

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET
POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE**

Ecole Normale Supérieure de Bouzareah

Support de cours

Module :

**Phonétique Articulatoire et Acoustique
- Semestre 1 -**

Filière : Informatique

Spécialité :

**Master 1 : TAL
« Traitement Automatique du Langage Naturel »**

Responsable du cours : Dr. Tebbi Hanane

Contenu de la matière

- 1. La Communication Parlée(CP) et la Communication Ecrite(CE)**
 - a. La CP
 - b. La CE
- 2. Généralités sur le signal vocal**
 - a. Introduction
 - b. Le son et la parole
 - c. La production de la parole
 - d. Les types des sons (Consonnes, Voyelles)
- 3. Introduction à la phonétique générale**
 - a. Phonétique articulatoire « la production du son »
 - b. Phonétique acoustique « la transmission des sons par l'air »
 - c. Phonétique perceptive « la réception des sons par l'oreille de l'interlocuteur»
- 3. Caractéristiques de la parole**
 - a. introduction
 - b. l'appareil phonatoire humain
 - b. Modes et lieu d'articulation
 - d. Opposition voisée /non voisé
 - e. Opposition oral /nasal
 - f. Opposition occlusives /fricatives
- 5. La Transcription Orthographique Phonétique TOP**
 - a. Introduction
 - b. Définition de TOP
 - d. Les ressources utilisées en TOP
- 6. Les techniques d'analyse acoustique**
 - a. Définition
 - b. l'analyse par spectrogramme
 - c. La segmentation
- 7. L'outil d'analyse PRAAT**

Chapitre 1 : La Communication Parlée et la Communication Ecrite (la CP / la CE)

1. La Communication Parlée (CP)

1.1. Introduction

On situation typique de CP deux interlocuteurs sont à portée de voix (quelque mètres l'un de l'autre), Lorsque A parle B reçoit non seulement l'information **verbale** mais aussi l'information **extra verbale** de A constituée par les gestes, la position du corps, les mimiques, etc.

De plus B décode l'information totale émise par A en fonction d'un triple contexte

- Contexte de situation lié à l'environnement commun aux deux interlocuteurs ;
- Le message envoyé (la parole) ;
- Contexte individuelle liée aux connaissances et aux aptitudes de B.

Enfin, il existe une rétroaction permanent et immédiate de B sur A. Le but de CP est de transmettre non pas des mots et des phrases mais des idées, des relations entre concepts, et cela en vue d'une certaine action sur le monde.

1.2. Quelques caractéristiques de la CP

Dans une situation de communication parlée l'émetteur et le destinataire communiquent en temps réel et si un message n'est pas compris la rétroaction est immédiate.

La CP est :

- Spontané ;
- Interactive ;
- S'inscrit dans un contexte multiple ;
- Floue imprécise sans que la communication elle même soit affectée.

2. La Communication Ecrite (CE)

2.1. Introduction

Le message écrit doit comporter en lui même toute l'information de message total. La CE doit être beaucoup plus standardisée et structurée

2.2. Caractéristiques de la CE

La communication écrite est presque toujours différée. Le récepteur lit le message sans l'émetteur et s'il ne le comprend pas, il ne peut pas y avoir explication en temps réel mais en temps différé. Il en résulte qu'on ne peut pas écrire comme on parle. A l'écrit tout métalangage est absent.

La CP est :

- Fortement structuré ;
- Généralement unidirectionnelle ;
- Contexte de situation pratiquement absent.

Le message écrit doit comporter tout l'information de message.

Pour conclure voici un schéma qui résume les différences entre les deux approches de communications (Figure 1.1 et Figure 1.2)

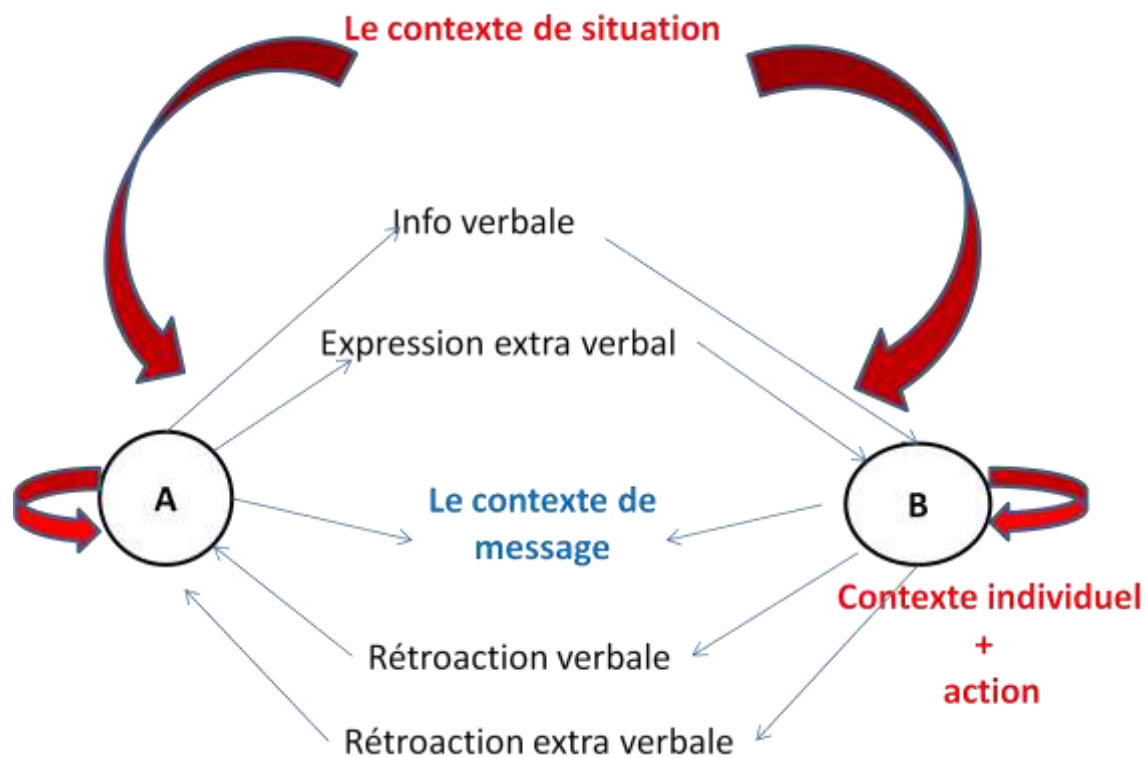


Figure 1. Environnement spatio-temporel



Figure 2. Environnement spatio-temporel différents

Chapitre 2 : Généralités sur le Signal Vocal

1. Introduction

La parole est le seul moyen qui permet de communiquer la pensée par un système de sons articulés. Les humains sont les seuls êtres vivants qui utilisent un tel type de système structuré.

2. Le son et la Parole

a- Qu'est ce qu'un son ?

Par définition, le son est ce que l'oreille perçoit de la vibration d'un corps. Cette vibration est une sorte d'onde (produite par un objet, guitare, piano, tambour, marteau, etc.), qui se propage par et à travers des corps physiques (air, eau, métal, bois, etc.) Figure 2.1. La parole se distingue des autres sons par des caractéristiques acoustiques ayant leurs origines dans le mécanisme de production



Figure 2.1 : Les trois étapes de la propagation d'un son [1].

3. La Production de la parole

La production de la parole réside dans les fluctuations de la pression de l'air engendrée, puis émis par l'appareil phonatoire.

Ces fluctuations constituent le signal vocal. Elles sont détectées par l'oreille qui procède à une certaine analyse et les résultats sont transmis au cerveau qui les interprète. La production de la parole n'est pas réalisée par un système propre, mais elle est assurée conjointement par les organes de la respiration (Figure 2.2).

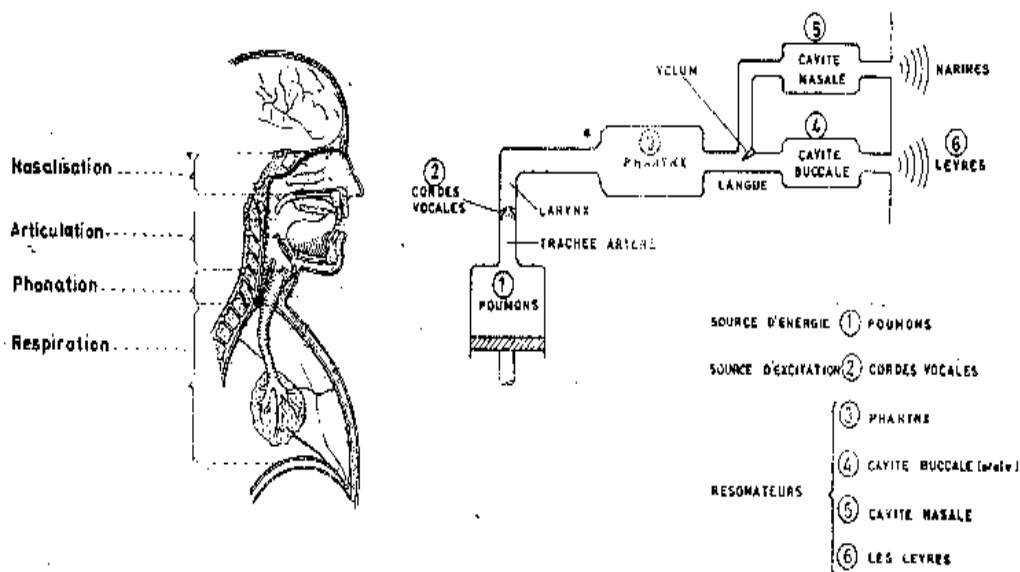


Fig. 1.2: L'appareil phonatoire en tant que système acoustique

Figure 2.2 : Modélisation de l'appareil phonatoire par un système acoustique [2]

4. Les types des sons

La parole est formée par des sons qui correspondent à *des phonèmes* tels que les voyelles, les semi-voyelles, les nasales, les liquides, les consonnes fricatives, les consonnes occlusives, et *des transitions entre ces phonèmes*, par exemple les diphtonges, polysons, etc. Les différents états de l'appareil phonatoire humain déterminent les différents types (nature) de sons produits qui sont :

4.1. Les Voyelles:

Les voyelles sont le résultat de l'ouverture du conduit vocal, les cordes vocales vibrent (sons voisés) et la forme des cavités (essentiellement la bouche) modifie le timbre. Les voyelles sont dites orales ou nasales selon que la cavité nasale n'est pas ou est mise en parallèle à la cavité buccale.

4.2. Les Consonnes:

Les consonnes se caractérisent par un rétrécissement de l'appareil phonatoire, les cordes vocales peuvent vibrer (sons voisés) ou laisser passer librement l'air (son non voisés). Ces consonnes sont fricatives si le rétrécissement est partiel ou occlusives (plosives) si une occlusion totale apparaît dans l'appareil phonatoire, causant une augmentation de la pression et un relâchement brutal de celle-ci lors de l'ouverture. Les consonnes sont caractérisé par une

constriction ou une fermeture soit momentanée soit complète du passage de l'air.

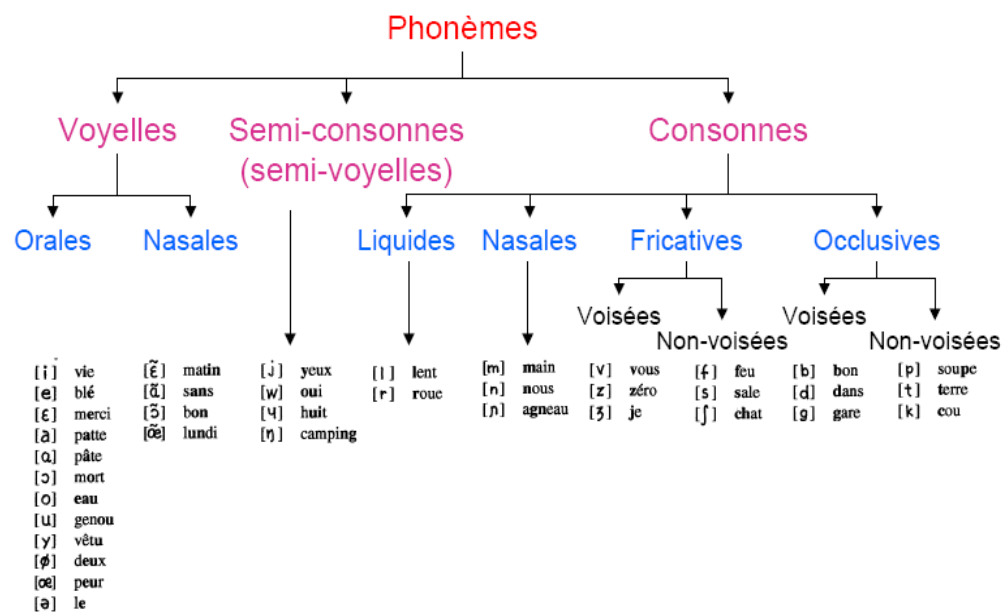


Figure 2.3 : Types de sons de la langue française.

Chapitre 3 : Introduction à la phonétique générale

1. La phonétique

La phonétique est la science des sons langagiers. Elle est l'étude scientifique des sons du langage humain. Cette science peut être abordée sous trois aspects différents:

- La production du son (phonétique articulatoire) ;
- La transmission des sons par l'air (phonétique acoustique) ;
- La réception de ces sons par l'oreille de l'interlocuteur (phonétique auditive).

1.1. La phonétique articulatoire

La phonétique articulatoire s'occupe de l'activité du conduit vocal (des cordes vocales, de la cavité buccale, de la mâchoire, des lèvres, etc.), qui rendent possible la parole. On peut dire aussi que c'est la configuration que prennent les organes phonatoires pour produire des sons langagiers.

Selon ce type on distingue deux classes de sons :

- Les voyelles [a, i, u, y, e, o, ε, ə ...], qui correspondent à un passage libre de l'air, et un conduit vocal suffisamment ouvert ;
- Les consonnes [p, t, b, g, m ...], qui sont caractérisées par une fermeture momentanée ou complète du conduit vocal, et

par un passage de l'air obstrué dans des endroits bien précis.
Ces derniers sont appelés lieux d'articulation.

1.2. La phonétique acoustique

La phonétique acoustique étudie des sons qui sont des ondes ou des vibrations qui se propagent dans l'air. On peut dire alors qu'elle s'occupe de la nature physique du message vocal (la fréquence fondamentale ou pitch F_0 , la durée ou le temps T , et l'énergie E), indépendamment de ses conditions de production et de réception. Par exemple le son produit par une consonne comme [s] a une fréquence plus élevée que le son produit par une consonne comme [z].

Selon cette classe, on peut établir deux catégories de sons

- Voisés [v, z, b, d, g...]: présence de la fréquence fondamentale ;
- Non voisés [p, t, k, f, s...] : absence de la fréquence fondamentale.

1.3. La phonétique auditive

La phonétique auditive ou perspective est la partie la plus subjective de la phonétique qui étudie la manière dont un son est perçu par l'oreille. Elle est rarement étudiée sauf pour l'élaboration de traitements orthophoniques.

Chapitre 4 : Les Caractéristiques de la parole

1. Introduction

Physiquement la parole est un phénomène vibratoire résultant de deux composantes :

- Le passage de l'air expiré à travers les Cordes Vocales « CV » (source d'excitation) ;
- Le système résonnant composé des 4 cavités : Pharyngale, buccale, nasale, et labiale.

2. L'appareil phonatoire humain

L'appareil vocal est un système phonatoire, qui comprend un ensemble d'éléments fonctionnant en étroite synergie pour produire les signaux acoustiques ; Les principaux organes phonatoires sont (Figure 4.1):

- Les poumons ;
- La trachée artère et son extrémité supérieur le Larynx qui support 2 muscles appelé les Cordes vocale l'ouverture variable séparant ces muscles est la Glotte ;
- Le conduit vocale est l'ensemble des cavités pharyngo-buccale (pharyngale et buccale) et nasale ;
- Le muscle mobile qui commande le couplage entre ces 2 cavités s'appel: le voile du palais (velum) ;

- La cavité nasale a une forme fixe et ne peut être obstruée (fermé) qu'en un seul point son extrémité postérieure ;
- La cavité buccale est susceptible de prendre des formes variées et de présenter des rétrécissements (blocage) en divers point (chaque lieu d'articulation → produit un son).

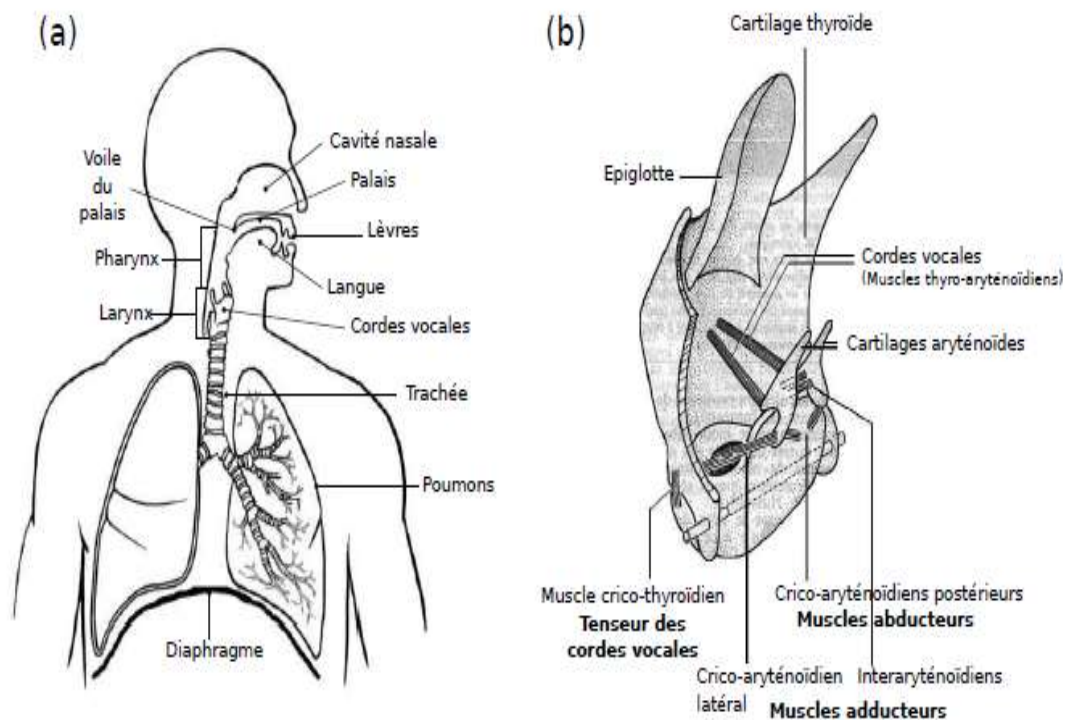


Figure 4.1: a) Les organes de la phonation b) schéma de la mécanique du larynx [3]

3. Modes et lieu d'articulation

Le lieu d'articulation est la zone du conduit vocal qui participe à la formation du son. Il présente la position de la constriction totale (cas des occlusives) ou partielle (cas des fricatives) d'une zone spécifique du conduit vocal lors du passage de l'air provenant des poumons. Le lieu d'articulation peut être bilabiale, glottale, labiodentale, etc. [4].

Exemple 1 :

	Occlusives	Emphatiques	Fricatives	Nasales	Liquides	Glides (semi-voyelles)
Labiales	ب b		ف f	م m		و w
Interdentales		ظ	ذ			
Dentales	د d	ت t	ص	ن n	ل l	ر r
Sifflantes		ص s	ز z			
Palatales	ج		ش			ي y
Vélaires	ك k		غ			
Uvulaire	ق q					
Pharyngales			ع			
Glottales	ء		ه h			

Exemple 2 :

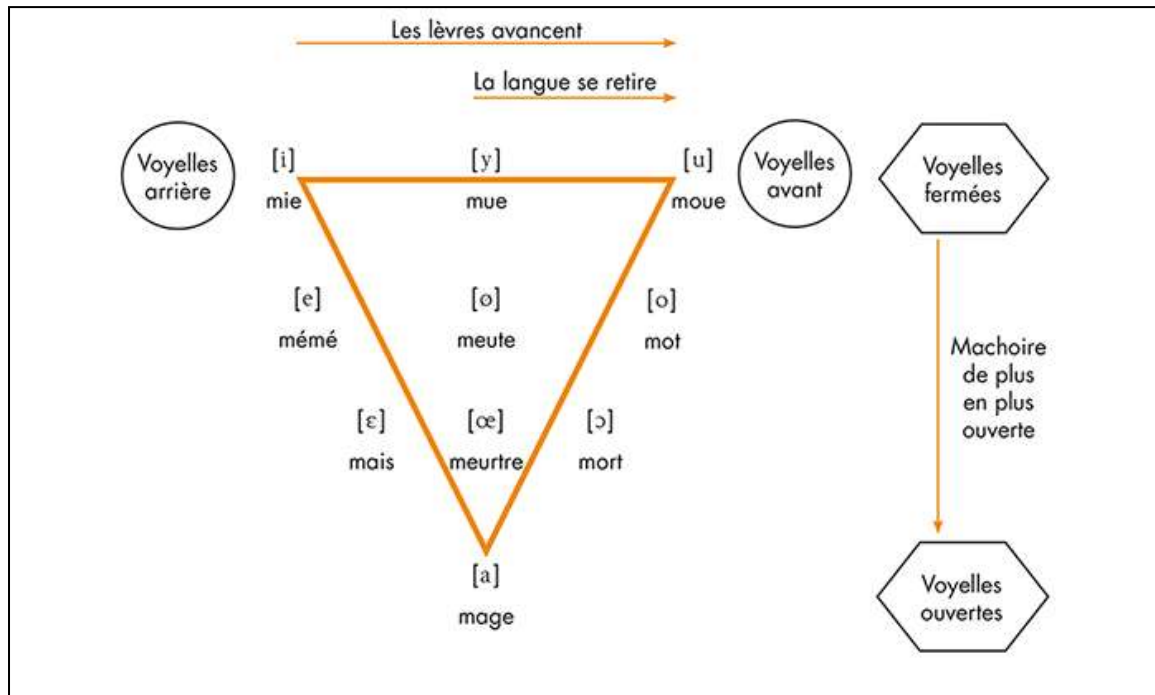


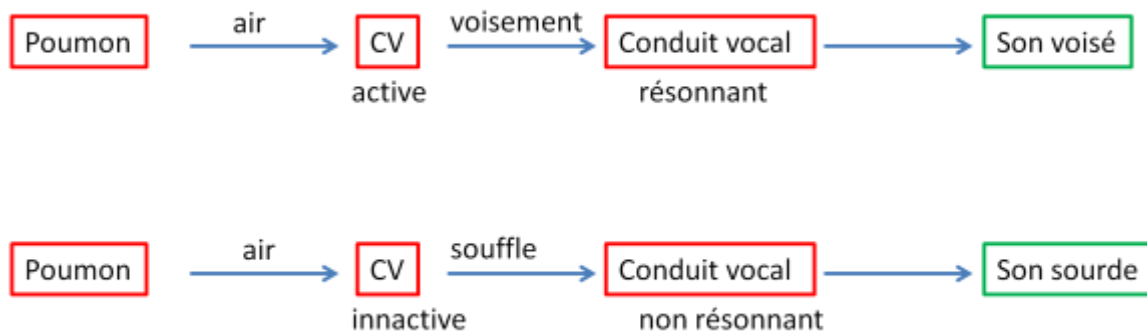
Figure 4.2. Le triangle vocalique [7]

3.1. Opposition voisée /non voisée (V/NV)

On peut établir deux catégories de sons selon qu'ils sont dus à :

- Une vibration laryngienne périodique : son voisé ou sonore ;
- Une génération de bruit à travers une constriction (fermeture) de conduit vocale : son non voisé ou sourde.

Ces deux sortes de sons peuvent être combinés, c'est-à-dire que la source de bruit vient s'ajouter à la source d'excitation vocale.



3.2. Opposition orale/ nasale

Le vélum peut se relever à l'horizontal interdisant toute communication entre les cavité nasale et pharyngo-buccale, on obtient un son orale au contraire si le vélum est abaissé, l'air passe en empruntant la cavité nasale: le son ainsi produit est dite nasale

3.3. Opposition occlusives/fricatives (plosives)/ (constrictives)

- Les occlusives sont caractérisé par une fermeture complète en un point de conduit vocale la détente de cette occlusion s'accompagne d'un bruit explosive typique de la consonne occlusive (p t k / b d g) ;
- Les fricatives sont générés par une constriction en un point de conduit vocale, cette dernière est accompagnée par un passage libre de l'air (f s ch / v z j).

Chapitre 5 : La Transcription Orthographique Phonétique

1. Introduction

Ce chapitre représente une étude plus approfondie des relations entre flexions orthographiques et phonétiques à travers le passage de la modalité graphique vers la modalité phonétique.

2. Définition La Transcription Orthographique Phonétique (TOP)

La Transcription Orthographique Phonétique ou phonétisation est une étape clé dans tout système de synthèse de la parole à partir du texte. Elle correspond au passage d'un texte écrit vers un texte lu. Elle fournit la prononciation associée au texte qu'on veut entendre.

La TOP est une tâche complexe, puisque les conventions adoptées lors de cette opération représentent souvent un compromis entre des choix théoriques et pratiques et cela pour traiter l'ensemble des exceptions et pour élaborer les règles de transcription.

3. Les ressources utilisées en TOP

La TOP est une étape qui peut être assurée sans se préoccuper du sens, ni de la signification. Cette étape se base sur l'utilisation de deux ressources qui sont la phonologie et la phonétique.

3.1. La phonologie

La phonologie est une science qui étudie les sons du langage du point de vue de leur fonction dans le système de communication linguistique [5]. Cette science se base sur l'étude de l'écriture.

Exemple: la différence fonctionnelle (dans l'écriture, et dans le sens) entre les deux mots /pas/ et /bas/.

3.2. La phonétique

La phonétique est une science qui étudie les sons du langage dans leur réalisation concrète, indépendamment de leur fonction linguistique [5]. Quand on fait de la phonétique on doit laisser à côté l'écriture de la langue, car ce n'est pas la forme orthographique qui influe sur la prononciation, mais plutôt le contraire.

Exemple : /les enfants / —————> [lezãfã] : s —————> z

C'est l'assimilation des sons.

RQ : Nous avons très généralement adapté la notation phonétique entre les 2 crochés par opposition la notation phonologique est représenté souvent entre les 2 barres obliques.

3.3. L'Alphabet Phonétique International «API »

L'Alphabet Phonétique International (API, IPA en Anglais : International Phonetic Alphabet) est né dans le cadre de la

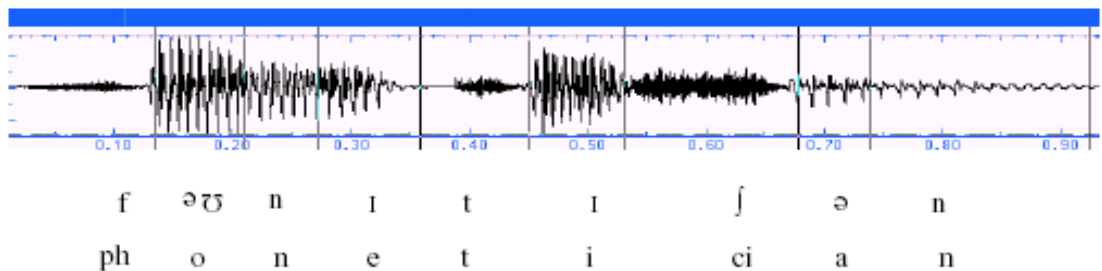
didactique des langues étrangères : il a été créé par une association de professeurs de langues (Association Phonétique Internationale) en 1888. Régulièrement révisé, sa dernière mouture date de 1993. L'objectif est clair : transcrire dans un même code de signes la prononciation de langues diverses [6].

Pour représenter les différents sons, on se base sur l'utilisation d'un mécanisme de transcription. L'alphabet normal convient assez mal à ce dernier :

- Puisqu'une seule lettre peut correspondre à plus d'un son ;
Pensez au /t / en Français qui peut être prononcé dans le mot /tata/ et il ne peut pas être prononcé dans le mot /chat/ ;
- Puisqu'un seul son peut se représenter au moyen de plus d'une lettre ;
Penser au son [s] en Français par exemple les deux mots /sérine/ et /français/ [6].
- Transcrire un énoncé, c'est noter le à l'aide d'une API ;
- Il faut savoir que plusieurs niveaux doivent être pris en compte durant le passage orthographique phonétique, Parmi ces niveaux on peut citer : Niveau phonétique / phonologique / lexicale syntaxique / et même sémantique.

4. Exemple de transcription

Voici le signal et la représentation phonétique résultante de la transcription de mot Anglais « phonetician » [6]



La liste ci-dessus représente d'autres exemples concrets de transcription des phrases écrites en Français

- c'est une question de style
[sɛtynə kɛstjɔ̃ də stil]
- le chef de gare ne trouve plus sa casquette
[lə ʃɛf də gar nə truvə ply sa kaskɛtə]
- notation
[notasjɔ̃]
- exemple de transcription
[ɛgzɑ̃mplə də trɑ̃skripsjɔ̃]
- dama est une voyelle Arabe
[dama ɛtynə vwajɛl arabə]

5. Les approches de la TOP [6]

La TOP peut se faire grâce à l'utilisation de lexiques et/ou de règles de réécriture. Cette possibilité donne naissance à deux grandes façons ou approches pour transformer un texte orthographique en un texte phonétique : la TOP à base de lexique et à base de règles.

5.1. L'utilisation du lexique

Dans ce cas on doit attribuer pour chaque mot en entrée la prononciation qui lui correspond sans se préoccuper de son contexte.

La rapidité, la souplesse et la simplicité représentent les principaux avantages de cette approche.

5.2. L'utilisation de règles

Dans cette approche chaque graphème est converti en phonème selon le contexte et cela grâce à l'utilisation d'un ensemble de règles de réécriture. Ces dernières sont bien détaillées dans les paragraphes suivants.

Le principal avantage de cette approche à base de règles réside dans la possibilité de modéliser les connaissances linguistiques des êtres humains par un ensemble de règles qui peuvent être intégrées dans des systèmes experts. *Flex* et *COMPOST* (qui est appliqué à la langue française et qui est développé à l'ICP-INP Grenoble :

France) sont les langages de programmation par règles de réécriture les plus utilisées. Par exemple pour *COMPOST* chaque règle est représentée comme suit :

$$C \longrightarrow R /G+D$$

C'est-à-dire transformer le caractère (le graphème) G en phonème R, s'il a comme contexte gauche G et comme contexte droit D.

L'outil de base pour ce type de transcription par règles qui assurent les relations existant entre le code orthographique et le code phonétique est *la grammaire contextuelle*. Chaque règle de cette grammaire a la forme suivante :

$$[\text{Phonème}] = \{\text{CG (Contexte Gauche)}\} + \{\text{C (Caractère)}\} + \{\text{CD (Contexte Droit)}\}$$

Le problème de cette approche réside dans la détermination et la gestion de nombre énorme de règles de transcription.

Chapitre 6 : Les Techniques d'Analyse vocale

1. Introduction

L'Analyse acoustique est une partie importante dans le traitement que subit le signal sonore [10], et ce pour pouvoir réaliser un système de haute qualité de synthèse, de compréhension, ou de reconnaissance de la parole.

2. L'étude acoustique de la parole [9]

On va étudier le signal vocal de point de vue acoustique en évaluant ses paramètres à savoir la fréquence fondamentale, la durée et l'énergie. Les modifications apportées à l'un de ces paramètres peuvent altérer indéniablement les autres paramètres. Cependant, si nous voulons étudier ces paramètres d'un point de vue acoustique, nous pouvons les considérer comme étant parfaitement indépendants.

2.1. La Fréquence Fondamentale (F_0)

La Fréquence Fondamentale est la fréquence de vibration des cordes vocales. Elle varie d'une personne à une autre en fonction de la longueur et de la masse des cordes vocales de chaque personne. Elle permet de diviser l'ensemble des sons de parole en trois grandes macros classes :

- Elle varie entre 70 -160 Hz : pour les hommes ;
- Elle varie entre 130 - 290 Hz : pour les femmes ;
- Elle est supérieure à 290 Hz : pour les enfants.

2.2. La durée

La durée est une mesure très variable. Elle représente le temps de la prononciation d'un phonème. Pour mesurer une durée quelconque, il faudrait au préalable désigner, d'une part, les unités à mesurer et d'autre part, leurs repères (les frontières) dans le signal parole. Elles peuvent concerner les phonèmes, distance entre voyelles, les pauses, etc. Les deux types de durées connus sont :

- La durée observée, qui correspond à la mesure objective du temps de l'activation des organes de phonation ;
- La durée perçue, liée au mécanisme de la perception et est fréquemment utilisée dans le cas des occlusives puisqu'elles sont caractérisés par une durée de réalisation non continue.

Généralement la durée d'une unité est mesurée par le nombre des trames (phonèmes) qu'elle contient, et pour calculer la durée de chaque trame il faut fixer deux événements sur le signal de parole qui délimitent les repères initial et final de cette trame.

2.3. L'Intensité ou l'énergie

L'intensité correspond au corrélat acoustique de la pression sous glottique. Autrement dit, elle représente l'amplitude des vibrations des cordes vocales. L'énergie contenue dans une portion du signal échantillonné est défini par :

$$E = \sum_{N=1}^T x_n^2 \quad \text{tel que } n=1, \dots, T \quad (1)$$

A l'échelle perceptive elle est représentée en déciBels (dB) par :

$$E_{dB} = 10 \times \log_{10} \left(\sum_{N=1}^T x_n^2 \right) \quad (2)$$

2.4. Le rythme

Le rythme est la caractéristique d'un phénomène périodique induite par la perception d'une structure dans sa répétition. Le rythme n'est pas le signal lui-même, ni même sa répétition, mais la notion de forme ou de « mouvement » que produit la répétition sur la perception et l'entendement. On reconnaît un même rythme dans des phénomènes de cadences, ou périodicités, différentes, lorsque l'ordre de succession et le rapport de durée entre ses moments de tension et de relâchement

est identique. Ces moments sont souvent décrits aussi comme moments d'*élévation* (ou *arsis*) et d'*abaissement* (ou *thesis*) ; Donc c'est le retour régulier dans la chaîne parlée ; Lorsqu'un son est doté de valeurs acoustiques constantes et il se répète à des intervalles égaux, il est prouvé qu'à partir d'un certain moment, l'oreille commence par s'habituer aux repères à intervalle régulier, ceci est appelée isochronie.

2.5. Le Timbre

C'est la richesse en fréquences d'un signal de parole. Il dépend des harmoniques présents dans le spectre du signal sonore. C'est la caractéristique du son qui permet de différencier deux sons de même hauteur et de même intensité.

2.6. La bande passante

En traitement de signal, la bande passante est l'intervalle de fréquences dans lequel l'amplitude de la réponse d'un système ne s'écarte pas trop d'un niveau de référence. Elle caractérise l'ensemble des fréquences admissibles plus elle est large plus grand sera le nombre d'informations.

2.7. La résonance

On parle de résonance, si la fréquence propre au milieu, est la même que celle de signal, ce qui va renforcer certaines fréquences du signal lui même.

3. Les techniques d'analyse acoustique

L'analyse acoustique ou l'analyse vocale consiste à tirer à partir du signal vocal un ensemble de paramètres pertinents, discriminants et robustes susceptibles de le représenter.

Plusieurs techniques d'analyse sont utilisées parmi lesquelles on peut citer l'analyse par :

- spectrogrammes ;
- codage prédictif linéaire (Linear Predictive Coding ou LPC) ;
- L'analyse Cepstrale (MFCC).

4. L'Analyse par spectrogrammes

Dans l'étude du phénomène acoustique, on peut réduire la description du son à trois grandeurs physiques : la fréquence (Hz), la durée (s) et l'amplitude (ou l'énergie) (dB).

Par exemple, un son de parole simple, c'est-à-dire sinusoïdal tel que le son qui correspond à la voyelle [fathatun] de l'Arabe Standard en milieu de mot est complètement décrit par les valeurs:

$$F_0 = 144.4488 \text{ Hz}, \quad t = 0.0865 \text{ s}, \quad A = 83.7061 \text{ dB}.$$

Cela signifie que les trois valeurs durées, fréquence et énergie sont les paramètres pertinents. Une meilleure analyse consiste à les représenter de manière claire et avec précision.

L'une des représentations possibles est d'associer deux à deux ces trois grandeurs et de tracer les graphes de ces associations, on obtient les trois plans suivants :

- dynamique (temps, amplitude) ;
- Du spectre (fréquence, amplitude) ;
- Mélodique (temps, fréquence).

Le spectrogramme est l'une des méthodes d'analyse qui assure une représentation tridimensionnelle de signal de parole tel que (Figure 6.1) :

- L'axe vertical représente la fréquence du son en Hz ;
- L'axe horizontal représente l'évolution temporaire du son ;
- Le degré de noircissement représente l'intensité (l'énergie) en dB du son.

L'objectif principal de spectrogramme est de connaître l'évolution temporelle du spectre de parole. Pour assurer cet objectif, il faut décomposer l'onde acoustique du son en ondes sinusoïdales de différentes fréquences au moyen d'une Transformée de Fourier.

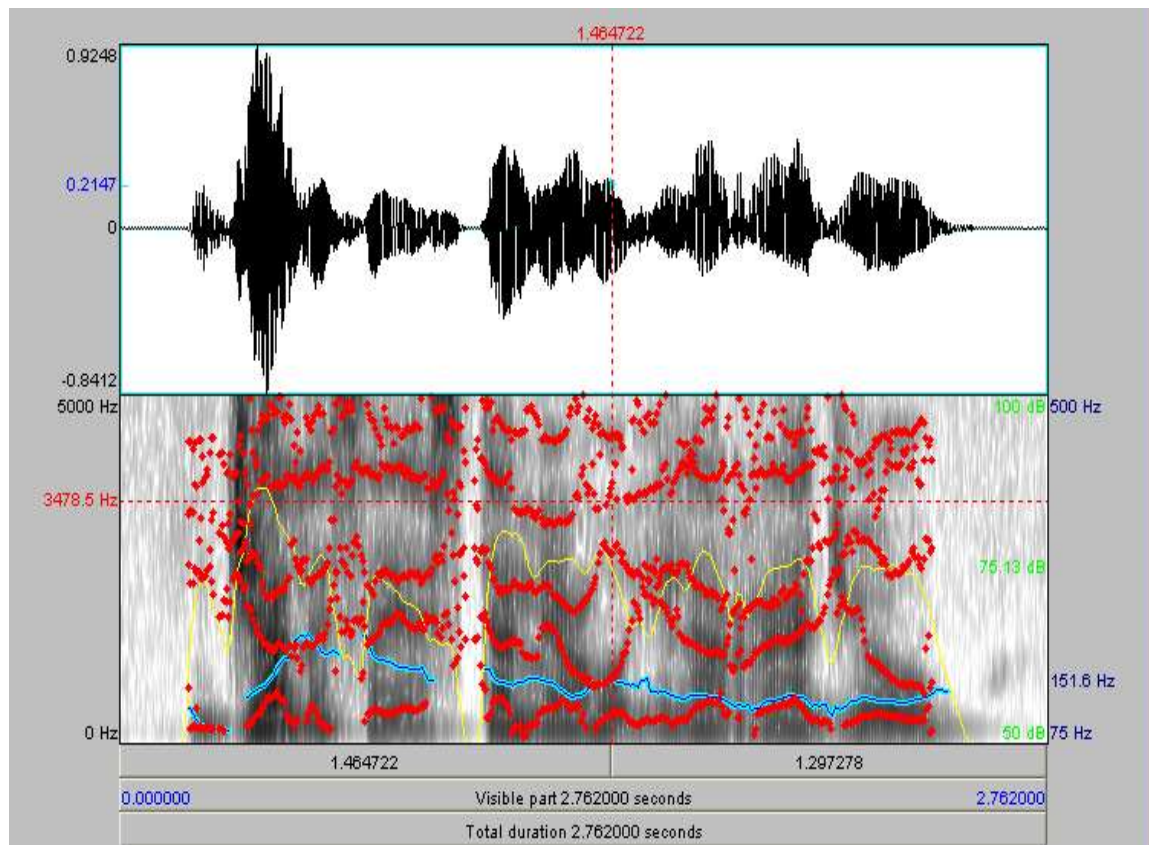


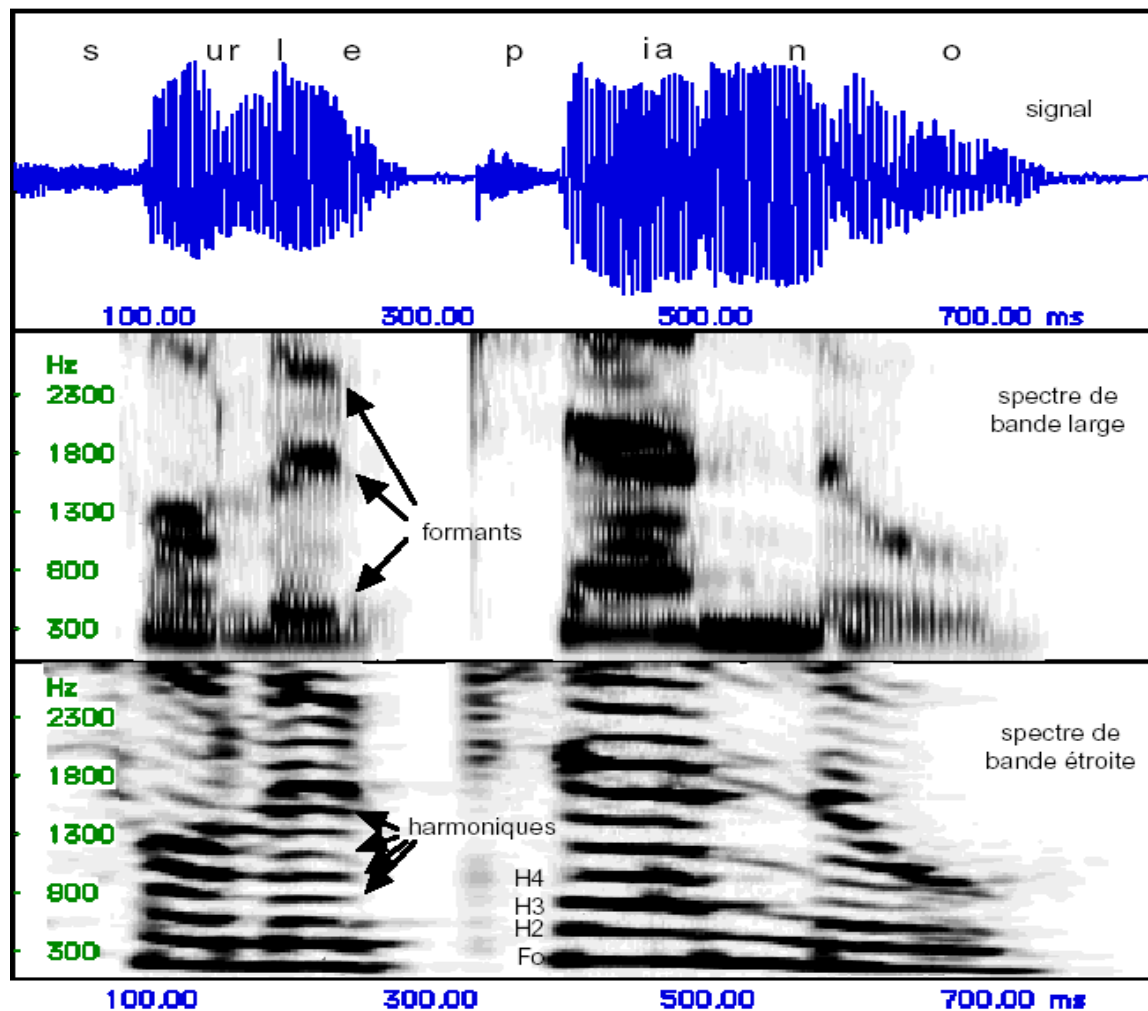
Figure 6.1. Spectrogramme de la phrase / جلس يستمع إلى الراديو / [zalasā yastamiʿu Eilaa arraadyuu].

Il existe deux représentations possibles pour un spectrogramme, la première en *bande étroite* et la deuxième en *bande large*. La différence essentielle entre les deux réside dans le choix des paramètres qui nous intéressent :

- un spectrogramme à Bande Large (BL) offre une meilleure résolution fréquentielle et permet de visualiser clairement l'évolution formantique des sons, mais il correspond à une

mauvaise analyse temporelle (Figure 6.2) La classe des occlusives représente le meilleur exemple adapté à cette présentation ;

- inversement, un spectrogramme à Bande Etroite (BE) offre une bonne résolution au niveau temporel, mais l'analyse fréquentielle est moins fine (Figure 6.2). Ce type de spectrogramme est souvent utilisé dans l'étude de l'intonation ainsi que dans l'analyse des fricatives.



**Figure 6.2. Différentes analyses en BL et BE de l'onde sonore
Correspondant à l'énoncé « Sur le piano » [8]**

5. La segmentation

La segmentation consiste à découper une parole naturelle et continue qui représente une source sonore en éléments acoustiques unitaires (phones, dipphones, mots, etc.).

Selon le dictionnaire Larousse, le terme de segmentation désigne la division d'un ensemble en portions bien délimitées. Autrement dit, c'est le processus de division d'une entité, généralement continue, en petites entités appelées segments. Chaque segment possède des propriétés propres qui permettent de le différencier des autres [6].

En parole la segmentation consiste à couper les séquences audio enregistrées en unités de tailles variables. Tel qu'on place des marqueurs temporels aux limites de ces unités phonétiques (Figure 6.3) ; et cela tout en mettant en correspondance le texte et l'audio.

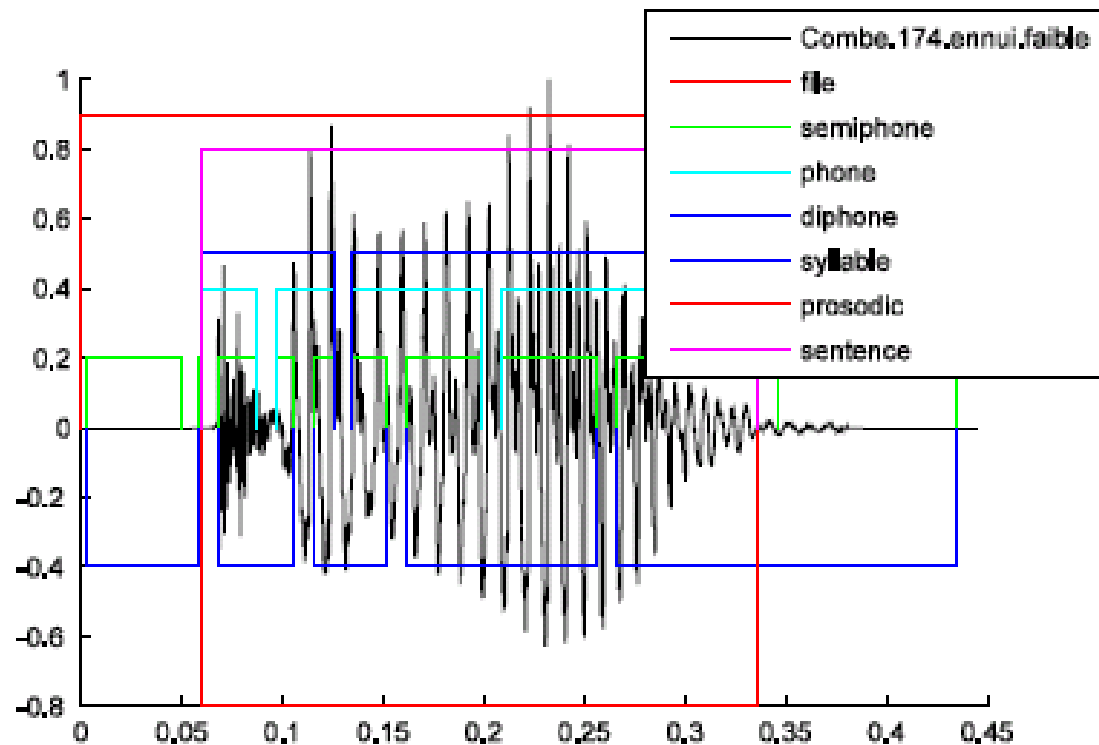


Figure 6.3 : Exemple de segmentation du mot « Comment ? »[6].

En ce qui concerne les modes de segmentation, nous distinguons deux catégories : Manuelle et Automatique

5.1. La segmentation manuelle

Est assurée par des experts phonéticiens de la langue, son inconvénient majeure est dû grâce à la difficulté de bien préciser les frontières des unités segmentales. De plus, une telle tâche nécessite un temps énorme et important durant l'annotation de grands corpus de parole. Il faut savoir qu'une segmentation manuelle effectuée

par plusieurs experts ne fournit pas nécessairement les mêmes résultats.

5.2. La segmentation semi automatique / automatique

Actuellement, la segmentation complètement automatique de parole est une tâche rarement possible [11]. En effet, étant donnée la complexité des phénomènes acoustico-phonétiques à traiter, cette tâche nécessite très souvent une intervention manuelle, que ce soit pour la préparation des données (étiquetage phonétique) du traitement automatique ou autre. Malgré l'existence des outils qui assurent cette opération, ils restent toujours non fiables puisqu'ils ne garantissent pas une très bonne qualité de parole synthétique. Pour cette raison, des vérifications manuelles faites par des experts humains sont indispensables à la segmentation de la parole.

Lors de la détection des frontières des unités extraites à partir du signal vocal, on peut utiliser l'une des deux techniques suivantes : la première est basée sur l'utilisation de la durée phonémique et la seconde se base sur l'évolution de la fréquence fondamentale :

- l'utilisation de la durée phonémique, dans ce cas l'unité prosodique représente la portion de signal de parole délimitée par deux pauses. Cette technique est pratiquement appliquée dans le traitement de la parole spontanée, puisque

cette dernière est caractérisée par un nombre très élevé de pauses.

L'utilisation de la fréquence fondamentale, les changements importants de la fréquence fondamentale marquent les frontières des unités à extraire.

Bibliographie

1. Tassart, « Traitement du signal ». [www.ircam.fr/ equipes/ analyse-synthèse / tassart](http://www.ircam.fr/equipes/analyse-synthese/tassart), 1998-1999.
2. Mhania. Guerti, « Contribution à la synthèse de la parole en Arabe Standard (Synthèse par diphtonges et techniques de Prédiction Linéaire), ILP-Alger, Algérie, 4 Mars 1984.
3. Louis delebecque, « étude, analyse et modélisation physique de la production de la parole avec application aux troubles liés à une surdité profonde », docteur de l'université Grenoble alpes, spécialité signal image parole télécoms, université Grenoble alpes, 21 septembre 2015
4. Faysal Ykhlef, « Modification de la fréquence fondamentale en vue de la synthèse de la parole à partir du texte de l'Arabe Standard ». Mémoire de Magister, département d'électronique, Université USD de Blida, Algérie, 2005.
5. A. René-Lévesque, & M. Guyart, «Glossaire de la terminologie toponymique ». Traduite par la Commission de toponymie de l'Institut Géographique National de France et par la Commission de toponymie du Québec Paris et Québec, Décembre 1997.
6. Hanane Tebbi, « La transcription orthographique phonétique en vue de la synthèse de la parole à partir d'un texte en Arabe Standard ». Mémoire de Magister,

département d'informatique, Université USD de Blida, Algérie, 2007.

7. Le Far, site : <https://www.le-far.fr/la-scene-normande/la-formation-des-artistes/kit-chanteurs/triangles-vocaliques>, dernier accès le 02/10/2021.
8. Maamar Hamadouche , « les techniques d'analyse en vue de la reconnaissance automatique de la parole », Mémoire de Magister, département d'informatique, Université USD de Blida, Algérie, 2007.
9. Hanane tebbi, « Modélisation de la Synthèse Vocale par un Système Expert », thèse de doctorat en sciences, informatique, Spécialité : Intelligence Artificielle et Base de Données Avancées, Faculté d'électronique et d'informatique, Université des Sciences et de la Technologie Houari Boumediene, Algérie, 16/01/2019.
10. Zeina Hamam, « Simulation numérique et analyse de l'émission acoustique due à la rupture de fibre et à la décohésion à l'interface fibre/matrice dans un micro composite », Matériaux. Université de Lyon, 2020. France.
11. Cédric Gendrot, « Traitement automatique et analyse de la variation dans la parole : des mesures phonétiques sur grands corpus aux réseaux de neurones profonds » Linguistique. Université Lumière Lyon 2, 2021.