentaires et certaines conditions de vie entraînent aussi des modifications de la voix par le forçage vocal qu'elles occasionnent [5], [40], ainsi:

- les mets trop épicés, les aliments trop chauds ou glacés et les excitants exercent un effet congestionnant sur les muqueuses pharyngées ;
- l'air sec et chaud des habitations et bureaux ou une température extérieure très froide perturbe le fonctionnement de l'appareil phonatoire (Fig.3.7) ;

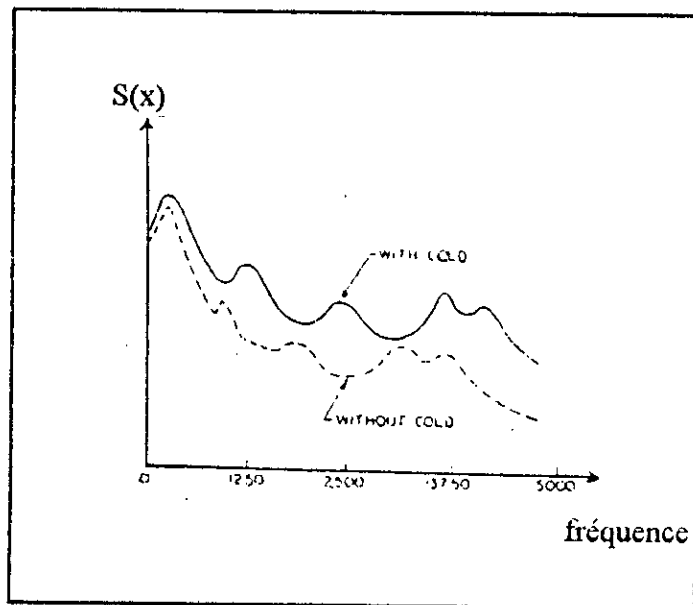


Fig.3.7- Influence d'un rhume sur le spectre du /n/ ( — ). (Sambur [41])

- les atmosphères empoussiérées ainsi que l'exposition à des gaz et vapeurs nocifs créent une irritation des voies aériennes supérieures ;
- le port de vêtements serrés peut constituer une contrainte mécanique à la dynamique respiratoire ;
- une mauvaise technique vocale peut entraîner une altération de la voix.

### **3.1.5 Variations pathologiques de la parole :**

Une anomalie des organes phonateurs peut être à l'origine de troubles dans la production de la parole, ainsi :

- les affections de la sphère O.R.L , en particulier du larynx, sont la cause de bon nombre de dysphonies transitoires. Beaucoup d'états inflammatoires des voies aériennes s'accompagnent de toux. Ce geste soumet les muqueuses laryngées à rude épreuve par l'irritation qu'il occasionne aux organes phonatoires [5];
- l'aphasie, résultant d'une lésion de la partie du cerveau régissant le langage, entraîne des troubles de l'articulation du mot et une altération du langage [16]. L'influence des émotions sur la voix est très marquée chez les aphasiques car les centres corticaux sont affaiblis [12];
- dans certains cas d'épilepsie, les malades ont une grosse difficulté à régler et à tenir leur voix sur une hauteur d'émission donnée [12];
- atteint d'hémiplégie, un locuteur ne peut moduler aucun son; la voix reste grave et monocorde [12];
- l'audition est un processus d'autocontrôle de la phonation. Son altération perturbe les locuteurs [8].

### 3.1.6 Effet de la coarticulation

Les différents exemples qui suivent illustrent l'effet de la coarticulation. Ainsi pour la syllabe "tou" (/tu/), dès la prononciation du /t/, les lèvres commencent à s'arrondir [10], et nous avons une labialisation de [i] dans la syllabe "lui" (/lyi/) [42]. Le phonème /k/, dans "qui", est articulé beaucoup plus vers l'avant du palais que dans "cou" [7].

La figure 3.8 montre la réduction de l'effort lors de l'articulation du /u/ dans le mot "doute" (/dut/) comparé à un /u/ isolé [7]. Pour ce mot, le locuteur part d'une articulation apico-dentale, recule rapidement la langue vers la zone vélaire, puis doit ramener la langue en position antérieure. En débit rapide, le locuteur dispose de peu de temps pour faire ce déplacement important. Il réduit donc la distance à parcourir en prononçant un /u/ plus antérieur. La voyelle est déplacée en direction de /y/ ;  $F_2$  augmente donc (fig.3.9).

La voyelle coarticulée doit rester assez proche de la cible articuloire pour éviter la confusion avec une autre voyelle.

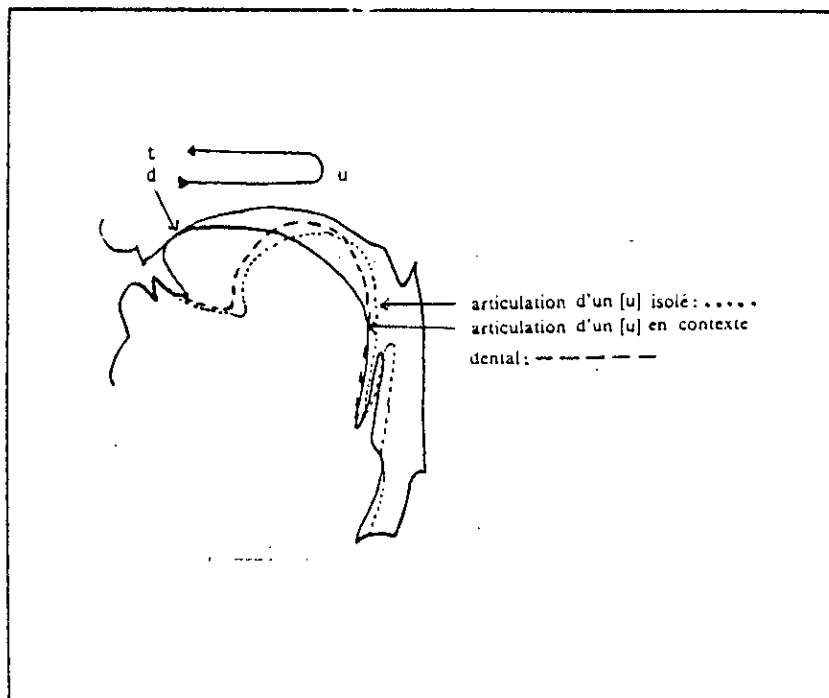
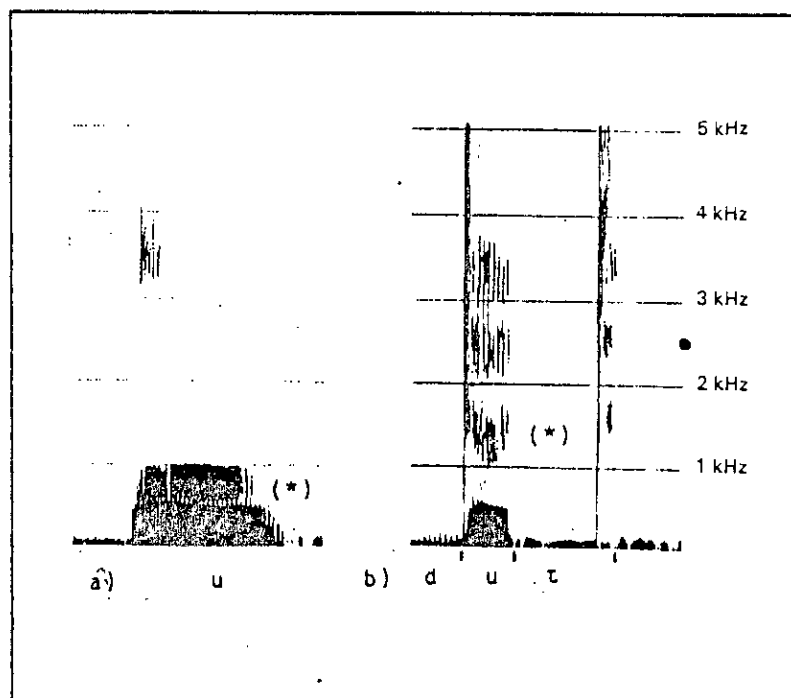


Fig.3.7 - Déplacement de l'articulation du /u/ en contexte dental (Calliope [7]).



(\*) : position de  $F_2$

Fig.3.9 - Spectrogrammes de la voyelle /u/ : hors-contexte (a) et en contexte dental (b).  
(Calliope [7])

## 3.2 Recherche des invariants et identification du locuteur

Malgré toutes les variations que nous avons énumérées, il faut toutefois souligner que l'on peut rechercher certains invariants propres à un locuteur. Ces derniers portent l'information phonétique du texte qui permet dans tous les cas d'identifier ce locuteur parmi d'autres.

### 3.2.1 Qualité des traits distinctifs

Le choix des traits est très délicat. Il s'agit de pouvoir extraire du signal vocal les caractéristiques de la voix de chaque locuteur par la sélection de traits acoustiques signifiants (chargés d'une quantité suffisante d'informations sur l'identité du locuteur).

Les paramètres servant à l'identification du locuteur doivent être [43] :

- aptes à représenter l'information utile sur l'identité du locuteur ;
- faciles à mesurer ;
- stables dans le temps ;
- d'apparition naturelle et fréquente dans la parole ;
- peu modifiés par un changement de l'environnement ;
- non imitables.

Un compromis devra être fait sur le choix des paramètres car aucun de ceux généralement retenus ne remplit pleinement les conditions de Wolf [43].

### 3.2.2 Le timbre

La caractéristique première de la voix d'un locuteur est son timbre qui n'est perceptible que pour les sons voisés les semi-voyelles et les voyelles.

### **3.2.3 L'articulation**

L'activité musculaire d'un locuteur aussi est un des traits distinctifs généralement retenus [44].

#### **3.2.3.1 La coarticulation**

En prononçant une phrase, le locuteur produit une suite de phonèmes qui sont enchaînés les uns aux autres de façon continue, en reliant les parties stables du signal par des zones de transition. Selon la qualité de l'articulation du locuteur, les transitions sont plus ou moins longues et les zones stables peuvent ne pas être atteintes. La dynamique du canal vocal représentée par les variations de la fonction de transfert, est donc un ensemble de traits distinctifs, lié à la musculature.

Le déplacement que font subir à une consonne les voyelles antérieures et postérieures est aussi caractéristique de l'élocution d'un individu.

#### **3.2.3.2 Durée d'établissement des plosives**

La durée du silence précédant l'explosion dans les plosives sert aussi d'identification. Car c'est un paramètre temporel comme la coarticulation régulé par des mécanismes plutôt réflexes.

#### **3.2.3.3 Enveloppe énergétique**

L'évolution de l'énergie du signal le long d'une phrase est aussi propre à un locuteur donné.

### 3.2.4 Les variations prosodiques

La prosodie est aussi un trait retenu car elle est fonction de la personnalité du locuteur. L'une de ses principales fonctions dans la perception de parole est d'organiser le signal acoustique émis par un locuteur en un objet auditif cohérent, structuré, pouvant être identifié en tant que tel, même en présence de sons similaires qui interfèrent.

Outre sa fonction de guide pour l'identification phonémique, la prosodie assure deux autres fonctions : une fonction de segmentation de l'énoncé et une fonction d'anticipation de la suite de l'énoncé. Ainsi un allongement de la dernière syllabe d'un mot associé à une forte montée de  $F_0$  est perçu comme l'indication d'une frontière syntaxique majeure [7], [8]. La prosodie sert à attirer l'attention de l'auditeur sur les parties potentiellement importantes du signal de parole. De plus, l'existence d'évolutions adéquates de la fréquence fondamentale au début d'une phrase permet à l'auditeur de deviner la structure syntaxique de la suite de la phrase et d'accélérer ainsi le traitement de son information phonétique .

La plus importante des composantes de la prosodie est le contour du pitch, c'est-à-dire l'évolution de la fréquence fondamentale  $F_0$  dans le temps. Cette dernière porte également une information sur l'identité du locuteur qui apparaît dans la distribution statistique de la fréquence (pitch moyen) et dans l'évolution temporelle, chaque locuteur ayant des variations intonatives favorites [7] .

La micro-mélodie des voyelles est peu importante, elle peut être négligée. Celle des consonnes, par contre, peut être très importante pour leur perception [7], [9]. Pour les consonnes, on adopte en général un schéma qui se raccorde au pitch des voyelles adjacentes.

Les règles de macro-mélodie interviennent au niveau du mot, des groupes de mots et au niveau de la phrase; des schémas intonatifs caractérisent chaque groupe syntaxique. Des règles spéciales interviennent aussi pour assurer les intonations propres aux phrases interrogatives ou impératives. Ainsi le pitch descend en fin de phrase dans un énoncé affirmatif et remonte dans un mode interrogatif. De même pour chaque syllabe, des règles précisent la fréquence du fondamental en fonction de sa place dans le mot, du nombre de syllabes qu'il

comporte, et de la place de ce dernier dans le groupe. Ainsi une remontée du pitch est le plus souvent associée à la dernière syllabe du groupe sujet.

## **Conclusion**

C'est sur la base des traits distinctifs énoncés dans ce chapitre que nous avons fait le choix des paramètres pour l'analyse de nos échantillons de parole. Toute la procédure est décrite en détail dans le chapitre suivant.



# CHAPITRE 4

## PROCEDURES D'ACQUISITION ET DE TRAITEMENT UTILISEES

### Introduction

L'étude des voix fait appel à des techniques spécialisées qui exigent un bon support matériel, un protocole de recueil de données rigoureux et l'utilisation de procédures mathématiques et statistiques du traitement de signal. Nous voyons, dans ce chapitre, comment nous avons traité le signal de parole depuis sa production par un locuteur jusqu'à l'étape finale d'analyse.

### 4.1 Génération du corpus

#### 4.1.1 Population

Pour recueillir nos échantillons de parole, nous avons pris une population variée de locuteurs. Celle-ci, composée d'enfants et d'adultes des deux sexes, se répartit en deux classes. Dans la première, nous avons étudié l'évolution de la fatigue le long de la journée. Nous traitons donc de l'influence d'une fatigue physique sur les muscles phonatoires. Parmi les 34 locuteurs étudiés, certains venaient de participer à une rencontre sportive et d'autres étaient des malades hospitalisés (Tableau 4.1).

Pour la deuxième classe de locuteurs, nous avons pris des locuteurs susceptibles de présenter une fatigue purement vocale. Pour cela, nous avons pris 4 animateurs radio et 2 acteurs que nous avons enregistrés avant et après une émission radiophonique d'assez longue durée (Tableau 4.2).

Age(ans) Sexe	8 - 10	11 - 13	15 - 16	18 - 21	22-24	29 - 33	≥ 50
	Masculin	8	1	4	4	0	0
Féminin	1	6	0	1	2	3	3

Tableau 4.1 - Population de locuteurs présentant une fatigue physique.

Sexe	Locuteur	Age (ans)	Durée de l'émission
Féminin	L 1	26	2 h 30'
	L 2	18	4 h
	L 3	58	2 h 30'
Masculin	L4	45	3 h
	L 5	57	3 h
	L 6	38	7 h

Tableau 4.2 - Population de locuteurs présentant une fatigue vocale.

De plus, pour éviter l'influence d'une saison chaude et humide, très éprouvante pour l'endurance des locuteurs, nous avons fait tous nos enregistrements en hiver.

### 4.1.2 Choix du corpus

Les transitions d'un phonème à un autre sont rarement nettes, car les déformations du signal, dues aux déformations de l'appareil phonatoire, sont continues [9], [10]. Les voyelles sont généralement stables et faciles à repérer grâce à leurs formants plus intenses. Mais les consonnes, plus fugitives, sont plus délicates à localiser et à identifier avec certitude [7]. Certaines consonnes (liquides, vibrantes) présentent des difficultés lors de la segmentation.

Le choix du corpus a été fait de façon à balayer les différents modes d'articulation des phonèmes (voyelles, liquides, vibrantes, fricatives, nasales et plosives). La liquide /l/ est de faible énergie par rapport aux voyelles [7]. Pour les fricatives comme /ʃ/ et /s/ qui sont des bruits, nous avons un ensemble de petites raies verticales très caractéristiques [7]. Les plosives sourdes, comme la bilabiale /p/ et l'apico-dentale /t/, présentent un silence suivi d'une barre d'explosion, appelée aussi burst, discrète pour la première et intense pour la seconde. Elles ont aussi des transitions discrètes et brèves [7].

Nous avons aussi choisi des emphatiques arabes comme la fricative /S/ et la plosive /D/ car elles allongent les voyelles [13]. Les variations d'intensité entre les voyelles et les consonnes telles que les nasales /m/ et /n/ et la vibrante /r/ aussi sont franches [7], [13]. De plus la durée d'une voyelle est allongée lorsqu'elle est suivie de la vibrante [7]. L'intérêt de l'allongement des voyelles est que nous évitons les zones de coarticulation lors de la segmentation.

Les locuteurs sur lesquels nous étudions la fatigue vocale ont prononcé le corpus du tableau 4.3. Nous avons fait prononcer aux autres, un même mot à différents instants : le matin, en milieu et en fin de journée. Celui-ci, de type CVC (C : consonne, V : voyelle), est père (/ peR /).

Corpus	Transcription phonétique
صام رمضان	/Saama ramaDaan/
فتح الباب	/fataHa el baab/
كتب رسالة	/kataba risaala/
Tu es parti	/ty ε paRti/
Ce choux est beau	/ sɔ̃ ʃu ε bɔ̃ /
C'est ta fleur	/ sɛ ta flɔ̃R /
Vas-y	/ va zi /
Le chat saute	/ lɔ̃ ʃa sɔ̃t /

Tableau 4.3 - Corpus des locuteurs fournissant un effort vocal.

### 4.1.3 Conditions d'enregistrement

Pour la première catégorie de locuteurs enregistrés, nous avons utilisé la chambre d'enregistrement de l'hôpital MUSTAPHA. Cela nous a permis d'enregistrer des locuteurs disponibles puisque hospitalisés. En effet, cela n'a pas été facile d'avoir, trois fois par jour, des locuteurs venant de l'extérieur.

Pour la seconde catégorie de locuteurs, nous avons utilisé les studios d'enregistrement de l'Entreprise Nationale de la Radio Sonore (ENRS). Cela nous a permis d'avoir des locuteurs coopératifs car habitués au microphone et articulant bien.

A tous ces locuteurs, nous prenons la précaution de donner des consignes qui sont destinées à nous aider lors de la phase d'analyse :

- donner le nom et l'heure;

- avoir un ton monocorde et surveiller l'intensité de la diction de façon à éviter les variations prosodiques pendant l'élocution.

## 4.2 Chaînes d'analyse utilisées

Nous avons utilisé deux chaînes d'analyse. La première est située au laboratoire d'Electronique Acoustique de l'Université de Bouzaréah (Fig.4.1). Celle-ci, basée sur le sonographe, nous a permis l'extraction des fréquences  $F_0$  pour tous nos signaux de parole. Elle nous a aussi donné les formants et les paramètres prosodiques. Nous avons monté la seconde au laboratoire de Traitement de Signal du Département d'Electronique de l'Ecole Nationale Polytechnique - Alger (Fig.4.2). Nous avons extrait, de la seconde, les formants par différentes méthodes d'analyse spectrale. La première chaîne est très commode d'utilisation car elle permet, par un système d'écoute des zones phonémiques choisies, un bon positionnement de façon à éviter la coarticulation. La seconde, ainsi que nous le verrons dans ce qui suit, nous permet d'agir sur les paramètres d'analyse.

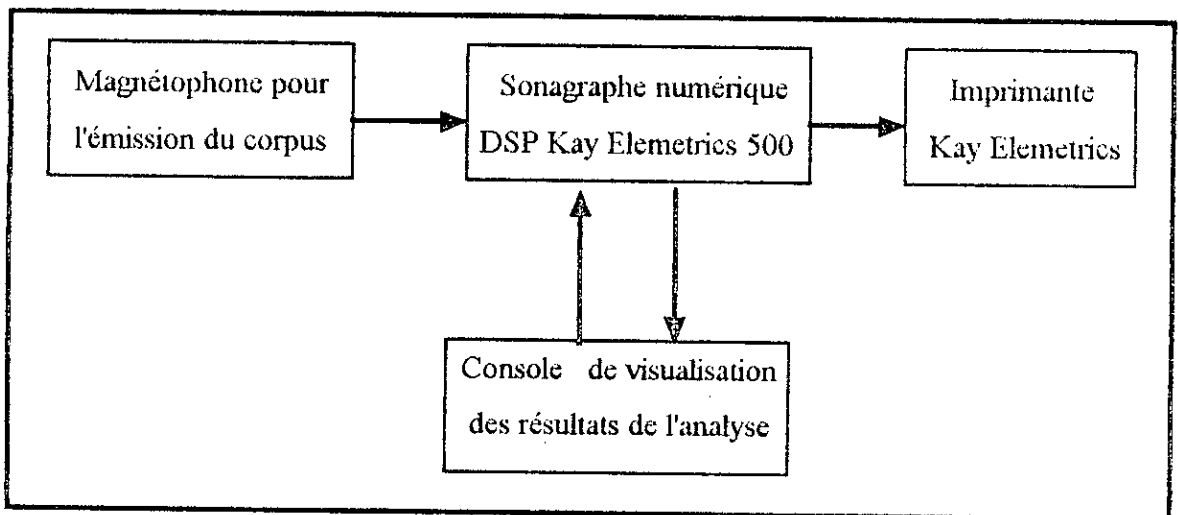
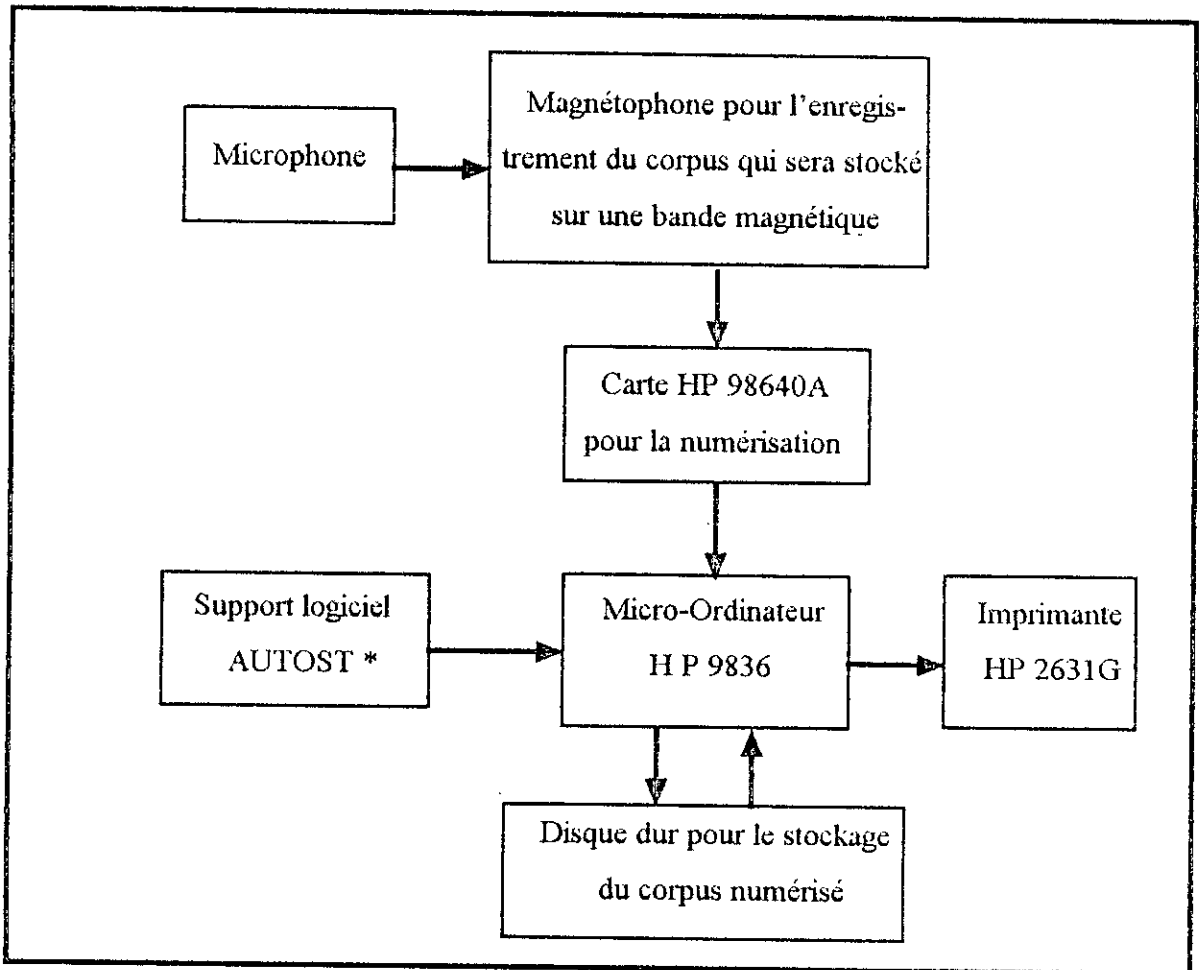


Fig.4.1- Synoptique de la chaîne d'acquisition du laboratoire de l'Université de Bouzaréah.



\* : logiciel AUTOSTART pour le calcul des densités spectrales

Fig.4.2 - Synoptique de la chaîne d'acquisition du laboratoire de Traitement du Signal (ENP-ALGER).

### 4.2.1 Première chaîne d'analyse

La chaîne d'analyse comprend (fig.4.1) :

- un magnétophone ;
- un sonographe numérique, le DSP Kay Elemetrics 500. Le calcul des fréquences  $F_0$  peut être obtenu en lisant la courbe de la fréquence fondamentale en fonction du temps (fig.4.3) ou en utilisant l'analyse à bandes étroites (Fig.4.4). Le sonographe permet

l'extraction des formants par lecture directe au milieu des bandes noires (Fig.4.5) ou par calcul de la DSP de n'importe quelle zone phonétique placée entre curseurs (Fig.4.6). Le sonographe permet aussi un affichage numérique en dB de la valeur de l'énergie à la position du curseur. L'axe des temps permet de mesurer la durée d'articulation des différents phonèmes ;

- une console de visualisation des graphes et des résultats avec un affichage direct de la fréquence et du temps à la position des curseurs;
- une imprimante.

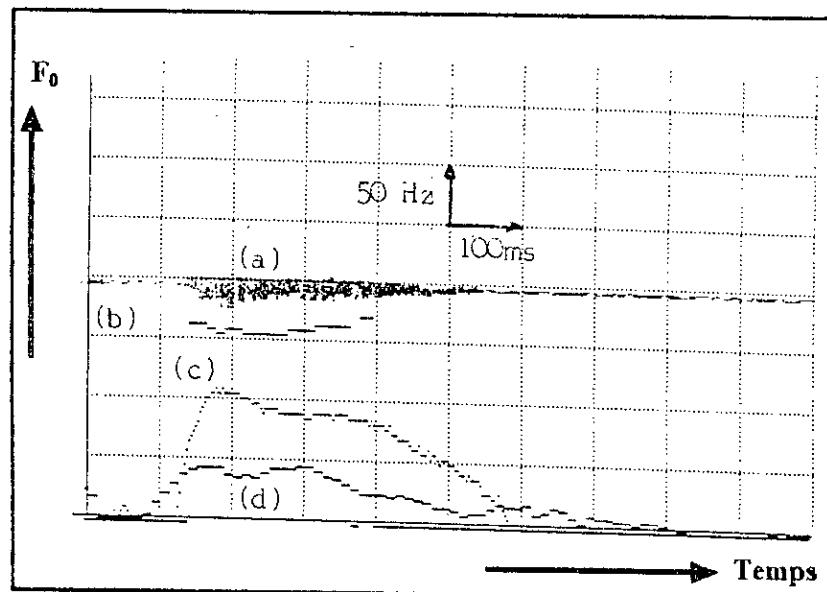
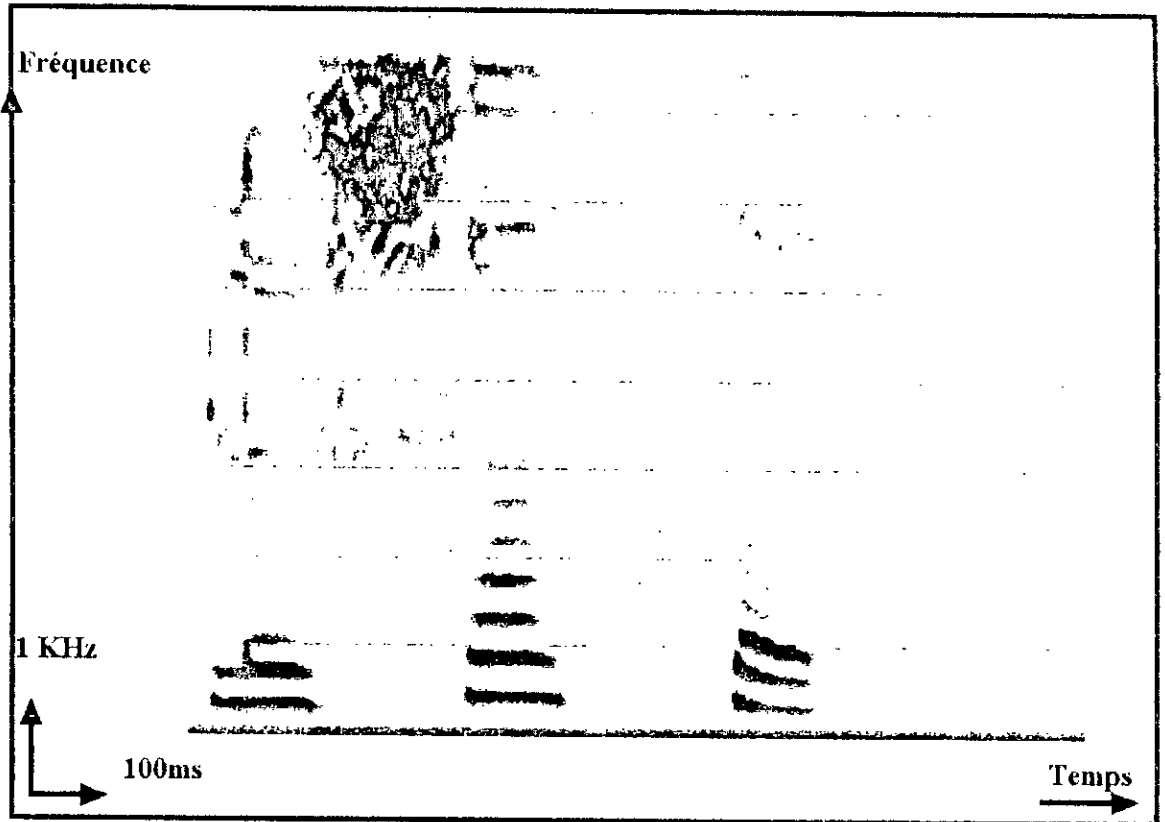
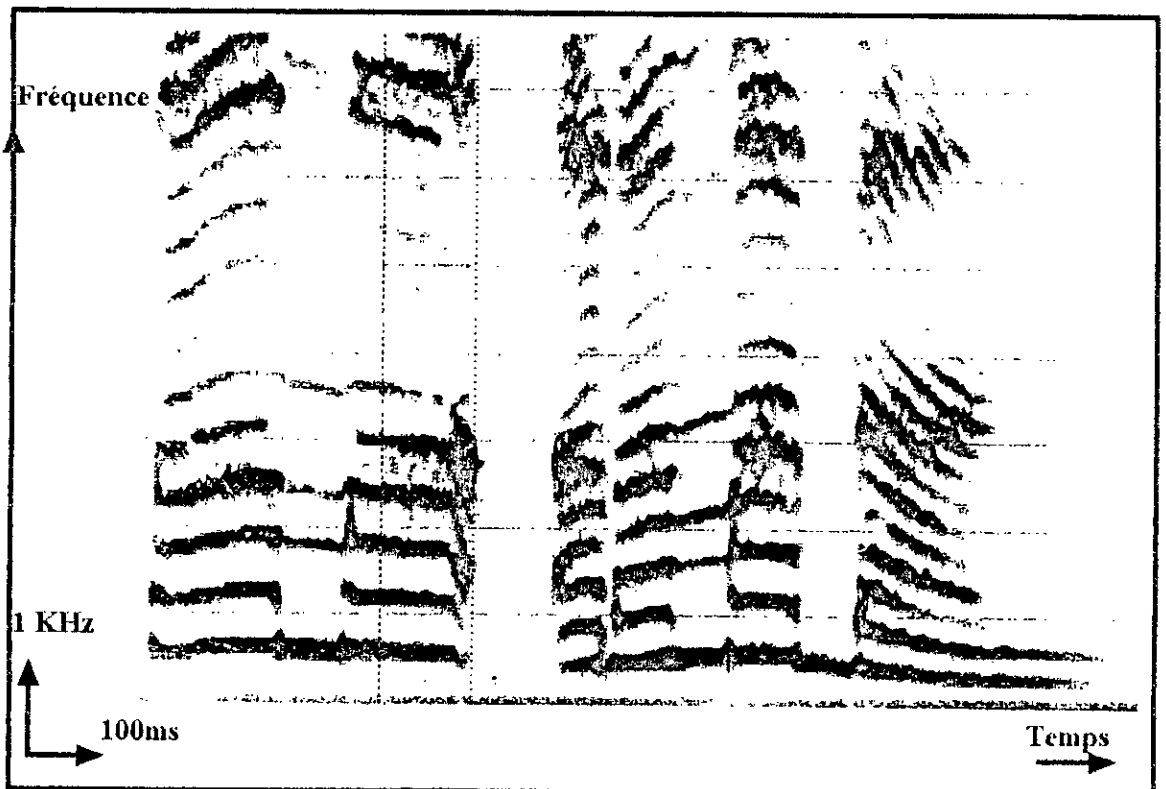


Fig.4.3 - Représentation temporelle de « père » sur le sonagramme (a). Nous pouvons aussi voir les courbes de  $F_0(t)$  en (b), de passage par zéro (c) et d'énergie (d).



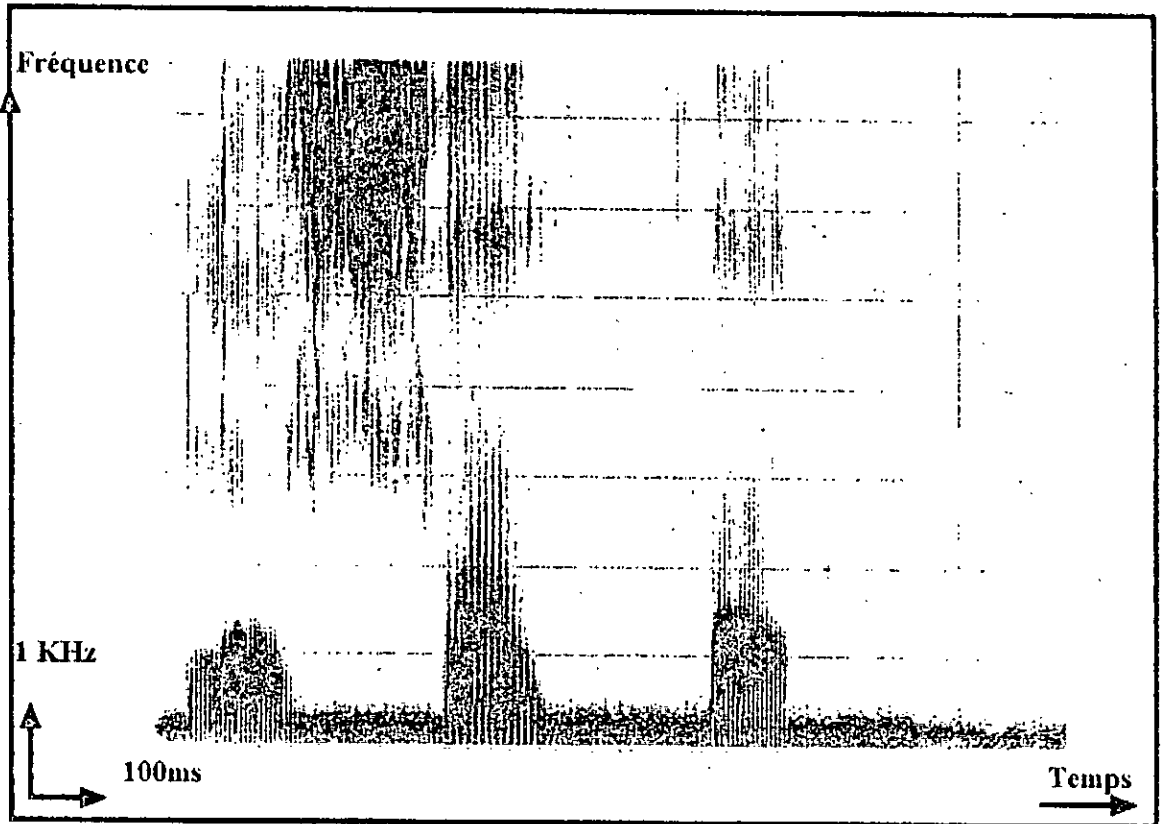
a- / lə ʃa sɔt /



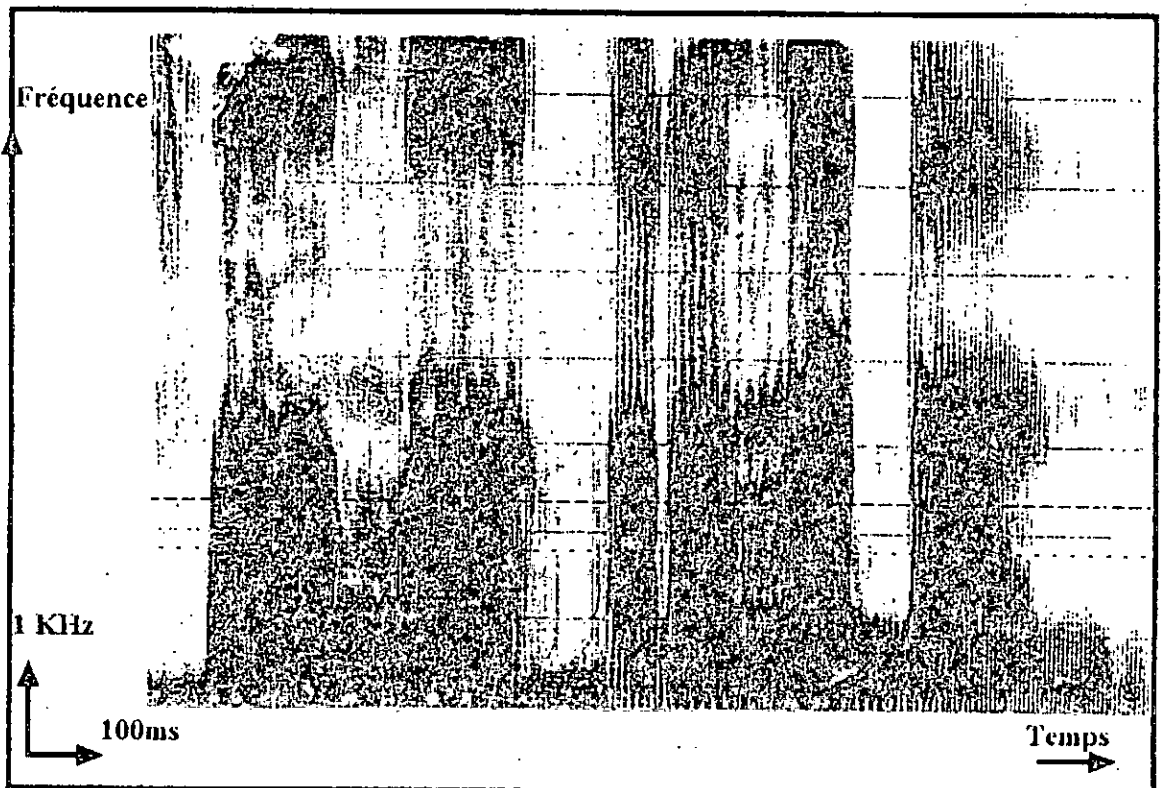
b- / Saama RamaDaan/

Fig.4.4 - Sonagramme à bandes étroites ( pour la lecture du pitch).





a- / lθ ja sɔt /



b- / Saama RamaDaan/

Fig.4.5 - Sonagramme à bandes larges ( pour la lecture des formants).

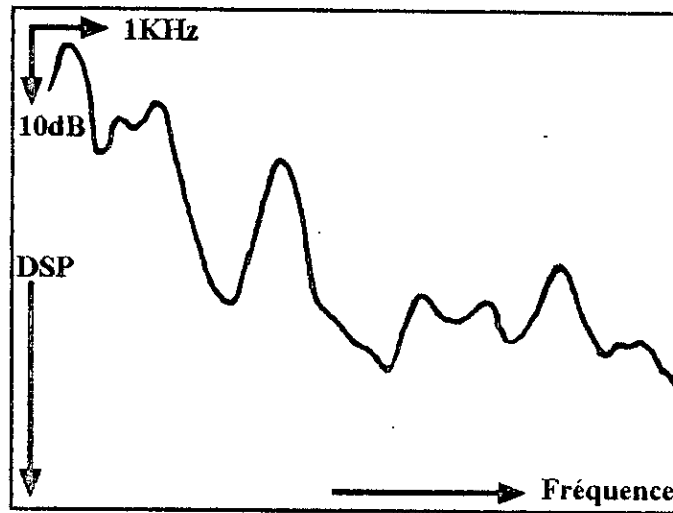


Fig. 4.6 - Densité spectrale de puissance du / a / obtenue sur le sonographe.

## 4.2.2 Seconde chaîne d'analyse

### 4.2.2.1 Chaîne pour l'extraction des formants

La chaîne comporte les éléments suivants (fig.4.2) :

- un magnétophone BRÜEL & KJAER de type 7005. C'est un enregistreur magnétique d'instrumentation à large bande;
- une carte d'acquisition HP 98640A. La numérisation du signal de parole en vue de son utilisation à l'ordinateur se fait au moyen de cette carte de conversion analogique numérique. Le programme d'acquisition et de conversion que nous avons élaboré est donné en annexe 2 ;
- un micro-ordinateur HP 9836 ;
- un disque dur dans lequel nous stockerons nos données numérisées en vue de leur utilisation future - notamment pour la phase d'analyse ;
- une imprimante HP 2631G.

#### 4.2.2.2 Le support logiciel Autostart (Autost)

Ce logiciel a été mis au point par Hewlett-Packard. Parmi les fonctions offertes par Autost nous avons :

- l'enregistrement et l'affichage sur l'écran du signal de parole pour contrôler et repérer les zones de parole utilisées (Fig.4.7) ;
- la segmentation et le zoomage du signal pour se positionner sur les frontières phonémiques désirées ;
- l'affichage sur l'écran des zones de parole qui seront soumises à l'analyse ;
- le calcul, par différentes méthodes d'analyse spectrale, des densités spectrales de puissance des zones vocaliques retenues (Fig.4.8).

Nous avons légèrement modifié le logiciel en y insérant un détecteur de crêtes de façon à avoir un affichage direct des formants et non une longue liste 256 valeurs de la DSP et des fréquences correspondantes (Tableau 4.4).

```

Pour I = 1 à 256
    Mettre dans une matrice A(I) les 256 valeurs de la DSP
    Dim A(256)
    I = 1
L1 I = I+1
    Si A(I) < A(I+1) aller à L1
    Si A(I) < A(I-1) aller à L1
    Sinon afficher I, DSP et valeur de la fréquence correspondant au formant
    Si I < 256 aller à L1
Fin
  
```

Tableau 4.4 - Algorithme du détecteur de crêtes

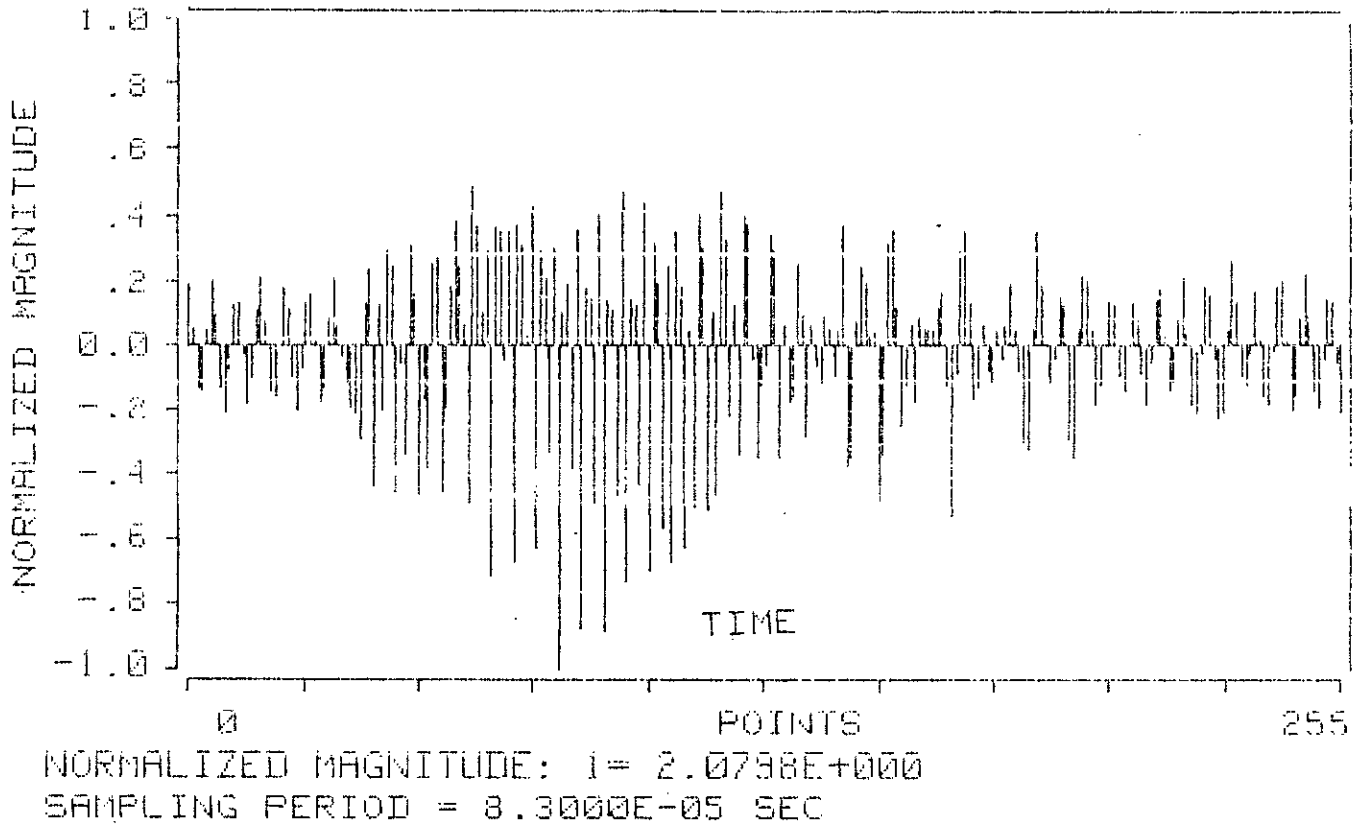


Fig.4.7 - Représentation temporelle du signal / pR /.

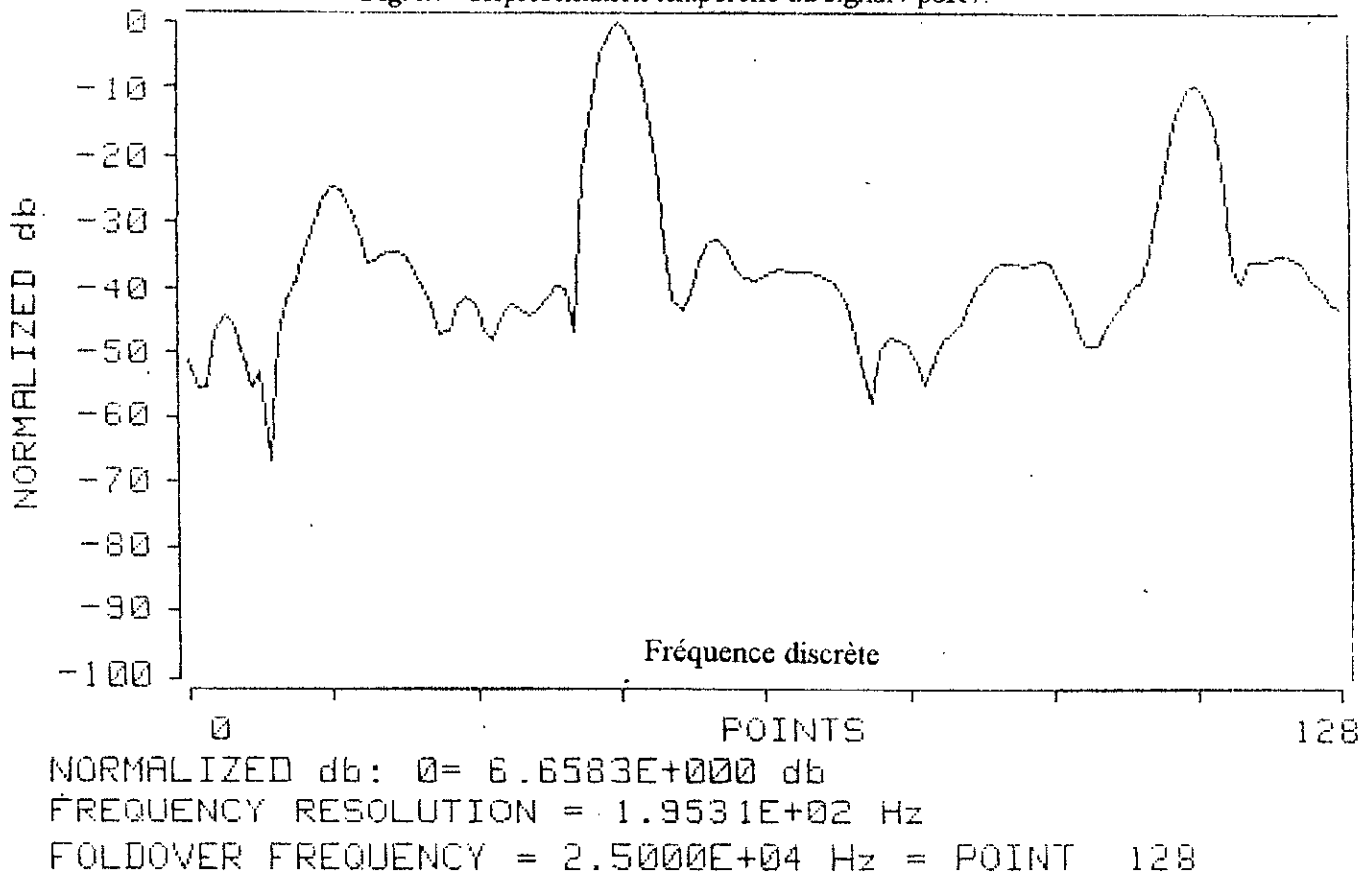


Fig.4.8 - Densité spectrale de puissance du phonème / ε / par la méthode du périodogramme.

### 4.2.2.3 Méthodes d'analyses retenues

Nous avons choisi d'analyser nos signaux, sur la seconde chaîne, au moyen des quatre méthodes de calcul de la DSP (le périodogramme et les trois modèles AR) que nous avons détaillées au chapitre deux. Puis nous comparons les résultats obtenus.

### 4.2.2.4 Choix des conditions d'analyse

Pour notre analyse, nous avons choisi :

- une fréquence d'échantillonnage de 12 KHz;
- pour tous nos signaux de parole, un nombre d'échantillons « N » égal à 51 que nous étendons ensuite à 250 par des zéros afin d'assurer une bonne résolution spectrale [9] ;
- la fenêtre de Kaiser ; elle est donnée par la formule :

$$W_k(n) = I_0(\alpha \sqrt{1 - [2n / (N - 1)]^2}) / I_0(\alpha) \quad (4.1)$$

avec :

$$- (N - 1) / 2 \leq n \leq (N - 1) / 2$$

Bien que la fenêtre de Hamming soit souvent utilisée en traitement de la parole, nous avons opté pour celle de Kaiser pour nos analyses [22], [24]. Le choix du paramètre «  $\alpha$  » égal à 8 et du nombre de termes dans la somme de la fonction de Bessel d'ordre zéro,  $I_0$ , égal à 10, permet un bon compromis entre la largeur du lobe principal et le niveau de l'amplitude du premier lobe secondaire (annexe 1).

- un ordre de prédiction « p » égal à 13 pour les différents modèles AR [9]. Mais pour certains échantillons (cf Ch5) il a fallu surestimer l'ordre pour pouvoir recueillir les formants.

## 4.3 Méthode de travail

Nous avons adopté des méthodes utilisées en identification de locuteurs. Ces dernières ont pour objectif l'extraction de paramètres propres à un locuteur donné (traits distinctifs). Nous ne cherchons à reconnaître ni un locuteur ni ce qu'il dit. Ces derniers sont connus. Notre but est relever les paramètres invariants de chaque locuteur avant l'effort. Ces derniers constitueront pour nous un repère. En effet, nous les comparerons avec les mêmes paramètres obtenus après l'effort. Et si une différence notable existe sur les traits obtenus avant et après l'effort, nous pouvons affirmer que la probabilité que le locuteur soit fatigué est élevée.

Divers facteurs font que la fréquence fondamentale d'un locuteur est très variable (cf Ch.3 § 3.1.2.4). Pour ne pas fausser l'interprétation des résultats, nous devons garder la constance de prononciation. De cette façon, les variations probables de  $F_0$  seront dues à l'influence de la fatigue et non à des variations naturelles de  $F_0$  qui relèvent de son émotion ou de traits relevant du code linguistique.

Nous insistons auprès des locuteurs pour qu'ils ne manifestent pas d'émotion et qu'ils contrôlent la hauteur, et l'intensité de leurs voix de façon à ne pas varier leur intonations. Nous leur avons aussi demandé de conserver le même rythme de prononciation des phrases à chaque enregistrement.

La première chaîne d'acquisition, basée sur le sonagraphe, permet l'obtention des caractéristiques articulatoires et prosodiques propres au locuteur. Ainsi, nous relèverons, pour chaque phonème, la durée, l'intensité et le pitch moyen. De même que nous mesurerons le temps de prononciation de la phrase entière. La seconde, par des méthodes spectrales, nous donne les formants et les énergies des échantillons de parole analysés. Il est vrai que la méthode du périodogramme n'a pas les performances d'un modèle AR mais, étant donné que nous comparons les résultats obtenus pour une même méthode, nous l'avons conservée pour notre étude.

## 4.4 Hypothèses formulées

Pour les langues française et arabe, les paramètres prosodiques réellement pertinents sont la hauteur du fondamental (qui définit l'intonation) et la durée des éléments acoustiques (qui fixe le rythme) [7]. Les indicateurs acoustiques retenus pour notre analyse sont la fréquence fondamentale, les formants, l'énergie ou la forme de son amplitude et la durée des phonèmes.

### 4.4.1 La fréquence fondamentale

La vitesse de fermeture et d'ouverture de la glotte conditionne la fréquence de vibration du son produit. La fréquence de la vibration dépend du degré de tension des cordes vocales. Plus elles sont tendues, et plus le rythme de fermeture et d'ouverture de la glotte s'accélère. Lors de chaque ouverture de celle-ci, les cordes vocales sont écartées l'une de l'autre. Leur écartement conditionne l'amplitude de la vibration. Quand la fatigue existe, la tension sur les cordes vocales est difficile à obtenir. La fréquence d'ouverture et de fermeture des cordes vocales devrait donc baisser.

### 4.4.2 Les formants

Nous n'émettons pas d'hypothèses particulières sur les variations des formants. Pour le mot saute (/sɔt/) , par contre, et par analogie avec le mot doute (/dut/) (cf Ch3 §3.1.6), la fréquence du deuxième formant devrait baisser [7].

### 4.4.3 Les énergies

L'intensité de la parole provient de l'énergie des poumons, et nous savons que lorsqu'un locuteur est fatigué, il y a une baisse de l'efficacité des contractions des muscles. Nous prévoyons une baisse de l'énergie lorsque le locuteur est fatigué.

#### 4.4.4 Durée des phonèmes

Nous avons mesuré le temps mis pour l'articulation des voyelles, des différentes consonnes, ainsi que le temps mis du silence de la plosive sourde bilabiale /t/ jusqu'à sa barre d'explosion. Pour nous le /t/ est final donc il n'y a pas de risque de confusion du silence de la plosive avec le silence pré-phonatoire (fig.4.5). Nous avons aussi mesuré le temps mis pour prononcer toute une phrase.

Nous nous attendons à ce que les durées de prononciation des voyelles soient réduites. Mais que celles des consonnes s'allongent. Le locuteur, sous l'effet de la fatigue, réduira la durée des premières qui sont faciles à émettre. Il éprouvera des difficultés pour les secondes, d'élocution un peu plus élaborée, et mettant en jeu plus de muscles.

## Conclusion

Nous avons pris pour la première classe de locuteurs un corpus composé d'un seul mot. Le nombre de locuteurs, par contre est important. Nous avons même inclus des enfants pour que l'étude sur la fatigue physique soit plus complète. L'enregistrement de ces derniers trois fois par jour nous donne donc une base de travail importante. Pour la seconde classe, le nombre de locuteurs est moins grand mais le corpus est plus élaboré. Nous avons utilisé, sur chaque chaîne, plusieurs méthodes d'analyse de façon à augmenter nos chances de détection des paramètres acoustiques sensibles à la fatigue.



# CHAPITRE 5

## RESULTATS OBTENUS ET INTERPRETATIONS

### Introduction

Nous présentons dans ce chapitre les résultats obtenus au moyen des différentes méthodes utilisées. Nous rappelons que nos locuteurs se répartissent en deux classes. L'une est composée de locuteurs susceptibles de présenter une fatigue résultant de l'activité physique, l'autre de ceux pouvant manifester une fatigue vocale.

### 5.1 Analyse des variations de la fréquence fondamentale sous l'effet de la fatigue

Toutes les mesures de pitch ont été faites sur le sonographe. Dans ce qui suivra, nous parlerons de fréquence  $F_0$  hachée ou coupée lorsque le graphe de celle-ci en fonction du temps ne donne aucune valeur pour des courts instants donnés (Fig 5.1). La coupure dans le graphe de  $F_0$  en fonction du temps est représentée par un astérisque sur le tableau de valeurs.

#### 5.1.1 Evolution du fondamental sous l'influence d'une fatigue non vocale

Les indices de  $F_0$ , que nous utilisons dans les tableaux qui suivent, correspondent aux différents instants de la journée 1 pour le matin, 2 pour la mi-journée et 3 pour le soir. Nous précisons que pour les besoins du graphe, et seulement lorsque le locuteur n'était pas présent pour l'enregistrement de la mi-journée, nous avons réutilisé les fréquences obtenues le matin en remplacement de celles de la mi-journée.

### 5.1.1.1 Variations du fondamental d'une locutrice sur dix jours d'enregistrement

Le graphe en figure 5.1.a et le tableau 5.1 donnent la variation de  $F_0$  d'une locutrice adulte sur dix jours d'enregistrement (J1 à J10). Celle-ci ne se dépensait pas beaucoup ; la fatigue existant en fin de journée correspond à une activité physique sans effort particulier.

Jours d'enregistrement	$F_{01}$	$F_{02}$	$F_{03}$
J1	204	219.4	194.12
J2	209.5	215	203.5
J3	205.8	211.87	194.5
J4	205.33	229.2	203.42
J5	214.6	218.16	201.6
J6	195.72	212.42	192.85
J7	208.4	224.83	199.57
J8	212	202	200
J9	194.5	207.33	177.8
J10	200	204	191

Tableau 5.1- Variation du fondamental (en Hz), le long d'une journée, du phonème /e/ en contexte /pε R/ pour dix jours d'enregistrement d'une locutrice.

### 5.1.1.2 Variations du fondamental pour des enfants

Le graphe de la figure 5.1.b et le tableau 5.2 donnent la variation de  $F_0$  au cours de la journée pour 15 enfants. Parmi ces derniers, 10 sont des sportifs (L1 à L5 et L9 à L13), les cinq autres (L6 à L8, L14 et L15) sont des malades hospitalisés. Le dernier enregistrement de la journée s'est fait juste après que les cinq locutrices L9 à L13 aient participé à une compétition sportive.

Locuteurs	Age (ans)	F <sub>01</sub>	F <sub>02</sub>	F <sub>03</sub>
L1	9	251.25	190.25	248.12
L2	10	252	240.50	231.66
L3	10	292.25	298.00	262.4 *
L4	10	232.88	208.50	223.00
L5	9	282.50	306.16	199.50
L6	12	228.33	232.57	243.85
L7	9	291.16	250.25	264.81 *
L8	10	231.00	258.25	200 *
L9	12	260.00	284.40	253.66 *
L10	11	280.00	268.50	177.42 *
L11	12	273.40	267.50	211.4 *
L12	12	266.40	202.75	109 *
L13	11	217.28	173.16	167.5 *
L14	13	250.71	228.66	247.93
L15	8	251.87 *	284.54	310.60

\* : hachure de F<sub>0</sub> (t)

Tableau 5.2- Variations du fondamental (en Hz), le long d'une journée, du phonème /ε/ en contexte / pε R/ pour des enfants.

### 5.1.1.3 Variations du fondamental de locuteurs adultes

Les graphes des figures 5.1.c et 5.1.d et les tableaux 5.4 et 5.5 donnent la variation de F<sub>0</sub> au cours de la journée pour 6 locuteurs (L17 à L22) et 8 locutrices (L27 à L34). Parmi ces derniers, 10 (L17 à L21) et (L27 à L31) sont des malades hospitalisés. Nous avons aussi enregistré 4 sportifs (L23 à L26) mais ils ne se sont présentés que deux fois : le matin et en milieu de journée. Nous n'avons donc, pour ces sportifs qui n'apparaissent pas sur le graphe, que la remontée de la fréquence fondamentale en milieu de journée par rapport à celle du matin.

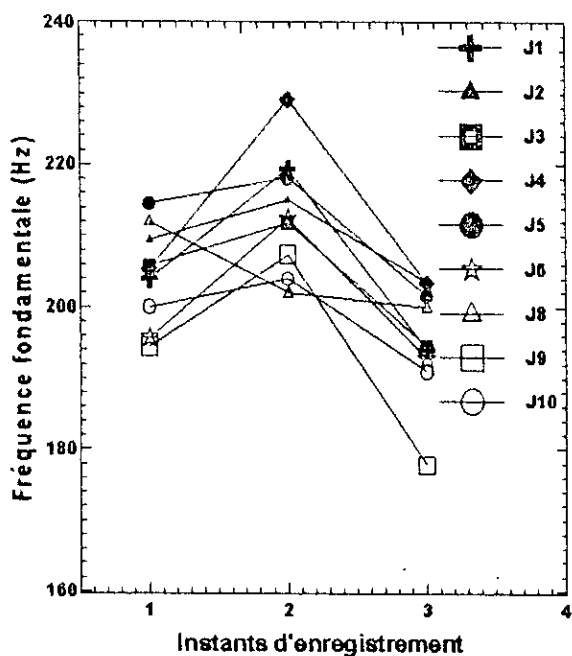
Locuteurs	Age (ans)	F <sub>01</sub>	F <sub>02</sub>	F <sub>03</sub>
L17	18	119.40	166.16	155
L18	51	89.5 *	93.85	99.88
L19	15	143.33	x	139.33 *
L20	15	245	197.87	223.33
L21	16	180.28 *	243.6	201.8 *
L22	61	131.44	133.77	122.25
L23	20	181.20	201.28	x
L24	21	163.57	165.2	x
L25	18	141.90	147.83	x
L26	16	149.33	131.77	x

Tableau 5.3 - Variations du fondamental (en Hz), le long d'une journée, du phonème /ε/ en contexte / pε R/, obtenues pour les locuteurs adultes.

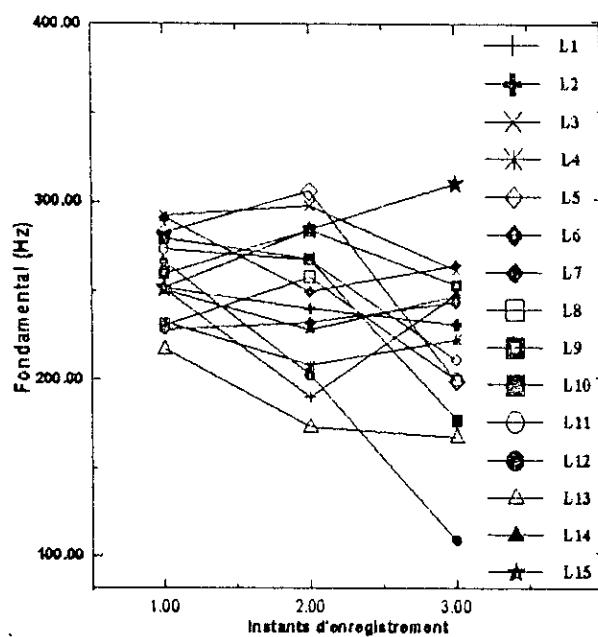
Locuteurs	Age (ans)	F <sub>01</sub>	F <sub>02</sub>	F <sub>03</sub>
L27	21	208.14	220.12	180.75 *
L28	24	251.9	229.85	285.72 *
L29	50	289.42	266	227.4 *
L30	22	200.42	214.44	175.25 *
L31	50	231.33 *	x	244.33 *
L32	62	226.33 *	206.33	206.25 *
L33	29	222 *	206	212.66 *
L34	33	281.9	275	x

X : le locuteur ne s'est pas présenté à l'enregistrement

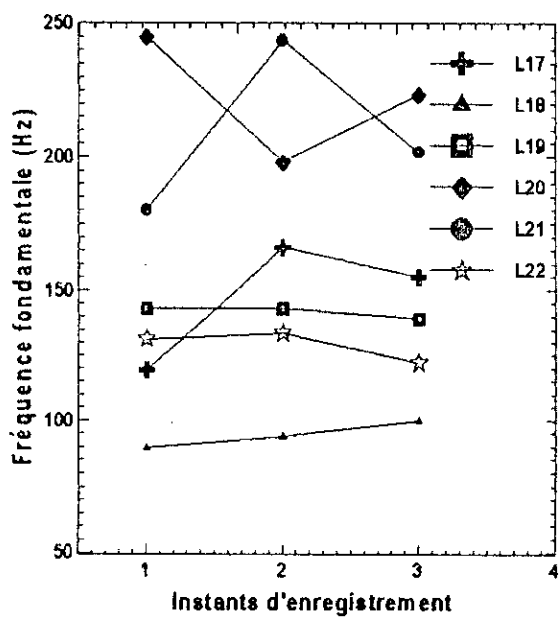
Tableau 5.4 - Variations du fondamental (en Hz), le long d'une journée, du phonème /ε/ en contexte / pε R/, obtenues pour les locutrices adultes.



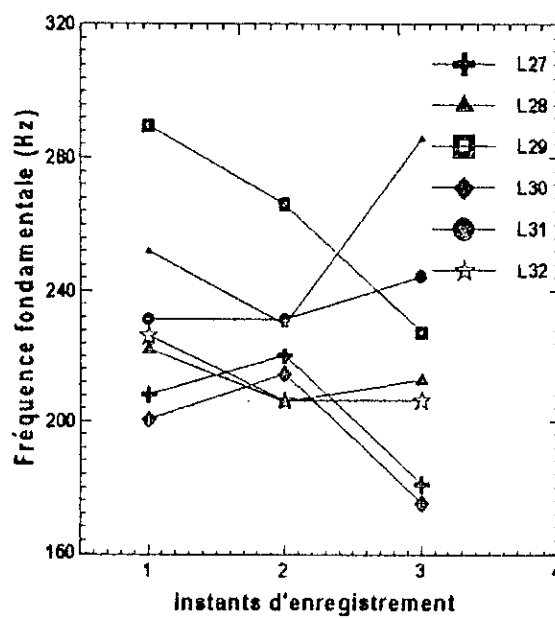
a - Variations pour une locutrice sur 10 jours



b - Variations pour 15 enfants



c - Variations pour 6 locuteurs



d - Variations pour 7 locutrices

Fig.5.1 - Variations du fondamental (en Hz), du phonème /ε/ en contexte /pεR/, au long d'une journée (1 : matin, 2 : mi-journée et 3 : soir).

### 5.1.1.4 Interprétation

Pour la majorité des locuteurs de la première classe, nous constatons une baisse de la fréquence fondamentale par rapport à celle du matin. En milieu de journée, la fréquence  $F_0$  est plus élevée que celle du matin. C'était prévisible car les enregistrements se faisaient après manger. Les muscles vocaux ont eu le temps de récupérer, de par l'apport de nourriture, l'énergie nécessaire à la contraction. Nous constatons que c'est pour les locutrices ayant fourni un effort physique important (L9 à L13) que nous avons le plus grand écart entre les pitches de début et de fin de journée. L'effort physique a donc entraîné une baisse de la fréquence fondamentale. De plus, quand la fatigue est prononcée, la courbe  $F_0(t)$  présente des hachures ou des coupures de l'ordre d'une vingtaine de millisecondes. Toutes les locutrices sportives ont présenté une hachure de  $F_0(t)$  de fin de journée.

## 5.1.2 Evolution du fondamental sous l'effet d'un effort vocal

Nous donnons deux valeurs différentes ; l'une obtenue au moyen de l'analyse en bandes étroites, l'autre en lisant la moyenne des valeurs obtenues sur la courbe de l'évolution de  $F_0$  en fonction du temps. Les résultats obtenus avant l'effort sont placés juste au dessus de ceux obtenus après.

### 5.1.2.1 Pitch des voyelles dans la phrase /lɔ̃ ʃa sɔ̃t/

Les résultats obtenus sont regroupés dans le tableau 5.5. Nous constatons que tous les locuteurs, sur la majorité des voyelles de la phrase, présentent une baisse du pitch après un effort vocal. De plus nous remarquons que certains présentent une plus grande fatigue que d'autres puisqu'il y a une hachure du graphe de  $F_0$  en fonction du temps. Pour certains locuteurs il y a aussi une remontée de la fréquence sur la voyelle /ɔ̃/ de la dernière syllabe. Ces derniers, dans ce cas, ont eu tendance à modifier leur ligne mélodique par habitude malgré les consignes d'une prononciation monocorde.

Méthodes Locuteurs	Lecture de la moyenne sur $F_0(t)$			Spectrogramme en bandes étroites		
	/ð/	/a/	/ɔ/	/ð/	/a/	/ɔ/
L1	266.6	305	178.33	260	300	200
	232 *	292 *	218.33	240	280	220
L2	176	173	184.25	180	160	200
	193	276	157	200	270	170
L3	227	187.25	221.7	240	180	220
	189 *	174.66	200.3*	180	170	200
L4	105	89	93.75*	100	100	80
	89 *	87 *	81 *	80	80	80
L5	117 *	142	110 *	117.5	142.5	110
	106.5*	106.6*	100 *	110	110	105
L6	95.33	118	121.25	90	120	120
	88	93.3 *	117.6*	80	90	120

Tableau 5.5 - Evolution du pitch (en Hz), sous l'effet d'un effort vocal, des voyelles de la phrase / lð ja sɔ/.

### 5.1.2.2 Pitch des voyelles de la phrase /Saama ramaDaan/

Nous avons éliminé l'enregistrement de L6 sur cette phrase ; il était de mauvaise qualité pour l'analyse. Pour la phrase en arabe aussi, nous relevons que le pitch diminue après un effort vocal et ce, pour la majorité des locuteurs. C'est pour le locuteur L4 que l'influence de la fatigue est la plus apparente. En effet,  $F_0$  diminue pour toutes les voyelles après l'effort et le graphe  $F_0(t)$  ne donnait plus aucune valeur sur les dernières voyelles bien que le signal soit parfaitement audible et compréhensible. Les résultats obtenus sont regroupés dans le tableau 5.6.

Méthodes Locuteurs	Lecture de la moyenne sur $F_0(t)$					Spectrogramme en bandes étroites				
	/aa/	/a/	/a/	/a/	/aa/	/aa/	/a/	/a/	/a/	/aa/
L1	249	243	222	276	160.33	250	240	200	280	160
	220	201	162	201	175.4	220	200	160	200	180
L2	182	199	199	182.33	151.66	180	200	200	180	150
	176	196.5	232	225.33	158.5	170	190	230	220	160
L3	204.25	250	262.25	200.33	134.25*	210	220	230	200	170
	189.2	220.4	210.5	197.66	173.83*	200	220	200	190	170
L4	99.66*	86	80	80	0	100	80	80	80	0
	92.66*	0	0	0	0	80	0	0	0	0
L5	120	110.3	102.33	97	93*	120	110	102.5	90	90
	111.66	103.66	98.5	102	101*	110	100	100	100	90

Tableau 5.6 - Evolution du pitch (en Hz), sous l'effet d'un effort vocal, des voyelles de la phrase /Saama ramaDaan/.

### 5.1.2.3 Pitch du phonème /m/ en contexte /Saama/

Les résultats montrant l'évolution du pitch du phonème /m/ en contexte montrent que 2 sur les 5 locuteurs ne présentent pas de baisse du pitch après l'effort. Un des deux pourtant ressent la fatigue puisqu'il y a une hachure du graphe  $F_0(t)$ . Le tableau 5.7 regroupe les résultats obtenus.

Locuteurs	$F_0$ (t)	Bandes étroites
L1	243	240
	301	280
L2	174.5	170
	176 *	170
L3	204	200
	185	180
L4	99	100
	87	80
L5	115	120
	107	100

Tableau 5.7 - Evolution du pitch (en Hz), sous l'effet d'un effort vocal, du phonème /m/ en contexte /Saama/.

### 5.1.2.4 Interprétation

Nous voyons que la fatigue vocale aussi entraîne une baisse de la fréquence fondamentale du locuteur. De plus, quand la fatigue est prononcée, la courbe  $F_0(t)$  aussi présente des hachures ou des coupures de l'ordre d'une vingtaine de millisecondes.

## 5.2 Influence de l'effort vocal sur les durées des phonèmes

Nous voyons sur les tableaux qui suivent la comparaison des temps mis pour la prononciation d'un phonème avant et après un effort vocal. Les résultats obtenus avant l'effort sont placés juste au dessus de ceux obtenus après.



### 5.2.1 Durées des phonèmes dans la phrase / lə ʃa sɔt /

Le tableau 5.8 récapitule l'évolution des durées d'élocution des différents phonèmes et de la phrase.

Locuteur	Durée de la phrase	/l/	/ə/	/ʃ/	/a/	/s/	/ɔ/	/t/	Pause
L1	1.30	0.13	0.0843	0.18	0.21	0.225	0.106	0.20	0.04
	1.40	0.11	0.084	0.17	0.19	0.18	0.112	0.17	0.20
L2	1.22	0.05	0.090	0.22	0.12	0.25	0.10	0.24	-
	1.00	0.11	0.08	0.14	0.11	0.19	0.09	0.17	-
L3	1.51	0.78	0.11	0.215	0.21	0.17	0.24	0.20	-
	1.74	0.15	0.12	0.2156	0.22	0.28	0.28	0.17	-
L4	1.10	0.11	0.097	0.170	0.130	0.18	0.14	0.17	-
	1.15	0.08	0.13	0.173	0.133	0.21	0.136	0.16	-
L5	1.55	0.087	0.093	0.15	0.20	0.21	0.13	0.20	0.3
	1.16	0.10	0.09	0.147	0.18	0.206	0.14	0.165	0.04
L6	1.27	0.11	0.06	0.19	0.112	0.206	0.122	0.125	0.4
	0.82	0.062	0.09	0.13	0.07	0.22	0.09	0.16	0.0

**Tableau 5.8** - Evolution de la durée des phonèmes (en secondes), sous l'effet d'un effort vocal, dans / lə ʃa sɔt /.

#### 5.2.1.2 Interprétation

Nous travaillons sur le nombre total de consonnes prononcées par les 6 locuteurs. Nous remarquons que la durée de 16 de celles-ci diminue alors que celle des 8 autres augmente. Donc les consonnes choisies, dans ce contexte particulier, diminuent lorsque l'effort vocal existe même lorsque la durée de prononciation de toute la phrase augmente. De la même façon, nous travaillons sur le nombre total de voyelles prononcées par les 6 locuteurs. Nous constatons que la durée de prononciation de 8 voyelles augmente alors que celle des 10 autres diminue.

### 5.2.2 Durées des phonèmes dans la phrase /Saama ramaDaan/

Nous n'avons pas pu délimiter avec certitude la durée d'élocution du phonème/n/ en fin de phrase. Les résultats obtenus sont regroupés dans le tableau 5.9.

Locuteurs Durée (s)	L1	L2	L3	L4	L5
	/S/	0.27 0.18	0.26 0.17	0.21 0.24	0.18 0.16
/aa/	0.13 0.17	0.15 0.11	0.31 0.33	0.12 0.16	0.32 0.25
/m/	0.10 0.096	0.11 0.09	0.10 0.09	0.078 0.071	0.073 0.078
/a/	0.14 0.15	0.22 0.09	0.24 0.22	0.037 0.12	0.087 0.078
/r/	0.087 0.080	0.07 0.02	0.13 0.16	0.035 0.034	0.024 0.021
/a/	0.081 0.084	0.056 0.072	0.084 0.10	0.031 0.21	0.057 0.078
/m/	0.078 0.078	0.09 0.084	0.096 0.075	0.07 0.062	0.06 0.059
/a/	0.093 0.091	0.11 0.08	0.10 0.13	0.071 0.09	0.07 0.09
/D/	0.047 0.075	0.075 0.09	0.122 0.07	0.056 0.053	0.073 0.078
/aa/	0.14 0.11	0.15 0.14	0.15 0.18	0.175 0.22	0.21 0.13
phrase	1.47 1.534	0.157 1.16	1.70 1.72	0.915 1.03	1.315 1.14

Tableau 5.9 - Evolution de la durée des phonèmes (en secondes), sous l'effet d'un effort vocal, dans /Saama ramaDaan/.

### 5.2.3 Interprétation

Nous constatons que, sur le nombre total de consonnes prononcées par nos 6 locuteurs, la durée d'élocution de 19 consonnes baisse alors que celle de 7 consonnes seulement augmente après un effort vocal. La diminution du temps de prononciation des consonnes se fait même lorsque celui de toute la phrase augmente. Sur le nombre total de voyelles prononcées, nous constatons que l'influence n'est plus aussi nette. Il n'y a en effet que 15 voyelles dont la durée augmente alors que celle des 12 autres diminue.

## 5.3 Evolution des énergies sous l'influence de l'effort vocal

Les valeurs des énergies des voyelles dans / l̥ɔ̃ʃa sɔ̃t / et / Saama / ont été obtenues au moyen du sonographe DSP 500. Nous précisons que toutes les valeurs des énergies des voyelles des tableaux 5.10 à 5.14 sont négatives et données en décibels. Dans chaque colonne, la barre sépare les résultats obtenus avant et après l'effort vocal. Les croix sur les tableaux (x) représentent les valeurs non obtenues.

### 5.3.1 Energies des voyelles dans la phrase / l̥ɔ̃ʃa sɔ̃t /

Locuteurs Formants	Locuteurs					
	L1	L2	L3	L4	L5	L6
F1	15/16	17/17	13/x	27/52	14/x	x
F2	35/38	42/43	32/40	43/51	39/45	36/42
F3	41/47	51/51	40/54	48/49	47/53	x
F4	45/47	52/54	54/57	52/54	53/50	48/49
F5	47/48	55/61	56/60	58/61	58/53	56/56
F6	x	x	56/57	x	x/58	57/60

Tableau 5.10 - Evolution de l'énergie (en dB), sous l'effet de l'effort vocal, de la voyelle /l̥ɔ̃/ des locuteurs.

Locuteurs Formants	Locuteurs					
	L1	L2	L3	L4	L5	L6
F1	28/28	20/22	17/27	41/35	26/29	18/21
F2	33/33	x/43	23/27	43/45	29/44	28/33
F3	50/x	50/54	39/36	46/48	43/53	44/49
F4	48/48	50/52	50/55	x/59	48/53	54/55
F5	55/56	51/x	53/x	x/63	63/58	52/52

Tableau 5.11 - Evolution de l'énergie (en dB), sous l'effet de l'effort vocal, de la voyelle /a/ des locuteurs.

Locuteurs Formants	Locuteurs					
	L1	L2	L3	L4	L5	L6
F1	23/24	20/20	14/18	33/38	22/23	x/19
F2	32/36	45/45	19/29	17/x	35/39	31/33
F3	42/46	53/57	45/x	60/62	55/56	49/52
F4	56/x	57/60	34/41	x/58	56/50	53/56
F5	63/51	61/63	41/44	x/58	x	58/60
F6			51/52		x/63	54/59
F7					61/63	55/58

Tableau 5.12 - Evolution de l'énergie (en dB), sous l'effet de l'effort vocal, de la voyelle /ɔ/ des locuteurs.

### 5.3.2 Energies des voyelles dans la phrase / Saama /

Locuteurs Formants	Locuteurs					
	L1	L2	L3	L4	L5	L6
F1	23/20	23/32	18/22	29/28	17/27	16/24
F2	25/28	29/40	18/22	35/29	23/41	23/31
F3	27/36	57/60	34/37	58/48	40/56	28/44
F4	40/41	56/66	42/43	64/54	37/48	52/59
F5	47/61	62/x	44/x		54/62	57/58
F6	54/61		44/49		50/x	63/57

Tableau 5.13 - Evolution de l'énergie (en dB), sous l'effet de l'effort vocal, de la voyelle longue /aa/ des locuteurs.

Locuteurs Formants	Locuteurs					
	L1	L2	L3	L4	L5	L6
F1	19/21	24/30	14/x	39/39	x/26	17/28
F2	27/30	30/38	x/21	47/40	28/43	x/22
F3	41/x	54/59	32/35	62/46	44/53	40/49
F4	34/50	58/63	47/49	x/54	43/50	53/53
F5	x/51	x	45/47	x	51/57	53/59
F6	63/52	61/x		x/58	x	x/58

Tableau 5.14 - Evolution de l'énergie (en dB), sous l'effet de l'effort vocal, de la voyelle brève /a/ des locuteurs.

### 5.3.3 Interprétation

Nous constatons, ainsi que nous le prévoyions, que pour tous les locuteurs et sur la majorité des phonèmes étudiés il y a une baisse de l'énergie des formants des voyelles étudiées.

## 5.4 Détermination et sélection des formants

La sélection des formants a été la partie la plus délicate et la plus longue de notre étude. En effet, il y a des formants dont la bande passante est trop large et d'autres dont le sommet d'énergie se partage en deux donnant naissance à deux formants de fréquences très proches. Nous nous sommes référés aux ellipses de dispersion des voyelles [7], [9], [13] et aux travaux antérieurs [45] pour éliminer les valeurs de formants aberrantes.

### 5.4.1 Formants de la première catégorie de locuteurs

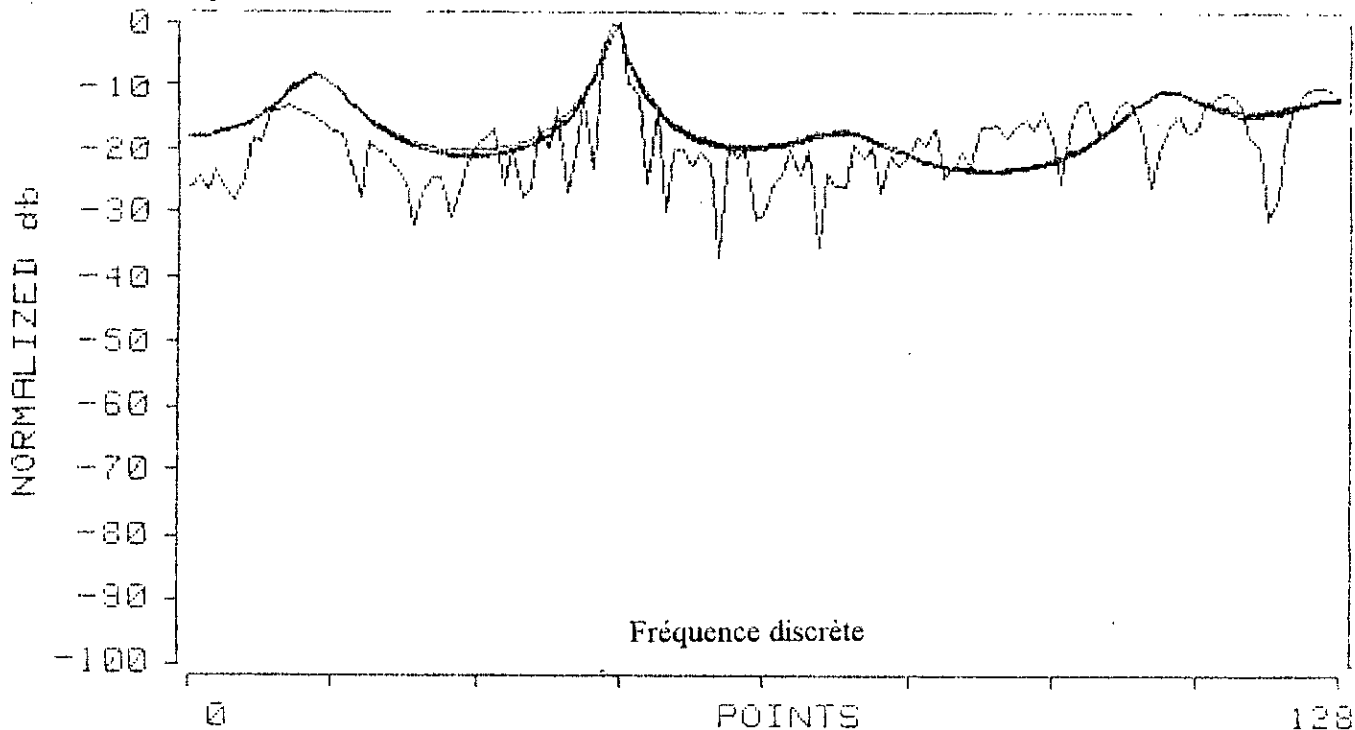
Les indices des fréquences, pour tous les tableaux qui suivent, correspondent aux différents instants de la journée : 1 pour le matin, 2 pour la mi-journée et 3 pour le soir.

#### 5.4.1.1 Comparaison entre les résultats obtenus par les différents modèles auto-régressifs

Pour nos calculs, nous avons utilisé des filtres AR d'ordre  $p = 13$ . Mais, pour certains échantillons de parole, il a fallu utiliser un ordre supérieur. Les indices, inscrits à côté des fréquences sur le tableau 5.15, indiquent l'ordre auquel il a fallu aller pour recueillir le formant. Nous voyons que les différences obtenues en utilisant les différents modèles ne sont pas significatives. Pour cela, dans la suite de notre travail, nous ne donnerons plus que les résultats obtenus pour un seul modèle AR : nous avons opté pour le modèle de Burg. Nous n'avons donné les résultats que pour une tranche de population et pour le premier formant seulement, mais la figure 5.7 montre bien que cela est valable pour tous les formants.

Locuteurs	Modèle auto-régressif		
	Morf	Burg	Levinson
<b>L9</b>	893.57 (14)	846.54	846.54
	987.63 (14)	846.54	987.63 (14)
	846.54	846.54	846.54
<b>L10</b>	846.54	846.54	846.54
	846.54 (16)	893.57 (15)	846.54(14)
	846.54	846.54	846.54
<b>L11</b>	940.60	940.60	940.6
	752.48 (14)	799.51	846.54(14)
	940.60	940.60	940.60
<b>L12</b>	846.54	846.54	846.54
	x	752.48	752.48
	799.51 (16)	846.54	799.51
<b>L13</b>	x	987.63	x
	846.54	846.54	846.54
	823.02	823.02	823.02
<b>L14</b>	846.54	846.54	846.54
	823.02	823.02	823.02
	917.08	917.08	917.08
<b>L15</b>	752.48	705.45	752.48
	863.57	893.57	893.57
	846.54 (14)	893.57 (14)	846.54(14)

**Tableau 5.15** - Variations de la fréquence F1 (en Hz), le long d'une journée, du phonème /ε/ en contexte /pε R/, de 10 enfants suivant le modèle AR utilisé.



NORMALIZED db: 0 = 5.3184E+000 db

FREQUENCY RESOLUTION = 2.1701E+02 Hz

FOLDOVER FREQUENCY = 2.7778E+04 Hz = POINT 128

**Fig.5.7** - Lissage du périodogramme, du phonème /ε/, par les modèles de Levinson, Morf et Burg.

### 5.4.1.2 Analyse des variations de la fréquence des formants sous l'influence d'une fatigue physique

Les tableaux ( 5.16 à 5.18) donnent les variations de la fréquence des formants par la méthode AR de Burg et par la méthode du périodogramme. Nous constatons que pour les deux méthodes, et pour toutes les tranches de population étudiées, nous n'observons plus le pic en mi-journée. Nous retrouvons pratiquement la même fréquence soit en début et en fin de journée, soit en début et en mi-journée ou alors carrément tout au long de la journée. Nous nous attendions donc à ce que la fatigue sur les muscles se répercutent sur la hauteur de la fréquence des formants des locuteurs. Or, d'après les résultats obtenus, la fréquence de ces derniers n'est apparemment pas altérée par ce type de fatigue.

Locuteurs	AR par Burg		Périodogramme	
	F <sub>1</sub> F <sub>2</sub> F <sub>3</sub>	F <sub>2</sub> F <sub>2</sub> F <sub>3</sub>	F <sub>1</sub> F <sub>2</sub> F <sub>3</sub>	F <sub>2</sub> F <sub>2</sub> F <sub>3</sub>
L17	799.51	2163.38	799.51	2210.41
	658.42	2304.47	705.45	2257.44
	658.42	2304.47	658.42	2257.44
L18	658.42	2304.47	564.36	2257.44
	681.93	2257.44	611.39	2210.41
	658.42	2257.44	564.36	2163.38
L19	799.51	2257.44	793.57	2210.41
	x	x	x	x
	799.51	2304.47	799.51	2210.41
L20	658.42	2210.41	658.42	2116.35
	752.48	2210.41	752.48	2257.44
	658.42	2257.44	705.45	2210.41
L21	658.42	2163.38	564.36	2210.41
	681.93	2163.38	611.39	2116.35
	658.42	2163.38	611.39	2210.41
L22	699.51	2304.47	746.54	2257.44
	699.51	2257.44	746.54	2210.41
	781.69	2210.41	787.63	2116.35

Tableau 5.16 - Variations des fréquences des formants (en Hz), le long d'une journée, du phonème /ε/ en contexte /pε R/, pour des locuteurs adultes.

Locuteurs	AR par Burg			Périodogramme	
	F1 <sub>1</sub>	F2 <sub>1</sub>	F1 <sub>1</sub>	F2 <sub>1</sub>	
	F1 <sub>2</sub>	F2 <sub>2</sub>	F1 <sub>2</sub>	F2 <sub>2</sub>	
	F1 <sub>3</sub>	F2 <sub>3</sub>	F1 <sub>3</sub>	F2 <sub>3</sub>	
L1	799.51	2210.41	793.57	2163.38	
	752.48	2539.62	699.51	2492.59	
	840.60	2069.32	793.57	2069.32	
L2	705.45	2304.47	752.48	2257.44	
	705.45	2304.47	752.48	2304.47	
	800.57	2163.38	893.57	2116.35	
L3	867.08	2492.59	846.54	2516.10	
	893.57	2398.53	887.63	2445.56	
	799.51	2445.52	840.60	2398.53	
L4	799.51	2257.44	893.57	2257.44	
	(15) 846.54	2304.47	846.54	2257.44	
	(15) 893.57	2257.44	846.54	2257.44	
L5	893.57	2304.47	846.54	2257.44	
	752.48	2163.38	752.48	2210.41	
	799.51	2210.41	799.51	2304.47	
L6	846.54	2163.38	893.57	2116.35	
	893.57	2257.44	840.60	2304.47	
	846.54	2257.44	893.57	2210.41	
L7	799.51	2210.41	799.51	2210.41	
	840.60	2304.47	847.63	2304.47	
	840.60	2351.50	847.63	2351.50	
L8	658.42	2163.38	752.48	2210.41	
	705.45	2163.38	752.48	2210.41	
	705.45	2163.38	752.48	2163.38	
L9	846.54	2163.38	803.57	2210.41	
	846.54	2351.50	803.57	2304.47	
	846.54	2351.50	658.42	2351.50	
L10	746.54	2163.38	752.48	2210.41	
	793.57	2210.41	740.60	2163.38	
	746.54	2233.92	793.57	2163.38	
L11	840.60	2163.38	840.60	2116.35	
	799.51	2304.17	799.51	2210.41	
	840.60	2163.38	840.60	2116.35	
L12	846.54	2445.56	799.51	2304.47	
	(15) 752.48	2304.47	799.51	2422.04	
	846.54	2304.47	846.54	2398.53	
L13	x	2539.62	840.60	2539.62	
	846.54	2351.50	846.54	2351.50	
	823.02	2539.62	846.54	2492.59	
L14	746.54	2210.41	793.57	2257.44	
	723.02	2398.53	7846.54	2398.53	
	767.08	2304.47	767.08	2304.47	
L15	705.45	2257.44	746.54	2257.44	
	793.57	2257.44	787.63	2257.44	
	793.57	2257.44	799.51	2257.44	

Tableau 5.17 - Variations des fréquences des formants (en Hz), le long d'une journée, du phonème /ε/ en contexte /pε R/, pour des enfants.



Locuteurs	AR par Burg		Périodogramme	
	F11	F21	F11	F21
	F12	F22	F12	F22
	F13	F23	F13	F23
L28	658.42	2304.47	681.93	2257.44
	(15) 752.48	2210.41	752.48	2116.35
	658.42	2210.41	658.42	2210.41
L29	890.60	2257.44	893.57	2257.44
	867.08	2304.47	890.60	2257.44
	893.57	(15)2022.29	846.54	2163.38
L30	611.39	2116.35	611.39	2257.44
	846.54	2116.35	799.51	2069.32
	893.57	(15)1975.26	752.48	2022.29
L31	705.45	2116.35	564.36	2163.38
	705.45	2210.41	752.48	2210.41
	705.45	2163.38	752.48	2257.44
L32	(17) 799.51	2163.38	846.54	2163.38
	x	x	x	x
	752.48	2163.38	752.48	2163.38
L33	658.42	2257.44	658.42	2210.41
	705.45	2445.56	705.45	2445.56
	705.45	2398.53	752.48	2351.50
L35	752.48	2092.83	775.99	2163.38
	846.54	2116.35	846.54	2116.35
	752.48	2069.32	752.48	2069.32

Tableau 5.18 - Variations des fréquences des formants (en Hz), le long d'une journée, du phonème /e/ en contexte /pɛ R/, pour des locutrices adultes.

### 5.4.2 Evolution des fréquences des formants sous l'influence de l'effort vocal

En utilisant le DSP sonograph, nous obtenons les formants au moyen de deux méthodes. La première se fait par lecture directe des valeurs au milieu des bandes noires. La seconde, permet l'obtention des pics des formants sur la courbe donnant la densité spectrale du signal étudié. Nous précisons que toutes les lectures de fréquences des formants ont été faites avec une précision de  $\pm 40$  dB. Dans chaque colonne, la barre sépare les résultats obtenus avant et après l'effort vocal. Pour les valeurs lues sur la seconde chaîne nous n'avons pas pu délimiter les différentes consonnes. Nous nous limiterons donc à l'étude des voyelles des phrases en arabe et en français. Toutes les croix sur les tableaux (x) montrent que la valeur n'a pu être déterminée.

### 5.4.2.1 Evolution des formants des voyelles dans la phrase

/lɔ̃ ʁa sɔ̃t/

a - par lecture directe

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	440/520	440/440	520/520	600/600	400/640	440/440
F2	1720/1760	1720/1840	1840/1800	1400/1600	1520/1480	1520/1440
F3	2560/2520	2640/2600	2640/2640	2200/2200	2520/2240	2040/2080
F4	3520/3600	3400/3400	3880/3840	3360/3360	3160/3200	3000/2960
F5	4560/4560	4200/4360	4960/4880	5440/4320	3800/3840	3920/3880
F6	x	x	6200/6120	x	5120/5120	4920/4840
F7	x	x	7240/7200	x	x	x/5880

Tableau 5.19 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /ɔ̃/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	840/840	600/600	960/840	600/760	560/680	520/600
F2	1480/1480	1800/1720	1440/1480	1400/1440	1320/1360	1240/1160
F3	2720/x	3080/3040	2960/2880	2200/2320	2480/2480	2000/2120
F4	3880/3880	4240/4200	4160/4560	3360/3520	3560/3520	3240/3440
F5	4640/4720	x	5120/x	4440/4440	5800/5480	4120/4200
F6	x	x	6600/6520	x	x	x

Tableau 5.20 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /a/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	520/600	560/560	600/520	520/600	440/480	480/520
F2	1400/1320	1480/1480	1240/1080	1160/1080	1040/1040	1080/1080
F3	x	2920/2760	2280/x	2160/2240	2680/2480	2120/2080
F4	4000/3840	3800/3800	3200/3240	3360/3320	3640/3360	3200/3080
F5	x/5160	x/4560	4000/3920	6040/6000	5840/x	3840/3840
F6	x	x	5240/5080	x	x	4960/4960
F7	x	x	x	x	x	6040/5880

Tableau 5.21 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /ɔ̃/ des locuteurs

## b - par lecture de la densité spectrale de puissance sur le sonographe

Locuteurs Formants	L1	L2	L3	L4	L5	L6
	F1	440 /360	360/440	440 / x	400 /760	360/ x
F2	1720/1760	1600/1760	1840/1720	1600/1600	1480/1480	1480/1480
F3	2560/2640	2720/2640	2600/2640	2160/2200	2480/2280	x
F4	3520/3560	3560/3520	3840/3760	3280/3280	3000/3280	2920/2920
F5	4560/4560	4160/4240	4960/4840	4240/4320	3760/3680	3840/3840
F6	x	x	6120/6080	x	x /5040	4920/4880
F7	x	x	x	x	x	x /5840

Tableau 5.22 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /a/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
	F1	840/880	600/600	1040/1000	560 /680	560/560
F2	1480/1480	x /1680	1440/1440	1360/1400	1280/1240	1120/1120
F3	2720/ x	2840/2920	2920/2880	2240/2280	2440/2480	1960/2160
F4	3680/3760	3760/4160	4040/4520	x /3520	3600/3520	3200/3360
F5	4640/4720	4880/ x	5120/x	x /5520	5720/5360	4120/4320
F6	x	x	6560/ 6480	x	x	x

Tableau 5.23 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /a/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
	F1	520 /560	560/560	600/ x	520 /800	440/400
F2	1400/1280	1360/1440	1160/960	1120/ x	1000/800	1040/1080
F3	x	2880/2720	2160/ x	2160/2240	2680/2480	2080/2080
F4	3200/3280	3760/3760	3240/3280	x /3280	3200/3280	3120/3160
F5	4000/ x	x	3920/3840	x	x	3840/3840
F6	5280/5120	4960/4520	5120/5040	x	x /5040	4960/4920
F7				x /6000	5880/6080	5840/5840

Tableau 5.24 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /ɔ/ des locuteurs

## c - par la méthode du périodogramme

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	799.51/705.45	x/705.45	799.51/752.48	752.48 / x -	376.24/ x	752.48/ x
F2	2022.29/1881.20	1599.02/1504.96	1881.20/1881.20	x	x	1928.23/2069.32
F3	2586.65/2633.68	2962.89/2915.86	2727.74/2633.68	2351.50/2492.59	2633.68/2351.50	x
F4	x/3433.19	3574.28/3715.37	3668.34/3433.19	4044.58/3809.43	3339.13/3480.22	3621.31/3480.22
F5	4608.94/ x	4185.67/4326.76	5690.43/5126.27	5032.21/4844.09	4514.88/4750.03	4703.00/4232.70
F6	x	x	x	x	5314.39/ x	x /5314.39

Tableau 5.25 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /ð/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	658.42 / x	x /611.39	x /611.39	752.48/ x	517.33/752.48	x /658.42
F2	1551.99/1269.81	2163.38/2304.47	1787.14/1410.90	1457.93/1175.75	x /1646.05	x /1410.90
F3	2633.68/2586.65	x	2304.47/2445.56	2962.89/2915.86	2304.47/2915.86	2304.47/2257.44
F4	3715.37/3950.52	3245.07/3151.01	3621.31/3386.16	3997.55/3997.55	3480.22/3997.55	3009.92/2962.89
F5	4797.06/4655.97	4373.79/4326.76	x /4608.94	4514.88/4985.18	x	4420.82/3950.52
F6	x	5455.48/5831.72	x /5314.39	5267.36/5549.54	5079.24/5126.27	x

Tableau 5.26 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /a/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	705.45/611.39	893.57/799.51	564.36/658.42	x /376.24	658.42/940.6	x /376.24
F2	1175.75/ x	x	x /1175.75	x /1034.66	x	1363.87/1316.84
F3	2539.62/2492.59	2774.77/2586.65	2586.65/2210.41	2445.56/2022.29	x /2398.53	x /2398.53
F4	3997.55/4185.67	3809.43/3386.16	3715.35/3292.10	x	x	x /3856.46
F5	x	4561.91/4420.33	4844.09/4514.88	4044.58/4185.67	4467.85/4420.82	4326.76/4703.00
F6	5408.45/ x	x	5690.63/5502.51	5549.54/5220.33	5267.36/5032.21	5220.33/5549.54

Tableau 5.27- Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /ɔ/ des locuteurs

## d - par le modèle auto-régressif

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	x /517.33	611.39/611.39	x	517.33/ x	658.42/ x	517.33/ x
F2	1975.26/1787.14	1363.87/ x	1975.26/1975.26	x	x	1975.26/ x
F3	2586.65/2633.68	2774.77/2962.89	x	2539.62/2727.74	2633.68/2539.62	x
F4	x /3339.13	3386.16/ x	3527.25/3574.28	x	x	3715.37/3574.28
F5	x	4232.70/4420.82	x	x	4514.88/4797.06	4703.00/ x
F6	x	x	5596.57/5220.33	5079.24/ x	5455.48/ x	5220.33/5267.36

Tableau 5.28 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /ə/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	611.39 /987.63	x / 893.57	x /517.33	x	470.30/705.45	423.27/ x
F2	1551.99/1363.87	x	1410.90/1551.99	x / 1128.72	x	x /1316.84
F3	x	2210.41/2257.44	2351.5/2680.71	x	2351.50/ x	2210.41/2257.44
F4	3715.37/3856.46	3245.07/3151.01	x /3433.19	x /3997.55	3574.28/3903.49	3009.92/3009.92
F5	x /4703.00	4326.76/4232.70	4279.73/4561.91	4467.85/ x	5032.21/5079.24	4514.88/5361.42
F6	5220.33	5455.48/ x	5173.30/ x	5173.30/ 5032.21		

Tableau 5.29- Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /a/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	x /564.36	x /752.48	517.33/658.42	* 564.36/517.33	329.21/940.6	x /329.21
F2	940.60/ x	x	x	x	x	x
F3	2351.50/2586.65	2727.74/2633.68*	2633.68/2116.35	x	2633.68/2445.56	x / 2445.56
F4	3997.55/4185.67	3668.34/3386.16	3668.34/3292.10	x	3997.55/ x	x
F5	x	4608.94/4420.82	4844.09/4608.94	4185.67/4232.70	4373.79/4326.76	x /4750.03
F6	5361.42/ x	x	x /5502.51	5314.39/5267.36	5220.33/5032.21	5126.27/ x

Tableau 5.30- Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle /ɔ/ des locuteurs

### 5.4.2.2 Evolution des formants des voyelles dans /Saama /

a - par lecture directe

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	760 /600	760/760	1080/960	640 /760	560 /520	680/600
F2	1240/1240	1120/1280	1440/1400	1280/1240	1160/1120	1160/1120
F3	x	2160/2600	3040/3000	2440/2400	2680/2680	2120/2280
F4	x /3520	3800/3720	3840/x	3600/3880	3640/3600	3200/3320
F5	4040/4120	x	x	x	x	4000/ x
F6	4800 / x	4960/ x	x /4600	x	5080/5160	5120/5080
F7	x / 5680	x	5280/ x	x	5880/5800	x

Tableau 5.31 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle longue /aa/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	560/320	760 /760	800 /520	600 /800	760/ 600	640/ x
F2	1440/1520	1440/1400	1360/1400	1120/1410	1240/1200	x
F3	3080/ x	3080/2800	3000/3000	2000/2240	2680/2680	2120/2240
F4	3760/3560	3840/3600	x	x /3800	3640/3200	3920/3320
F5	x /4240	x	4320/4520	x	x /5320	x /4560
F6	x /5800	5000/ x	5360/5440	x	5960/5880	x /5920

Tableau 5.32 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle brève /a/ des locuteurs

b - par lecture de la densité spectrale de puissance sur le sonographe

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	720/320	720 /760	1080/960	560 /720	560/640	640/600
F2	1460/1240	1120/1280	1440/1400	1200/1200	1120/1120	1120/1080
F3	x	2240/2660	x	2360/2400	2680/2680	2080/2160
F4	3520/3520	3760/3760	3840 / x	3640/3840	3600/3520	3200/3360
F5	4800/ x	4880/ x	x /4640	x	x	4000/4640
F6	x	x	5200/ x	x	x /5440	5120 5080
F7	x /5680	x	6200/6160	x	5840/ x	

Tableau 5.33 - Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle longue /aa/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	640/320	720 /760	960 / x	560 /800	x /640	640/600
F2	1440/1520	1160/1320	x /1200	1160/1360	1120/1160	x /1000
F3	3040/ x	3040/2800	3040/2960	2080/2160	2600/2600	2120/2240
F4	3680/3600	3840/3600	x	x /3840	3520/3520	3960/3320
F5	x /4240	x	4320/4480	x	x	4960/4560
F6	x /5760	5000/ x	5360/5480	x /5600	5880/5560	x /5920

Tableau 5.34- Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle brève /a/ des locuteurs

c - par la méthode du périodogramme

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	564.36/ x	423.27/705.45	x	658.42 /893.57	658.42/799.51	x /658.42
F2	1646.05/ x	1599.02/ x	1975.26/1928.23	1551.99/1410.90	1457.93/ x	x /1128.72
F3	2304.47/2022.29	2304.47/2821.80	x	2445.56/2586.65	2304.47/2116.35	x /2398.53
F4	3480.22/3245.07	3997.55/3386.16	3103.98/3103.98	3668.34/3386.16	3856.46/3715.37	3056.95/3151.01
F5	4091.61/3856.46	4985.18/4373.79	4703.00/4750.03	4514.88/4420.82	x /4279.73	4232.70/4797.06
F6	4608.94/4467.85	5502.51/5784.69	5502.51/ x	x	5831.72/ x	

Tableau 5.35- Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle longue /aa/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	893.57/ x	987.63/658.42	470.30/x	893.57/752.48	658.42/658.42	x
F2	x /1457.93	x /1599.02	1559.02/1551.99	x	1363.87/1316.84	x /1128.72
F3	2586.65/2774.77	x	2304.47/ x	2351.50/2304.47	2210.41/2351.50	2445.56/ x
F4	3527.25/ x	3903.49/3621.31	3198.04/3856.46	x /3151.01	3245.07/ x	3856.46/3856.46
F5	4232.70/4655.97	4514.88/4373.79	x /4750.03	4514.88/4514.88	5596.57/5032.21	4420.82/4938.15
F6	5267.36/5267.36	5126.27/5267.36	x	5596.57/5502.51	x /5831.72	5267.36/ x

Tableau 5.36- Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle brève /a/ des locuteurs

## d - par le modèle auto-régressif

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	564.36/ x	423.27/ x	940.6/987.63	658.42/846.54	517.33/846.54	x
F2	1551.99/ x	1175.75/ x	1975.26/1740.11	1504.96/1740.11	1457.93/ x	1457.93/ x
F3	2398.53/2163.38	x	x	x	2304.47/2116.35	x /2257.44
F4	3386.16/3245.07	3292.10/3386.16	3056.95/3056.95	3480.22/3339.13	3950.52/ x	3009.92/ x
F5	x	4138.64/4373.79	4467.95/4467.95	4561.91/4326.76	x /4750.03	4373.79/4655.97
F6	x /5079.24	x	5455.58/ x	x	5173.30/ x	

Tableau 5.37- Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle longue /aa/ des locuteurs

Locuteurs Formants	L1	L2	L3	L4	L5	L6
F1	893.57/ x	x	470.3/ x	893.57/705.45	611.39/658.42	x
F2	x /1504.96	x /1693.08	1599.02/1316.84	x	1363.87/ x	1363.87/1175.75
F3	2680.71/2868.83	3009.92/2774.77	2304.47/2962.89	2445.56/2445.56	x /2539.62	2445.56/2962.89
F4	3621.31/ x	x /3668.34	x	3856.46/ x	3292.10/ x	3527.25/3856.46
F5	4326.76/4561.91	4750.03/4279.73	x	4938.15/4561.91	x	4420.82/4938.15
F6	5220.33/ x	x	5361.42/5549.54	5549.54/5455.48	5596.57/4891.12	x

Tableau 5.38- Evolution de la fréquence des formants (en Hz), sous l'effet d'un effort vocal, de la voyelle brève /a/ des locuteurs

## 4.2.2.3 Interprétation

Pour certains formants la lecture a été impossible du fait de leur grande baisse d'énergie par rapport aux autres de même enregistrement. De plus, souvent la lecture du premier formant n'a pu être faite de par sa fusion avec le formant le plus proche  $F_0$  ou  $F_2$ .

Les deux moyens d'obtention des formants sur le sonographe donnent des résultats approchants. Ces résultats sont de meilleure qualité que ceux obtenus par le modèle AR ou par le périodogramme. En effet, le système d'écoute de l'appareil permet de bien se positionner sur le phonème désiré.



La méthode du périodogramme a donné les plus mauvais résultats. On y trouve les plus grands écarts dans la variation des formants. Le modèle AR, qui permet une bonne analyse même quand le nombre des échantillons n'est pas important, réduit cet écart existant entre les valeurs obtenues avant et après l'effort. Si la deuxième chaîne permettait l'écoute du phonème analysés, ce serait par le modèle AR que l'on aurait les meilleurs résultats.

Les différents moyens d'obtention des formants montrent que pour la majorité des locuteurs, pour tous les formants et pour les différents phonèmes, la valeur des formants ne semble pas varier de façon notable. Au vu de tous ces résultats nous ne pouvons affirmer, avec certitude, une influence de la fatigue sur la valeur des fréquences des formants.

## Conclusion sur l'étude entreprise

Pour notre travail, nous recommandons aux locuteurs de contrôler leurs voix de façon à éviter les variations macroprosodiques et de réduire au possible les variations dans la ligne de déclinaison en fin de phrase. Nous constatons, au terme de ce travail, que la fatigue introduit des variations microprosodiques au niveau des différents phonèmes que nous avons étudiés.

Ainsi, la fatigue influe surtout sur les consonnes - dans les contextes que nous avons choisis ; elles voient leur durée diminuer. Ni la durée de la phrase ni celle des voyelles ne montrent une si grande influence. Les travaux précédents [1], [2] ont montré que la durée de la phrase augmente sous l'effet de l'exercice puis baisse quand l'exercice devient plus complexe. Dans notre étude, nous ne pouvons apprécier le degré de difficulté du travail de chacun des locuteurs. Les phonèmes les plus sensibles à un effort vocal, dans les contextes que nous avons choisis, sont surtout /m/ /r/ /l/ /j/ et le silence avant le burst du phonème /t/. En effet la durée de ces derniers a diminué pour pratiquement tous les locuteurs même lorsque la durée de la phrase augmentait.

Nous voyons aussi qu'après une fatigue, apparaissant après une activité physique ou un effort vocal, les locuteurs présentent une baisse de la fréquence fondamentale. Quand la fatigue est prononcée il semble, et ce pour les deux classes de locuteurs, que les cordes vo-

cales ne fonctionnent plus normalement pour de très courts instants. En effet la courbe de l'amplitude du pitch en fonction du temps ne donne aucune valeur pour de très courts instants comme si le son était non voisé. Il faudrait confirmer cela avec un laryngoscope. Nous remarquons que dans ce cas là, cela ne nuit pas à la compréhension des locuteurs. La fatigue atténuant les contractions jusqu'à les arrêter quand il n'y a pas de récupération d'énergie, le comportement de  $F_0$  était prévisible.

De plus et ainsi que nous le prévoyions, la fatigue entraîne une baisse dans l'énergie des formants des voyelles dans les contextes que nous avons choisis. L'analyse par le sonagramme a montré, pour les deux catégories de locuteurs et contre toute attente, que la fatigue n'avait aucune influence sur la position des formants.

Nous pensons qu'une compensation se fait au niveau des résonateurs du locuteur pour que l'interlocuteur perçoive bien le son désiré par le locuteur même s'il a été mal produit. Nous pensons que la représentation mentale des sons a suffi pour atténuer les défaillances dues à la fatigue qui peuvent apparaître au niveau des muscles vocaux.

Ainsi, si la redondance contenue dans la parole est une protection contre les perturbations environnant le locuteur, la représentation mentale des sons est une protection contre les défaillances éventuelles - quelles qu'en soient les causes - du locuteur lui-même. Nous voyons là les formidables possibilités du corps humain et pourquoi une synthèse de qualité est difficile à obtenir. Cette étude, nous permet d'apprécier toute la richesse et l'efficacité de la représentation mentale.

## Conclusion générale

La fatigue a fait l'objet de nombreux travaux scientifiques. Mais les recherches se sont limitées aux analyses biomédicales. Par ce travail, nous voyons une approche différente dans le traitement du sujet. Notre but est de rechercher les paramètres acoustiques de la parole sensibles à la fatigue. Pour le mener à bien, nous avons été amenés à créer une base de données de sons dont le laboratoire manquait. Celle-ci, assez conséquente est composée de voix d'enfants et d'adultes des deux sexes. Elle est mise à la disposition de tout futur utilisateur.

Nous avons recommandé aux locuteurs de ne pas manifester d'émotion et de contrôler la hauteur et l'intensité de leurs voix et de ne pas varier leur intonation. Nous leur avons aussi demandé de conserver le même rythme de prononciation des phrases à chaque enregistrement. De cette façon nous évitons les variations macroprosodiques. Nous avons ensuite comparé les différents phonèmes prononcés avant et après un effort.

Puis nous avons analysé ces derniers au moyen de différentes méthodes : le spectrogramme, le périodogramme et le modèle autorégressifs de Morf, Levinson et Burg [7]. Pour analyser nos échantillons de parole, nous avons travaillé en comparant des paramètres invariants qui sont les traits distinctifs propres à chaque locuteur. Nous avons pris comme paramètres d'étude la fréquence fondamentale, les formants, les énergies et les durées de prononciation de différents phonèmes et de phrases entières.

Les différents modèles AR ont donné des résultats sensiblement approchants et meilleurs que ceux obtenus par la par la méthode du périodogramme. C'est par la méthode DSP sur le sonagraphe que nous voyons le mieux que les fréquences des différents formants tendent à rester stables. Les résultats des analyses ont montré une baisse des valeurs des paramètres prosodiques. L'influence de la fatigue sur les formants est moins marquée.

La sélection des formants, la partie la plus délicate et la plus longue de notre étude, a été un des problèmes que nous avons rencontrés dans notre travail. De plus, nous avons

fortement ressenti la nécessité d'un système d'écoute sur la chaîne d'acquisition du laboratoire de Traitement de Signal de l'ENP-Alger.

Les résultats de notre étude profitent aux locuteurs exécutant de tâches à haut risque comme les aviateurs ou les chirurgiens. Cela peut aussi être étendu aux conducteurs de véhicules automobiles. Nous pouvons améliorer le travail en choisissant un corpus plus complexe qui pourrait mettre en difficulté le locuteur. Puis nous verrons le taux de rejet par un système de reconnaissance de locuteurs. En effet le nombre des omissions ou des élisions de phonèmes devrait être plus important. Une autre approche du problème consisterait en l'analyse des électro-encéphalogrammes de la région du cortex régissant la parole.

Pour une meilleure étude, nous préconisons un travail d'équipe, en milieu hospitalier, entre des médecins phoniâtres et des spécialistes du traitement de la parole et ce sur une année au minimum. Le travail sera plus complet car à tout moment nous pourrions voir l'influence de toute émotion, fatigue ou encore maladie sur la voix. Nous aurons, par la même occasion, la possibilité de comparer avec les résultats obtenus par le biais d'analyse médicales et d'étendre notre étude aux locuteurs ayant des troubles pathologiques de la voix.

La fatigue apparaît moins vite chez un individu entraîné à l'effort ou si le travail est agréable. Le travail peut être complété en étudiant l'influence de la fatigue selon le degré d'entraînement des locuteurs à l'effort et nous pourrions essayer d'apprécier le temps qu'elle met à apparaître selon la perception psychologique du travail par ces derniers.

## BIBLIOGRAPHIE

- [1] R. Ruiz, C. Legros, A. Guell, *Voice analysis to predict the psychological or physical state of a speaker*, Aviation, Space and Environmental Medecine N°61, pp. 266-271, March 1990.
- [2] R. Ruiz, C. Legros, A. Guell, *Voice analysis : application to the study of the influence of a workload*, Journal d'acoustique, Vol. 3, N°2, pp. 153-159, Juin 1990.
- [3] C. Frankish, D. Jones, K. Hapeshi, *Decline in accuracy of automatic speech recognition as a fonction of time on task : fatigue or voice drift?*, International journal of man-machine studies, Vol. 36, N°6, pp. 797-816, 1992.
- [4] B. Malmberg, *La phonétique, Que-sais-je ? N° 637*, Presses Universitaires de France, Paris 1984.
- [5] P. Dejonckere, *Précis de pathologie et de thérapeutique de la voix*, Editions Universitaires J. P. Delarge, Paris 1980.
- [6] E. Emerit, *Cours de phonétique acoustique*, Collection Sibawayh SNED, Alger 1977.
- [7] Calliope, *La parole et son traitement automatique*, CNET-ENST, Masson, Paris 1989.
- [8] J. S. Liénard, *Les processus de la communication parlée- Introduction à l'analyse et la synthèse de la parole*, Masson, Paris 1977.
- [9] R. Boite et M. Kunt, *Traitement de la parole*, Presses Polytechniques Romandes, Lausanne 1987.
- [10] J. Guibert, *La parole compréhension et synthèse par les ordinateurs*, Presses Universitaires de France, Paris 1979.
- [11] B. Frachet, *La communication - modalités, technologies et symboles*, Arnette, Paris 1991.
- [12] E. Garde, *La voix, Que sais-je? N°627*, Presses Universitaires de France, Paris 1970.
- [13] S. H. Al ani, *Arabic phonology - An acoustical and physiological investigation*, Mouton, 1970.
- [14] J. Cantineau, *Cours de phonétique arabe*, Librairie C. Klincksieck, Paris 1960.

- [15] A. H. Moussa, *Statistical Study of Arabic Roots in Mojam Al-Sehah*, Kuwait University, 1973.
- [16] A. Domart et J. Bourneuf, *Nouveau Larousse médical*, Larousse, Paris 1990.
- [17] *Le dictionnaire des médecines naturelles*, Editions Marabout, Verviers 1980.
- [18] M. Rieu, *L'adaptation à l'effort*, In: Science et vie Le sport au quotidien, N°147, pp. 84-94, Juin 1984 Paris.
- [19] F. Péronnet, G. Thibault, G. Brisson et M. Ledoux, *Le marathon Equilibre énergétique - Endurance et alimentation du coureur sur route*, Editions Vigot, Montréal 1983.
- [20] G. A. Brooks and T. D. Fahey, *Fatigue during muscular exercise*, In: Physiology of Muscular Exercise, N° 25 pp. 701-724, New-York 1983.
- [21] M. Bellanger, *Traitement numérique du signal Théorie et pratique*, CNET-ENST Masson, Paris 1987.
- [22] L. R. Rabiner and B. Gold, *Theory and application of Digital Signal Processing*, Prentice-hall, Inc. , New Jersey 1975.
- [23] R. Miquel, *Le filtrage numérique par microprocesseurs*, Editests, Paris 1985.
- [24] M. Kunt, *Traitement numérique des signaux*, Presses Polytechniques Romandes, Lausanne 1984.
- [25] M. Laberrere, J. P Kiref et B. Gimonet, *Le Filtrage et ses applications*, Cepadues-Editions, Toulouse 1988.
- [26] A. V. Oppenheim and R. W. Schafer, *Digital signal processing*, Prentice-Hall, Inc. , New Jersey 1975.
- [27] S. M. Kay and S. L. Marple Jr. , *Spectrum Analysis - A modern perspective*, Proc. IEEE, Vol. 69, N°11, pp. 1380-1419, Nov 1981.
- [28] A. A. Giordano and F. M. Hsu, *Least square estimation with applications to digital signal processing*, Willey Interscience publication, New york 1985.
- [29] J. D Markel and A. H Gray Jr, *Linear Prediction of Speech*, Communications and cybernetics 12, Editeurs K. S. Fu, New- York - Berlin 1982.
- [30] J. P. Haton et coauteurs, *Reconnaissance automatique de la parole*, Dunod informatique, Paris 1991.

- [44] Y. Grenier, *Identification du locuteur et adaptation au locuteur d'un système de reconnaissance phonémique*, Thèse de docteur ingénieur, Octobre 1977.
- [45] N. Tiziri, *Reconnaissance automatique de la parole dans un contexte multilocuteur - Application aux voyelles de l'arabe standard et du français*, Thèse de magister, Université d'Alger, Alger 1985.

Vu l'amplitude assez importante de leurs lobes secondaires, les fenêtres rectangulaires et de Hanning ne sont généralement pas utilisées. Celle de Hamming est souvent préférée en parole. Elle est donnée par la relation :

$$W_H(n) = W_R(n) (\alpha + (1 - \alpha) \cos(2\pi n / N)) ; \quad 1 \leq n \leq N \quad (1)$$

$$= 0 \quad ; \quad \text{autrement}$$

$W_R(n)$  étant la fenêtre rectangulaire,  $\alpha$  un paramètre invariable -  $\alpha = 0.54$  - et «  $N$  » le nombre d'observations.

La fenêtre de Kaiser est donnée par la formule [22] :

$$W_k(n) = I_0(\alpha \sqrt{1 - [2n / (N - 1)]^2}) / I_0(\alpha) \quad (2)$$

avec :

$$-(N - 1) / 2 \leq n \leq (N - 1) / 2$$

Dans ce cas, «  $\alpha$  » est un paramètre variable.  $I_0(x)$ , la fonction modifiée de Bessel d'ordre zéro, est donnée par la formule [22] :

$$I_0(x) = 1 + \sum_{k=1}^{\infty} (x/2)^k / k! \quad (3)$$

La fenêtre de Kaiser présente une amplitude des lobes secondaires bien moins élevée que celle de Hamming et une plus grande souplesse d'utilisation. En effet, elle laisse à l'utilisateur le choix par le paramètre «  $\alpha$  », d'une taille permettant un bon compromis entre la largeur du lobe principal et le niveau de l'amplitude du premier lobe secondaire [22], [24].



## ANNEXE 2

```

1  **PROGRAMME D'ACQUISITION ET DE CONVERSION DES DONNEES
5  COM /Reapcon/ INTEGER Repeat(1:1500)
10  INTEGER Select_code,Gain
11  INTEGER Nb_pts
15  Name$="ADC"
20  Model$="98640A"
25  Select_code=18
30  Gain=1
36  Pace=.000083
40  Error$="NO"
45  Unit$="Standard"
46  Nb_pts=256
49  DIM A(256)
51  Meas_lib_init
55  Config_0(Name$,model$,Select_code,Gain,Pace,Error$,Unit$)
58  Init(Name$)
62  Calibrate(Name$,0,Pace,256)
63  PRINT "A"
65  FOR I=1 TO Nb_pts
66  Input(Name$,I,Volts)
77  A(I)=Volts
87  NEXT I
90  INPUT Fichier$
91  CREATE BDAT Fichier$,Nb_pts,256
92  ASSIGN @Path1 TO Fichier$
94  FOR I=1 TO Nb_pts
95  OUTPUT @Path1,A(I)
97  NEXT I
98  END

162  CSUB Meas_lib_init
163  CSUB System_init
164  CSUB Init(Name$)
165  CSUB Input(Name$,INTEGER Channel,REAL Var,OPTIONAL INTEGER Gain,REAL P
166  TIONAL INTEGER Repeat)
167  CSUB Random_scan(Name$,INTEGER Addr_array(*),REAL Data_array(*),OPTIO
168  NTEGER Repeat,REAL Pace_array(*),INTEGER Gain_array(*))
169  CSUB Set_gain(Name$,INTEGER Gain)
170  CSUB Set_units(Name$,Units$,OPTIONAL Multiplier,Offset)
171  CSUB Enable_incr(Name$)
172  CSUB Disable_incr(Name$)
173  CSUB Calibrate(Name$,INTEGER Chan,REAL Pace,INTEGER Number)
174  CSUB Config_0(Name$,OPTIONAL Model$,INTEGER Select_code,Gain,REAL Pace
175  rt_err$,Units$,Multiplier,Offset)

```