Université de Nancy1
UFR S.T.M.I.A
GFD : Informatique

Centre de Recherche en
Informatique de Nancy
INRIA Lorraine

Contribution à l'étude et à la reconnaissance automatique de la parole en Arabe standard

THÈSE

présentée et soutenue publiquement le **12 Novembre 1991**

pour l'obtention du

Doctorat de l'université de Nancy I

(Spécialité Informatique)

par

Mahieddine DJOUDI

Composition du jury :

Président :	Jean-Marie Pierrel
Rapporteurs :	Monique Grandbastien
	Henri Méloni
Examinateurs :	Jean-Paul Haton
	Marie-Christine Haton
	Dominique Fohr

Une bonne parole est comme un bon arbre
dont la racine est solide et dont les branches vont jusqu'au ciel.
Il donne ses fruits en toute saison avec la permission de son Seigneur.
Le Saint Coran

à la mémoire de mon père,
à ma mère,
à mon oncle Amor.

Je tiens à exprimer mes vifs remerciements à Monsieur Jean-Paul Haton, professeur à l'université de Nancy 1 et Directeur de recherche à l'INRIA qui a bien voulu m'accueillir au sein de son équipe et me guider tout au long de ce travail. Je le remercie pour tous les conseils et les encouragements qu'il a su prodiguer à mon égard. Je lui serai toujours reconnaissant.

Mes remerciements vont également à Monsieur Jean-Marie Pierrel, Professeur à l'université de Nancy l, qui m'a fait découvrir ce domaine passionnant de la reconnaissance de la parole et qui a aimablement accepter de présider ce jury.

Je remercie Madame Monique Granbastien, Professeur à l'université de Nancy 1 qui a bien voulu consacrer de son précieux temps pour juger ce travail.

Je remercie également Monsieur Henri Méloni, Professeur à l'université d'Avignon d'avoir accepté d'être rapporteur de cette thèse et de se déplacer spécialement pour faire partie du jury.

Je remercie tout particulièrement Monsieur Dominique Fohr, Attaché de recherche au CNRS, pour ses idées et propositions qui ont enrichi nos discussions et apporté un soutien considérable au présent travail.

Je remercie Madame Marie-Christine Haton Professeur à l'université de Nancy I, qui m'a honoré de sa présence au jury.

Je n'oublierai pas de remercier les personnes qui m'ont prêté leur voix, en participant à la réalisation des corpus. Je remercie aussi Mohamed Debyeche, Hadj Aouizerat et El Djoher Remaki avec qui j'ai travaillé sur les caractéristiques phonétiques propres à la langue Arabe.

Enfin, je remercie tous mes amis et mes collègues pour leur soutien et leurs encouragements de tous moments.

Table des matières

Table des figures

Liste des tableaux

Introduction

Contexte du travail

La parole est certainement le moyen le plus direct et le plus naturel utilisé par l'homme pour échanger l'information. Le progrès enregistré dans le domaine du traitement du signal, le développement des moyens informatiques (matériels et logiciels) et l'apport de l'intelligence artificielle permettent d'envisager l'utilisation de la parole pour communiquer et dialoguer avec une machine.

La réalisation d'un système de reconnaissance de la parole continue, dans un contexte multilocuteur, posent beaucoup de problèmes. Les caractéristiques phonétiques et linguistiques de la langue sont largement impliquées dans le processus.

L'Arabe est une langue sémitique. Elle comprend un standard compris par l'ensemble de la communauté arabophone et une multitude de dialectes différents les uns des autres. La langue arabe a fait l'objet de plusieurs études anciennes et récentes, mais jusqu'à présent peu de travaux ont été effectués concernant la reconnaissance automatique.

Sur le plan phonétique, l'Arabe standard présente la particularité d'être une langue essentiellement consonantique qui se caractérise par la présence des consonnes pharyngales, glottales et emphatiques et l'existence d'une opposition temporelle brève-longue des voyelles.
Au niveau morphologique, l'Arabe possède un système complet basé sur la notion de racine qui constitue le pilier du dictionnaire.

Objectifs

Les objectifs fixés dans cette thèse sont la réalisation d'un système de décodage acoustico-phonétique en parole continue et dans un contexte multilocuteur pour l'Arabe standard et son intégration dans un système général de reconnaissance de phrases. Pour ce faire, nous nous inspirons des travaux déjà réalisés. et nous tenons

compte des caractéristiques propres de la langue. Les méthodes utilisées ont été adaptées du système développé au CRIN pour le décodage phonétique du Frančais, dans le cadre du projet APHODEX.

Plan de la thèse

Au chapitre un, nous exposerons les points qui touchent de près notre travail, à savoir les systèmes experts, les différentes techniques de raisonnement approximatif, la reconnaissance de la parole et le décodage acoustico-phonétique de la parole continue.

Dans le chapitre deux, nous présenterons une étude phonétique de l'Arabe standard basée essentiellement sur des visions de spectrogrammes de mots et de phrases. Cette étude nous a permis de définir les caractéristiques acoustiques des phonèmes nécessaires à tout système de reconnaissance analytique.

Le chapitre trois sera consacré à la présentation de l'architecture et les différents modules du système SAPHA que nous avons développé pour le décodage acoustico-phonétique de l'Arabe standard.

Dans le chapitre quatre nous présenterons les algorithmes de segmentation du signal en grandes classes phonétiques, les procédures d'extraction des indices phonétiques et les techniques d'étiquetage.

Pour chaque phonème de la langue nous donnerons les valeurs des valeurs des indices caractéristiques calculés automatiquement. Nous exposerons les deux méthodes d'identification utilisées, l'une dite procédurale et l'autre basée sur un système expert à base de règles de production. L'évaluation de la segmentation et de l'étiquetage a été faite à partir du corpus DJOUMA de 50 phrases phonétiquement équilibrées pour trois locuteurs masculins. Les performances du système et l'ensemble des résultats seront donnés et commentés.

Nous développerons au cours du chapitre cinq des idées directrices pour la conception d'un système linguistique et l'intégration du décodeur acoustico-phonétique dans un système de reconnaissance de phrases de l'Arabe moderne.

Chapitre 1

L'IA et la reconnaissance de la parole

1 Introduction

Le travail que nous développerons tout le long de cette thèse fait appel à plusieurs aspects d'intelligence artificielle (IA), de raisonnement et de reconnaissance automatique de la parole. Nous allons décrire dans ce chapitre l'ensemble de ces aspects en mettant l'accent sur les points qui touchent de plus près le contenu de notre travail.

2 L'intelligence artificielle

L'intelligence artificielle (IA) s'est donnée pour objectif l'étude et l'analyse des comportements humains dans les domaines de compréhension, de perception, de résolution des problèmes et de prise de décision afin de pouvoir ensuite les reproduire à l'aide d'une machine (ordinateur).

La différence entre l'intelligence artificielle et l'informatique classique réside dans le type de problèmes à résoudre. Les programmes traditionnels de gestion ou d'analyse numériques sont basés sur une démarche précise, décrite par un algorithme et qui conduit nécessairement au résultat. En revanche, les programmes "intelligents" s'appuient sur des méthodes heuristiques et travaillent le plus souvent sur des connaissances sous forme symbolique plutôt que sur des données numériques.

2.1 La représentation des connaissances

L'intelligence artificielle connue pour ses applications aux systèmes experts, à la compréhension du langage naturel, à la commande de robots et la résolution des problèmes nécessite un volume important de connaissances. Celles-ci doivent être mises sous une forme assimilable par l'ordinateur et structurées de manière à se prêter au traitement informatique. Suivant le type du problème à résoudre et le point de vue du concepteur, il existe différents formalismes de représentation des connaissances que l'on peut classer en trois types : [Kayser 84] [Laurière 82a]

— La représentation procédurale qui inclut les automates d'états finis et qui concerne la programmation classique [Meyer 78] [Markov 54].
— La représentation déclarative qui comprend les règles de production et le calcul des prédicats et qui permet de résoudre des problèmes de nature déductives ou inductives [Colmerauer 77] [Shortliffe 76].
— La représentation structurée sous forme de réseaux sémantiques, frames, scripts, objets ou modèles connexionistes [Schank 77] [Minsky 75].

2.2 Les systèmes experts

Les systèmes experts [Laurière 82b] représentent une des applications prometteuses de l'intelligence artificielle. Un système expert (S.E) peut être défini comme étant un ensemble de programmes structurés, qui a pour but la modélisation du comportement d'un expert humain face à une tâche intellectuelle de son domaine d'expertise. Là ou' il est implanté, un S.E a pour rôle de remplacer l'expert humain, d'être un support de travail pour les utilisateurs du domaine et pourquoi pas un bon moyen pour les amateurs de devenir eux-mêmes des experts.

Les principaux domaines dans lesquels les premiers S.E ont été développés sont la médecine, la géologie, les finances, l'aéronautique, l'agriculture et la gestion. Actuellement, ils commencent à percer et toucher des domaines à la pointe des recherches en informatique, comme par exemple la conception des systèmes d'information, le traitement d'image ou la reconnaissance de la parole.

Architecture classique d'un système expert

L'architecture d'un S.E fait apparaître plusieurs parties, l'idée de départ était de séparer la connaissance du domaine du programme informatique qui permet d'utiliser cette connaissance et ensuite de créer un environnement permettant aux non

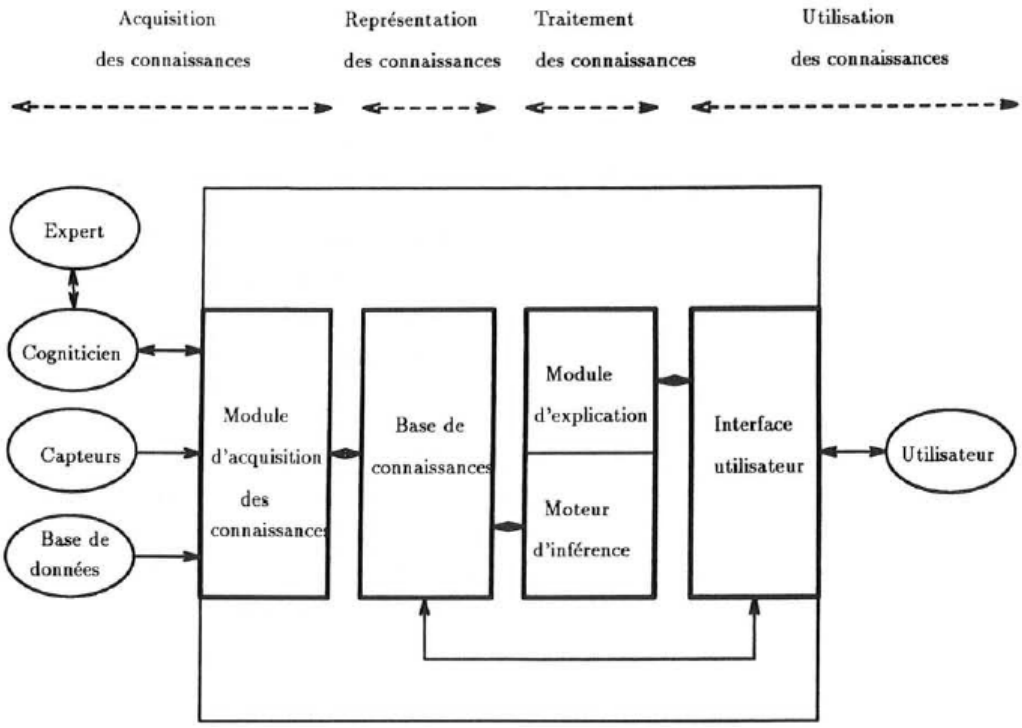


FIGURE 1.1 – Architecture d'un système expert

informaticiens d'utiliser et d'interagir avec le système de manière aisée. Les composantes essentielles d'un système expert sont : voir figure 1.1.

La base de connaissances : elle contient les connaissances concernant le domaine d'expertise en question et les méthodes de recherches des solutions. Autrement dit, la base de connaissances (B.C) contient tout le savoir qui fait d'un homme un expert dans son domaine.

Il n'est pas question - du moins pour l'instant - de livrer ces connaissances à l'état brut dans la mémoire de l'ordinateur. Avec la complicité d'un cogniticien, l'expert doit décrire son expertise dans un formalisme accessible par la machine appelé langage de représentation des connaissances.

Le moteur d'inférence : appelé aussi interpréteur. C'est un programme qui est chargé de résoudre un problème spécifié par les données en utilisant les informations et les méthodes contenues dans la base de connaissances. En principe, un moteur d'inférence doit être indépendant du domaine d'expertise sur lequel il travaille. Nous reviendrons avec plus de détail sur le principe de fonctionnement des moteurs d'inférence dans le cas des systèmes à règles de production.

Le module d'acquisition des connaissances : c'est une interface avec l'ex-

pert qui permet à ce dernier de tester le fonctionnement du système lors de la mise au point de la base de connaissances. Il sert à détecter les incohérences et les redondances qui peuvent exister et à aider à faire évoluer la base par des insertions, de modifications ou de suppressions de parcelles de connaissances.

L'interface utilisateur : permet aux utilisateurs et aussi aux experts d'utiliser le système déjà mis au point en consultation pour résoudre un problème donné du domaine d'expertise du système. Le système aura besoin de quelques informations supplémentaires pour trouver la solution et l'utilisateur doit avoir la possibilité de répondre en langage naturel.

Le module d'explication : souvent la ligne du raisonnement est plus importante que la solution même. Le rôle du module d'explication est de fournir à l'expert comme à l'utilisateur une trace du raisonnement du moteur du début jusqu'à l'aboutissement au résultat ainsi que le pourquoi du choix d'une solution intermédiaire parmi tant d'autres. Dans des cas particuliers, le module d'explication peut être intégré dans le moteur d'inférence.

Le rôle du cogniticien

Si la réalisation de systèmes experts passe par l'écriture de moteurs d'inférence, la définition de la base de connaissances est une activité complexe que les spécialistes doivent maitriser. Le rôle du cogniticien est de rendre explicite la connaissance généralement implicite d'un expert qui sait ce qu'il fait mais pas comment il y parvient et de trouver une représentation de ces informations qui puisse se mettre sous une forme informatique acceptable par un moteur d'inférence.

2.3 Les systèmes à règles de production

La plupart des systèmes experts actuels utilisent des règles de production (SRP). Une règle est de la forme :

$SI < conditions > ALORS < conclusions >$

Dans un SRP, la base de connaissances est composée d'une base de règles et d'une base de faits. La base de règles contient l'ensemble des parcelles de connaissances concernant le domaine d'application du système et les méthodes de recherches des solutions (métarègles). La base de faits contient les connaissances (ou données) initiales du traitement. Ces connaissances correspondent à une situation précise à partir de laquelle le système pourra faire des déductions. La base de faits correspond en quelque sorte à une mémoire à court terme, elle est modifiée au fur et à mesure de

la progression du raisonnement et vidée à la fin du traitement. La recherche de solutions se fait par l'exécution d'une séquence de cycles de base. Le moteur d'inférence possède deux techniques de fonctionnement selon une stratégie en chaînage avant ou en chaînage arrière.

Le chaînage avant

A partir des faits connus, le moteur cherche à déduire d'autres faits en utilisant les règles dont les prémisses sont entièrement contenues dans la base de faits. Le processus continue jusqu'à ce que le but soit atteint ou qu'aucune règle ne puisse s'appliquer. Lorsque plusieurs règles sont applicables, le moteur choisit l'une d'elles en appliquant des règles de choix ou métarègles.

Le chaînage arrière

Le moteur part du but (ensemble fini de faits à vérifier) et cherche les règles qui le contiennent dans la partie conclusion. Si l'une des règles a toutes ses prémisses dans la base de faits on s'arrête. Dans le cas contraire, on considère les prémisses comme de nouveaux buts et on réitère le processus.

Le chaînage mixte

Une telle stratégie consiste à considérer qu'une partie des faits sont à établir (chaînage arrière) et d'autres sont déjà établis (chaînage avant). Les règles peuvent donc être déclenchées simultanément par les faits et par les hypothèses.

Avec la forme simple des règles, les SRP présentent l'avantage d'être facilement incrémentales et permettent l'intégration de stratégies expertes. Leurs inconvénients résident dans le fait qu'il est parfois très difficile de formuler des connaissances sous forme de règles. Le système utilisé pour l'étiquetage phonétique de l'Arabe est un système à base de règles de production qui fonctionne en chaînage avant et qui peut être activé en chaînage arrière pour confirmer ou infirmer une hypothèse.

2.4 Le raisonnement approximatif

Le raisonnement peut être défini comme étant une technique qui permet d'obtenir de nouvelles connaissances. Il existe deux classes de raisonnement : le raisonnement certain et le raisonnement approximatif. Le raisonnement certain est basé sur la logique des propositions et le calcul des prédicats. Dans un système de raisonnement en présence d'incertitude et/ou d'imprécision, l'objectif consiste à évaluer à quel

point une proposition donnée peut être considérée comme valide. Il est donc nécessaire de disposer d'outils et de méthodes qui permettent de représenter l'incertitude et l'imprécision d'une proposition. Nous allons décrire dans ce qui suit les techniques de modélisation numérique du raisonnement approximatif. [Prade 87]

Le calcul des probabilités

La notion de probabilité est la façon la plus ancienne et la mieux connue de quantifier l'incertain. Etant donné l'ensemble fini de propositions Ω, une mesure de probabilité p est une application de Ω dans $[0,1]$ telle que :

— $p(\emptyset) = 0$; $\emptyset$: l'ensemble vide.

— $p(\Omega) = 1$.

— $\forall A \in \Omega, \forall B \in \Omega$, Si $A \bigcap B = \emptyset$ alors $p(A \bigcup B) = p(A) + p(B)$.

Sachant que

— $\forall A \in \Omega, \neg A \in \Omega$ et

— $\forall A \in \Omega, \forall B \in \Omega A \bigcup B \in \Omega$

des axiomes précédents, on peut déduire la relation :

$$\forall A \in \Omega p(A) + p(\neg A) = 1. \tag{1.1}$$

Cette relation traduit le fait que dès que l'on connaît la probabilité d'une proposition, il est possible d'en déduire la probabilité de la proposition contraire.
La dépendance du type *"Si A Alors B"* est interprétée dans l'approche probabiliste en terme de conditionnement. Il s'agit de déterminer la probabilité d'un événement sachant que l'on dispose déjà d'une certaine information sur le résultat. Cette idée peut être formulée par la règle dite de conditionnement de Bayes :

— Si $p(A) > 0$ Alors $p(B/A) = p(A \bigcap B)/p(A)$.

Ou' $p(A)$ s'interprète alors comme le degré conditionnel de croyance dans B sachant A.
L'application de la règle de Bayes suppose la connaissance à priori de chaque événement. Le raisonnement probabiliste basé sur la règle de Bayes est utilisé dans le système expert PROSPECTOR [Duda 82].

Les mesures de crédibilité et de plausibilité

En théorie des probabilités, lorsqu'on affecte une partie de la masse totale de probabilité à un événement A, la masse restante doit obligatoirement être répartie sur $\neg A$ au niveau de l'ensemble référentiel Ω sur lequel on se place ; ce qui implique une distribution de la masse totale sur des singletons. Pour assouplir cette contrainte, les chercheurs ont proposé des mesures d'incertain différentes des probabilités mais ayant en commun des propriétés minimales.

Shafer [Shafer 76] a autorisé la répartition de la masse totale sur n'importe quel sous ensemble de Ω (à l'exception de l'ensemble vide).
Soit m une fonction de pondération de Ω dans $[0, 1]$ telle que :

$$m(\emptyset) = 0; \tag{1.2}$$

$$\sum_{A \in \Omega} m(A) = 1. \tag{1.3}$$

La mesure de crédibilité cr basée sur m s'exprime alors par :

$$\forall A \in \Omega cr(A) = \sum_{B \subset A} m(A). \tag{1.4}$$

Par dualité, on peut définir une mesure de plausibilité pl à partir de cr comme :

$$\forall A \in \Omega, pl(A) = 1 - cr(\neg A); \tag{1.5}$$

Ce qui exprime le fait qu'une proposition est d'autant plus crédible que la proposition contraire n'apparaît pas plausible. Ainsi dans le cas d'une ignorance totale :

$$cr(A) = cr(\neg A) = 0 \tag{1.6}$$

et

$$pl(A) = pl(\neg A) = 1. \tag{1.7}$$

$$cr(A) = cr(\neg A) = 0 \tag{1.8}$$

et

$$pl(A) = pl(\neg A) = 1. \tag{1.9}$$

Le calcul des possibilités et des nécessités

Si pour toute proposition A et B

$$cr(A \bigwedge B) = min(cr(A), cr(B)), \tag{1.10}$$

la mesure de crédibilité est appelée alors mesure de nécessité, et elle sera notée N. De même, si pour toutes propositions A et B

$$pl(A \bigvee B) = max(pl(A), pl(B)), \tag{1.11}$$

la mesure de plausibilité est appelée alors mesure de possibilité, et elle sera notée Π. La relation qui lie les mesures de nécessité et de possibilité est, pour toute proposition A

$$N(A) = 1 - \Pi(\neg A). \tag{1.12}$$

Cette relation exprime le fait que la nécessité d'une proposition correspond à l'impossibilité de la proposition contraire. Une vérité nécessaire est donc vraie dans tous les mondes possibles, contrairement, à une vérité contingente qui est vraie dans le monde réel seulement.

Les ensembles flous

La théorie des ensembles flous [Zadeh 78] est, une généralisation du raisonnement basé sur les mesures de nécessités et possibilités [Zadeh 65]. Elle concerne la définition et la structure d'ensembles aux limites imprécises. L'appartenance à un tel ensemble n'est pas soit vrai soit faux, mais elle s'exprime suivant un continuum de valeurs (degrés d'appartenance) échelonnées dans l'intervalle $[0, 1]$, ou‘ le 0 correspond à la non appartenance absolue et 1 à l'appartenance totale. Les fonctions d'appartenance permettent de représenter et de manipuler aisement des ensembles flous en se servant du formalisme mathématique.
Les règles de calcul des valeurs de possibilités et nécessités "conditionnelles" s'expriment avec les fonctions floues de la manière suivante :
Soit A un ensemble flou sur le référentiel S et μ_A la fonction d'appartenance.

— Si F est un sous-ensemble non flou de S ;

$$\Pi(F/A) = sup_{s \in F} \mu(s) \tag{1.13}$$

et

$$N(F/A) = inf_{sinF}(1 - \mu(s)) \tag{1.14}$$

— Si F est flou, les formules se généralisent en :

$$\Pi(F/A) = sup_{s \in S} min(\mu_F(s), \mu_A(s)); \tag{1.15}$$

$$N(F/A) = inf_{s \in S} max(\mu_F(s), 1 - \mu_A(s)). \tag{1.16}$$

Les techniques des ensembles flous ont été utilisé en décodage phonétique pour affecter un score de reconnaissance à un phonème en fonction de la valeur de l'indice phonétique.

Cas de MYCIN

L'utilisateur en MYCIN attache, explicitement ou par défaut à chaque fait fourni un degré de certitude DC compris entre -1 et $+1$, ou‘ -1 correspond à toujours faux, $+1$ à totalement vrai et le 0 correspond donc à l'ignorance totale.
Par ailleurs, les experts qui fournissent les règles de l'application associent à chacune un cœfficient d'atténuation CA compris lui aussi entre -1 et $+1$. La combinaison d'informations est faite en quelque sorte par extension de la formule des probabilités composées au cas négatif. Ainsi, si les conditions spécifiées par la prémisse sont vérifiées à un degré DC alors la conclusion d'une règle dont le coefficient est CA est certaine au degré :

$$C = DC * CA \tag{1.17}$$

Lorsqu'une prémisse est une simple conjonction de faits élémentaires pouvant figurer dans la base de faits, le degré de certitude de la prémisse, vu comme un tout est calculé comme le minimum des degrés de certitude des faits élémentaires impliqués.

$$DC = min(DC_{fi}) \tag{1.18}$$

ou‘ DC_{fi} désigne le DC du i_{ieme} fait.
De même, si lors du raisonnement, deux règles différentes donnent la même conclusion avec des degré C1 et C2, elles se renforcent mutuellement. Le degré de certitude cumulé C associé à la conclusion après exploitation des deux règles est défini par les formules suivantes :

— Si $C1 > 0$ et $C2 > 0$

$$C = C1 + C2 - C1 * C2 \tag{1.19}$$

— Si $C1 < 0$ et $C2 < 0$

$$C = C1 + C2 + C1 * C2 \tag{1.20}$$

— Si $C1 * C2 <= 0$ et $min(|C1|, |C2|) \neq 1$

$$C = \frac{C1 + C2}{1 - min(|C1|, |C2|)} \tag{1.21}$$

— Si $C1 * C2 <= 0$ et $min(|C1|, |C2|) = 1$

$$C = 1 \tag{1.22}$$

Tout fait dont le cœfficient C est inférieur en valeur absolue à 0.2 est considéré comme peu crédible et éliminé par le système général de la base de faits. La combinaison des scores dans le moteur d'inférence utilisé dans APHODEX et SAPHA est largement inspirée de celle utilisée dans MYCIN.

3 La reconnaissance automatique de la parole

La parole est certainement le moyen le plus direct et le plus naturel utilisé par l'homme pour échanger l'information. Depuis longtemps, la parole passionne les chercheurs qui pensent pouvoir arriver un jour à utiliser ce même moyen pour communiquer avec une machine.
Nous présentons dans cette section, les éléments qui interviennent dans l'émission et la perception de la parole chez l'homme, ainsi que les techniques utilisées en reconnaissance automatique.

3.1 Les organes de la parole chez l'homme

L'appareil phonatoire

Les trois niveaux de production de la parole chez l'homme sont :

Les poumons qui sont le siège des échanges gazeux de la respiration. Pendant l'inspiration les muscles thoraciques et le diaphragme se relâchent, c'est alors que sont produits la plupart des phonèmes. le rôle du souffle respiratoire est de fournir l'énergie nécessaire à la production de la parole. La trachée artère est un conduit approximativement cylindrique de 1.5 cm à 2 cm de diamètre et d'environ 12 cm de long qui communique avec les poumons par les bronches et se termine par le larynx.

Le larynx suspendu à l'os hyoide, il est essentiellement constitué de 5 pièces cartilagineuses reliées par des ligaments et par des muscles. La muqueuse laryngée tapisse l'intérieur et comporte deux paires de replis : l'une inférieure forme les cordes vocales et l'autre supérieure ; les bandes ventriculaires. Entre les deux, se situe le ventricule de Morgani.
Le pharynx peut être mobilisé lors de la production de la parole de deux manières différentes soit pour produire des sons chuchotés soit pour produire des sons voisés.

Le conduit vocal qui comporte deux parties : le conduit oral auquel est connecté le conduit nasal lorsque le velum est baissé.

Le conduit oral est constitué du pharynx qui communique avec le larynx au niveau de l'épiglotte et de la cavité buccale dont la forme est déterminante pour produire la spécificité de chaque phonème.

La cavité nasale est séparée en deux longitudinalement par le septum qui délimite ainsi deux fosses nasales débouchant sur les narines.

Les grammairiens arabes ont donné une description assez précise des organes de l'appareil phonatoire. Ils font notamment la distinction entre différentes parties du larynx dont l'une semble désigner le pharynx ; le palais et surtout la langue sont divisés en plusieurs parties (voir la figure 1.2 tirée de [Meloni 82]).

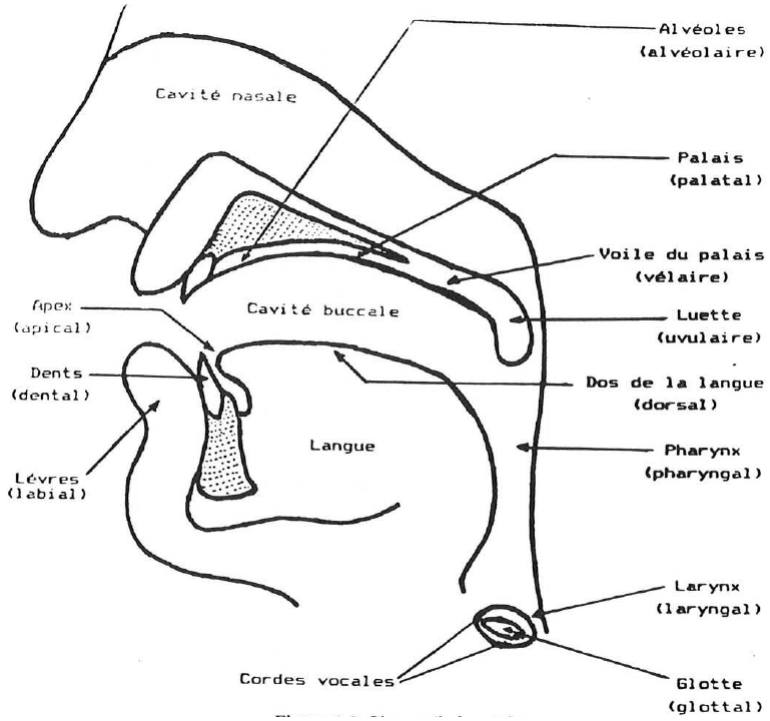


FIGURE 1.2 – L'appareil phonatoire

L'appareil auditif

L'audition met en jeu deux types de structures anatomiques :

L'oreille qui permet la transmission et l'analyse des sons. elle est composée de trois parties :

— L'oreille externe comprend le pavillon et le conduit auditif externe qui se termine par le tympan. L'oreille externe canalise les ondes acoustiques et

le tympan les transforme en vibrations mécaniques.

— L'oreille moyenne est un ensemble de cavités qui communiquent entre elles. Du fait de la contraction des muscles, l'oreille moyenne s'adapte au bruit et joue ainsi le rôle de protecteur de la partie interne de l'oreille.

— L'oreille interne est une cavité osseuse de forme très complexe qui comprend le vestibule, les canaux semi-circulaire et la cochlée.

L'appareil neuro-sensoriel elle est composée essentiellement de :

— L'organe de Corti qui comporte deux sortes de cellules : les cellules de soutien et les cellules auditives sensorielles. C'est dans ces dernières que naît l'influx nerveux.

— Les fibres nerveuses auditives qui se rassemblent dans le nerf cochléaire pour former avec le nerf vestibulaire l'un des nerfs de la huitième paire de nerfs craniens.

3.2 Les avantages de la communication parlée

Les avantages de la communication orale homme-machine sont multiples. Ce mode de relation libère complètement l'usage de la vue et des mains et laisse l'usager libre de ses mouvements. La vitesse de transmission des informations est supérieur dans le sens homme-machine à celle qui permet l'usage du clavier. De plus la voix informe sur l'identité du locuteur et peut être par ailleurs transportée par des moyens simples existants, comme le réseau téléphonique. Ses avantages se traduisent par l'existence d'une grande variété d'applications liées à la reconnaissance automatique de la parole, à titre d'exemple on peut citer :

CAO : Commandes des machines.

Robotique : Guidage d'un robot.

Transport : Réservation automatique.

Postes : Demande de renseignements.

Médecine : Aide aux handicapés.

Bureautique : Machine à dicter à commande vocale.

Le développement de telles applications vise à améliorer le confort de l'utilisateur, à augmenter l'efficacité de la communication et à offrir de nouvelles possibilités.

3.3 ... Et les difficultés

La reconnaissance automatique de la parole est encore un problème difficile à résoudre. Cette situation résulte de la complexité du langage parlé. La parole présente un certain nombre de caractéristiques spécifiques qui demeurent très difficiles à analyser. En effet, si une phrase, en général, ne s'écrit que d'une seule manière, elle possède une infinité de prononciations toutes différentes. L'absence de silence entre les mots, la très grande variabilité interlocuteur et intralocuteur rendent délicate la définition d'invariants. La présence du bruit et des perturbations apportées par le microphone compliquent davantage le problème.

3.4 Les techniques de reconnaissance

Pour réaliser des systèmes de reconnaissance, deux approches principales sont utilisées, l'une consiste à reconnaître globalement des mots séparés par des intervalles de silence, et l'autre permet en revanche d'aborder le problème de la reconnaissance de la parole continue éventuellement dans un contexte multilocuteur. Nous allons décrire ces deux approches.

La méthode globale

Dans cette approche, l'unité de base est le plus souvent le mot considéré comme une entité globale. Cette méthode est caractérisée par le fait qu'une phase d'apprentissage est nécessaire pendant laquelle l'utilisateur prononce la liste des mots du lexique de son application.
Lors de la phase de reconnaissance, le mot à reconnaître est comparé à tous les mots de références du lexique. Le mot ressemblant le plus au mot prononcé est alors reconnu. Les avantages d'une telle approche sont d'une part l'indépendance vis-à-vis des particularités de la langue du fait de la phase d'apprentissage, et d'autre part, l'excellente capacité de reconnaissance pouvant atteindre les 99%. Néanmoins, le vocabulaire est assez limité et les systèmes sont le plus souvent monolocuteur, de plus la prononciation en mots isolés est peu naturelle.

La méthode analytique

Cette approche tente de détecter et d'identifer des unités élémentaires (phonèmes, diphonèmes, syllabes) puis de reconnaître la phrase effectivement prononcée. La méthode fait apparaître plusieurs modules qui communiquent entre eux.
Le module acoustique a pour rôle d'extraire les caractéristiques du signal de parole

destinées aux module phonétique et prosodique. Le module prosodique sert a trouver les informations sur le rythme et l'intonation de la phrase et le module phonétique traduit la liste des indices en une suite d'unité phonétiques. Le module phonologique porte sur les phénomènes de la langue dont le contenu phonétique est modifié par les articulations rapides, les liaisons et les variétés dialectales.
Au niveau du lexique, interviennent les informations sur les mots qui composent la langue. Le module syntaxique renferme les règles de la grammaire qui permettent de décrire et d'analyser la langage en termes grammatical et fonctionnel et permet donc de définir toutes les séquences de mots acceptables.
Le niveau sémantique permet de donner la signification de l'énoncé et le rejet des phrases syntaxiquement correctes n'ayant aucune interprétation. Le module pragmatique, utilisé en dialogue, permet de déterminer le sens de la phrase dans le contexte de l'application et de gérer l'historique du dialogue [Pierrel 87].

4 Le décodage acoustico-phonétique

De toutes les opérations décrites par les différents modules de l'approche analytique, la transformation du signal vocal en une suite d'étiquettes phonétiques est la plus fondamentale. Toute erreur à ce niveau augmente considérablement l'indéterminisme des traitements ultérieurs. Le décodage acoustico-phonétique est lié à deux aspects importants en reconnaissance de la parole : la représentation paramétrique et l'identification phonétique.

4.1 La représentation paramétrique

Elle consiste en la conversion du signal de parole en une représentation linguistique structurée. Son rôle est la réduction des données redondantes et le calcul des paramètres qui permettent de distinguer les phonèmes [Gong 85]. Les principales représentations sont :

La représentation temporelle

Le signal sonore qui constitue le message est préalablement converti par un microphone en un signal électrique, le résultat est une tension alternative dont l'amplitude varie de façon continue en fonction du temps (signal analogique). Pour être exploité par ordinateur, le signal passe par une phase de digitalisation qui le transforme en une suite de nombres mesurant son amplitude à des intervalles de temps successifs très brefs (on parle d'échantillonnage du signal) à partir de quoi on calcule les pa-

ramètres : énergie, nombre de passage par zéro, fréquence fondamentale, etc. C'est une représentation sensible au bruit et dépend de la fréquence fondamentale et du déphasage de la voie de transmission.

La représentation fréquentielle

Elle est obtenue en mesurant pendant des intervalles de temps plus larges, les éléments des différentes fréquences qui composent le signal et leurs amplitudes. Elle est assurée en exprimant le signal (de forme compliqué) en une combinaison linéaire de fonctions de base (sinusoïdes ou exponentielles) dont les propriétés sont bien connues et qui sont facilement manipulables.
La représentation fréquentielle la plus utilisée est obtenu par l'application d'une transformée de Fourier sur le signal :
Si le signal $x(n)$ est discret, la transformée de Fourier est définie par :

$$X(w) = \sum_{n=-\infty}^{\infty} x(n) - jwn \tag{1.23}$$

et elle se calcule par un algorithme de transformée de Fourier rapide (FFT) [Gong 83].
Un spectrogramme est un graphique de ce type de représentation, appelée parfois représentation temps-fréquence-amplitude :

-Abcisse : temps.
-Ordonnée : fréquence.
-Niveau de gris : amplitude.

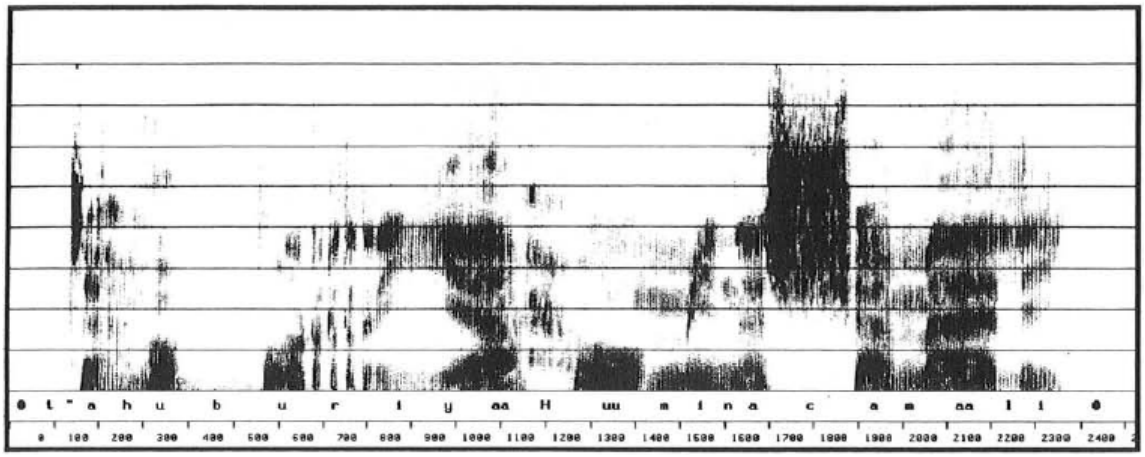


FIGURE 1.3 – Exemple de spectrogramme

La représentation fréquentielle permet de réduire d'un facteur dix environ le flux

d'informations représenté par le signal vocal et d'éliminer les redondances présentes dans celui-ci, elle est d'ailleurs utilisée par le système auditif humain.

L'analyse par prédiction linéaire

Abrégée en LPC, comme "Linear Prediction Coding", la méthode a été utilisée pour la transmission de la parole par compression importante de données. Elle fait intervenir le modèle de production de la parole, elle est donc bien adaptée à sa représentation paramétrique. L'extraction des paramètres se fait de manière très rapide dans le domaine temporel. L'idée est d'exprimer le signal à un instant donné comme combinaison linéaire de son passé, tout en essayant de régler les cœfficients de combinaisons pour que l'énergie de l'erreur entre le signal réel et le signal prédit soit minimale. Une transformée de Fourier à partir de cœfficients LPC permet de rendre plus précises la largeur de bande et la position des formants, mais la LPC présente l'inconvénient de représenter mal les creux dans le spectre et ne convient donc pas bien pour les sons nasalisés.

4.2 Les techniques de décodage phonétique

Les techniques utilisées en décodage phonétique peuvent être classées selon trois approches :

1. Une approche de classification automatique avec ou sans apprentissage utilisée pour définir des prototypes.
2. Une approche de reconnaissance des formes qui consiste à affecter des étiquettes à des segments grâce à des critères de proximité.
3. Une approche basée sur la reconnaissance de traits ou‘ l'on retrouve les approches système expert.

Nous allons décrire brièvement les méthodes qui sont amplement utilisées pour l'identification.

La quantification vectorielle

Elle consiste à utiliser les probabilités statistiques des sons dans leur espace de représentation (spectre d'amplitude, cœfficients cepstraux ou de prédiction). Cette technique rentre dans la problématique plus générale de la classification automatique. Elle part du postulat que deux formes proches dans leur espace de représentation sont aussi proches en soi. L'utilisation de cette technique pose plusieurs difficultés :

— La représentation correcte des formes de références.
— Le découpage de l'espace en classes pertinentes.
— Le choix de la métrique dans cet espace.

Cette technique nécessite une phase d'apprentissage pour constituer le dictionnaire de référence. L'autre inconvénient de cette méthode est qu'elle permet de faire une série de décisions locales mais sans référence aux échantillons passés sans tenir compte des phénomènes de coarticulation. Son principal avantage est de réduire considérablement le débit par référence à une liste de prototypes connue à priori.

La comparaison dynamique

En utilisant le principe de mise en correspondance optimale, la programmation dynamique permet de tenir compte des distorsions temporelles entre deux formes à comparer. Chaque mot du lexique est représenté par une suite de vecteurs $r(1), ..., r(j)$, chaque forme à reconnaître est représentée par $t(1), ...t(i)$. Il faut trouver un chemin de recalage qui à chaque vecteur de T, fait correspondre un vecteur R. Ce chemin devra prendre en compte les contraintes naturelles de la parole. Ensuite, parmi tous les chemins de recalages possibles, il faut choisir celui dont la somme des distances le long du chemin est minimale. L'algorithme permet d'éliminer rapidement des références lorsque des différences notables apparaissent au cours d'une étape quelconque de l'examen de la référence à reconnaître. Les avantages de la méthode sont d'une part son excellente capacité de reconnaissance et son indépendance vis à vis des particularités de la langue. Mais dans ce cas le problème de l'apprentissage et de la segmentation se posent de manière plus aigue" que pour la quantification vectorielle, c'est pourquoi elle est souvent utilisée en aval de la quantification vectorielle et sert à la reconnaissance des mots isolés ou enchaînés [Sakoe 8].

Les modèles stochastiques

Dans les modèles stochastiques, les formes acoustiques des références sont représentées par un graphe sous forme d'une chaîne de Markov ou plus précisément par des modèles de Markov cachés HMM (Hidden Markov Model). Le graphe est composé d'un nombre fini d'états représentant les segments stables du signal, tandis que les variations spectrales sont modélisées par les arcs de transition. Ce graphe peut être vu comme un modèle de production d'un mot, ou‘ chaque transition est accompagnée de l'émission d'un vecteur de paramètres spectraux. A chaque état S_j sera associé une distribution de probabilité $P(a_i/S_j)$, probabilité de produire l'évé-

nement a_i sur une transition d'origine S_j, et à chaque arc sera associé une probabilité de transition $P(S_j/S_i)$, probabilité que le modèle passe de l'état S_i à l'état S_j en une seule transition. Les paramètres du modèle sont obtenus au cours de la phase d'apprentissage à partir d'un nombre important d'énoncés du même mot. Ce qui donne à la méthode un avantage majeur de prise en compte de la variabilité du signal vocal. En revanche, s'appuyant sur une modélisation purement mathématique, ils ne permettent pas d'introduire de façon explicite des connaissances phonétiques [Baker 75].

Les modèles connexionistes

Ils sont fondés sur une modélisation des réseaux de neurones. Ces derniers possèdent des avantages forts intéressants tels que le parallélisme, le raisonnement à partir de données incomplètes et la capacité de généralisation. Nous assistons actuellement à un regain d'intérêt pour l'utilisation des modèles connexionistes en reconnaissance automatique de la parole même s'ils n'ont pas encore prouvé leur supériorité par rapport aux autres méthodes.

Les méthodes décrites ci-dessus sont plutôt des techniques de reconnaissance de formes qui s'avèrent mal adaptées à la reconnaissance d'unités phonétiques instables. L'identification phonétique nécessite la prise en compte du contexte. Un problème qu'on ne peut résoudre avec les méthodes de reconnaissance de formes. La solution généralement adoptée est d'utiliser des techniques d'IA basées sur la reconnaissance par des traits phonétiques.

La reconnaissance de traits

Cette approche met en jeu plusieurs types de connaissances et cherche à représenter les entités phonétiques (phonèmes) en terme de traits et contexte. Ces connaissances peuvent être mises sous formes de règles et l'on peut employer à partir de là une stratégie de système expert. La reconnaissance s'effectue en général en deux étapes :

— la segmentation du signal en grandes classes phonétiques. Cette phase a pour but de délimiter des segments sur la signal de parole,

— l'étiquetage des segments en utilisant les traits et les contextes. Cette phase consiste à identifier les segments obtenus lors de la segmentation en affectant à chaque segment une suite d'étiquettes phonétiques. Le résultat est le plus souvent un treillis phonétique utilisé comme entrée aux modules du niveau supérieur.

Parmi les systèmes utilisant cette approche nous pouvons citer APHODEX [Fohr 86], SERAC [Gillet 84] et le système SAPHA [Djoudi 89c] que nous avons développé pour le décodage phonétique de l'Arabe et qui fait l'objet de cette thèse.

5 Conclusion

Nous avons présenté dans ce chapitre de généralités, les notions les plus importantes sur lesquelles reposent l'essentiel de notre travail de décodage acoustico-phonétique de l'Arabe standard. Nous allons présenter dans le chapitre suivant une étude phonétique de la langue avant d'aborder par la suite le décodage phonétique et les problèmes qu'il faut résoudre.

Chapitre 2

L'étude phonétique

1 Introduction

Le développement d'un système de décodage phonétique de l'Arabe nous impose de bien étudier la composante phonétique de la langue afin de dégager les caractéristiques et de bien cerner les difficultés qui existent. Nous présentons dans ce chapitre une étude phonétique faite à partir d'un corpus de phrases et de mots et enrichie par les études anciennes et nouvelles.

2 Passé et présent de l'Arabe

L'Arabe est une langue sémitique, historiquement formée de deux branches : l'Arabe du Sud et l'Arabe du Nord. La branche sud est née au sud de l'Arabie, elle inclut entre autre les langues sabaienne et himyarite. Elle est similaire à l'Arabe du nord aussi bien dans ses formes grammaticales que dans son vocabulaire. L'Arabe du sud devient une langue morte après la chute de l'empire himyarite au 6^{eme} siècle après J.C. Depuis ce temps, la branche nordique est devenue la langue sémitique la plus importante. On raconte que le premier homme a avoir parlé en Arabe s'appelait yarub ben qahtan dans la région du yemen, d'ou‘ le nom donné à la langue. Les premiers spécimens de la langue arabe sont des inscriptions dites lihyanites et thamoudéennes datant du deuxième siècle de notre ère. Le plus ancien texte manifestant l'existance de l'Arabe a été trouvé à Annemara près de Damas. Il date de l'an 326 après J.C et orne le tombeau d'un roi arabe, mais l'écriture arabe apparaît pour la première fois dans deux inscriptions, celle de Zabad près d'Alep (en l'an 512) et celle de Huran au sud de Damas (vers 568 après J.C). Ces textes épigraphiques sont malheureusement trop courts et d'un contenu trop mince pour nous renseigner sur la génèse de l'arabe classique.

Pendant la période préislamique, la langue arabe était développée dans une très riche langue poétique et devient la langue commune des médias de l'Arabie préislamique.

La diffusion du Coran par l'essor de l'Islam a entraîné la langue dans une extension telle que n'en a connue aucune autre langue du monde. Elle a véhiculé la civilisation musulmane sur l'immense espace qu'elle a conquis.
L'étude de la langue par les philologues et les grammairiens arabes répondait à un besoin pressant [Fleisch 61], d'une part la récitation du Coran n'admettait pas la moindre faute de prononciation ni le moindre doute sur son interprétation et d'autre part, la conversion à l'Islam de population non arabes exigeait l'enseignement de la langue aux nouveaux adhérents.

L'étude de l'Arabe a commencé au huitième siècle dans deux grandes écoles, Baṣra et Koufa. Le premier homme à ouvrir la voie à l'étude de la grammaire fut Abu Asswad Edduali. On raconte que le calife Ali aurait donné à Abu Asswad une indication sur les divisions du mot, qu'il convenait d'établir dans la grammaire ; Ali aurait ajouté : / ?unħu/ "engage toi dans cette route", d'ou‘ le nom de /naħw/ donné à la grammaire. A ce grammairien succéda El Khalil Ibn Ahmed et Sibawayh. Ce dernier a laissé un remarquable ouvrage le kitab [Sibawayh 89], ou‘ l'essentiel des faits grammaticaux se trouvent réunis. Ensuite vinrent d'autres grammairiens parmi lesquels nous citons particulièrement Qoutrob, Al Asmai, Al Mazini et Ibn jinni [Jinni 54].

La fortune de l'écriture arabe a dépassé celle de la langue littéraire, puisqu'elle sert à noter diverses autres langues telle que le turc et le persan. Cette écriture s'est développée au 7^{eme} siècle en une cursive rapide ou‘ les lettres sont généralement jointes. C'est une écriture consonantique qui se lit horizontalement de droite à gauche.

L'Arabe standard moderne est une continuité linguistique de l'Arabe classique. Elle est la langue commune à plus de 100 millions d'arabophones, et une langue liturgique de l'Islam pour près d'un milliard de musulmans. Elle est aussi la langue de la science, de l'enseignement et de la littérature, celle du théatre, de la presse, de la radio et de la télévision.

En dépit de l'acceptation quasi-unanime de l'Arabe standard contemporain et son adoption générale comme le moyen commun de communication à travers le monde arabe, il n'est pas la langue du quotidien du peuple. Les citoyens d'un pays ou d'une région quelconque trouvent qu'il est plus facile et plus convenu de se parler dans leur propre et particulier dialecte. Les différences dans la phonologie, la morphologie et la syntaxe de ces dialectes sont souvent si grandes qu'une commu-

nication verbale entre deux illettrés de deux pays différents est très difficile sinon impossible. Par conséquent une connaissance de l'Arabe standard contemporain devient impérative lorsque les locuteurs des différentes régions dialectales sont obligés à communiquer.

Dans le cadre de nos recherches sur la reconnaissance de la parole arabe, nous présentons dans ce qui suit une étude phonétique et phonologique de l'Arabe moderne standard. Pour des raisons de commodité, nous utiliserons tout le long de la thèse, la notation interne pour la codification des phonèmes.

3 L'outil d'analyse

Snorri [Laprie 88] est un outil d'observation destiné aux spécialistes comme aux non spécialistes de la parole. Il possède les fonctionnalités classiques pour acquérir et restituer le signal de parole, le visualiser et calculer le spectrogramme, ainsi qu'un certain nombre de fonctions permettant d'analyser plus finement la parole et de manipuler des corpus. Snorri fonctionne sur poste de travail concurrent 5600 ou sur SUN. Il utilise une carte d'acquisition 12 ou 16 bits. Tous les calculs de traitement du signal sont effectués sur un processeur vectoriel qui multiplie par un facteur de 10 la vitesse d'exécution. L'affichage a lieu sur une console graphique couleur. Les copies d'écran sont faites sur une imprimante postcript avec une résolution de 400 points par pouce. Snorri est programmé en C et fait appel au multifenêtrage et aux primitives graphiques sous unix et Xwindows.

4 La méthode d'analyse

La description acoustique des sons de parole s'appuie sur la représentation fréquence-temps (spectrogramme). L'intensité des composants spectraux est donné par le degré de noirceur du tracé. Après l'acquisition du signal de parole, nous calculons la transformée de Fourier par un algorithme FFT sur le signal numérique. L'affichage se fait sur une console graphique. A partir de cette représentation, nous avons effectué une analyse acoustique des phonèmes de l'Arabe standard, analyse que nous avons enrichie par notre recherche bibliographique sur les différents aspects de la phonétique arabe. Cette étude a porté sur des séquences consonne-voyelle, des paires minimales et des phrases lues.

5 Structure acoustique des voyelles

Les voyelles se caractérisent principalement par la présence de zones de fréquences ou‘ les harmoniques sont particulièrement intenses (formants) qui apparaissent sur le spectrogramme sous la forme de bandes noires plus ou moins parallèles à l'axe des temps. Les expériences ont montré que la position fréquentielle des trois premiers formants caractérisait le timbre vocalique. L'explication qu'on donne aux trois formants notés F1, F2 et F3 est la suivante :

— F1 naît dans la cavité résonnante comprise entre le larynx et le dos de la langue.

— F2 naît dans la cavité résonnante située entre le dos de la langue et les lèvres.

— F3 dépend de l'arrondissement des lèvres.

5.1 Le timbre vocalique

L'Arabe standard comporte trois voyelles /ħarakaat/ qui s'opposent phonologiquement par le timbre. Un texte arabe ne les note pas habituellement. Pour les besoins de l'enseignement ou la nécessité de fixer la prononciation exacte d'un mot ou d'un texte on est amené à les exprimer par des signes extérieurs écrits au dessus ou en dessous des consonnes. Ces voyelles sont : /a/ : /fatħa/, /i/ : /kasra/ et /u/ : /ḍamma/. De même, on note le /sukuun/ pour désigner l'absence de voyelle après la consonne. La description que l'on donne à ces voyelles est la suivante :

/a/ est une voyelle centrale ouverte, elle se prononce en ouvrant largement la bouche et en conservant la langue dans une position horizontale.

— elle est prononcée comme /a/ moyen franĉais au voisinage des consonnes emphatiques /ṭ/, /ḍ/, /ṣ/, /ð̣/, vélaires /χ/, /γ/, /q/ et pharyngales /ħ/ et /ε/.

— elle est prononcée entre /a/ moyen et /e/ ouvert ailleurs.

/i/ est une voyelle antérieure fermée qui se prononce en portant le devant de la langue en avant et en l'étalant largement tandis que l'arrière frôle presque le palais.

— elle est prononcée entre /i/ et /é/ fermé au voisinage des consonnes emphatiques, vélaires et pharyngales.

— elle est prononcée comme /i/ fermé au voisinage des autres consonnes.

/u/ est une voyelle postérieure fermée qui se prononce en contractant la langue au fond de la bouche et en avanc̃ant les lèvres qui s'arrondissent jusqu'à presque se joindre.

— elle est prononcée entre /ou/ et /o/ fermé franc̃ais au voisinage des emphatiques, des vélaires et des pharyngales.

— elle est prononcée comme /ou/ franc̃ais ailleurs.

Dans le plan F1-F2, les voyelles /a/, /i/ et /u/ seront disposées aux extrémités d'un triangle pointé vers le bas. Ce triangle représente grossièrement la position moyenne de la langue dans la cavité buccale selon deux axes dits "antérieur-postérieur" et "ouvert-fermé" (voir figure 2.1).

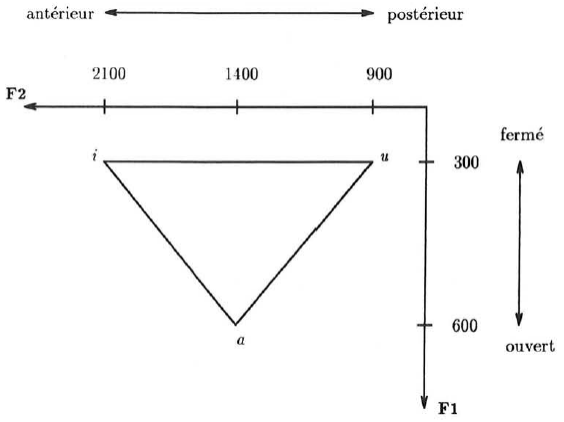


FIGURE 2.1 – Les voyelles courtes dans le plan (F1,F2)

Nous pouvons donc interpréter l'augmentation de F1 comme le résultat d'une ouverture articulatoire et une augmentation de F2 comme une antériorisation de l'articulation. L'arrondissement des lèvres se traduit par une baisse de F3.

5.2 La quantité vocalique

Le système vocalique comprend deux quantités phonologiques pour chaque timbre. A chaque voyelle brève /a/, /i/ et /u/ s'oppose respectivement une voyelle longue /mad/ /aa/, /ii/ et /uu/.

Les voyelles longues sont représentées dans un texte arabe par des signes d'allongement /ʔ/ dépourvu de /hamza/, /j/ et /w/. Elle sont toujours considérées comme des monophtongues et non pas des diphtongues même si dans la tradition des grammairiens arabes, une voyelle longue est perc̃ue comme deux voyelles brèves.

C'est ainsi qu'on note par V une voyelle brève et par VV une voyelle longue. En Arabe, la durée des voyelles et l'opposition temporelle brève/longue sont fondamentales aux niveaux grammatical et sémantique.
La durée relative d'une voyelle dépend de son environnement et de la vitesse d'élocution. Nous donnons dans le tableau suivant les valeurs moyennes de la durée en milliseconde des voyelles.

Voyelle	Durée en ms
a	75
u	65
i	70
aa	160
uu	140
ii	150

TABLE 2.1 – La durée moyenne de voyelles

Toutefois, en position finale, les voyelles brèves sont caractérisées par une durée plus grande et les voyelles longues par une durée moins importante.
L'étude de l'organisation temporelle de la quantité vocalique ne peut se limiter à la mesure de la durée de la phase de manifestation la plus caractéristique de la voyelle, elle nécessite la prise en compte des manifestations des consonnes adjacentes.

5.3 Les fréquences formantiques

Nous donnons dans le tableau suivant les valeurs moyennes des formants des voyelles de l'Arabe standard portant sur cinq locuteurs masculins. Le corpus est constitué de l'ensemble des voyelles contenues dans le corpus DJOUMA [Djoudi 89b].
Il existe une différence entre ces valeurs et celles données par [Ani 70] qui a calculé

formant	a	i	u	aa	ii	uu
F1	550	340	350	590	330	350
F2	1400	1900	1070	1460	2020	890
F3	2600	2700	2230	2570	2800	2270

TABLE 2.2 – Les valeurs moyennes des formants des voyelles

les moyennes pour des voyelles supposées prononcées en isolés. Les valeurs données

par [Belkaid 84] sont sensiblement différentes des nôtres. Le calcul étant fait sur un échantillon de mots prononcés par un seul locuteur.

L'analyse du corpus révèle une grande variabilité dans les fréquences formantiques des voyelles. Nous pouvons distinguer deux sources principales de variabilités, l'une liée à des différences physiologiques entre les locuteurs, et l'autre aux effets de la coarticulation et l'influence du contexte. En plus de ces variantes dans la prononciation des voyelles, il existe un phénomène phonétique dit /?imaala/ qui concerne le timbre /a/. Le phonème est antériorisé, son point d'articulation se rapproche de celui du /i/. D'après Sibawayh [Sibawayh 89], le phonème /aa/ subit le /?imaala/, si la syllabe précédente ou suivante comporte un /i/, exemple /miθel/ (exemple). La présence des consonnes pharyngalisées empêche le /?imaala/. De même, les travaux sur la synthèse de la parole confirment, l'existence d'une différence spectrale importante entre des voyelles en fonction de la présence des consonnes emphatiques ou non [Guerti 87] [Mouradi 87].

5.4 Les diphtongues

L'Arabe comporte deux diphtongues : /aw/ et /ay/ comme dans /θawb/ (vêtement) et /$z$ayb/ (poche).

6 Les consonnes

Le système consonantique de l'Arabe standard est généralement décrit comme étant composé de 28 consonnes /$\hbar$uruuf/ : ce sont les 28 lettres de l'alphabet. Les grammairiens arabes comptent 29 consonnes, la consonne surnuméraire étant le Alif. Le Alif n'est en fait qu'un signe graphique servant de support à la plosive glottale /?/ et par conséquent à la voyelle qui peut l'accompagner. Au plan de l'écriture, la plupart de ces consonnes prennent quatre formes, en apparence différentes, selon qu'elles sont isolées, initiales, médiales, ou finales. Certaines consonnes ne diffèrent entre elles que par l'absence ou l'existence d'un, deux ou trois points diacritiques suscrits ou souscrits. Les points diacritiques fixent la valeur de 15 consonnes sur 28. L'omission, l'addition ou le déplacement d'un seul de ces points suffit à modifier profondément l'aspect et la nature d'un mot.
Pour mettre en évidence l'importance des points diacritiques, voici un exemple de phrase ou‘ 4 mots ont la même graphie et seuls les points diacritiques permettent de les différencier :

قيل قتل فيل قبل طلوع الشمس

traduction : "Il est dit qu'un éléphant a été tué avant le lever du soleil."
Sur le plan acoustique, les consonnes forment une classe très hétérogène que l'on peut décomposer en sous classes ayant des caractéristiques distinctes.

6.1 Qualités des consonnes

Nous distinguons plusieurs qualités de consonnes, les plus importantes sont :

1. Le /*z*ahr/ et son contraire le /hams/ qui correspond à l'opposition sourde/sonore.

2. Le /ʔitbaaq/ (emphase) et son contraire le /ʔinfita*ħ*/. L'emphase s'applique aux consonnes pour lesquelles, lors de leur réalisation, la langue se plie et s'incurve pour former un creux dans lequel le son est pressé.

3. Le /taf*χ*iim/ dont le contraire est le /tarqiiq/, s'applique aux consonnes et aux voyelles. Il traduit une expression acoustique grasse.

4. Le /ʔisti*ɛ*laʔ/ : cette qualité décrit le mouvement articulatoire que fait la langue quand elle se meut vers la partie postérieure de la cavité buccale.
5. Les consonnes /mudlaqa/ sont /l/, /r/, /m/, /n/, /b/ et /f/.
6. Les consonnes /ʔal-qalqala/ sont /q/, /*z*/, /ṭ/, /d/ et /b/.
7. Les consonnes /ʔas-saffir/ sont les sifflantes /s/, /ṣ/ et /z/.
8. Les consonnes /ʔal-liin/ sont les consonnes douces /w/, /y/ et /ʔ/, appelées aussi lettres de prolongation (/*ħ*uruuf al-mad) car elles servent à représenter les voyelles longues.
9. Le /ʔin*ħ*iraaf/ définit le caractère latéral du /l/.
10. Le /takriir/ caractérise la consonne /r/ dans l'articulation apicale à battements.
11. Le /tafa$\int$ $\int$i/ est une propriété du /$\int$/ dont la constriction est médiane.
12. Le /ʔistiṭaala/ est une qualité du /ḍ/ : l'allongement du fait de son appendice latéral.

Les consonnes sont donc caractérisées par leurs modes et leurs lieux d'articulation.

Transcription API	t	k	?	b	d	q	t̪	d̪	z	f	θ	s	ʃ	χ	h	ɣ	Z	ʕ	ɛ	h	s̪	ð̪	m	n	l	r	w	y
Equivalent Arabe	ت	ك	ء	ب	د	ق	ط	ض	ج	ف	ث	س	ش	خ	ح	ذ	ز	غ	ع	ه	ص	ظ	م	ن	ل	ر	و	ي
Code informatique	t	k	A	b	d	q	t̪	d̪	J	f	T	s	c	X	H	D	Z	G	E	h	s̪	ð̪	m	n	l	r	w	y
Plosive شديد	+	+	+	+	+	+	+	+																				
Fricative رخو									+	+	+	+	+	+	+	+	+	+	+	+	+	+						
Nasale أغن																							+	+				
Vibrant منكرر																									+			
Latéral منحرف																										+		
Semi voyelle لين																											+	+
Sourde مهموس	+	+	+			+	+			+	+	+	+	+	+						+							
Sonore مجهور				+	+			+	+						+	+	+	+	+	+		+	+	+	+	+	+	+
Non emphatique منفتح	+	+	+	+	+	+			+	+	+	+	+	+	+	+	+	+	+	+			+	+	+	+	+	+
Emphatique مطبق							+	+													+	+						

FIGURE 2.2 – Les consonnes et leurs modes d'articulation

Nous décrivons ci-dessous les consonnes par grandes classes phonétiques et nous reviendrons plus tard sur la description des consonnes particulières à la langue.

6.2 Les plosives

Physiologiquement, une plosive /∫adiid/ est caractérisée par :

1. la formation d'une fermeture à l'intérieur de la cavité vocale par un ou plusieurs articulateurs à l'endroit ou‘ le conduit de pression est bloqué et qui

Transcription API	t	k	?	b	d	q	ṭ	ḍ	z	f	θ	s	ʃ	χ	h	ð	Z	ɣ	ɛ	h	ṣ	ḍ	m	n	l	r	w	y
Equivalent Arabe	ت	ك	ء	ب	د	ق	ط	ض	ج	ف	ث	س	ش	خ	ح	ذ	ز	غ	ع	ه	ص	ظ	م	ن	ل	ر	و	ي
Code informatique	t	k	A	b	d	q	tˤ	dˤ	J	f	T	s	c	X	H	D	Z	G	E	h	sˤ	Dˤ	m	n	l	r	w	y
Bilabial شفوي				+																			+				+	
Labiodental شفوي سني										+																		
Dental سني											+					+						+						
Alvéodental سنخي سني	+				+		+	+				+					+				+			+				
Alvéolaire سنخي																									+	+		
Palatal حنكي									+				+															+
Vélaire لهوي		+												+														
Uvulaire لهوي						+												+										
Pharyngal حلقي															+				+									
Glottal حنجري			+																	+								

FIGURE 2.3 – Les consonnes et leurs lieux d'articulation

apparaît comme un vide sur le spectrogramme.

2. la brusque libération de cette pression qui apparaît comme une barre d'explosion ou burst sur le spectrogramme.

/t/ : /ta/ est une plosive aspirante alvéodentale non voisée. Le /t/ apparaît sur le spectrogramme comme un burst de durée relative de 20 à 40 msec, plus long avec les voyelles longues. Le burst du /t/ est sous la forme d'une barre verticale suivi par un intervalle de faible bruit. La concentration du burst, avec /a/ et /aa/ est vers 4500 Hz. Avec /u/ et /uu/, le burst est plus long et est concentré vers 3800-4300 Hz. Avec /i/ et /ii/, le burst apparaît plutôt comme une fricative de courte durée concentrée à une fréquence d'environ 5000 Hz. Devant /t/, les transitions des voyelles suivantes sont en général montantes pour le formant F1, et descendantes pour F2 et F3. Les débuts de F2 du /a/ sont approximativement à 1500 Hz et pour le /aa/ ils sont légèrement plus bas autour de 1300 Hz. Les débuts de F2 du /i/ et /ii/ devant /t/ se situent vers 2000 Hz. Les débuts de F2 du /u/ et /uu/ devant /t/ montent brusquement à 1100 Hz.

/k/ : /kaf/ est une plosive postpalatale non voisée et aspirante. Le phonème possède un allophone palatalisé /k'/ qui se produit devant /i/ et /ii/. Le /k/ apparaît sur le spectrogramme comme un burst sous forme d'une barre verticale suivi d'un petit bruit de friction, le tout d'une durée 60-80 msec. et parfois le burst est double. Avec /a/ et /aa/, la concentration du burst est entre 2300 et 2500 Hz, elle est vers 2800 Hz avec /i/ et /ii/ et vers 2100 Hz avec /u/ et /uu/. Le /k/ influe sur les voyelles suivantes en augmentant sensiblement le formant F2 et en abaissant F3, ce qui donne une transition de F2 descendante et celle de F3 montante. Les formants de /u/ et /uu/ semblent ne pas être affectés.

/b/ : /ba/ est décrit comme une occlusive bilabiale voisée non aspirante mais en réalité, il est en variation libre c'est-à-dire qu'il peut être voisé ou non. Le voisement du /b/ apparaît sur le spectrogramme , en basse fréquence vers 50 Hz, avec une durée de 60-110 msec. Le /b/ influe sur les débuts du deuxième formant du /i/ et /ii/ par leur abaissement à 1600-1700 Hz. Les débuts du 2^{eme} formant du /a/ et /aa/ sont légèrement affectés et ceux du /u/et /uu/ ne le sont pas du tout.

/d/ : /dal/ est une plosive alvéodentale voisée non aspirante de durée de 80-100 msec. Il apparaît sur le spectrogramme comme un /t/ excepté pour le voisement. De même l'influence du /d/ sur les voyelles voisines est semblable à celle du /t/.

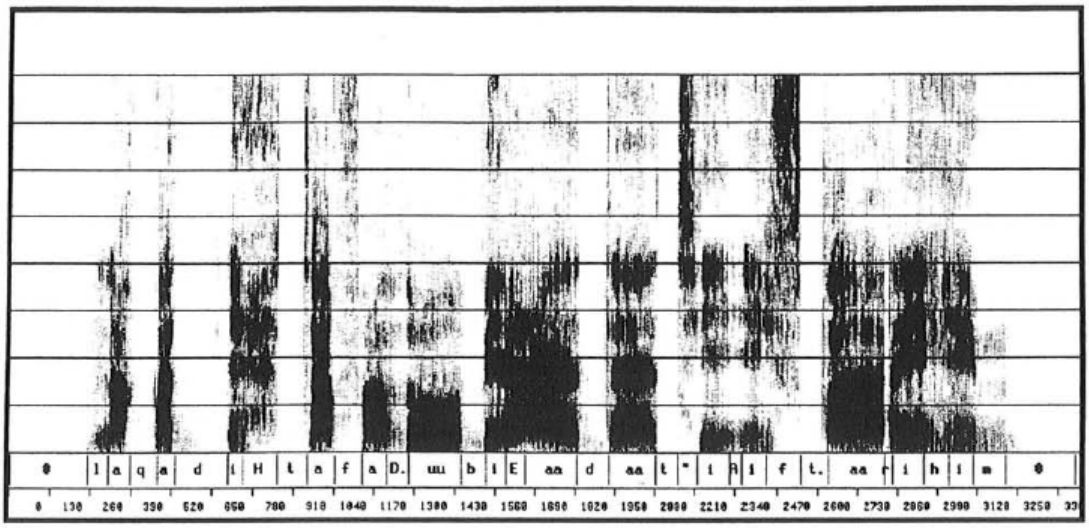


FIGURE 2.4 – Spectrogramme d'une phrase contenant des plosives

/q/ : /qaf/ est une plosive uvulaire non voisée et non aspirante qui apparaît comme un burst important qui commence en basses fréquences et qui monte à 3000 Hz. Celui-ci est suivi d'un petit intervalle de silence de durée moyenne de 30-40 msec sans bruit, ce qui indique l'absence d'aspiration. Le /q/ exerce une influence sur les débuts de F2 du /i/ et /ii/ en les abaissant à 1600 Hz, par contre les débuts de F1 sont légèrement relevés sous l'influence du /q/. Les débuts de F2 du /u/ et /uu/ sont relevés à 900 Hz. Le /q/ produit un début de F2 du /a/ très bas, autour de 1200 Hz, celui du /aa/ est à environ 1150 Hz. En moyenne, la durée relative du silence du /q/ est d'environ 130 msec. Par son caractère vélaire, le /q/ est considéré par certains phonéticiens comme étant une consonne emphatique dont l'homologue serait le /k/.

6.3 Les fricatives

Les fricatives /riχw/ sont produites dans la cavité vocale par une constriction étroite qui rend la circulation d'air turbulente. Acoustiquement, les fricatives non voisées possèdent en général un haut bruit aléatoire et les fricatives voisées possèdent des structures de résonance faibles qui apparaissent comme des ombres de formants faibles avec un léger bruit.

/*z*/ : /*z*im/ est un phonème voisé alvéopalatal et affriqué qui peut être non voisé en position finale. Phonétiquement, le /*z*/ est une combinaison de deux phones [d] et [z]. Cette combinaison plosive-fricative apparaît sur le spectrogramme comme une plosive suivie d'une fricative voisée. Initialement, le [*d*] peut avoir

ou non un burst et il est immédiatement suivi par un bruit aléatoire en hautes fréquences à partir de 2200 Hz avec une continuation de la barre de voisement du [*d*]. Le plus souvent, avec /a/ et /aa/, le bruit est tellement court qu'il apparaît sur le spectrogramme comme un burst. Le [*z*] semble avoir une influence immédiate sur /u/ et /uu/ qui est indiquée par la transition montante des débuts de F2 vers 1000 Hz. Son effet sur les autres voyelles, /*z*/ a tendance à abaisser le formant F2. La durée du /*z*/ est de 100-160 msec. Le [d] tient généralement la moitie de cette durée.

/f/ : /fa/ est une fricative labiodentale non voisée qui apparaît comme un faible bruit aléatoire, sa durée varie entre 80 et 120 msec. Le bruit commence vers 3000 Hz avec /a/ et /aa/, vers 3200 Hz avec /i/ et /ii/et vers 3500 Hz avec /u/ et /uu/ et s'étale en hautes fréquences.

/θ/ : /tha/ est une fricative interdentale non voisée, qui apparaît comme un bruit aléatoire plus fort que le /f/ et de durée de 80 à 120 msec.

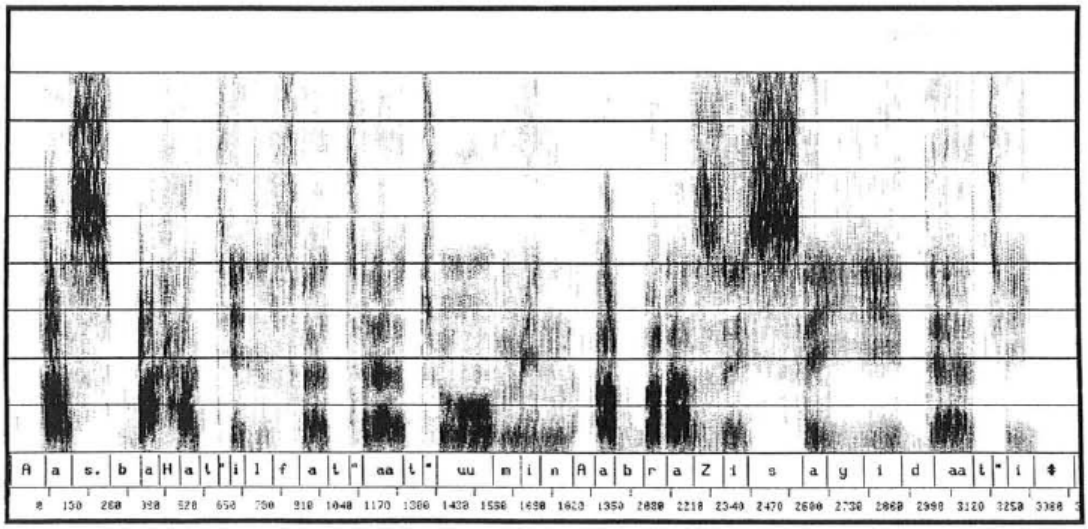


FIGURE 2.5 – Spectrogramme d'une phrase contenant des fricatives

/s/ : /sin/ est une fricative sifflante dentale non voisée qui elle aussi apparaît comme un bruit aléatoire de durée 100-170 msec, en hautes fréquences. Avec /a/ et /aa/, le bruit commence à partir de 2700 Hz, avec /u/ et /uu/ à partir de 2300 Hz et à partir de 2800 Hz avec /i/ et /ii/. Les débuts de F2 du /i/ et /ii/ avec /s/ sont légèrement abaissés à 2000-2100 Hz. Pour le /u/ et /uu/ des transitions aigue"s des débuts de F2 montent à 1000 Hz. Les débuts du /a/ et /aa/ semblent ne pas être affectés avec le /s/ ceci s'explique peut être par le fait que F2 du /s/, du /a/ et du /aa/ sont dans la même région

(1300-1500 Hz).

/$\int$/ : /**chin**/ est une fricative chuintante palatale non voisée qui apparaît comme un bruit aléatoire de durée 100-150 msec, en hautes fréquences. Le bruit commence vers 2000 Hz avec /a/ et /aa/, vers 1900 Hz avec /u/ et /uu/ et vers 1700 Hz avec /i/ et /ii/. Le /$\int$/ possède un bruit aléatoire plus haut en fréquence que la plupart des autres fricatives. Son influence sur les voyelles voisines est presque identique à celle du /s/.

/χ/ : /**kha**/ est une fricative vélaire non voisée qui apparaît comme un bruit aléatoire regroupé sous forme de structure de formant dans la bande 1000-5000 avec /a/ et /aa/, dans la bande 700-4800 avec /u/ et /uu/ et dans la bande 1900-4900 Hz avec /i/ et /ii/. La durée relative du /χ/ est de 100-160 msec. Avec /χ/, les débuts de F2 du /a/ et /aa/ sont à 1500 Hz, les débuts de F2 du /i/ et /ii/ sont abaissés à 1800-1900 Hz, ceux du /u/ et /uu/ légèrement soulevés à 1000 Hz.

/∂/ : /∂**al**/ est décrit comme une fricative interdentale voisée, de durée relative 100-160 msec qui possède une résonance apparaissant comme des formants faibles : F1 vers 275 Hz, F2 vers 1500 Hz et F3 à 2350 Hz. Dans l'espace interformantique, il y a un certain bruit faible. Le /∂/ influe sur les débuts de F2 du /i/ et /ii/ en les abaissant vers 1900 Hz et semble influer aussi sur les débuts de F1 en les soulevant légèrement. Les débuts de F2 du /u/ et /uu/ avec /∂/ sont vers 1000 Hz, ceux du /a/ vers 1500 Hz et ceux du /aa/ à 1400 Hz.

/**z**/ : /**zi**/ est une fricative sifflante dorsoalvéolaire voisée qui semble avoir trois faibles structures de formant F1 vers 250 Hz, F2 à 1600 Hz et F3 à 2400 Hz. En hautes fréquences, /z/ contient un bruit aléatoire à partir de 3000 Hz avec une durée relative de 100-160 msec. Le /z/ influe sur les débuts de F2 du /u/ et /uu/ en les soulevant à 950-1000 Hz. Avec /z/ les débuts de F2 du /i/ et /ii/ sont légèrement abaissés à 2000 Hz et ceux du /a/ et /aa/ à 1500 Hz. F3 semble dans tous les cas montant.

/γ/ : /**Ghayn**/ est décrit comme une fricative grasseyée voisée qui possède deux allophones, l'un est uvulaire près de /a/, /aa/, /u/ et /uu/ et l'autre vélaire près du /i/ et /ii/. Sur le spectrogramme, /γ/ apparaît comme un bruit à structure de formants étalé en basses fréquences et dont la limite supérieure se situe vers 6400 Hz avec /a/ et /aa/, vers 6000 avec /u/ et /uu/ et vers 7000 Hz avec /i/ et /ii/. La durée du /γ/ est de l'ordre de 100-1600 msec. Avec /γ/, les débuts de F2 du /i/ et /ii/ sont abaissés à 1900 Hz, ceux du /u/ et /uu/ sont légèrement soulevés à 850 Hz, ceux du /a/ sont à 1300

Hz et ceux du /aa/ sont à 1250 Hz.

6.4 Les nasales

La nasalité est définie en terme physiologique comme étant la formation d'une ou plusieurs fermetures orales et le passage de l'air à travers le nez. Au cours de la production des nasales les 2 cavités orale et nasale sont donc normalement utilisées. En Arabe, il y 2 consonnes nasales /ghounna/ le /m/ et le /n/ décrites comme suit :

/m/ : /mim/ est une nasale bilabiale voisée de durée de 70-90 msec. Elle possède des résonances faibles qui apparaissent comme des formants F1 à 250 Hz, F2 à 1000 Hz et F3 à 2700 Hz et parfois, d'autres résonances plus faibles juste au dessus de F1. Le /m/ exerce une influence sur les voyelles antérieures fermées /i/ et /ii/ en abaissant les débuts du 2^{eme} formant vers 1850 Hz. Sur les voyelles centrales /a/ et /aa/, l'influence du /m/ se manifeste par un léger abaissement des débuts du second formant, par contre aucune influence sur les voyelles postérieures fermées /u/ et /uu/. En gros, le /m/ semble être similaire au /b/, excepté pour la qualité nasale caractérisée par un modèle de formants nasaux.

/n/ : /nun/ est une nasale alvéodentale voisée de durée 80 à 100 msec. Comme le /m/, le /n/ possède des résonances faibles apparaissant comme des formants F1 à 250 Hz, F2 à 1500-1600 Hz et F3 vers 2800-3000 Hz. Le /n/ relève les débuts de F2 du /u/ et /uu/ à 1300-1400 Hz, abaisse légèrement ceux de /i/et /ii/ à 1950 Hz et n'influe pas du tout sur /a/ et /aa/, ceci est dû au fait que le /n/ possède une chaîne de fréquences similaire à celle de /a/ et /aa/.

6.5 Le vibrant

/r/ : /ra/ est un vibrant /munkarar/ apicoalvéolaire lingual voisé qui comporte une vibration accentuée de la pointe de la langue. La durée du /r/ est de l'ordre de 80-100 msec, il possède des structures de formants qui sont interrompus par des intervalles verticaux très courts de silence de l'ordre de 10 msec qui peuvent être interprétés physiologiquement comme le résultat de la frappe du bout de la langue contre le palais. F1 du /r/ est de l'ordre de 260 Hz. Avec /i/, /ii/, /u/ et /uu/ F2 est d'environ 1400 Hz, avec /a/ et /aa/, il est autour de 1200 Hz. Le /r/ exerce une influence sur /i/ et /ii/ en abaissant les débuts de F2 de 1250 Hz à 1700 Hz, les débuts de F3 sont aussi

abaissés. Les débuts de F2 du /u/ et /uu/ sont soulevés à 1050 Hz. Enfin F2 du /a/ est environ à 1300 Hz et celui du /aa/ autour de 1200 Hz.

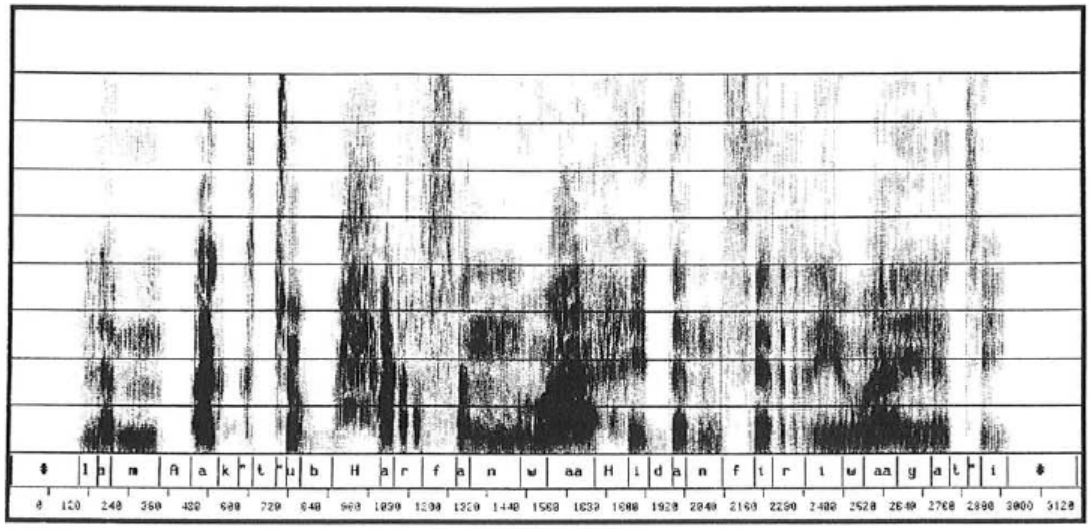


FIGURE 2.6 – Spectrogramme d'une phrase contenant des sonnantes

6.6 Le latéral

/l/ : /lam/ est un phonème lingual qui possède deux allophones : le plus commun de ces allophones est une latérale (/munħarif/) dentale voisée de durée 80-120 msec, qui possède des structures de formants similaires à celles des voyelles F1 à 300 Hz, F2 à 1500-1600 Hz et F3 à 2400-2500 Hz. Le /l/ abaisse les débuts de F2 du /i/ et /ii/ à 1850 Hz, soulève ceux de /u/ et /uu/ vers 1300 Hz ; aucune influence remarquable n'est observée sur /a/ et /aa/. L'autre allophone est une consonne emphatique latérale postdentale /l/ qui se produit dans un environnement extrêmement limité et seulement devant /a/ et /aa/. Dans le langage littéraire, on le trouve dans le mot /Allah/ (Dieu) et ses dérivés mais dans les dialectes il est plus commun. Le contraste phonémique entre /l/ et /l/ est donné par la paire minimale :
/waḷḷah/ : "jurer par Dieu" et /wallah/ : "il l'a nommé walï

6.7 Les semivoyelles

/w/ : /waw/ est une semivoyelle bilabiale de durée 80-100 msec. Il possède des structures de formants similaires à celles du /u/ et /uu/ avec F1 à 350 Hz,

F2 à 950 Hz et F3 à 2100 Hz. Les débuts et les fins de F2 de /w/ glissent vers les phones précédent et suivant.

/j/ : /ya/ est une semivoyelle palatale de durée 80-100 msec. Le /j/ possède des formants similaires à ceux du /i/ et /ii/ avec F1 à 275 Hz, F2 à 1900 Hz et F3 à 2650 Hz. Généralement, les débuts et les fins de /j/ glissent vers les phones précédent et suivant.

7 Les consonnes glottales et pharyngales

Les consonnes glottales et pharyngales se distinguent des autres consonnes par le fait qu'elles ont des lieux d'articulations verticaux. Un lieu d'articulation vertical est défini comme un ensemble de localisations anatomiques qui vont du palais jusqu'à la glotte inclusivement par opposition au lieu horizontal ou‘ les emplacements sont entre les lèvres et la luette.
Ces consonnes sont plus difficiles à étudier parce que leurs points et leurs manières d'articulation sont dans la région laryngale et pharyngale qui ne sont pas facilement accessibles. [Ghazali 87b], [Ani 70]
L'Arabe comporte deux consonnes glottales /h/ et /?/ et deux consonnes pharyngales /$\hbar$/ et /ε/ que nous allons décrire.

7.1 Les consonnes glottales

/h/ : (ha) est décrit comme une fricative glottale non voisée de durée relative 100-160 msec qui apparaît le plus souvent comme un bruit à structure de formant et qui devient voisée en milieu intervocalique. Devant /i/ et /ii/ le bruit est concentré dans la région 4000-4700 Hz devant /u/ et /uu/ la concentration du bruit est basse en fréquence vers 4200 Hz. Quand le /h/ est devant /a/ et /aa/ la concentration du bruit est autour de 4500 Hz. Il semble que le /h/ soit instable et par conséquent les voyelles qui l'avoisinent jouent un rôle très important dans la détermination de la zone de concentration du bruit. Les débuts de F2 du /i/ et /ii/ sont abaissés à 2000 Hz, ceux du /u/ et /uu/ sont soulevés à 900 Hz, enfin le F2 du /a/ avec /h/ est à 1500 Hz et celui du /aa/ à 1400 Hz.

/?/ : (hamza) est décrit comme une plosive glottale non voisée dont la structure acoustique est très dépendante du contexte de production et de sa position à l'intérieur du mot. Initialement, le /?/ apparaît sur le spectrogramme

sous forme variée. Dans quelques cas, il est sous forme d'un burst suivi d'un petit intervalle de silence de durée 15-20 msec ou bien suivi d'un faible bruit. En milieu non intervocalique, le / ?/ apparaît comme un intervalle de silence de durée 65-85 msec suivi d'un burst de durée environ 15 msec. En position finale, le / ?/ est en variation libre et apparaît comme un burst qui peut être suivi ou non d'un bruit faible, ce burst est précédé par un intervalle de silence de durée 80-120 msec. Avec /a/ et /aa/, le / ?/ apparaît comme un burst de durée 20-30 msec.

La concentration du burst avec /a/ et /aa/ est dans la région 1500-1700 Hz. Les mesures des formants du /a/ et /aa/ sont F1 : 575-650, F2 : 1180-1300 et F3 : 2300-2400.

Avec /u/ et /uu/, il n'y a pas de burst mais un très faible glissement de début de la voyelle, spécialement le long de F1-F2. Les mesures des formants de ces voyelles sont F1 : 380-400 Hz, F2 850-950 Hz et F3 : 2100-2300 Hz. Avec /i/ et /ii/, la concentration du burst lorsqu'il existe, est dans la bande 5200-5600 Hz et les mesures des formants du /i/ et /ii/ devant / ?/ sont F1 : 280-300 Hz, F2 : 1900-2100 Hz et F3 : 2700-2900 Hz.

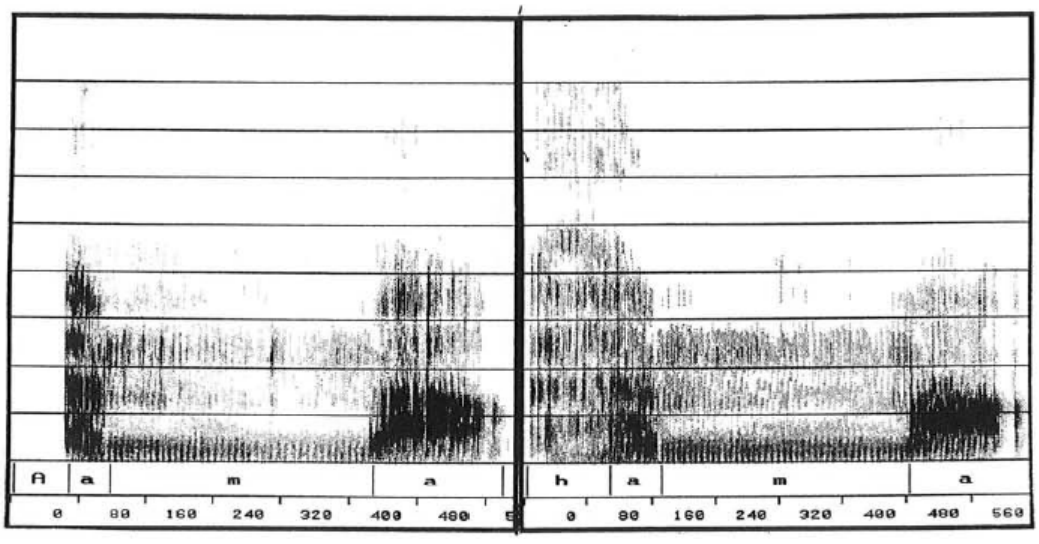


FIGURE 2.7 – Le / ?/ et le /h/

7.2 Les consonnes pharyngales

/ħ/ : (**ħa**) est une fricative pharyngale non voisée de durée 100-150 msec qui devient voisée en milieu intervocalique. Lors de la production du /ħ/ une constriction est formée par le dorsum de la langue contre la paroi postérieure du pharynx et c'est cette constriction qui différencie principalement le /ħ/ du /h/. Acoustiquement, le /ħ/ apparaît comme un bruit plus fort que celui du /h/. Avec /i/ et /ii/, la concentration du bruit est vers 3700-4500 Hz, et la limite du bruit est vers 1800 Hz. Avec /u/ et /uu/, la concentration est vers 4000 Hz et avec /a/ et /aa/ elle est dans la bande 3400-4200 Hz. Avec /ħ/, les débuts de F2 du /i/ et /ii/ sont abaissés à 1800 Hz, ceux du /u/ et /uu/ sont légèrement relevés à 900 Hz. Le F2 du /a/ est à 1400 Hz et celui du /aa/ à 1300 Hz.

/ε/ : (**εayn**) est décrit comme une fricative pharyngale voisée dont la structure est très dépendante du contexte de production : en position initiale, le /ε/ apparaît sur le spectrogramme comme une sorte de "burst" - durée 40-50 msec - dont l'intensité est quelque part entre 1450 et 1550 Hz. Le /ε/, en cette position affecte les débuts de F1, F2 du /i/ et /ii/ ou' F1 est relevé à 400 Hz et même encore plus et F2 est abaissé à moins de 1500 Hz. Cette transition est graduelle pour une durée qui varie de 50 à 100 msec. Le /ε/ exerce aussi une influence sur les débuts de F2 du /u/ et /uu/ en les relevant à 950 Hz et parfois un peu plus. Devant /ε/, le formant F2 du /a/ est vers 1300-1350 Hz et celui du /aa/ est dans la région 1250-1300 Hz. En milieu intervocalique, il a la forme d'une sorte de continuation des formants des voyelles précédente et suivante. En position finale, le /ε/ est généralement aspiré, il apparaît sur le spectrogramme comme un son de très faible énergie dans un intervalle de durée 130-160 msec précédé par un glissement de F1 et F2 du son précédent (le plus souvent une voyelle), cet intervalle est terminé par un souffle ou un bruit faible.

8 Les consonnes emphatiques

Le système phonétique de l'Arabe tire son originalité de la présence des consonnes emphatiques : la langue arabe est souvent appelée la langue du /ḍad/. Un phonème qui n'existe qu'en Arabe et qui est d'ailleurs difficile à prononcer. Les quatre articulations définies traditionnellement comme emphatiques sont /ṭ/, /ṣ/, /ḍ/ et /ð̣/. Leur nombre varie d'un auteur à un autre. Le sentiment des non linguistes est gé-

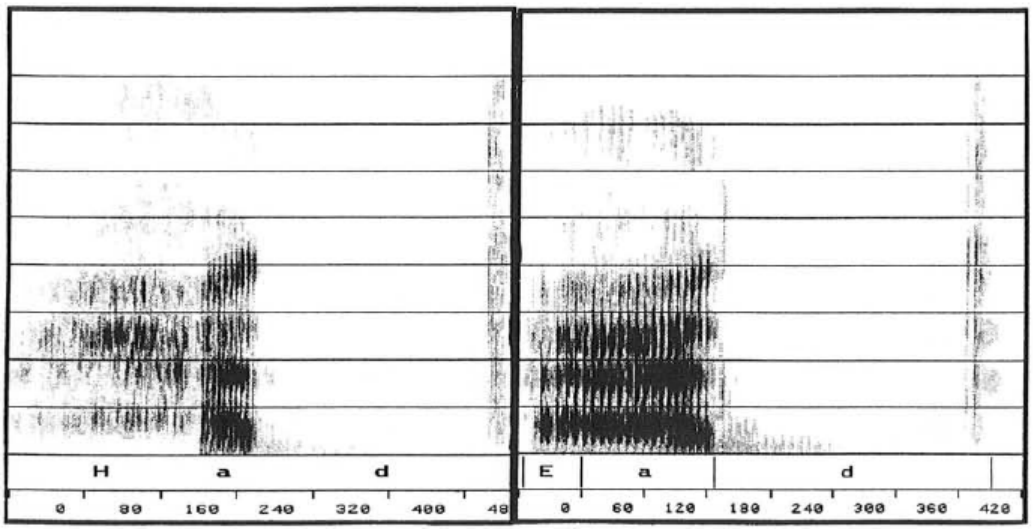


FIGURE 2.8 – Le /ħ/ et le /ε/

néralement que les sons emphatiques sont prononcés avec fermeté et possède, donc une autre tonalité. Essayons de voir ce que pensent les linguistes.

8.1 Définition traditionnelle de l'emphase

Sibawayh, l'un des chefs de file de l'école de Basra donne pour les consonnes emphatiques la description suivante : "les consonnes sont couverts /muṭbaq/ ou découvert /munfatiħ/. Sont /muṭbaq/ les consonnes /ṣ/, /ḍ/, /ṭ/ et /ẓ/ et /munfatiħ/ toutes les autres, car pour aucune d'elles l'on ne dispose la langue comme un couvercle relevé vers le palais" et déclare que "la langue, placée pour chacune de ces quatre consonnes emphatiques à son lieu se dispose à partir de ce lieu, sur toute sa surface, comme un couvercle relevé vers la partie correspondante du palais. La langue ainsi disposée, le son est comprimé entre la langue et le palais jusqu'au lieu de chacune de ces consonnes". Il ajoute que "dans le cas des consonnes comme le /d/ et le /z/ etc, le son n'est pas comprimé pour chacun qu'à son lieu, là ou' la langue est placée" et confirme que "ces quatre consonnes emphatiques ont deux lieux intéressant la langue, fait qui a été mis en évidence par la compression particulière du son". Enfin, pour Sibawayh "s'ils n'étaient pas emphatiques, le /ṭ/ serait un /d/, le /ṣ/ un /s/ le /ḍ/ un /ð/ et le /ḍ/ sortirait, lui de la langue arabe, car il est seul dans son lieu [Sibawayh 89].

De Sacy : il définit l'emphase ou l'articulation emphatique comme étant une

sorte de renflement qu'il n'est pas aisé de définir mais qui fait en quelque sorte entendre un o sourd après la consonne, ainsi le mot /ṣaad/ se prononce presque comme /soad/ sans cependant que cet o se fasse entendre distinctement [Sacy 10] .

Marčais : quant à lui déclare que l'emphase impliquerait essentiellement un exhaussement du dos de la langue vers le palais, l'air étant pressé dans l'espace compris entre la langue et son couvercle : la cavité palatale (d'ou' le terme couverture) et il ajoute que l'emphase vue à l'écran radioscopique comporte une extension da la langue de l'avant vers l'arrière avec affaissement du milieu du dos et donc élargissement de la cavité palatovélaire [Marçais 48].

Obrecht : il a étudié l'opposition vélarisées / non vélarisée des consonnes à l'aide de la synthèse de la parole et de tests de perception. L'étude porte sur un dialecte arabe libanais, dont le système comprend les consonnes /b/, /d/, /t/, /z/, /s/, /m/, /n/, /l/ non vélarisées et les consonnes vélarisées correspondantes. Pour l'auteur, les consonnes emphatiques sont /ḅ/, /ḍ/, /ṭ/,/ẓ/, /ṣ/, /ṃ/, /ṇ/, /ḷ/ et /ħ/, /γ/, /q/, /χ/ et /∂/. Il dégage une zone de vélarisation du formant F2 allant de 1000 à 1400 Hz qui lui permet d'indiquer un rapprochement de F1 et F2 des voyelles suivant les consonnes emphatiques. [Obrecht 68]

Bonnot : dans une recherche expérimentale sur la nature des consonnes emphatiques de l'Arabe, l'auteur confirme que les consonnes emphatiques possèdent deux lieux d'articulation, l'un antérieur et l'autre postérieur mais que la corrélation d'emphase est limitée à six consonnes (/ṭ/-/t/, /ḍ/-/d/ et /ṣ/-/s/) et que l'opposition /k/-/q/ doit être exclue [Bonnot 77] .

Ghazali : considère les consonnes emphatiques comme une des trois classes de consonnes arrières qui sont articulées dans la cavité postérieure du conduit vocal. Pour lui ce sont des consonnes coronales pharyngalisés, qui en plus de leur articulation antérieure primaire, manifestent une articulation secondaire résultant de la rétraction du dos de la langue vers la paroi postérieure de l'oro-pharynx au niveau de la deuxième vertèbre cervicale [Ghazali 87b]

Al Ani : en abordant l'emphase, il estime que la région entraînée est pharyngale et non pas vélaire comme le prétendent d'autres phonéticiens. L'emphase selon lui se limite aux consonnes /ṭ/, /ḍ/, /ṣ/ et /∂̣/ ainsi que le /ḷ/ dans un contexte particulier [Ani 70]. Dans une étude plus détaillée, il confirme que la principale caractéristique des consonnes emphatiques est l'élévation du formant F1 et la baisse du formant F2. La différence $F1-F2$ est par conséquent, une donneé très importante pour reconnaître les consonnes emphatiques qui

sont qualifiées par le terme arabe /maffa$\chi\chi$ma/ [Ani 83].

Troubetzkoy : selon lui les phonèmes porteurs de la marque de vélarisation emphatique sont au nombre de neuf. En plus du /ṭ/, /ḍ/, /ṣ/ et /ð̣/, il fait de /q/ l'emphatique de /k/, de /γ/ l'empahtique de /z/, du /$\hbar$/ l'emphatique /h/ et considère /χ/ et /ε/ comme des emphatiques [Troubetzky 70].

S. A. Adem : a effectué une étude des consonnes emphatiques /ṭ/ et /ṣ/ et leurs homologues non emphatiques /t/ et /s/ de l'égyptien parlé sur cinq paires minimales en contexte /a/, /aa/ et /ii/. Il trouve que la présence de l'emphase se manifeste par la baisse de formants F2, F3 et F4 [Adem 83].

R. Jakobson : présente une analyse du système consonantique du dialecte druze à partir d'une série de traits (pharyngalisé/non pharyngalisé, compact/diffus, nasal/oral, sourd/sonore, ...). C'est au moyen du trait pharyngalisé/non pharyngalisé (ou flat/plain) que l'auteur pratique une dichotomie de tout le système en assimilant à l'opposition emphatique/ non-emphatique, la distinction que font les grammairiens arabes entre /mufa$\chi\chi$ama/ et /muraqaqa/ ou‘ le phonème /q/ est considéré comme le terme marqué par rapport à /k/ dans l'opposition pharyngalisé/non pharyngalisé. Pour l'auteur, sont pharyngalisés ou "flat" les consonnes /ε/, /$\hbar$/, /ṭ/, /ð̣/, /ṣ/, /q/, /ḅ/, /ṃ/ et /ḷ/. Le reste des consonnes sont non pharyngalisées ou "plain" mais le /y/ et /w/ ne sont ni "flat" ni "plain". [Jakobson 72]

Cohen : en dehors du voisement, les qualités les plus importantes des consonnes sont / ?iṭbaaq/, le /tafχiim/ et le / ?istiεla ?/. Cohen propose un tableau présentant le rapport d'inclusion de ces trois faits phonémiques [Cohen 69]. Pour lui, les consonnes /mufa$\chi\chi$ama sont /ṭ/, /ḍ/, /ð̣/, ṣ/, /$\hbar$/, /γ/

En dépit des différentes descriptions des consonnes emphatiques et les définitions de l'emphase, le processus articulatoire qui se produit lors de la réalisation de ces articulations reste peu clair comme le résultat de l'utilisation du mot ambigu : l'emphase, la couverture ou l'itbaaq.

Les auteurs sont presque unanimes à l'exception de Bonnot [Bonnot 77] sur le fait que chaque fois qu'une consonne emphatique se produit à l'intérieur d'une syllabe, la syllabe toute entière est phonétiquement emphatique. Ceci met tous les phonèmes allophoniquement conditionnés par cet environnement. De même, selon ces auteurs, le phénomène d'emphase n'est pas enfermé dans la limite de la syllabe mais peut ou non avoir une influence sur la syllabe voisine et au delà.

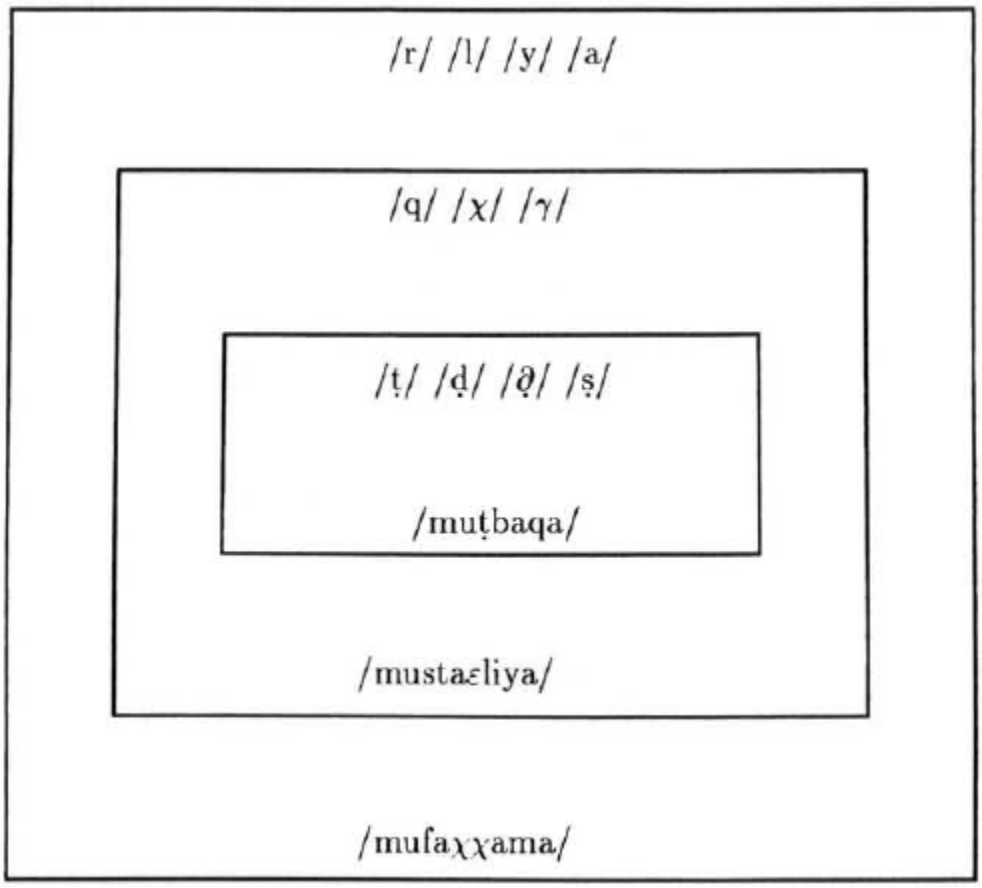


FIGURE 2.9 – Le faits phonémiques selon Cohen

8.2 Description acoustique

Les consonnes emphatiques de l'Arabe comportent une forte tension des différents organes articulatoires, elle relèvent d'un phénomène buccal qui consiste en un report en arrière de la racine de la langue et en un abaissement et un creusement du dos de la langue. Acoustiquement, elles se reconnaissent en général, par l'élévation de la transition de F1 et la baisse de la transition de F2 de la voyelle précédente et suivante. Les emphatiques de l'Arabe moderne sont au nombre de quatre : deux plosives et deux fricatives que nous décrivons.

/ṭ/ : /ṭa/ est une plosive alvéodentale emphatique non aspirante et non voisée qui apparaît sur le spectrogramme comme un burst de durée relative 20-30 msec, plus long avec les voyelles longues et qui est sous forme d'une barre verticale plus forte que celle du /t/ suivi d'un intervalle sans bruit remarquable, donc plus en durée. En position finale, la durée relative de l'intervalle avant l'explosion est de 100-180 msec, la concentration du burst est généralement plus basse en fréquence que celle du /t/. Avec /i/ et /ii/, la concentration est dans la région de F1-F2 à 1700-2400 Hz, avec /u/ et /uu/ entre 1400 et 2200 Hz et avec /a/ et /aa/ entre 1500-2400 Hz. Devant /ṭ/, F1 et F2 du /i/ et /ii/ sont tous les deux influencés par abaissement des débuts de F2 vers 1700 Hz et le soulèvement des débuts de F1 à 4500 Hz. La transition de F2 du /i/ prend à peu près le tiers de la durée de la voyelle alors qu'avec /ii/la transition prend environ le cinquième. Les débuts de F2 du /u/ et /uu/ devant /ṭ/ se situent vers‘ 900 Hz, F1 de /u/ et /uu/ est vers 459 Hz. Le F2 du /a/ est vers 1150-1250 Hz et celui du /aa/ est à 1050-1150 Hz, ce qui différencie sensiblement le /a/ du /aa/. Le /a/ et /aa/ ont un formant F1 vers 650 Hz.

/ḍ/ : /ḍad/ est une plosive alvéodentale emphatique voisée et non aspirante de durée 80-100 msec.

C'est un phonème très difficile à prononcer et on le confond presque toujours avec /ð̣/. La paire minimale qui distingue ces deux phonèmes est :
/ḍalla/ : s'égarer et /ð̣alla/ : rester.

/ṣ/ : /ṣad/ est une fricative dorsoalvéolaire sifflante emphatique non voisée. Le /ṣ/ apparaît comme un bruit aléatoire de durée relative 100-170 msec en hautes fréquences à partir de 2800 Hz. Le /ṣ/, comme /ṭ/, influe sur les voyelles, les débuts de F2 du /i/ et /ii/ sont abaissés à 1700 Hz et on voit donc une transition montante. Le /ṣ/ affecte aussi les débuts de F1 du /i/ et /ii/ qui sont relevés de 300 Hz à 450 Hz L'influence du /ṣ/ sur F1 et F2 du

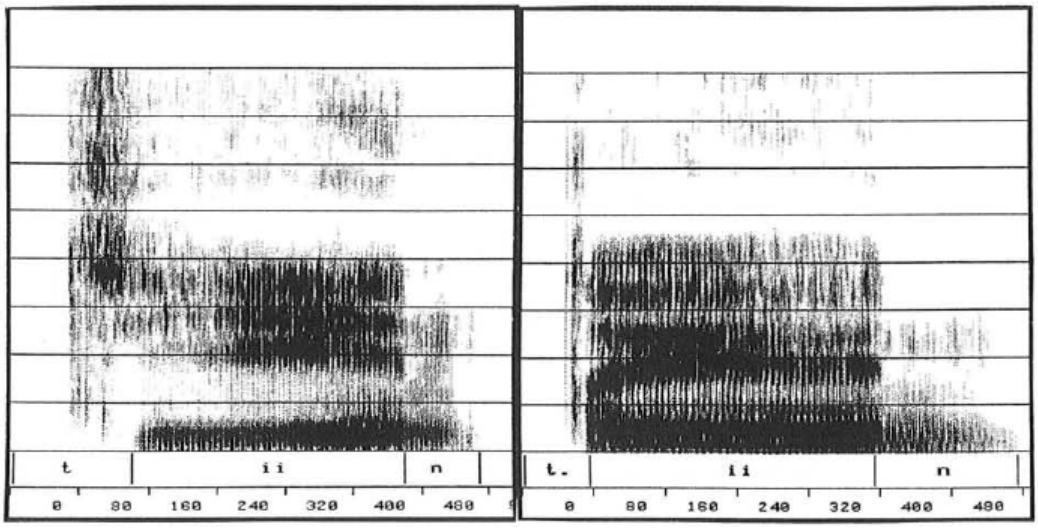


FIGURE 2.10 – Le /t/ et le /ṭ/

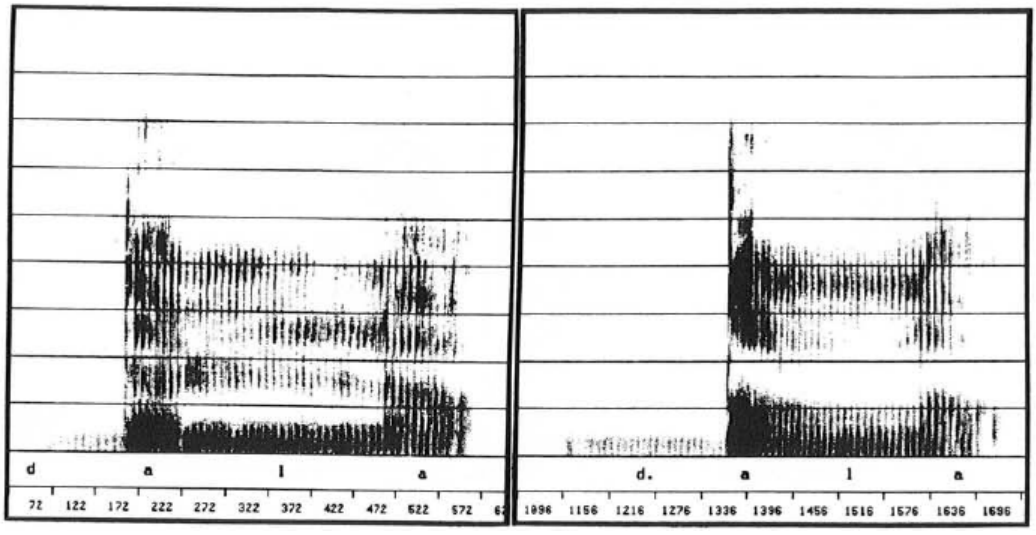


FIGURE 2.11 – Le /d/ et le /ḍ/

/i/ est plus remarquable que sur ceux du /ii/. Sur l'axe du temps, l'influence semble atteindre le tiers de la durée du /i/. Les débuts de F2 du /u/ et /uu/ avec /ṣ/ sont graduellement relevés vers 1050 Hz. Les débuts de F1 du /u/ et /uu/ se situent vers 400 Hz. Les fréquences du second formant du /a/ avec /ṣ/ sont vers 1200-1250 Hz et celle du /aa/ vers 1100 Hz. F1 du /a/ et /a/ est vers 650 Hz.

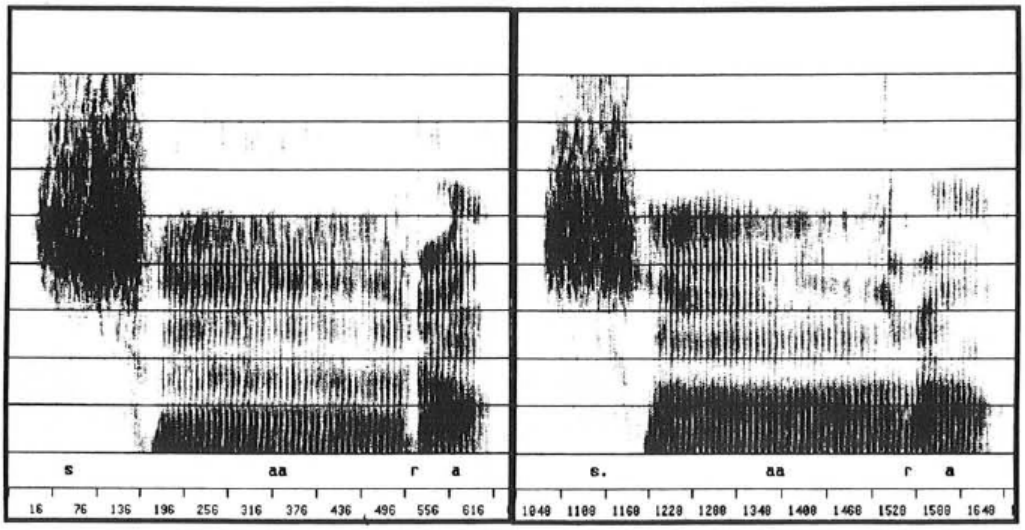


FIGURE 2.12 – Le /s/ et le /ṣ/

/*ḍ*/ **: /Dha/** est décrit comme une fricative interdentale emphatique et non voisée de durée relative 900-120 msec. Il semble avoir une résonance qui apparaît comme des formants faibles F1 à 400 Hz, F2 à 1200 Hz et F3 à 2350 Hz et un certain bruit, mais les valeurs des structures formantiques sont très influencées par le contexte vocalique adjacent. /*ḍ*/ influe sur les débuts de F2 du /i/ et /ii/ par leur abaissement à 1700-1800 Hz. Les débuts de F1 du /i/ et /ii/ montent à 400 Hz. Au voisinage du /*ḍ*/, les débuts de F2 du /u/ et /uu/ sont abaissés à 900 Hz, ceux de F1 sont soulevés à 450 Hz. Enfin, le F2 du /a/ est à 1100-1200 Hz et celui du /aa/ à 1050-1150 Hz. F1 du /a/ et /aa/ se sitent vers 600 Hz.

Nous signalons aussi que le phonème /l/ possède un allophone emphatique dans le mot /ʔaḷḷah/ et que le /r/ s'emphatise souvent devant /a/ et /aa/. De même, l'influence des dialectes sur la prononciation de l'arabe standard provoque la pharyngalisation de certaines autres consonnes, particulièrement les labiales /b/, /m/ et /f/ [Puech 87].

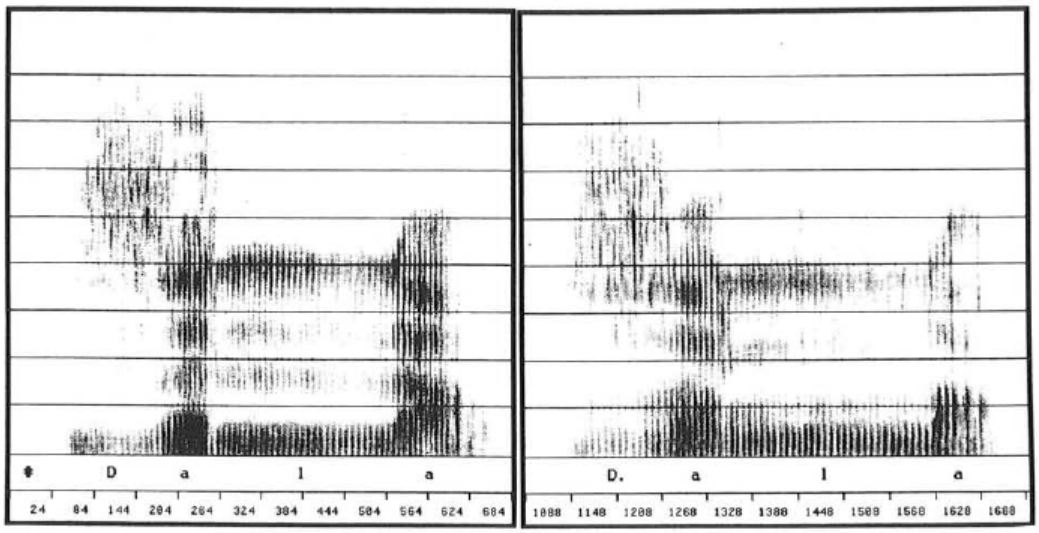


FIGURE 2.13 – Le /∂/ et le /∂̣/

9 Données complémentaires

9.1 Les séquences phonologiques

En Arabe, la longueur des groupes consonantiques est égale à 2, ainsi une séquence de 3 consonnes est toujours cassée par l'insertion d'une voyelle courte, le plus souvent /i/. Notons aussi que ces groupes ne se produisent jamais en début de mot mais uniquement au milieu et en fin de mot.
Les analyses statistiques [Ani 70] [Mrayati 87] sur la production des groupes de consonnes en Arabe ont permis de construire un tableau des séquences phonologiques irréalisables en Arabe standard.

Cette étude permet de dégager deux groupes : les consonnes antérieures et les consonnes postérieures. Les consonnes postérieures incluent les consonnes vélaires, uvulaires, pharyngales et glottales. Les consonnes antérieures incluent les consonnes restantes qui forment des groupements avec les consonnes postérieures.
Nous pouvons conclure donc que les séquences impossibles ne sont pas réalisées pour des raisons évidentes de proximité articulatoire.

Consonne 1	Consonne 2
t	d, ṭ, ḍ, θ, ∂, $\partial\!\!\!\!\!\!\cdot$
k	?, q, γ
?	χ, $\hbar$, γ, ε, h
b	f
d	t, ṭ, ḍ, θ, ∂, $\partial\!\!\!\!\!\!\cdot$
q	?, k, χ, γ
ṭ	t, d, ḍ, θ, ∂, z, $\partial\!\!\!\!\!\!\cdot$
ḍ	t, d, ṭ, θ, s, $\int$, ∂, z, ṣ, $\partial\!\!\!\!\!\!\cdot$
z	-
f	b, θ, m
θ	ṭ, ḍ, s, $\int$, ∂, z, ṣ, $\partial\!\!\!\!\!\!\cdot$
s	θ, $\int$, ∂, z, ṣ, $\partial\!\!\!\!\!\!\cdot$
$\int$	θ, s, ∂, z, ṣ, $\partial\!\!\!\!\!\!\cdot$
χ	$\hbar$, γ, ε, h
$\hbar$	χ, γ, ε, h
∂	t, d, ṭ, ḍ, θ, s, $\int$, z, ṣ, $\partial\!\!\!\!\!\!\cdot$
z	ṭ, θ, s, $\int$, ∂, ṣ, $\partial\!\!\!\!\!\!\cdot$
γ	?, χ, $\hbar$, ε, h
ε	?, χ, $\hbar$, γ
h	χ, $\hbar$, γ
ṣ	θ, s, $\int$, ∂, z, $\partial\!\!\!\!\!\!\cdot$
$\partial\!\!\!\!\!\!\cdot$	t, d, ṭ, ḍ, θ, s, $\int$, ∂, z, ṣ
m	b, f
n	l
r	n, l
l	n, r
w	-
j	-

TABLE 2.3 – Les séquences impossibles des consonnes

9.2 La nounation

Les trois signes qui représentent graphiquement les voyelles sont quelquefois redoublés à la fin des noms et les voyelles finales se lisent alors comme si elles étaient suivies du son /n/. Ce redoublement de voyelles s'appelle la nounation ou /tanwin/ ; il indique l'indéterminisme. Par le fait qu'on prononce un son qui ne se transcrit pas, la nounation a des incidences au niveau morphologique. Sur le plan de la prononciation, le son /n/ de la nounation est souvent omis, lorsqu'il est en fin de phrase précédé par /a/ sinon, il a bien la forme du /n/ mais de faible énergie.

9.3 La gémination

La gémination /Idgham/ ou /tachdid/ entraîne le renforcement de l'articulation et l'accentuation des propriétés de la consonne, elle provoque ainsi une prolongation de la fermeture de la plosive ou du continuant des autres consonnes et ne peut jamais être considérée comme le dédoublement de la consonne. Phonétiquement, il n'existe pas de consonnes géminées : une géminée est considérée comme une consonne longue. La caractéristique la plus évidente est donc la différence très importante de durée, en effet le rapport géminée / simple est de l'ordre de 2. Selon Bonnot [Bonnot 79], au contact d'une consonne géminée, les transitions et les parties stables des formants F1 et F2 de la voyelle sont plus basses alors que le troisième formant est plus haut. La gémination en Arabe a une fonction différenciative. Les consonnes géminées apparaissent dans les positions ou‘ un groupe de deux consonnes est admis. La gémination joue par ailleurs un rôle structurel dans le développement morphologique du nom et du verbe. Il y a deux sortes de gémination : la gémination nécessaire et la gémination euphonique.

1. La gémination nécessaire apparaît toujours après une voyelle. C'est de la gémination nécessaire que dépend la signification même du mot. Elle peut donc à elle seule différencier les mots exemple /ʒamaal/ : (beauté) et /ʒammal/ : (chamelier) ou bien /ħamaam/ : (pigeons) et /ħammaam/ : (bain).

2. La gémination euphonique est employée pour remplacer par une consonne facile à prononcer, les sons ou groupes de sons qui paraissent durs à l'oreille. Elle a lieu dans les noms déterminés par l'article / ʔal/ commenc̃ant par une consonne assimilante.

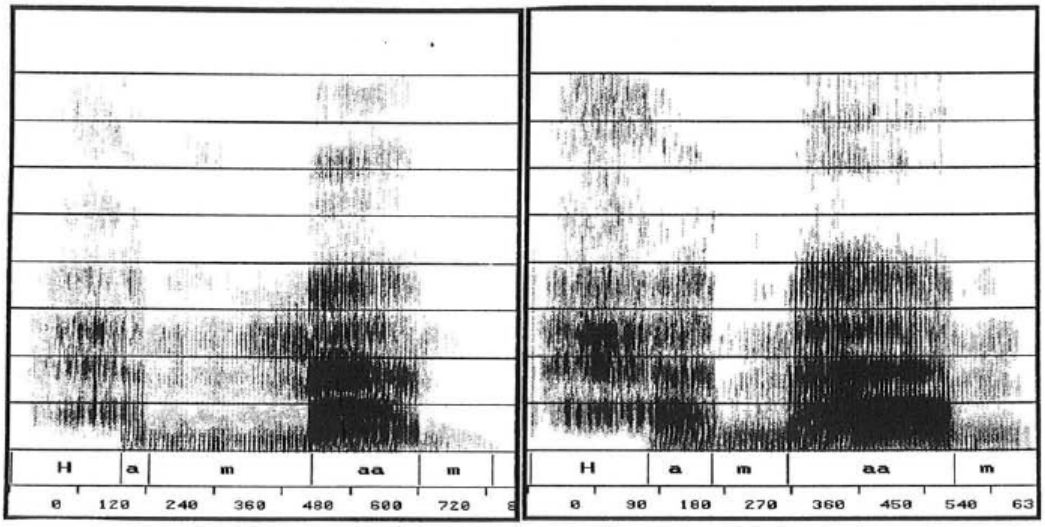


FIGURE 2.14 – Opposition géminée-simple

9.4 L'assimilation et la dissimilation

Ce sont des changements phonétiques qui interviennent à l'intérieur des mots entre les éléments stables. L'assimilation est le phénomène par lequel deux phonèmes différents tendent à devenir semblables. Elle peut être totale ou partielle. L'assimilation totale rend les phonèmes identiques. L'assimilation partielle laisse une différence, elle pousse simplement les phonèmes a' acquérir des caractères communs ; il peut y avoir ainsi assimilation du mode d'articulation, du lieu d'articulation (sonorité, sourdité, vélarisation ou nasalisation).
La dissimilation est le processus inverse de l'assimilation : par la dissimilation deux phonèmes identiques ou présentant des caractères communs tendent à se différencier. La raison essentielle de ces changements phonétiques est la recherche d'une prononciation plus facile. Un exemple d'assimilation est celui du /l/.

L'assimilation du /l/

Les consonnes assimilantes initiales d'un nom assimilent le /l/ de l'article défini / ?al/ qui le précède et reçoivent la gémination euphonique. Les consonnes assimilantes sont au nombre de 14, ce sont /t/ /θ/, /d/, /∂/, /r/, /z/, /s/, /$\int$/, /ṣ/, /ḍ/, /ṭ/, /$\partial\!\!\!^{.}$/, /l/ et /n/.
Au lieu de / ?al$\int$ams/ (soleil) on prononce / ?a$\int$ $\int$ams/. La consonne /l/ de l'article est assimilée dans la prononciation par la consonne assimilante qui la suit c'est à dire /$\int$/.

Ces 14 consonnes sont aussi appelées consonnes solaires, parce que l'une d'elles /ʃ/ est la première du mot /ʃams/ : soleil et a servi comme exemple d'assimilante.
Les 14 autres consonnes sont appelées non assimilantes parce qu'elles n'assimilent pas l'article /ʔal/ qui le précède.
Elles sont aussi appelés lunaires parce que l'une d'elles /q/ est la première du mot /qamar/ : lune et a servi comme exemple de non assimilante : /qamar/ et /alqamar/.

9.5 La prothèse

Pour éviter de commencer certains mots par deux consonnes (c'est à dire, éviter à l'initiale, un groupe combiné explosif), on leur prépose une /ʔalif/ prosthétique affectée d'une voyelle instable sans /hamza/ lorsque ces mots se trouvent en isolés ou au début d'une phrase. Ex. /ʔiftaraqnaa/ (nous nous séparâmes).
La voyelle disparaît lorsque les mots commenc̃ant par deux consonnes se trouve au milieu d'un discours. Ex /thumma ftaraqnaa/ (puis nous nous séparâmes). De même, nous prononc̃ons /uktub/ (écris) et /wa ktub/ (et écris).
Remarquons qu'il ne faut jamais faire de pause dans la lecture avant une /ʔalif/ prosthétique, la première consonne du mot qui la suit faisant partie de la dernière syllabe du mot précédent.

9.6 La syllabe

La syllabe en Arabe débute toujours par une consonne et une seule. Elle peut se terminer par une voyelle (syllabe ouverte) ou par une consonne (syllabe fermée). Ceci exclut, à l'initiale du mot, les groupes combinés explosifs, et à l'intérieur des mots, les groupes de plus de deux consonnes. Il existe donc cinq types de syllabes

1. la syllabe CV ex. /ka/ : "comme"

2. la syllabe CVV ex. /laa/ : "non"

3. la syllabe CVC ex. /hum/ : "ils"

4. la syllabe CVVC ex /riiħ/ : "vent"

5. la syllabe CVCC ex /baħr/ : "mer"

De très nombreux travaux considèrent cependant six types, en admettant la structure CVVCC qui ne pourrait apparaître qu'en position finale, en forme pausale et à la condition que les deux consonnes finales soient identiques. Les analyses spectrographiques que nous avons effectuées montrent qu'une consonne géminée en position pausale est réalisée acoustiquement comme une seule et unique consonne. En principe, si deux consonnes se suivent à l'intérieur du mot, la première sera rattachée à la syllabe précédente et la seconde à la suivante. Il y a donc autant de syllabes que de voyelles, la voyelle étant le noyau de la syllabe. Les quatre premiers types de syllabes se produisent en début, au milieu et en fin de mot. Le plus fréquent étant le type CV. Le cinquième ne se produit qu'en fin de mot ou en isolé. Une syllabe peut être courte (type CV) ou longue (le reste). Cette structure est totalement différente de celle des dialectes, ou un grand nombre d'autres combinaisons est possible [Korichane 87], [Benkirane 87]

9.7 L'accent d'intensité

Les syllabes ne sont pas produites avec la même intensité. Il y a dans les mots arabes polysyllabiques, une syllabe qu'on met en relief. On dit de cette syllabe qu'elle porte l'accent d'intensité. L'existence de l'accent en arabe a constitué un sujet d'intérêt et de contreverse avant que la réponse par l'affirmative ne soit admise. Les anciens grammairiens l'ont complètement ignoré, de même que Fleisch qui admet que l'accent ne joue aucun rôle distinctif dans la langue [Fleisch 61]. Parmi les études récentes qui attestent l'existence de l'accent, nous citerons [Ani 70], [Belkaid 84] et [Rajouani 87]. D'une façon générale, la place de l'accent dépend de la nature des syllabes (ouvertes ou fermées) et de la quantité des voyelles. Elle est déterminée par les règles suivantes :

1. L'accent d'intensité n'est jamais sur la dernière syllabe.

2. Dans un mot polysyllabique, l'accent se place sur la première syllabe longue en partant de la fin du mot.

3. Le mot qui n'a que des syllabes brêves, a l'accent sur la première.

4. Les conjonctions copulatives /wa/ et /fa/ ne changent pas la place de l'accent.

10 Conclusion

Nous avons présenté dans ce chapitre une étude phonétique de l'Arabe standard basée essentiellement sur des observations spectrographiques de phrases. Cette étude nous a permis d'acquérir une expérience dans le domaine de la lecture de spectrogrammes et ainsi de construire une base de connaissances et développer des algorithmes de reconnaissance phonétique.

Chapitre 3

Le système SAPHA et les outils d'analyse

1 Introduction

Le système SAPHA permet de faire la reconnaissance analytique des phonèmes de l'Arabe standard en parole continue et dans un contexte multilocuteur. Ce système peut être considéré comme un étage d'un système de reconnaissance de l'Arabe qui intègre des informations linguistiques ou bien comme un module d'une machine à dicter vocale.

2 Architecture du système

Le système SAPHA, qui utilise les fonctions d'édition et de traitement de la parole du logiciel Snorri [Laprie 88] est structuré en modules (voir figure 3.1). Il reçoit en entrée le signal de parole et renvoie comme résultat un treillis phonétique. Autour des modules de reconnaissance, nous avons développé des procédures pour l'analyse phonétique et l'affichage graphique ainsi que des modules d'évaluation des performances du système. L'évaluation nécessite un corpus de phrases équilibrées prononcées par plusieurs locuteurs et étiquetées manuellement.

3 Acquisition et représentation paramétrique

3.1 Le module d'acquisition

Ce module est composé d'un ensemble de fonctions :

— Acquérir de la parole, le système demande le temps d'acquisition qui est limité

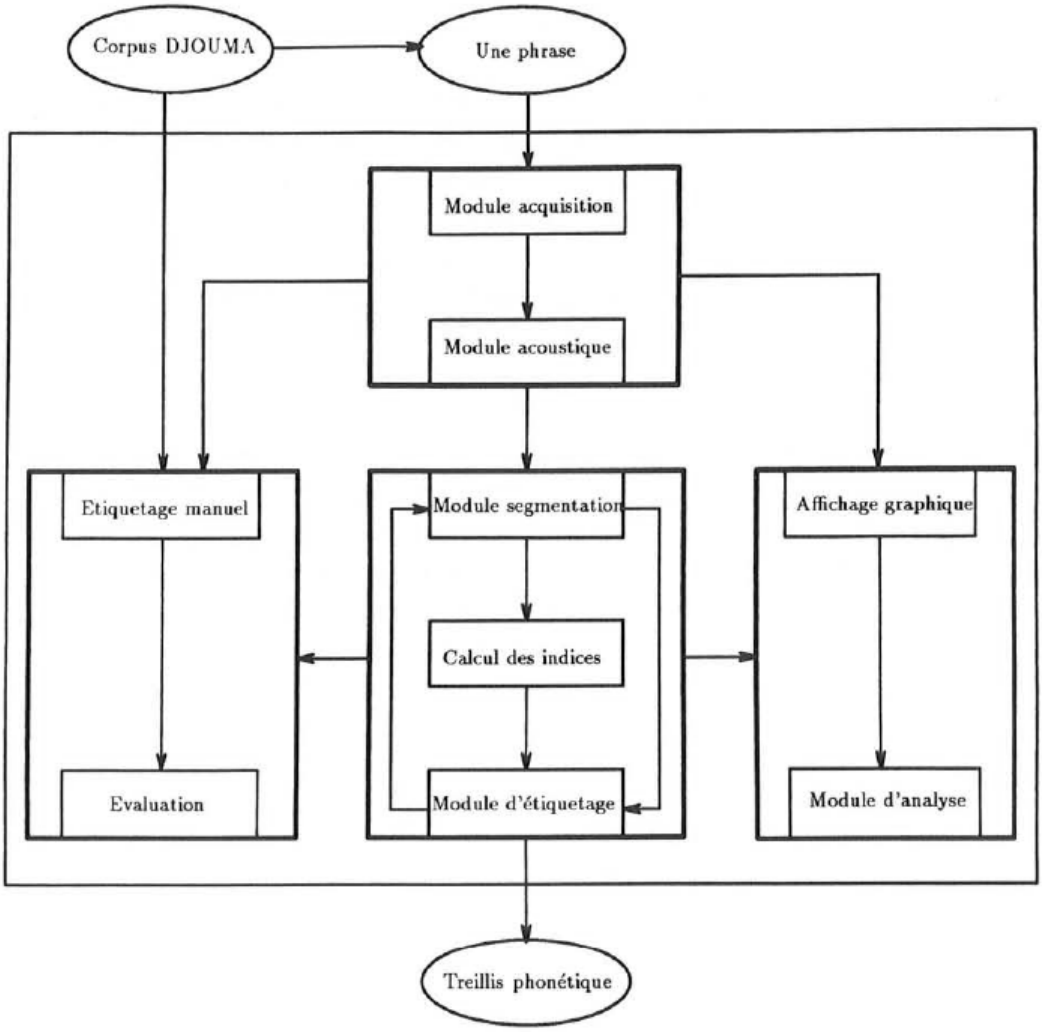


FIGURE 3.1 – Architecture de SAPHA

à 4 secondes. Le signal est stocké sur fichier disque contenant les échantillons de taille 16 bits.

— Ecouter une partie ou l'ensemble du signal, le système demande le début et la fin de la zone à restituer dans le signal temporel ; l'utilisateur choisit le nombre de restitutions.

— Choisir la fréquence d'échantillonnage lors de l'acquisition et la restitution. Cette fréquence est utilisée par plusieurs modules et elle est fixée par défaut à 16kHz, mais il est possible de l'augmenter jusqu'à 33 kHz.

— Lire un fichier de parole dans l'un des corpus existants (DJOUMA, Paires minimales, test ou privé).

— Sauvegarder une partie ou l'ensemble du signal temporel. Il suffit de donner le nom du fichier et ensuite le début et la fin du signal pour délimiter la partie de la phrase à sauvegarder.

— Afficher le signal temporel sur une fenêtre de la console graphique.

— Faire un zoom sur le signal temporel.

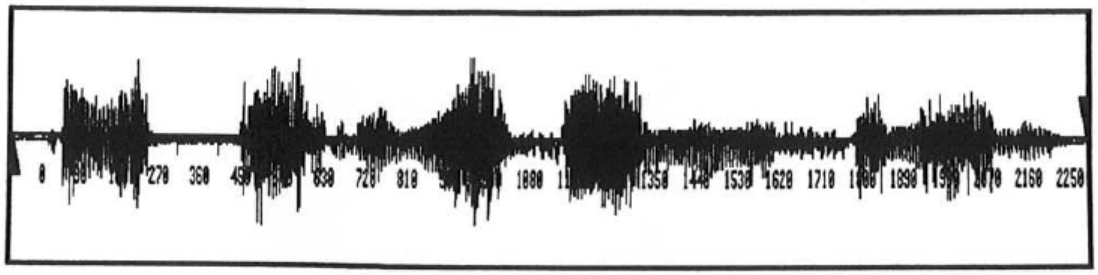


FIGURE 3.2 – Signal temporel

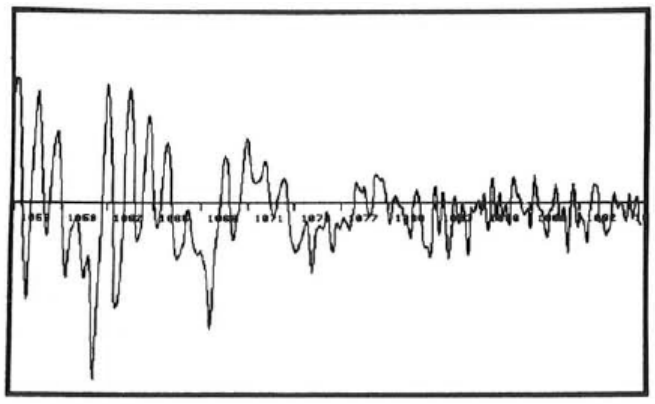


FIGURE 3.3 – Zoom du partie du signal

3.2 Le module acoustique

Ce module se charge d'extraire les paramètres acoustiques à partir du signal temporel. Il permet en particulier de :

— Calculer et afficher un spectrogramme en 13 niveaux de couleurs. Le spectrogramme normal est un spectrogramme à bande large avec une fenêtre de Hamming de 4 ms et un déplacement de 2 ms mais il est possible de calculer les spectrogrammes suivants :
  — Spectrogramme calculé sur les cœfficients LPC qui renforce les fréquences formantiques.
  — Spectrogramme à bande étroite pour séparer les harmoniques de la fréquence fondamentale.
  — Spectrogramme lissé cepstralement.

  Les algorithmes utilisent le processeur vectoriel, ce qui permet d'obtenir un temps de calcul d'environ 1 seconde pour un spectrogramme de 4 secondes de parole.

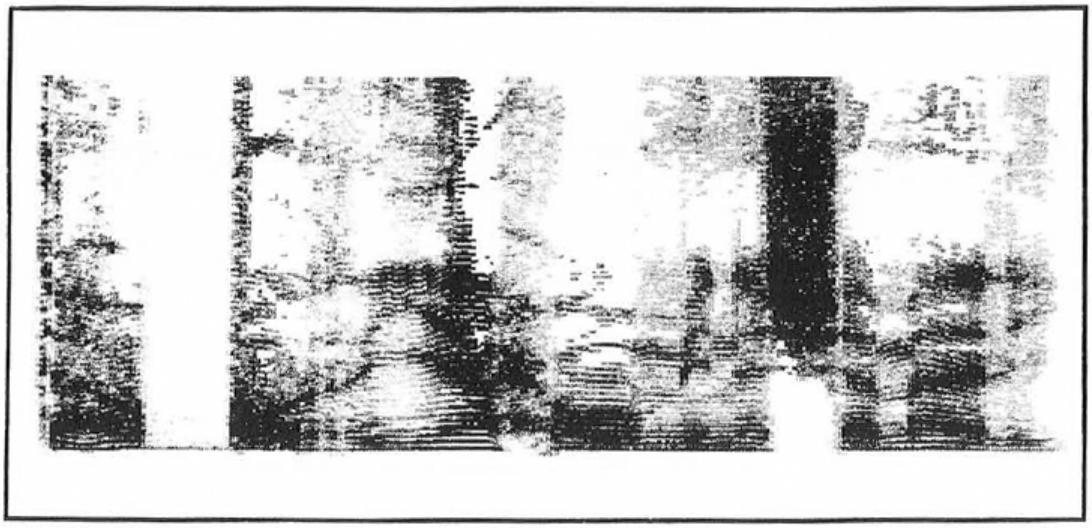


FIGURE 3.4 – Spectrogramme bande étroite

— Réafficher un spectrogramme déjà calculé en modifiant sa présentation (lissage et nombre de canaux). L'utilisateur possède plusieurs palettes de couleurs pour l'affichage (gris, rouge, bleu, vert, orange). C'est la couleur orange qui est prise par défaut.
— Calculer l'amplitude du signal comme étant :
  $E_a = \frac{1}{N} \sum_{n=0}^{N-1} |x(n)|$
  Dans la table 3.1, nous donnons les valeurs moyennes des amplitudes segments

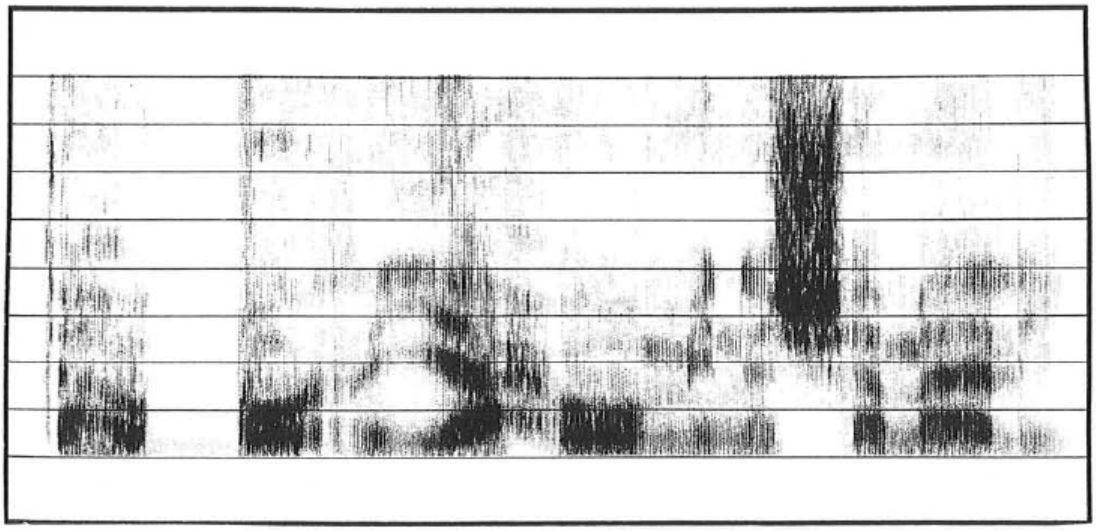


FIGURE 3.5 – Spectrogramme bande large

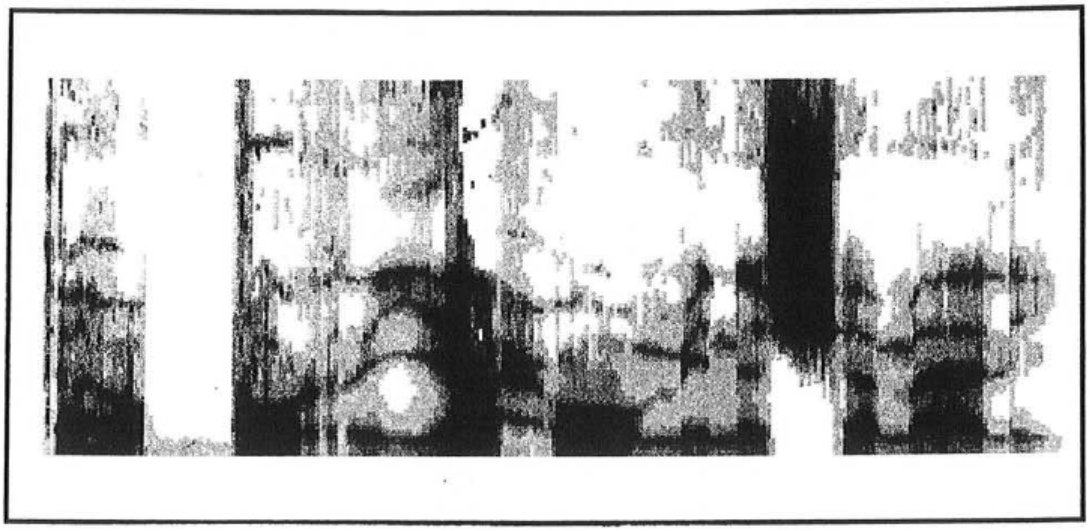


FIGURE 3.6 – Spectrogramme sur le coefficients LPC

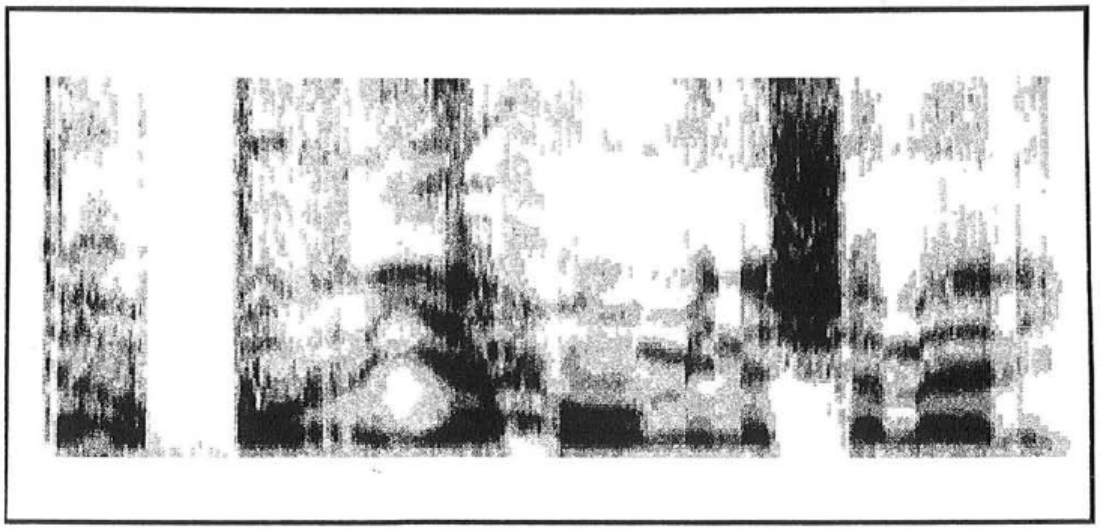


FIGURE 3.7 – Spectrogramme lissé cepstralement

phonétiques de la langue.

— Calculer le nombre de passages par zéro du signal temporel par seconde. Ce paramètre est utilisé pour distinguer entre parole et non parole et permet de différencier les sons voisés des sons non voisés. Nous donnons dans la table 3.2 les valeurs moyennes de la densité de passages par zéro par seconde des segments phonétiques :

— Calculer la fréquence fondamentale ou pitch, qui correspond aux vibrations des cordes vocales. La méthode utilisée est celle de l'autocorrélation [Rabiner 76].

4 Les modules de décodage

4.1 Le module de segmentation

Il consiste à segmenter le signal de parole en grandes classes phonétiques en utilisant des algorithmes non contextuels et reposant sur des critères simples. Nous avons retenu trois grandes classes : les voyelles, les plosives et les fricatives.
Les fonctions prévues à ce niveau sont : le calcul, l'affichage et la sauvegarde de la segmentation.
Nous détaillons dans le chapitre 4 les procédures de segmentation.

Phonème	Amplitude
a	396
i	216
u	320
aa	433
ii	262
uu	315
t	46
k	33
ʔ	41
b	51
d	50
q	45
ṭ	65
ḍ	83
z	86
f	40
θ	46
s	43
$\int$	86
χ	88
$\hbar$	62
∂	111
z	143
γ	106
ε	346
h	302
ṣ	53
$\partial\!\!\!^{\cdot}$	119
m	203
n	199
r	133
l	167
w	264
j	301

TABLE 3.1 – Amplitude moyenne du signal

Phonème	Densité
a	887
i	1041
u	918
aa	940
ii	725
uu	782
t	219
k	287
?	362
b	225
d	343
q	352
ṭ	170
ḍ	859
z	3431
f	3160
θ	593
s	6295
$\int$	5458
χ	2312
$\hbar$	2044
∂	950
z	1846
γ	1015
ε	1433
h	468
ṣ	6125
$\partial\!\!\!\!\cdot$	967
m	650
n	538
r	1254
l	718
w	784
j	971

TABLE 3.2 – Densité de passages par zéro

4.2 Le calcul d'indices

L'extraction des indices phonétiques pertinents est une étape très importante dans le processus de décodage phonétique. Nous avons développé une procédure pour chaque indice phonétique. Ces indices sont :

— la durée d'un segment,
— le degré de voisement,
— la barre d'explosion et ses paramètres,
— les valeurs des formants,
— les transitions formantiques,
— le centre de gravité énergétique,
— la limite inférieure du bruit de friction.

Nous reviendrons avec plus de détail dans le chapitre suivant sur les procédures d'extraction de ces indices.

4.3 Le module d'étiquetage

C'est à ce niveau que se fait le décodage proprement dit. En partant des segments fournis par le module de segmentation, le module tente de trouver les bons phonèmes prononcés en utilisant les indices extraits lors de l'étape précédente. Deux méthodes d'étiquetage ont été utilisées, l'une procédurale et l'autre utilisant un système expert à règles de production. Nous exposons ces deux méthodes dans le chapitre suivant.

5 Les outils d'analyse

5.1 L'étiquetage manuel

L'étiquetage manuel permet d'affecter des étiquettes phonétiques à des segments de parole à partir de la représentation spectrographique de la phrase. L'étiquetage n'est possible que sur console graphique et il permet de :

— Insérer une étiquette phonétique qui apparaît toujours avant la marque que l'on pose avec la souris. Le système demande ensuite l'étiquette à mettre grâce à des menus portant sur les phonèmes de l'Arabe standard répartis en grandes classes.
— Effacer une étiquette et la marque correspondante. Il suffit de cliquer le segment à détruire.
— Changer l'étiquette d'un segment. Pour cela, il suffit de cliquer sur l'étiquette

à modifier et de choisir ensuite la bonne étiquette.
— Déplacer la limite d'un segment. L'utilisateur clique une première fois sur cette limite et une seconde fois pour indiquer ou‘ la déplacer.
— Calculer et afficher un spectrogramme haute définition.
— Ecouter un morceau de signal autant de fois que l'on veut dans le but de faciliter l'étiquetage.

Le résultat de l'étiquetage manuel est sauvegardé dans un fichier qu'il est ensuite possible de lire pour le consulter ou le modifier. Il servira en particulier à évaluer les performances du système tant au niveau segmentation qu'au niveau reconnaissance.

5.2 L'affichage graphique

Afin de pouvoir faire une analyse détaillé sur les phonèmes, nous avons prévu des fonctions d'affichage des courbes qui permettent d'afficher :

— Les deux formants d'un segment de parole le plan F1-F2. L'utilisateur choisit les limites avec la souris. Ce sont les deux premières racines du calcul de la LPC.
— Les pics des cepstres les plus visibles jusqu'à 5000 Hz. Les cepstres sont calculés sur une fenêtre de 256 échantillons avec un recouvrement de 128 échantillons.
— Les racines de LPC dont la largeur de bande est inférieur à 700 Hz.
— Les pics de cepstres qui ne correspondent pas à une racine de la LPC.
— La FFT, la FFT lissée, la LPC et le cepstre correspondant au signal que l'utilisateur clique sur le spectrogramme. Une marque rouge apparaît sur le spectrogramme et permet de savoir à quel instant correspond les courbes calculées.
— Le suivi des pics du cepstre sur le spectrogramme.
— La courbe d'énergie dans une bande de fréquences.

5.3 L'analyse phonétique

Il s'agit d'analyser phonétiquement une phrase à partir de sa représentation spectrographique. Cette analyse consiste à calculer les valeurs des paramètres phonétiques des segments. Le processus peut être activé de deux manières :

— L'une manuelle, qui consiste à préciser la nature (voyelle, plosive, fricative ou autre) et les limites (début et fin) du segment à analyser en cliquant sur le spectrogramme. Le résultat est ensuite affiché sur la console.
— L'autre automatique qui consiste à calculer les valeurs des indices de tous les

segments. La nature et les limites de chaque segment sont obtenues soit par l'étiquetage manuel soit par le module de segmentation.

Les indices à extraire sont ceux utilisés lors du décodage phonétique et qui sont calculés par le module d'extraction des indices (voir chapitre 4).

6 L'évaluation des performances

Le module d'évaluation a pour tâche de calculer les performances du système par rapport à l'étiquetage manuel. Deux évaluations sont prévues, l'une concerne la segmentation et l'autre l'étiquetage.

6.1 Evaluation de la segmentation

Elle consiste à rapprocher les résultats de la segmentation automatique avec la segmentation manuelle effectuée sur les phrases du corpus en utilisant un algorithme de programmation dynamique. Pour chaque phonème, le système rend le nombre d'occurrences dans chaque classe phonétique ainsi que les taux de bonne segmentation, d'insertion et d'omission.

6.2 Evaluation du décodage

Elle consiste à calculer le taux de reconnaissance phonétique du système par comparaison du résultat du décodage avec la transcription correcte des phrases obtenue lors de l'étiquetage manuelle. Le résultat est une matrice de confusion. Pour chaque phonème est donné le nombre de fois ou‘ il a été confondu avec un autre. La mise en correspondance entre le résultat du système et la transcription de la phrase utilise une matrice initiale de confusion qui fixe au préalable les confusions possibles entre les phonèmes de la langue. Le remplissage de la matrice de confusion tient compte du mode d'articulation et de la proximité entre les lieux d'articulation des phonèmes.

7 Le corpus DJOUMA

Pour le besoin de l'étude phonétique et/ou la reconnaissance de l'Arabe moderne standard, les phonéticiens et les informaticiens ont été contraints à la réalisation de corpus. Ces corpus étaient dans leur grande majorité des corpus de mots ou de type CV et CVC et ne répondaient pas donc au critère de prise en compte réelle du

contexte de production des phonèmes et des phénomènes de coarticulation. Nous avons estimé nécessaire d'avoir dans un premier temps un corpus de phrases lues que nous avons baptisé DJOUMA /DJOUmal MAqroua/ qui veut dire en Arabe phrases lues.

7.1 Constitution du corpus

En l'absence de phrases arabes phonétiquement équilibrées dans les revues des phonéticiens comme c'est le cas du frančais [Combescure 81], nous avons entamé une recherche dans les journaux, les magazines et les livres arabes pour trouver des phrases qui peuvent faire l'objet de corpus. Dès le départ nous avons fixé trois critères.

1. La simplicité de la forme syntaxique de la phrase.
2. La diversité dans la forme et le contenu des phrases.
3. La longueur raisonnable de la phrase.

En se basant sur ces critères, nous avons construit 78 phrases. Par la suite, nous avons effectué un sondage au près de trois personnes qui ont lu l'ensemble des phrases. Notre but est d'éliminer les phrases difficiles à comprendre ou à prononcer et corriger éventuellement les autres phrases. Après cette phase, nous avons retenu 50 phrases.

Après que les phrases aient été construites, nous les avons entièrement voyellées de fačon à ce que la prononciation soit exacte selon les règles grammaticales de l'Arabe standard, ensuite nous avons réparti ces phrases en 5 séries de 10 phrases chacune en essayant dans la mesure du possible, d'équilibrer phonétiquement chaque série.

7.2 Enregistrement

L'enregistrement s'est effectué dans un milieu calme (au bureau après 18 heures) sur une cassette au chrome.
Onze locuteurs (7 hommes et 4 femmes) ont prononcé chacun les 50 phrases du corpus par série de 10 phrases à un rythme naturel d'élocution. Un temps d'environ deux secondes sépare la prononciation de deux phrases consécutives.
Les phrases ont été échantillonnées à une fréquence de 16 kHz sur 12 bits et stockées sur disque magnétique, chacune dans un fichier dont le nom est un code alphanumérique de la forme ANNAA.par ou'

— A désigne la série (A, B, C, D ou E).
— NN est le numéro de la phrase (ex 08)

— AA sont les initiales du locuteur (ex DM).

En moyenne, une phrase contient 30 phonèmes et la durée moyenne d'une phrase est de l'ordre de 3.8 secondes. De même nous avons étiqueté manuellement l'ensemble de ces phrases. L'évaluation du système de reconnaissance est faite par comparaison à cet étiquetage manuel.

7.3 Analyse statistique

Une analyse statistique de notre corpus a été faite afin d'avoir une idée sur la répartition fréquentielle des phonèmes dans le corpus. Nous donnons dans la table 3.3 la répartition des consonnes et des voyelles dans le corpus.

Série	Nb. consonnes	Nb. voyelles	Nb. total
A	159	131	290
B	172	134	306
C	168	135	303
D	169	138	307
E	166	138	304
Total	834	676	1510
Pourcentage	55%	45%	100%

TABLE 3.3 – Répartition des phonèmes dans le corpus

Une étude statistique faite sur un corpus de 10000 mots arabes a donné une répartition de 57% de consonnes et 43% de voyelles [Datta 90].

Les voyelles

La répartition fréquentielle des voyelles dans le corpus DJOUMA entièrement voyellé est résumée dans la table 3.4. D'une manière générale, selon le timbre, la répartition est donnée par la table 7.3. La table 3.6 donne la répartition des voyelles selon la quantité vocalique. enfin, selon la présence théorique du caractère emphatique (une voyelle est considérée comme emphatisée, si la consonne précédente est emphatique), la répartition des voyelles est résumé dans la table 3.7

Les consonnes

La répartition des consonnes sur les 5 séries du corpus DJOUMA est donnée par la table 3.8. De ce tableau, on peut déduire la répartition selon le mode d'articulation

Voyelle	A	B	C	D	E	Total	Pourcentage
a	58	63	53	49	65	288	43%
i	37	29	31	37	23	157	23%
u	11	17	19	22	21	90	13%
aa	19	16	22	21	19	97	15%
ii	5	8	7	7	8	35	05%
uu	1	1	3	2	2	9	01%
Total	131	134	135	138	138	676	100%

TABLE 3.4 – Répartition des voyelles dans le corpus

Timbre	Nombre	Pourcentage
a	386	57%
i	192	28%
u	98	15%

TABLE 3.5 – Répartition des voyelles selon le timbre

Durée	Nombre	Pourcentage
Brève	535	79%
Longue	141	21%

TABLE 3.6 – Répartition des voyelles selon la quantité

Milieu	Nombre	Pourcentage
Non emphatique	637	94%
Emphatique	39	06%

TABLE 3.7 – Répartition des voyelles selon l'emphase

Consonne	A	B	C	D	E	Total	Pourcentage
t	19	20	17	14	15	85	10%
k	03	07	03	05	07	25	03%
?	10	10	08	09	07	44	05%
b	11	10	04	06	07	38	05%
d	08	07	06	08	10	39	05%
q	06	07	08	06	03	30	04%
ṭ	04	04	06	01	03	18	02%
ḍ	01	01	01	03	03	09	01%
z	03	04	02	04	03	16	02%
f	07	05	05	06	07	30	04%
θ	01	01	01	01	02	06	01%
s	04	05	07	04	04	24	03%
$\int$	02	01	04	03	03	13	02%
χ	01	01	01	01	01	05	01%
$\hbar$	06	03	06	08	06	29	03%
∂	00	01	01	01	01	04	00%
z	01	01	01	01	02	06	01%
γ	02	03	01	02	01	09	01%
ε	04	06	08	04	07	29	03%
h	04	02	02	04	01	13	02%
ṣ	03	01	02	03	02	11	01%
∂̣	01	01	01	01	01	05	01%
m	10	10	13	15	10	58	07%
n	11	20	15	18	15	79	09%
r	10	13	12	08	12	55	07%
l	19	20	23	18	18	98	12%
w	04	04	04	07	05	24	03%
y	04	04	06	08	10	32	04%
Total	159	172	168	169	166	834	100%

TABLE 3.8 – Répartition des consonnes dans le corpus

de consonnes (voir table 3.9). Selon la présence théorique du voisement, le résultat

Consonnes	Nb consonnes	Nombre	Pourcentage
Plosives	8	288	35%
Fricatives	14	200	24%
Autres	6	346	41%

TABLE 3.9 – Répartition des consonnes selon le mode d'articulation

est résumé dans la table 3.10. Selon la présence de l'emphase, la répartition des

Consonnes	Nb consonnes	Nombre	Pourcentage
Voisées	12	320	38%
Non voisées	17	514	62%

TABLE 3.10 – Répartition des consonnes selon le voisement

consonnes est donnée par la table 3.11.

Consonnes	Nb consonnes	Nombre	Pourcentage
Non emphatiques	24	791	95%
Emphatiques	4	43	05%

TABLE 3.11 – Répartition des consonnes selon l'emphase

L'étiquetage manuel du corpus DJOUMA nous a permis d'effectuer une évaluation du système SAPHA. Pour affiner notre analyser phonétique nous avons deux autres types de corpus :

- Un corpus contenant un nombre important de phonèmes spécifiques à la langue (consonnes emphatiques, pharyngales et vélaires).
- Un corpus de paires minimales pour comparer des sons proches. Les principales oppositions sont :
  - de durée pour chaque timbre vocalique,
  - d'emphase (la consonne emphatique et son homologue non emphatique),
  - de gémination,
  - de mode d'articulation,
  - de lieu d'articulation,
  - entre pharyngales, entre vélaires et entre glottales.

8 Conclusion

Nous avons présenté dans ce chapitre l'architecture générale du système et les outils d'analyse automatique de la parole. Nous n'avons pas détaillé lors de cette présentation les étapes de segmentation, de calcul des indices phonétiques et d'étiquetage proprement dit. Ces trois points feront l'objet du chapitre 4.

Chapitre 4

Le décodage phonétique

1 Introduction

La mise en œuvre d'un système de décodage phonétique est un problème très délicat. Quelles connaissances et dans quel formalisme doit-on structurer ces connaissances pour avoir un système performant ? La réponse à ces questions, nous amène à développer des algorithmes des segmentation, des procédures d'extraction des indices phonétiques et des techniques d'étiquetage de l'Arabe standard [Djoudi 90c].

2 La segmentation du signal

Elle consiste à segmenter le signal de parole en grandes classes phonétiques, en utilisant des algorithmes procéduraux non contextuels et reposant sur des critères simples [Carbonell 86]. Le but essentiel de la segmentation est de :

— réduire l'explosion combinatoire lors de la reconnaissance,
— permettre un cadrage pour l'étiquetage automatique.

Nous avons retenu trois grandes classes :
— les voyelle { /a/, /i/, /u/, /aa/, /ii/ et /uu/ },
— les plosives { /t/, /k/, /ʔ/, /b/, /d/, /q/ et /ṭ/ },
— les fricatives { /z/,/f/,/θ/, /s/, /ʃ/,/χ/, /ħ/,/z/, /ṣ/ } et la barre d'explosion fricative.

Autour de ces grandes classes, nous avons ajouté :
— une classe pour les segments ayant des caractéristiques communes aux voyelles et aux fricatives,
— une classe pour les segments ayant des caractéristiques communes aux plosives et aux fricatives.

L'affectation d'un segment à l'une des classes s'effectuera lors du traitement des intersections et des inclusions.
Les algorithmes de segmentation qui ont été développés pour le Frančais s'adaptent bien pour l'Arabe. Il suffit d'enlever de la classe des fricatives les phonèmes qui présentent des structures formantiques et de prendre en considération le fait que certaines fricatives de l'Arabe présentent une limite inférieure du bruit plus basse et donc un centre de gravité énergétique plus bas que celui du Frančais.

2.1 La segmentation des voyelles

Une vision rapide d'un spectrogramme permet de dégager la caractéristique essentielle des voyelles, à savoir une forte énergie particulièrement dans la bande de fréquences ou‘ sont situés les deux premiers formants. Pour délimiter les voyelles, nous utilisons donc :

— la courbe de l'énergie dans une bande de fréquences comprise entre 250 et 3500 Hz (zone ou‘ sont localisés les deux premiers formants),
— la courbe de l'énergie totale.

La première courbe est obtenue en sommant parmi les canaux correspondants aux fréquences comprises entre 250 et 2500 Hz ceux qui atteignent le seuil de visibilité sur le spectrogramme [Carbonell 86]. La seconde est obtenue en calculant l'énergie du signal temporel.
Nous recherchons sur ces deux courbes les pics qui vérifient :

— une intensité au moins égale à la moitié de l'énergie du pic précédent,
— une vallée droite et gauche suffisante en fonction de la hauteur du pic,
— un degré de voisement suffisamment important.

Cette procédure permet en outre de calculer une durée vocalique moyenne, qui nous donne une indication sur la vitesse d'élocution.

2.2 La segmentation des plosives

Les plosives apparaissent sur le spectrogramme comme des zones temporelles blanches (absence d'énergie) suivies éventuellement par une barre d'explosion. Une barre de voisement apparaît le long de l'axe du temps, si la plosive est sonore. C'est ainsi que, pour segmenter les plosives, nous calculons une courbe d'énergie sur le signal temporel préaccentué et filtré par un filtre passe haut dont la fréquence de coupure a pour valeur 600 Hz. Les plosives correspondent à un minimum local sur cette courbe.

2.3 La segmentation des fricatives

Les fricatives sont caractérisées par une énergie importante sous forme de bruit en hautes fréquences et l'absence d'énergie visible en basses fréquences. Ce qui fait que le centre de gravité énergétique est élevé pour les fricatives par rapport aux autres phonèmes. De même, le bruit aléatoire engendre un nombre élevé de passages par zéro.
Pour segmenter les fricatives, nous calculons deux courbes :

- une courbe des passages par zéro sur un signal filtré par un passe haut dont la fréquence de coupure est de 800 Hz,
- une courbe du centre de gravité, calculé sur les parties du spectre visibles sur nos spectrogramme numériques :
  G=($\sum$ i*S(i) /($\sum$ S(i)) pour i de 1 à nombres de canaux (128),
  avec S(i)= intensité en dB du *eme* canal s'il est visible, 0 sinon.

Une fricative est détectée si on met en évidence un maximum local sur ces deux courbes.

2.4 Traitement des intersections et des inclusions

Chacune des trois procédures décrites ci-dessus rend une liste des segments détectés (début, fin, position de l'extremum, cœfficient de certitude). Il faut maintenant fusionner ces trois listes pour obtenir une première segmentation sous forme de treillis. Trois cas sont possibles :

- Deux segments sont disjoints, un segment existe donc entre les segments. Si la durée de ce segment est supérieure à la moitié de la durée vocalique moyenne, il est étiqueté comme une sonnante et va contenir les phonèmes qui n'appartiennent pas aux classes précitées. Dans le cas contraire, le segment est rattaché à ses voisins, et les nouvelles limites sont calculées par différence spectrale. La classe des sonnantes comporte donc le vibrant /r/, le latéral /l/, les nasales /m/ et /n/, les semivoyelles /w/ et /y/, ainsi que les fricatives ayant une structure formantique à savoir le /γ/, /h/, /∂/, /∂̣/ et /ε/.
- Un segment est inclus dans l'autre. Dans ce cas, on garde le segment englobant qui appartiendra à une nouvelle classe ayant des caractéristiques communes aux deux segments (par exemple fricatif et voyelle pour un /i/).
- Il y a intersection entre les deux segments. Si le rapport d'inclusion entre les deux segments est important, on se rapporte au cas précédent. Dans le cas contraire, on génère deux segments (par exemple plosif puis fricatif lorsqu'il s'agit d'un /*f*/ qui se manifeste comme étant un plosive suivie d'une fricative).

les limites des segments sont calculées par différence spectrale.

3 L'extraction des indices

3.1 La durée du segment

La durée relative d'un phonème dépend de son environnement, de la vitesse d'élocution et d'autres facteurs. Cette durée est significative dans la langue Arabe ; ainsi une différence dans la longueur de la voyelle provoque une différence de sens du mot, par exemple : /sin/ (âge) et /siin/ (lettre s).
La durée relative des consonnes dépend du fait qu'elles se trouvent en début, au milieu ou en fin de mot. Elle dépend aussi du voisement, de l'aspiration et de la gémination de la consonne.
Les fricatives sont souples dans leurs prononciations, elles peuvent être allongées aussi longtemps que la circulation de l'air le permet, les durées relatives des fricatives sont donc très variables.
Le calcul des valeurs moyennes de la durée nous a permis de regrouper les phonèmes en grandes classes. Nous donnons, dans le table 4.1, les valeurs moyennes de la durée des phonèmes selon le contexte de production. A partir de ce tableau, nous pouvons

phonèmes	début	milieu	milieu géminé	fin	fin géminé
voyelles brèves	-	75	-	80	-
voyelles longues	-	155	-	140	-
plosives voisées	100	80	180	80	130
plosives sourdes	?	80	150	90	160
fricatives	105	105	180	100	180
vibrant	60	50	120	50	120
nasales	75	70	160	105	160
latéral	80	70	135	85	130
semi-voyelles	90	80	160	80	160

TABLE 4.1 – Durée moyenne des phonèmes

faire les remarques suivantes :

— un mot en Arabe ne commence jamais par une voyelle et une voyelle n'est jamais géminée,
— la durée d'une voyelle longue est approximativement le double d'une voyelle courte,
— en fin de mot, une voyelle courte devient plus longue et une voyelle longue plus courte,

— la durée d'une consonne géminée est environ le double de son homologue simple,
— le silence d'une plosive sourde au début d'une phrase est confondu avec le silence préphonatoire, il ne peut donc être calculé,
— le /r/ est le phonème le plus court.

3.2 Le degré de voisement

C'est le rapport entre le nombre de prélèvements voisés et le nombre total des prélèvements d'un segment. Le voisement d'un prélèvement est déterminé lors du calcul de la fréquence fondamentale. Les voyelles ont le plus souvent un degré de voisement égal à 1. Le degré de voisement des consonnes est donné par le table 4.2.

Le calcul du degré de voisement, nous amène à faire un certain nombre de remarques :

— le degré de voisement des consonnes dépend du fait qu'elles se trouvent en début, au milieu ou en fin de mot (le /h/ est toujours voisé excepté en début de mot). Il dépend aussi du voisement des phonèmes adjacents. Nous observons des segments avec une partie voisée et une autre non voisée,
— le voisement dépend beaucoup des locuteurs, par exemple, le /b/ est complètement sourd pour des locuteurs, il est parfois sonore pour d'autres.
— rares sont les consonnes qui sont tout à fait sourdes ou tout à fait sonores,

3.3 La présence de la barre d'explosion

La barre d'explosion est un paramètre porteur d'informations sur l'existence et la nature des plosives. C'est une explosion d'énergie qui suit généralement le silence de la plosive qui apparaît comme une barre verticale sur le spectrogramme. La barre d'explosion est souvent d'une durée très brève, ce qui fait que l'algorithme de sa détection est difficile à concevoir. La méthode que nous avons mise au point est la suivante :

— A partir de l'énergie totale du signal obtenue par FFT, nous calculons les courbes d'énergie partielle dans les bandes de fréquences [0-1000], [1000-2000], [2000-3000], [3000-4000], [4000-5000], [5000-6000], [6000-7000] et [7000-8000] Hz.
— La détection de la barre d'explosion dans chaque bande de fréquence comme étant des pics (maxima d'énergie) visibles. Le résultat est un ensemble éven-

Phonème	Degré de voisement
t	0.1
k	0.0
?	0.1
b	0.4
d	0.4
q	0.1
ṭ	0.0
ḍ	0.7
z	0.5
f	0.2
θ	0.1
s	0.1
$\int$	0.1
χ	0.5
$\hbar$	0.1
∂	0.7
z	1.0
γ	0.7
ε	1.0
h	0.9
ṣ	0.2
$\dot{\partial}$	0.6
m	1.0
n	0.9
r	0.7
l	0.8
w	0.9
j	0.8

TABLE 4.2 – Degré de voisement des consonnes

tuellement vide de prélèvements correspondants au pics détectés.

— La localisation de la barre d'explosion en retourant le numéro de prélèvement qui contient le plus de pics sur l'ensemble des 8 bandes de fréquences et qui correspond bien à la visibilté de la barre sur le spectrogramme. Il se peut qu'il ait deux barres d'explosion, dans ce cas la procédure rend les deux numéros de prélèvements correspondants.

L'évaluation de la procédure sur un corpus de 232 plosives contenues dans 56 phrases d'un corpus en Franĉais a donné un pourcentage de 87% de bonne localisation, avec un taux d'insertion égal à 13% [Djoudi 86]. Les tests effectués sur 293 plosives arabes du corpus DJOUMA ont donnés un pourcentage de 85%. Le taux d'insertion est d'environ 10%.

3.4 L'analyse de la barre d'explosion

L'analyse de la barre d'explosion (si elle existe) consiste à extraire les informations suivantes ;

— sa durée,

— la fréquence et l'énergie des meilleurs pics ou concentrations d'énergie,

— le degré de compacité qui nous renseigne sur la caractère compact ou diffus du burst.

— le centre de gravité énergétique.

— les fréquences début et fin du burst.

Le degré de compacité ou de concentration dc est calculé par la formule suivante :

$$dc = (max/moy) - 1; \tag{4.1}$$

ou‘ max et moy sont respectivement le maximum et la moyenne de l'énergie du prélèvement du burst. Nous avons opté pour cette formule après avoir remarqué que le maximum d'énergie n'est jamais supérieur au double de la moyenne. Donc le degré de compacité prend ses valeurs dans l'intervalle [0,1]. Il tend vers zéro lorsque l'énergie est uniformément répartie sur toute la bande de fréquences. Dans ce cas le burst est parfaitement diffus. L'autre cas limite, le degré de compacité est égal à 1 et le burst est vraiment compact.

Les autres paramètres sont facilement calculables. La durée n'est prise en compte que dans le cas d'un burst fricatif. Dans le cas contraire, nous considérons qu'elle est égale à la durée d'un prélèvement. La figure 4.1 donne une description schématique des paramètres du burst.

L'ensemble des ses indices nous permettent de connaître le lieu d'articulation de la plosive (labial, dental, vélaire, uvulaire ou glottal). Actuellement, seuls, la durée,

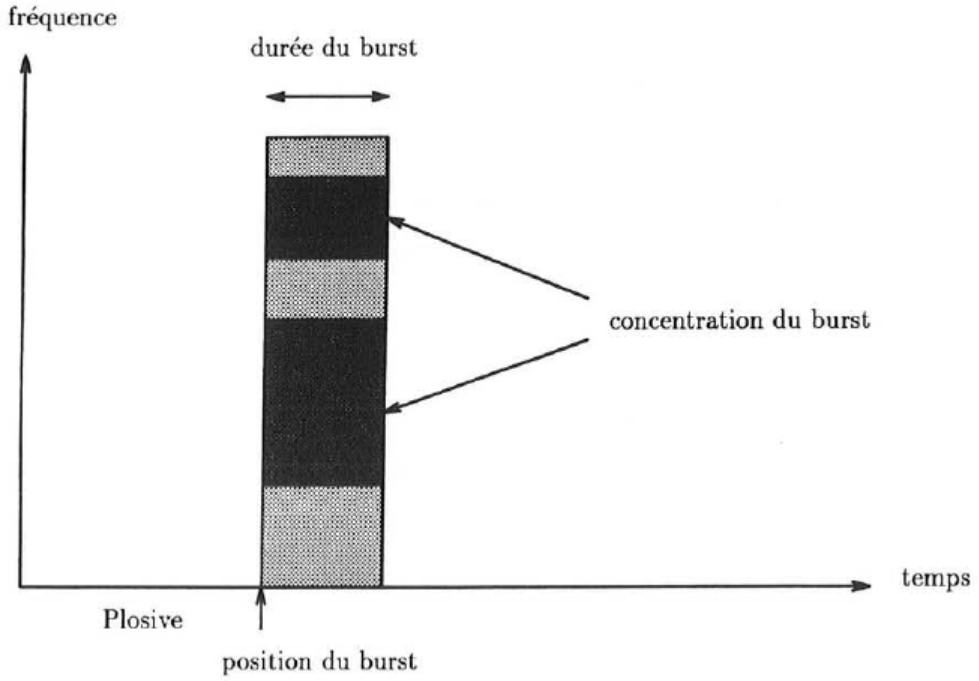


FIGURE 4.1 – Les paramètres du burst

la fréquence et le degré de concentration sont utilisés lors de l'étiquetage phonétique. Une modélisation de la barre d'explosion est nécessaire. Nous donnons dans la table 4.3 les valeurs du degré de concentration (dc) et la fréquence du burst (fb) des plosives dans les différents contextes vocaliques.

Contexte	*a/aa*		*i/ii*		*u/uu*	
gauche	fb	dc	fb	dc	fb	dc
t	4600	0.46	4900	0.48	4250	0.47
k	1800	0.66	2000	0.62	2500	0.60
?	4200	0.59	3700	0.56	-	-
b	2000	0.34	2100	0.31	2150	0.29
d	4140	0.42	3960	0.37	3180	0.57
q	1400	0.61	2120	0.67	2300	0.63
ṭ	3800	0.54	3000	0.66	2500	0.45

TABLE 4.3 – Degré de compacité et fréquence du burst

D'une façon générale, le burst est diffus pour les consonnes labiales et dentales et compact pour les vélaires et glottales.

3.5 Les valeurs des formants

Les formants sont importants pour deux raisons, d'une part, il décrivent les cibles vocaliques correspondant aux zones stables et d'autre part, ils permettent de prendre en compte les phénomènes de coarticulation. L'extraction des formants peut se faire à partir des pics ou des cœfficients de prédiction linéaire mais aussi à partir d'un lissage cepstral. Le codage LPC a l'intérêt de conduire assez directement aux formants quand le segment de parole correspond à une zone vocalique non nasalisée qui est bien le cas des voyelles de l'Arabe standard. Pour chaque prélèvement du segment vocalique et à partir des cœfficients LPC (ordre 20) nous calculons les premiers pics significatifs qui seront des candidats à des positions de formants. Ensuite, nous calculons les quatre premiers formants comme étant des pics visibles dans les bandes de fréquences [250-850], [750-2300], [1800-3000] et [2900-4000] Hz respectivement pour F1, F2, F3 et F4.

Ensuite, nous devrons sortir la valeur de chaque formant pour un segment donné. Pour cela plusieurs approches sont possibles ;

— prendre la valeur au centre du segment,
— calculer la moyenne sur tous le segment,
— calculer la valeur médiane.

Nous avons retenu la troisième méthode parce qu'elle échappe aux erreurs de calcul des pics de la première approche et elle évite aussi les effets de la coarticulation néfastes de la seconde méthode.

Les valeurs des formants sont plus influencées par le lieu d'articulation de la consonne précédente. Nous avons retenu les contextes suivants :

— labial : /b/, /f/, /m/ et /w/ ;
— dental : /t/, /d/, /θ/, /∂/, /n/, /s/, /z/, /r/ et /l/ ;
— palatal : /$\int$/, /z/ et /y/ ;
— vélaire : /k/, /χ/ et /γ/ ;
— emphatique : /ṭ/, /ḍ/, /ṣ/, /∂̣/ et /q/ ;
— pharyngal : /$\hbar$/ et /ε/ ;
— glottal : /ʔ/ et /h/.

Le /q/ est en réalité une consonne uvulaire dont l'influence sur la voyelle suivante est la même que celle des consonnes emphatiques. Nous l'avons considéré, dans ce cas comme emphatique. Nous donnons dans le tableau 4.4 les valeurs des formants des voyelles dans différents contextes. Il est clair qu'à l'intérieur d'une même classe apparaît des différences plus ou moins sensibles.

Contexte	Formant	a	i	u	aa	ii	uu
	F1	630	430	420	590	280	370
Labial	F2	1250	1720	1000	1200	2100	780
	F3	2420	2620	2170	2580	2700	2370
b	F4	3290	3450	3560	3510	3360	3020
	F1	490	420	420	500	280	390
Dental	F2	1550	1970	1090	1330	2020	990
	F3	2500	2790	2290	2550	2880	2220
t	F4	3330	3450	3000	3370	3480	3230
	F1	520	280	300	570	280	300
Palatal	F2	1530	2030	1100	1330	2050	1080
	F3	2490	2900	2650	2380	2750	2300
$\int$	F4	3400	3380	3120	3490	3400	3360
	F1	530	370	310	590	370	380
Vélaire	F2	1230	1570	1070	1510	1980	920
	F3	2170	2530	2270	2430	2780	2440
k	F4	3700	3330	3320	3370	3480	3150
	F1	650	440	440	650	380	440
Emphatique	F2	1200	1680	900	1250	2050	890
	F3	2600	2300	2650	2740	2700	2700
ṭ	F4	3650	3250	3350	3700	3400	3500
	F1	620	300	400	580	331	340
Pharyngal	F2	1570	2000	880	1620	2100	810
	F3	2470	2500	2100	2610	2700	2300
$\hbar$	F4	3830	3600	3380	3540	3410	3130
	F1	660	370	440	670	250	380
Glottal	F2	1270	1860	1130	1130	2020	830
	F3	2390	2680	2340	2430	3070	2130
?	F4	3360	3840	3490	3550	3420	3100

TABLE 4.4 – Valeurs des formants des voyelles en contexte

Les formants caractérisent aussi les consonnes à structure formantique ou sonnantes. Nous donnons dans le tableau 4.5 les valeurs des deux premiers formants de ces consonnes en contexte vocalique.

Consonne	*moyen*		*a/aa*		*i/ii*		*u/uu*	
	F1	F2	F1	F2	F1	F2	F1	F2
/m/	330	1250	340	1270	230	1230	290	1070
/n/	330	1450	370	1240	340	1420	340	1020
/r/	450	1530	600	1300	500	1530	350	1110
/l/	350	1600	330	1620	350	1670	380	1580
/w/	450	950	440	990	460	1030	390	1050
/j/	300	2050	310	1980	280	2070	300	2040
/∂/	300	1400	330	1450	300	1400	310	1380
/∂̣/, /ḍ/	460	1320	450	1500	500	1400	450	1200
/γ/	480	1270	550	1240	470	1550	450	840
/h/	410	2050	670	1560	430	2000	550	1340
/ε/	600	1450	650	1300	470	1250	460	1030

TABLE 4.5 – Valeurs des formants des sonnantes

3.6 Les transitions formantiques

Pour prendre en compte les phénomènes de coarticulation, on doit étudier les transitions formantiques CV ou VC aux frontières entre la voyelle et la consonne adjacente et dire pour chaque formant si la transition est montante, descendante ou plate (voir figure 4.2). Pour se faire, nous prenons un intervalle à la frontière de la voyelle et nous passons les valeurs du formant par une procédure de régression linéaire qui approxime un ensemble de points par une droite en utilisant la méthode des moindres carrées ; le signe du cœfficient directeur (la pente) de la droite permet de déterminer la nature de la transition.

Les transitions formantiques des formants F1, F2 et F3 des voyelles sont souvent utilisées pour l'identification des segments consonnantiques. La table 4.6 résume les transitions les plus remarquées des consonnes aux voisinage des classes des voyelles [1] Nous tenons à signaler, que pour certains phonèmes les transitions ne sont pas nettes daans la plupart des cas. Elles sont données ici qu'à titre indicatif. Les transitions VC des voyelles précédentes sont souvent les mêmes.
En règles générales :

1. dans la table lire m pour transition montante, d pour descendante et p pour plate

Contexte gauche	*a/aa*			*i/ii*			*u/uu*		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
t	m	d	d	m	d	d	m	d	d
k	m	d	m	d	d	p	p	d	m
?	p	m	m	p	p	p	p	p	p
b	m	m	m	p	m	d	m	m	p
d	m	d	d	m	d	d	m	d	d
q	m	d	p	p	m	p	p	p	p
ṭ	m	d	p	m	m	p	p	m	m
ḍ	m	p	p	m	m	p	p	d	m
z	m	d	d	m	d	d	m	d	d
f	m	p	m	p	m	d	m	m	p
θ	m	d	d	m	d	d	m	d	d
s	m	d	d	m	d	d	m	d	d
ʃ	p	m	d	p	m	p	d	d	d
χ	m	m	p	p	p	p	d	d	m
ħ	p	p	m	p	m	d	p	d	m
ð	p	p	d	p	d	d	m	p	p
z	m	p	p	m	p	p	p	p	p
γ	p	m	m	p	m	p	m	m	m
ε	d	m	m	d	m	m	d	d	m
h	p	m	p	p	d	d	p	m	m
ṣ	m	p	p	m	m	p	p	d	m
ð̣	m	p	p	m	m	p	p	d	m
m	m	p	m	p	m	d	m	m	p
n	m	d	d	m	d	d	m	d	d
r	p	d	d	m	d	d	p	d	d
l	m	d	d	m	d	d	m	d	d
w	m	p	m	p	m	d	p	m	p
j	m	d	p	p	d	p	p	m	m

TABLE 4.6 – Les transitions formantiques

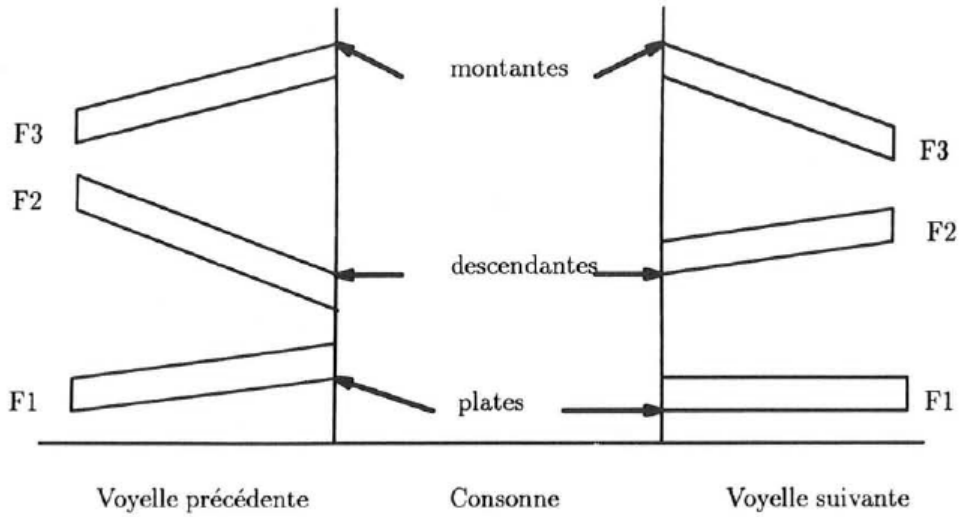


FIGURE 4.2 – Les transitions formantiques

— pour les consonnes labiales F2 et F3 sont montants,
— pour les dentales F2 et F3 sont descendants,
— pour les vélaires F2 est descendant et F3 montant,
— pour les emphatiques F1 est descendant et F2 montant.

3.7 La limite inférieure du bruit

Ce paramètre est très important pour différencier les fricatives. Pour calculer cette limite, nous évaluons sur chaque prélèvement du segment le seuil de visibilité inférieure et puis sur l'ensemble du segment nous calculons la moyenne. Les valeurs de la limite inférieure de bruit de friction des fricatives sont données par la table 4.7.

Les remarques que nous faisons sur la limite inférieure du bruit sont les suivantes :
— nous n'avons pas considéré les fricatives qui présentent des structures formantiques.
— pour la même fricative, il existe une différence entre entre les valeurs de la limite selon que la voyelle suivante est postérieure, antérieure ou centrale.
— la fricative /θ/ est celle qui a la limite la plus élevée,
— la fricative emphatique /s̩/ possède un bruit légèrement plus bas que son homologue non emphatique /s/,
— le /$\hbar$/ possède la limite de bruit la plus basse, elle est le plus souvent au dessous de 1000 Hz.

D'une façon générale, plus le lieu d'articulation est à l'arrière, plus la limite du bruit

Fricative	Limite du bruit
z	1750
f	1380
θ	3500
s	3300
$\int$	1800
χ	1000
$\hbar$	900
z	3450
$\dot{s}$	3200

TABLE 4.7 – Limite inférieure du bruit des fricatives

est basse.

3.8 Le centre de gravité énergétique

Il s'agit pour chaque prélèvement d'un segment donné de calculer la position fréquentielle du centre de gravité de l'énergie et de calculer la moyenne sur tout le segment. Ce paramètre utilisé déjà lors de la segmentation sert à l'identification des fricatives.
Les valeurs du centre de gravité des fricatives sont résumées dans le tableau 4.8 :

Fricative	Centre de gravité
z	3160
f	4550
θ	4820
s	4760
$\int$	4570
χ	3580
$\hbar$	3860
∂	2880
z	4900
γ	3420
ε	1570
h	2880
$\dot{s}$	4940
$\dot{\partial}$	2910

TABLE 4.8 – Centre de gravité énergétiques des fricatives

4 L'étiquetage procédural

Le module de segmentation rend une suite de segments délimités par des frontières et qui appartiennent à l'une des grandes classes. L'étiquetage procédural consiste à développer une procédure pour chaque grande classe afin d'affecter à chaque segment le ou les étiquettes phonétiques les plus plausibles en utilisant les indices pertinents spécifiques à chaque classe.

4.1 Etiquetage des voyelles

Afin de reconnaître les segments vocaliques, nous utilisons :
— Un rapport de distance entre les formants F1, F2 et F3 du segment et les valeurs de références de ces formants pour chaque voyelle de la langue,
— une distance entre la durée du segment et la durée vocalique moyenne de la phrase.

4.2 Etiquetage des plosives

Pour l'identification des plosives, nous utilisons les paramètres :
— le degré de voisement du segment,
— la présence du burst [Djoudi 86],
— la valeur en Hz du centre de gravité du burst,
— le degré de concentration du burst,
— les transitions des formants F1, F2 et F3 de la voyelle suivante.

4.3 Etiquetage des fricatives

Pour l'étiquetage des fricatives, on se base sur les indices :
— le degré de voisement,
— la limite inférieure du bruit de friction,
— le centre de gravité du bruit,
— les transitions des formants F1, F2 et F3 de la voyelle suivante.

4.4 Etiquetage des sonnantes

Pour reconnaître les segments à structure formantique, nous utilisons :
— un rapport de distance entre les formants F1 et F2 du segment et les valeurs de références de ces formants pour chaque consonne,
— le degré de voisement.

4.5 Procédures d'identification

Pour l'identification phonétique nous avons fixé un score initial pour chaque paramètre phonétique en fonction de sa pertinence et nous avons développé une procédure par classe phonétique en utilisant les indices caractéristiques décrites ci-dessus. La combinaison des scores pour obtenir le score final, se fait en utilisant les fonctions logiques classiques ainsi que les fonctions floues suivantes (voir figure 4.3) :

super(x,d,f) : qui rend 1 si $x > f$; 0 si $x < d$ et linéairement une valeur entre 0 et 1 si x est entre d et f.

infer(x,d,f) : qui rend 1 si $x < d$; 0 si $x > f$ et linéairement une valeur entre 0 et 1 si x est entre d et f.

aumilieu(x,d,f) : qui rend 1 si x est le milieu du segment df ; 0 si $x < d$ ou $x > f$; et linéairement une valeur entre 0 et 1 si x est entre d et milieu ou x est entre milieu et f.

A la fin nous choisissons les 3 meilleures étiquettes qui totalisent les plus grands scores.

Lors de l'étiquetage procédural nous avons rencontré deux problèmes :

— la difficulté de prendre en compte le phénomène de coarticulation et l'influence des phonèmes voisins sur le segment à étiqueter,

— la mise à jour des connaissances phonétiques qui nécessite la modification des algorithmes d'étiquetage.

Ces inconvénients nous ont amené à adopter une démarche utilisant un système à base de connaissances.

5 Le décodage par le système à base de connaissances

Le module d'étiquetage, dans la deuxième version de SAPHA consiste, en une base de connaissances d'identification de phonèmes sous forme de règles de production et un moteur d'inférence. Le système se charge d'affecter à chaque segment détecté par le module de segmentation une liste de un ou plusieurs phonèmes. Ce système a été déjà utilisé pour le décodage phonétique du Franĉais dans le cadre du projet APHODEX [Carbonell 86], [Fohr 86].

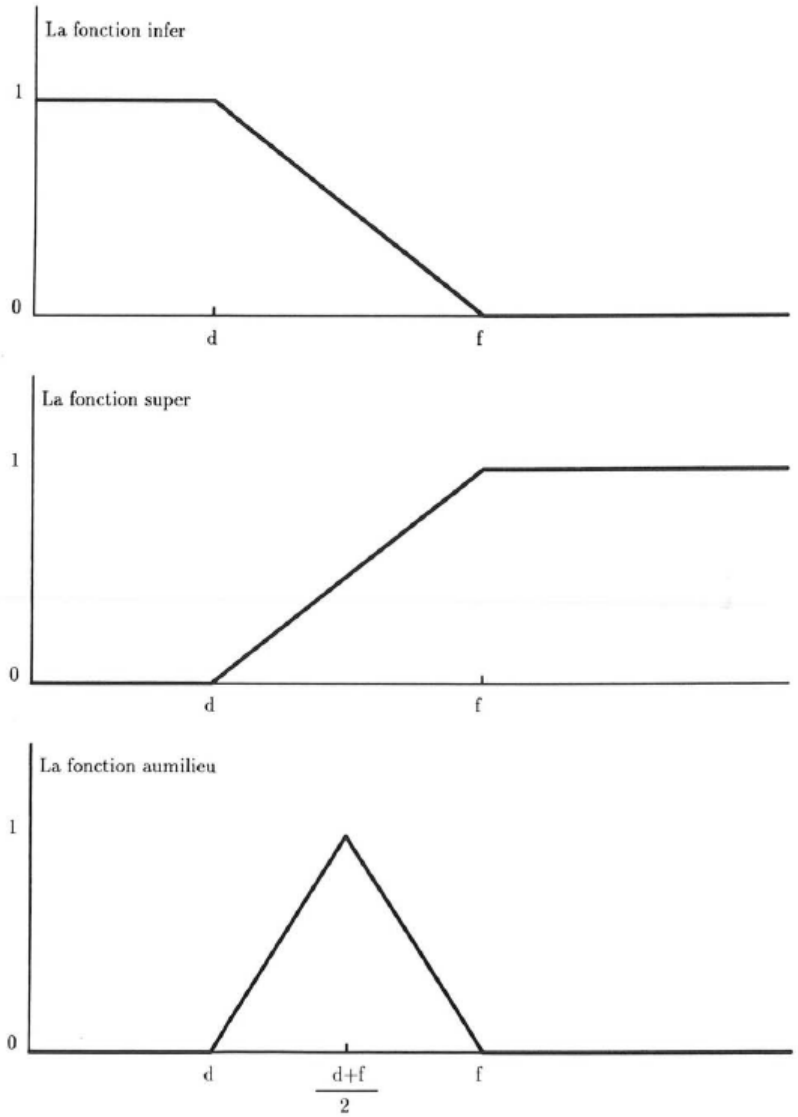


FIGURE 4.3 – Les fonctions floues

5.1 La base de connaissances

A défaut d'un expert phonéticien de l'Arabe, nous avons nous même construit une base de connaissances en s'appuyant sur l'analyse phonétique faite sur des phrases lues du corpus DJOUMA

La base de règles

Les connaissances acquises sont actuellement formalisées sous forme de règles de production [Carbonell 86].
Une règle se compose de plusieurs parties, certaines pouvant être facultatives. La syntaxe générale d'une règle est la suivante :

```
R Numero_de_regle
C commentaire_en_claire C
CONTEXTE_DROIT [liste de phonemes]
CONTEXTE_GAUCHE [liste de phonemes]
SI
    premisse
ALORS
    conclusion
FIN
```

Le numéro de la règle sert à repérer les règles. Il est utilisé pour construire la trace de raisonnement du système.

Le commentaire sert a donné une indication sur la signification de la règle.

Les contextes droit et gauche constituent une partie des conditions d'application d'une règle. Chaque contexte est une liste éventuellement vide de phonèmes. Les contextes gauche et droit limitent donc l'application de la règle à une situation particulière.

La partie prémisse contient des conditions sur des faits utilisant les opérations mathématiques usuelles ($+$, $-$, $*$, $/$, log, exp, sinus, cosinus, tangente ...), les opérateurs booléens ($=$, $>$, $>=$, $<$, $<=$, $\neq$, et) et les opérateurs flous ($>>$, $<<$ et $\circ$) qui sont respectivement l'équivalent des fonctions *super*, *infer* et *aumilieu* utilisé lors de l'étiquetage procédural. Un fait peut être une constante numérique ou bien une variable correspondant à une valeur d'un indice phonétique du segment en cours, du segment précédent ou du segment suivant. Pour savoir à quel segment appartient la variable, on la fait suivre d'un suffixe PRE, ACT ou SUC, respectivement pour le segment précédent,

actuel ou suivant. Chaque fois qu'un indice inconnu est rencontré dans la partie prémisse, on active une procédure pour l'extraire.

La conclusion est une liste de phonèmes pondérés par des cœfficients de vraisemblance ou une action à exécuter. Les pondérations vont de −100 (complètement faux) à +100 (totalement vrai). Les valeurs positives servent à affirmer un phonème, les valeurs négatives à l'infirmer.

Si deux prémisses sont séparées par la fonction "&" (et) on calcule le minimum de chacun de leurs cœfficients de vraisemblance.
Si une prémisse contient des variables qui sont constituées d'une liste de valeurs (par exemple "burst-freq") on instancie ces variables avec toutes les valeurs possibles de la liste et la prémisse est affectée d'un cœfficient de vraisemblance correspondant au maximum de tous ceux calculés.

La base de connaissances contient 192 règles. Il existe différents types de règles :

— règles de phonétique générale qui traduisent des connaissances générales sur la segmentation et/ou le voisement. Les numéros vont de 1 à 100. Voici un exemple :

```
R008
   C Non voisement et segment plosif C
   SI
      NON(pitch_ACT) &
      pret-plos_ACT
   PHONEMES [ t 50 k 50 A 50 q 50 t. 50 # 50 ]
FIN
```

Certaines règles sont à pondération négative, la présence d'un indice permet d'éliminer un ou plusieurs phonèmes. Exemple :

```
R008
   C Non voisement  C
   SI
      NON(pitch_ACT)
   PHONEMES [ b -50 d -50 J -50 h -50 E -50 m -50 n -50 l -50 r -50 w -50
-50 G -50 D -50 d. -50 D. -50 ]
FIN
```

— règles sur les voyelles qui utilisent la durée du segment et une distance entre les valeurs des formants F1, F2 et F3 du segment et des valeurs de références de ces formants pour chaque voyelle et dans différents contextes. Les numéros des règles vont de 101 à 200.

Une règle sur la durée :

```
R102
   C duree super a 1.2 dvm donc voyelle longue  C
   SI
      pret_voy_ACT &
      >>(duree_ACT (dvm_T * 0.8) (dvm_T * 1.2))
   PHONEMES [ aa 80 ii 80 uu 80 ]
FIN
```

Une règle sur les formants :

```
R107
   C a en contexte emphatique dist infer des formants C
   CONTEXTE_GAUCHE  [ t. d. s. D. q ]
   SI
     pret-voy_ACT &
     <<( (DIST(formant1_ACT 650) +
          DIST(formant2_ACT 1200) +
          DIST(formant3_ACT 2600) * 0.6  ) 1 21)
   PHONEMES [ a 100 ]
FIN
```

— règles sur les plosives dont les numéros vont de 201 à 300 et qui comportent des règles sur la présence, la forme et la fréquence du burst et sur les transitions formantiques. Voici quelques exemples :

règle sur l'absence du burst :

```
R210
   C burst absent donc /A/ C
   SI
     NON(burst-present_ACT)
   PHONEMES [ A 50 ]
FIN
```

règle sur la forme du burst :

```
R223
  C burst etendu donc /A/ ou /b/ C
  SI
    burst-present_ACT &
    burst-etendu_ACT
  PHONEMES [ A 50 b 50 ]
FIN
```

règle sur la fréquence du burst :

```
R223
  C freq burst contexte i ii  donc t C
  SI
    burst-present_ACT &
    ^(burst-freq_ACT 4800 5200 )
  PHONEMES [ t 50 ]
FIN
```

règle sur les transitions formantiques :

```
R231
  C Contexte a aa et  F2  montant alors k C
  CONTEXTE_DROIT [ a aa ]
  SI
    pret-plos_ACT &
    pret-voy_SUC &
    f2-g-montant_SUC
  PHONEMES [ k 50 ]
FIN
```

— règles sur les fricatives qui s'appuyent sur le seuil du bruit de friction, le centre de gravité énergétique et les transitions formantiques . Les règles vont de 301 à 400, en voici des exemples :
règle sur la limite inférieure du bruit :

```
R304
  C seuil friction  fricative c C
  SI
    seuil-fric > 1500 &
    seuil-fric < 2500
  PHONEMES [ c 70 ]
FIN
```

règle sur le centre de gravité

```
R306
  C centre de gravite super donc  fricative s s. Z C
  SI
    centre-g > 4700
  PHONEMES [ s 70 s. 70 Z 70 ]
FIN
```

La base de mesures

A chaque segment correspond une base de mesures qui contient les résultats des procédures de traitement du signal appliquées sur ce segment. Une mesure comprend son nom, sa valeur (booléenne ou liste de valeurs numériques possibles) et le numéro de début et de fin de prélèvement ou‘ a été calculée la mesure.

La base de faits

Au départ, à chaque segment correspond un nœud du treillis de décodage. A chaque nœud est associée une base de faits qui contient la liste des règles déjà appliquées, la liste des contextes droit et gauche supposés à ce stade, les faits qui ont été utilisés et la liste pondérée des phonèmes déduits et la liste des buts à atteindre. A chaque nouveau contexte, on affecte un nouveau nœud.

5.2 Le moteur d'inférence

Le moteur du système se charge d'affecter à chaque segment détecté par le module de segmentation une liste de un ou plusieurs phonèmes. Il est capable de remettre en cause la segmentation à tout moment, de dérouler en parallèle une analyse sur plusieurs segments et de fournir une trace de son raisonnement. Le moteur d'inférence fonctionne en chaînage avant et en chaînage arrière en effectuant une analyse segment par segment en fonction de la stratégie choisie. L'activation d'une règle est subordonnée à des conditions de compatibilité des contextes gauche et droit, de la conclusion et de la segmentation. Une plausibilité est affectée à chaque phonème hypothétique. Pour traiter l'incertitude et l'imprécision contenues dans les règles, le moteur utilise un raisonnement approximatif basé sur la logique floue et les cœfficients de vraisemblance.

Choix de la stratégie

Le moteur d'inférence doit choisir à chaque étape le nœud à étudier et la règle à appliquer.
Pour le premier point, on peut décider pour un décodage gauche droite, c'est à dire dans l'ordre de production des phonèmes ou bien on opère pour un décodage par ilots de confiance ou‘ la priorité sera donnée aux classes phonétiques les plus faciles à reconnaître [Djoudi 91].

Pour le second point, on applique déjà les règles les plus sûres puis les règles dont les prémisses sont satisfaites avec un grand degré de confiance. Si aucune conclusion ne peut être déduite, on applique toutes les règles restantes.

Le treillis phonétique

La sortie du système de décodage est un treillis de phonèmes. Chaque segment est composé d'un ou de plusieurs nœuds. Un nœud étant une interprétation contextuelle du segment. Chaque nœud du treillis contient :

— les limites du segment auquel correspond le nœud,
— les règles activées,
— les valeurs des indices utilisés,
— une liste de phonèmes avec leurs pondérations,
— les contextes gauche et droit.

6 Résultats et commentaires

6.1 Résultats de la segmentation

Nous avons testé les algorithmes de segmentation sur le corpus DJOUMA, segmenté manuellement. Les résultats que nous indiquons sont calculés par comparaison avec cet étiquetage manuel (voir table 4.9).

Les résultats de la segmentation par classe sont résumés dans la table 4.10.

A partir de ces résultat, les remarques que nous pouvons faire sont les suivantes :

— Le corpus étant équilibré, certains phonèmes sont peu présents, leur résultat n'est pas très représentatif. Une évaluation avec d'autres corpus est donc nécessaire pour pouvoir en tirer des conclusions. Nous revenons avec plus de détails sur les résultats des consonnes emphatiques, pharyngales, glottales et vélaires.
— Dans l'ensemble, les algorithmes fournissent un bon score de segmentation (supérieur à 90%). Les voyelles sont souvent omises à la fin de la phrase. Les omissions les plus fréquentes concernent la classe des sonnantes : lorsque deux sonnantes sont adjacentes, le système rend un seul segment, l'autre segment est automatiquement omis. Le /n/ de la nounation possède une énergie faible, il est donc souvent omis à la fin de la phrase.
— Les insertions sont relativement peu nombreuses dans les classes des voyelles et des plosives. Par contre, elles sont importantes dans les fricatives et les sonnantes. Les phonèmes responsables des insertions dans la classe des fricatives

Phon	Nb	Plo	Voy	Fri	Aut	FriVoy	PloFri	Ommis	Taux
t	170	165	1	1	1	0	0	2	97 %
k	51	49	0	0	0	0	0	2	96 %
?	93	66	1	0	23	0	0	3	70 %
b	77	73	0	0	3	0	0	1	94 %
d	80	75	0	1	4	0	0	0	93 %
q	61	60	0	0	1	0	0	0	98 %
ṭ	36	33	0	0	1	0	0	2	91 %
a	623	1	552	2	14	15	0	39	91 %
i	339	3	197	2	5	106	0	26	89 %
u	149	1	127	0	4	1	0	16	85 %
aa	181	0	163	0	1	9	0	8	95 %
ii	58	0	27	2	0	24	0	5	87 %
uu	42	1	32	0	2	1	0	6	78 %
z	32	2	0	28	1	1	0	0	96 %
f	60	4	0	55	0	0	1	0	93 %
θ	13	0	0	12	1	0	0	0	92 %
s	48	0	0	47	0	0	0	1	97 %
$\int$	26	0	0	25	0	1	0	0	100 %
χ	10	0	0	10	0	0	0	0	100 %
$\hbar$	58	1	0	41	12	1	0	3	72 %
z	12	0	0	12	0	0	0	0	100 %
ṣ	22	0	0	22	0	0	0	0	100 %
$\partial\!\!\!\cdot$	18	6	0	1	10	0	0	1	55 %
ḍ	10	2	2	0	5	0	0	1	50 %
∂	8	1	0	0	5	1	0	1	62 %
γ	19	2	0	3	12	1	0	1	63 %
ε	59	1	2	0	43	2	0	11	72 %
h	29	0	2	0	17	0	0	10	58 %
m	117	3	5	0	84	0	0	25	71 %
n	161	10	2	0	108	1	0	40	67 %
l	205	1	4	26	119	2	0	53	58 %
r	112	1	4	2	78	1	0	26	69 %
w	49	1	1	0	41	0	0	6	83 %
j	65	1	1	19	17	2	0	25	26 %

TABLE 4.9 – Résultat de la segmentation

Classe	Nombre présents	Nombre trouvés	Nombre insérés
Plosives	720	672 (93%)	63 (9%)
Voyelles	1392	1254 (90%)	124 (9%)
Fricatives	281	256 (91%)	99 (35%)
Sonnantes	852	539 (63%)	193 (21%)
Total	3245	2721 (84%)	479 (15%)

TABLE 4.10 – Résultats de la segmentation en grandes classes

sont le /j/ et à un degré moindre le /l/. Le /j/ est présent à plus de 50% dans la classe des fricatives. Les insertions dans la classe des voyelles sont dûes aux consonnes à structure formantique. Le /n/ présente parfois une occlusion, il est ainsi segmenté comme plosive. Le /?/ en milieu intervocalique présente de faibles formants, il est donc segmenté comme une sonnante.

6.2 Résultats de la reconnaissance

Nous présentons les résultats de la reconnaissance phonétique sous forme d'une matrice de confusion, en prenant à chaque fois les trois meilleurs étiquettes. Deux évaluations ont été faites, l'une dans un cadre monolucuteur et une deuxième dans un contexte multilocuteur.

Résultats en monolocuteur

Avec le corpus DJOUMA (50 phrases équilibrées), nous avons effectué une première évaluation pour un seul locuteur masculin sans aucune adaptation. Le pourcentage d'étiquetage correct est donné par la matrice de confusion de la figure 4.4 :

Résultats en multilocuteur

Sur le même corpus, nous avons effectué une évaluation du système pour trois locuteurs masculins, le résultat est donné par la matrice de confusion de la figure 4.5.

Le pourcentage de reconnaissance par grandes classes phonétique est résumé par la table 4.11.

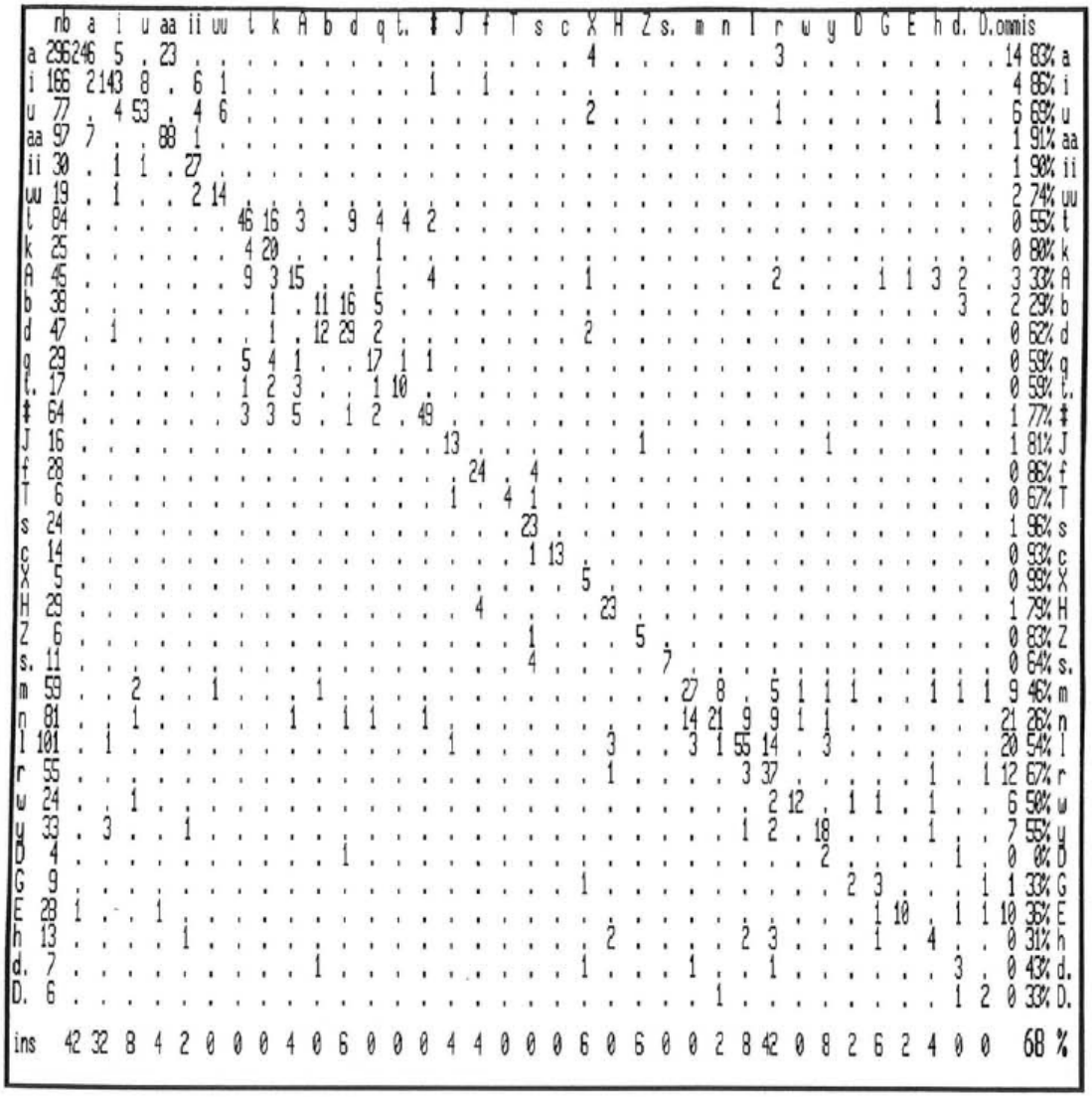


FIGURE 4.4 – Matrice de confusion en monolocuteur

Classe	Présents	Reconnus	Insérés
Voyelles	2052	1805 (78%)	270 (13%)
Plosives	847	450 (53%)	72 (9%)
Fricatives	422	292 (69%)	156 (37%)
Sonnantes	1263	510 (40%)	284 (22%)
Total	4584	3057 (65%)	782 (17%)

TABLE 4.11 – Résumé des résultats de la reconnaissance

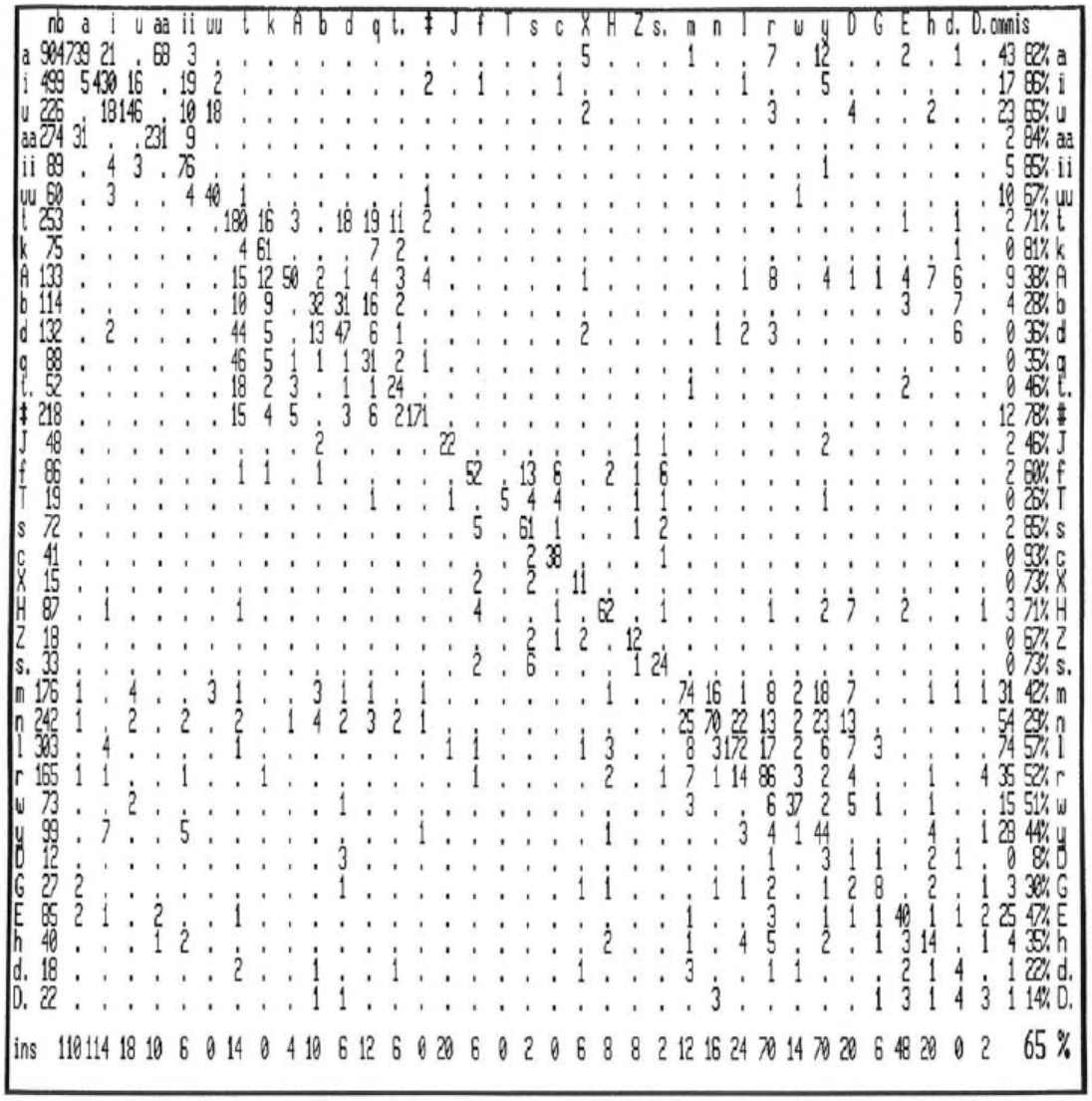


FIGURE 4.5 – Matrice de confusion en multilocuteur

A l'analyse de la matrice de confusion, nous pouvons remarquer la constitution de blocs autour de la diagonale principale. Les confusions les plus importantes sont entre les phonèmes de la même classe phonétique (voyelle, plosive, fricative ou sonnante). L'analyse des résultats du décodage phonétique de l'Arabe nous a permis d'expliquer certaines erreurs. Les principales causes sont :

— Les erreurs de segmentation, en particulier lorsque deux phonèmes appartenant à une même classe, se suivent dans la phrase, le système rend un seul segment. Ce cas est relativement fréquent dans la classe des sonnantes.
— Les procédures de calcul des indices phonétiques ne fournissent pas toujours les valeurs correctes.
— La base de règles est insuffisante. Il reste un certain nombre de règles à écrire ou à modifier. Une expérience dans le domaine de la lecture des spectrogrammes s'impose pour vérifier la cohérence de la base.

Comparaison avec APHODEX

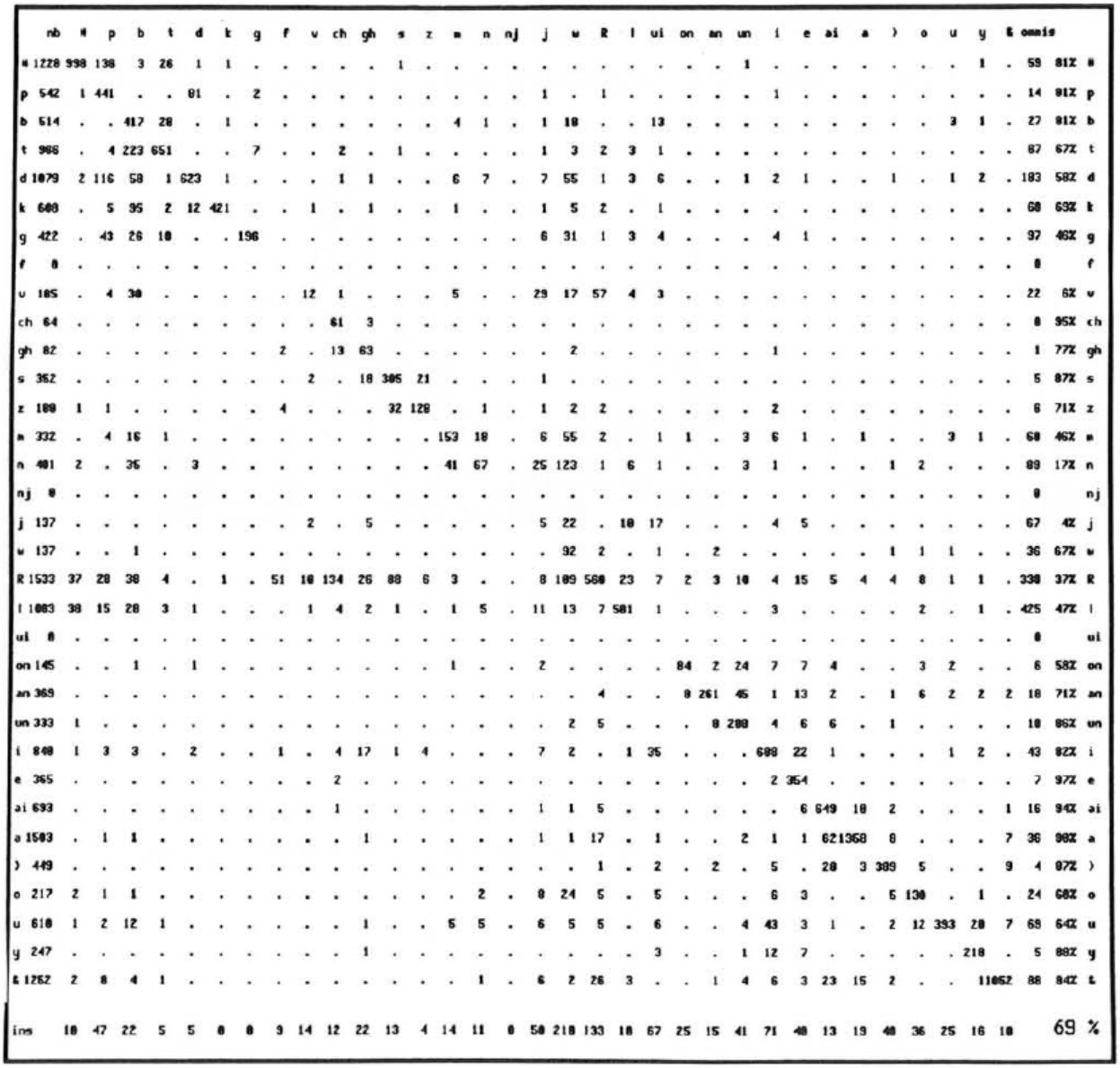
APHODEX est un système expert qui a été développé au sein de l'équipe RFIA du CRIN pour les décodage acoustico-phonétique multilocuteur de la parole continue du Franĉais. Il modélise le savoir faire d'un expert phonéticien en lecture de spectrogrammes. L'expertise acquise (règles et stratégies) est formalisée sous la forme d'un système à règles de production. L'architecture générale du système fait apparaître quatre partie : un module de segmentation, un ensemble de procédures d'extractions de paramètres, une base de règles et un moteur d'inférence. Les principales caractéristiques du système sont :

— remise en cause de la segmentation possible à tout moment,
— déroulement en parallèle de l'analyse sur plusieurs segmentations,
— prise en compte des phénomènes contextuels,
— gestion de l'incertitude en ce qui concerne l'interprétation des mesures (raisonnement incertain),
— fabrication d'un véritable treillis phonétique,
— échange d'informations avec les niveaux supérieurs (utilisation d'APHODEX en vérification d'hypothèses lexicales).

L'architecture de SAPHA est largement inspirée de celle d'APHODEX.

L'évaluation du système APHODEX a été faite sur un corpus contenant 16848 phonèmes. Les résultats sont présentés par la matrice de confusion de la figure 4.6 calculée par recalage entre l'étiquetage manuel et le décodage automatique.

Même s'il est difficile de comparer les deux systèmes, (les deux langues présentent un nombre de phonèmes différent et des caractéristiques phonétiques différentes),



	nb	H	p	b	t	d	k	g	f	v	ch	gh	s	z	m	n	nj	j	w	R	l	ui	on	an	un	i	e	ai	a	)	o	u	y	&	omis	
H	1228	998	138	3	26	1	1	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.	59	81%	H
p	542	1	441	.	.	81	.	2	.	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	.	1	.	.	.	.	.	.	.	.	14	81%	p
b	514	.	.	417	28	.	1	.	.	.	.	.	.	.	4	1	.	1	18	.	.	13	.	.	.	.	.	.	.	.	.	3	1	.	27	81%	b
t	986	.	4	223	651	.	.	7	.	.	2	.	1	.	.	.	.	1	3	2	3	1	.	.	.	.	.	.	.	.	.	.	.	.	67	67%	t
d	1079	2	116	58	1	623	1	.	.	.	1	1	.	.	6	7	.	7	55	1	3	6	.	.	1	2	1	.	.	1	.	1	2	.	183	58%	d
k	608	.	5	95	2	12	421	.	.	1	.	1	.	.	1	.	.	1	5	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	68	69%	k
g	422	.	43	26	18	.	.	196	.	.	.	.	.	.	.	.	.	6	31	1	3	4	.	.	.	4	1	.	.	.	.	.	.	.	97	46%	g
f	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8		f
v	185	.	4	30	.	.	.	.	.	12	1	.	.	.	5	.	.	29	17	57	4	3	.	.	.	.	.	.	.	.	.	.	.	.	22	6%	v
ch	64	.	.	.	.	.	.	.	.	.	61	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	95%	ch
gh	82	.	.	.	.	.	.	.	2	.	13	63	.	.	.	.	.	.	2	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	1	77%	gh
s	362	.	.	.	.	.	.	.	.	2	.	18	305	21	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	87%	s
z	188	1	1	.	.	.	.	.	4	.	.	.	32	128	.	1	.	1	2	2	.	.	.	.	.	2	.	.	.	.	.	.	.	.	8	71%	z
m	332	.	4	16	1	.	.	.	.	.	.	.	.	.	153	18	.	6	55	2	.	1	1	.	3	6	1	.	1	.	.	3	1	.	68	46%	m
n	481	2	.	36	.	3	.	.	.	.	.	.	.	.	41	67	.	25	123	1	6	1	.	.	3	1	.	.	.	1	2	.	.	.	89	17%	n
nj	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8		nj
j	137	.	.	.	.	.	.	.	.	2	.	5	.	.	.	.	.	5	22	.	18	17	.	.	.	4	5	.	.	.	.	.	.	.	67	4%	j
w	137	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	92	2	.	1	.	2	.	.	.	.	.	1	1	1	.	.	36	67%	w
R	1533	37	28	38	4	.	1	.	51	18	134	26	88	6	3	.	.	8	189	588	23	7	2	3	18	4	15	5	4	4	8	1	1	.	338	37%	R
l	1083	38	15	28	3	1	.	.	.	1	4	2	1	.	1	5	.	11	13	7	581	1	.	.	.	3	.	.	.	.	2	.	1	.	425	47%	l
ui	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8		ui
on	145	.	.	1	.	1	.	.	.	.	.	.	.	.	1	.	.	2	.	.	.	.	84	2	24	7	7	4	.	.	3	2	.	.	6	58%	on
an	369	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	.	8	261	45	1	13	2	.	1	6	2	2	2	18	71%	an
un	333	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	5	.	.	.	8	200	4	6	6	.	1	.	.	.	.	18	86%	un
i	848	1	3	3	.	2	.	.	1	.	4	17	1	4	.	.	.	7	2	.	1	35	.	.	.	688	22	1	.	.	.	1	2	.	43	82%	i
e	365	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	354	.	.	.	.	.	.	.	7	97%	e
ai	693	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	1	5	.	.	.	.	.	.	6	649	18	2	.	.	.	1	16	94%	ai
a	1503	.	1	1	.	.	.	.	.	.	.	1	.	.	.	.	.	1	1	17	.	1	.	.	2	1	1	621	368	8	.	.	.	7	36	98%	a
)	449	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	2	.	2	.	5	.	28	3	389	5	.	.	9	4	97%	)
o	217	2	1	1	.	.	.	.	.	.	.	.	.	.	.	2	.	8	24	5	.	5	.	.	.	6	3	.	.	5	130	.	1	.	24	68%	o
u	618	1	2	12	1	.	.	.	.	.	.	1	.	.	5	5	.	6	5	5	.	6	.	.	4	43	3	1	.	2	12	393	28	7	69	64%	u
y	247	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	3	.	.	1	12	7	.	.	.	.	.	218	.	5	88%	y
&	1262	2	8	4	1	.	.	.	.	.	.	.	.	.	.	1	.	6	2	26	3	.	.	1	4	6	3	23	15	2	.	.	1106	2	88	84%	&
ins		18	47	22	5	5	0	0	9	14	12	22	13	4	14	11	0	58	218	133	18	67	25	15	41	71	48	13	19	48	36	25	16	18		69 %	

FIGURE 4.6 – Matrice de confusion du Francˇais

l'analyse des matrices de confusion nous permet de faire les remarques suivantes :

- le taux global de reconnaissance est meilleur pour APHODEX que pour SAPHA. Les connaissances dans APHODEX émane d'un expert phonéticien, par contre celles de SAPHA sont les résultats de notre propre étude spectrographique,
- APHODEX fourni un excellent résultat sur les voyelles (environ 90 %) en utilisant uniquement les valeurs des formants. En Arabe, nous avons vu que la durée des voyelles a une importance capitale et que cette durée dépend du contexte de production des phonèmes. Nous ne savons pas comment séparer exactement les voyelles longues des voyelles courtes,
- les corpus d'évaluation du francˆais est nettement plus important mais il est loin d'être équilibré, par contre, les phonèmes sont uniformément répartis dans le corpus DJOUMA.

Afin d'améliorer le taux global de reconnaissance, nous préconisons :

- l'utilisation de l'énergie du bruit de friction comme paramètre pour la reconnaissance des fricatives. Jusqu'à présent seul la limite inférieure du bruit est utilisée.
- le remplacement de la valeur au centre des formants par la valeur médiane. Ceci permet d'éviter les valeurs aberrantes dues à un mauvais calcul des pics.
- l'utilisation conjointe des contextes droit et gauche dans l'écriture des règles sur les formants des sonnantes ainsi que la visibilté des formants.

6.3 Résultats du système sur les consonnes arrières

L'originalité du système phonétique de l'Arabe se fonde sur la présence des consonnes emphatiques, pharyngales, glottales et uvulaires. La caractéristique commune à ces consonnes est l'existence pour chacune d'elles d'un point d'articulation à l'arrière de l'appareil phonatoire. Elles sont dites donc consonnes arrières. Nous nous sommes intéressé à leur reconnaissance automatique. Nous signalons que ces consonnes se présentent en nombre réduit dans la langue, par conséquent, nous étions amenés, pour l'analyse, comme pour la reconnaissance, à construire des corpus de mots et de phrases avec un nombre très important des consonnes à étudier. Nous avons donc un corpus pour les consonnes emphatiques, un autre pour pour les consonnes pharyngales et glottales et enfin, un dernier pour les consonnes vélaires et uvulaires.

Résultats du système sur les consonnes emphatiques

Il s'agit pour nous, de reconnaître les consonnes : /ṭ/, /ḍ/, /ṣ/ et /ð̣/ à partir des connaissances que nous avons acquis de notre propre expérience dans la lecture des spectrogrammes de parole [Djoudi 90a]. Ainsi, pour la reconnaissance de la plosive /ṭ/ nous nous sommes fondés sur :

— La présence du burst, son centre de gravité ;
— Le degré de voisement ;
— Les transitions formantiques des voyelles qui précèdent et celles qui suivent les consonnes en question.

Concernant la fricative /ṣ/ nous avons utilisé :

— Le voisement ;
— La limite inférieure du bruit de friction ;
— Les transitions formantiques des voyelles adjacentes.

Pour les consonnes /ḍ/ et /ð̣/, en plus du degré de voisement, des transitions formantiques des voyelles adjacentes. Nous avons utilisé les valeurs des formants F1 et F2 des consonnes eux mêmes quand elles apparaîssent comme sonnantes. Pour les quatre consonnes, nous avons pu écrire, jusqu'à présent, 37 règles de production. Nous avons testé les algorithmes de segmentation du système sur le corpus contenant un grand nombre de consonnes emphatiques. Les résultats obtenus pour trois locuteurs et calculés par comparaison avec la segmentation manuelle de ce corpus, se présentent comme indiqués dans la table 4.12. Bien que défini comme fricative,

Phon	Nb	Plo	Voy	Fri	Son	FriVoy	PloFri	Omis	Taux
ṭ	64	59	0	1	2	0	0	2	92 %
ṣ	22	0	0	21	0	0	0	1	95 %
ð̣	38	11	0	6	19	0	0	2	50 %
ḍ	28	8	2	3	13	0	0	1	46 %

TABLE 4.12 – Résultat de la segmentation des consonnes emphatiques

/ð̣/ apparaît plutôt comme une sonnante avec des structures formantiques .

Pour l'identification, les trois phonèmes qui totalisent le meilleur score sont pris pour étiquettes du segment. Le pourcentage de reconnaissance se présente comme nous le montre la table 4.13. Les résultats obtenus pour /ṭ/ et /ṣ/ sont encourageants par rapport à ceux de /ḍ et /ð̣/, néanmoins, pour réduire les insertions pour tous les phonèmes d'autres règles sont nécessaires.

Phonèmes	Présents	Reconnus	Insérés
/ṭ/	64	41 (64%)	7 (10%)
/ṣ/	57	52 (91%)	5 (8%)
/ð̣/	38	18 (47%)	4 (10%)
/ḍ/	28	13 (46%)	3(10%)

TABLE 4.13 – Résultat de la reconnaissance des consonnes emphatiques

Résultats du système sur les consonnes glottales et pharyngales

Il s'agit de reconnaître les consonnes : / ?/, /ε/, /$\hbar$/et /h/, à partir des connaissances acquises de notre expérience en lecture des spectrogrammes de parole.
Ainsi, la reconnaissance de la plosive / ?/ est fondée sur :

— la présence de la barre d'explosion et sa fréquence ;
— le degré de voisement ;
— les transitions formantiques des voyelles qui suivent la consonne en question.

Concernant la consonne /h/ et /ε/ nous avons utilisé :

— le degré de voisement ;
— les valeurs des formants F1 et F2 de la consonne,
— les transitions formantiques des voyelles adjacentes (gauche et droite).

Concernant la fricative /$\hbar$/ nous avons utilisé :

— le degré de voisement ;
— la limite inférieure du bruit de friction ;
— les transitions formantiques des voyelles adjacentes.

Pour l'ensemble des quatre consonnes étudiées, nous avons écrit 39 règles de production. Voici par ailleurs un exemple de règle :

```
R322
C Regle sur /h/ contexte a/aa et u/uu
CONTEXTE_DROIT[u,uu]
CONTEXTE_GAUCHE[a,aa]
SI
  ^(formant1_ACT 400 480) &
  ^(formant2_ACT 1100 1200)
PHONEMES[h 80]
FIN
```

Cette règle veut dire : Si le phonème précédent est une voyelle centrale (/a/ ou /aa/) et le phonème suivant une voyelle postérieure (/u/ ou /uu/) et si pour le segment

étudié F1 est vers 440 Hz et F2 vers 1150 Hz alors le phonème en question est fort possible qu'il soit un /h/. Nous signalons la présence simultanée de conditions sur les contextes gauches et droit.

Nous avons testé les algorithmes de segmentation sur un corpus de phrases contenant un grand nombre de de consonnes glottales et pharyngales. Les résultats que nous indiquons sont calculés par comparaison avec l'étiquetage manuel de ces phrases pour 3 locuteurs masculins.
L'évaluation de la segmentation est résumée par le table 4.14.

Phon	Nb	Plo	Voy	Fri	Son	FriVoy	PloFri	Omis	Taux
/?/	93	66	1	0	23	0	0	3	70 %
/ħ/	58	1	0	41	12	1	0	3	72 %
/ε/	59	1	2	0	43	2	0	11	73 %
/h/	29	0	2	0	17	0	0	10	58 %

TABLE 4.14 – Résultat de la segmentation des consonnes glottales et pharyngales

L'analyse des résultats nous permet de dresser quelques remarques :
— le phonème /?/, lorsqu'il n'est pas omis volontairement dans la prononciation apparaît presque toujours comme une plosive. Il se présente comme une sonnante à structure formantique en milieu intervocalique.
— la consonne /ħ/ est le plus souvent classée comme fricative. Elle est segmentée comme une sonnante, si le bruit présente pour certains locuteurs des structures formantiques.
— contrairement à la classification traditionnelle, le /h/ et le /ε/ sont considérés comme des sonnantes avec des structures formantiques. La segmentation le confirme bien.

En appliquant les règles ainsi écrites, nous obtenons les résultats de l'identification phonétique des consonnes pharyngales et glottales que nous résumons par la table 4.15.

Phonèmes	Présents	Reconnus	Insérés
/?/	93	47 (50%)	6 (6%)
/ħ/	58	46 (79%)	7 (11%)
/ε/	59	27 (46%)	5 (8%)
/h/	29	14 (49%)	4 (12%)

TABLE 4.15 – Résultat de la reconnaissance des consonnes pharyngales et glottales

Résultats du système sur les consonnes vélaires et uvulaires

Il est question, cette fois-ci de reconnaître les consonnes : /k/, /q/, /χ/ et /γ/ à partir des connaissances que nous avons acquises de notre propre expérience dans la lecture des spectrogrammes de parole. Ainsi, pour la reconnaissance des plosives /k/ et /q/ nous nous sommes basés sur :

— la présence du burst, son centre de gravité ;
— le degré de voisement ;
— les transitions formantiques des voyelles qui précèdent et celles qui suivent les consonnes en question.

Concernant la fricative /χ/ nous avons utilisé :

— le voisement ;
— la limite inférieure du bruit de friction ;
— les transitions formantiques des voyelles adjacentes.

Pour la consonne /γ/, en plus du degré de voisement, des transitions formantiques des voyelles adjacentes. Nous avons utilisé les valeurs des formants F1 et F2 des consonnes eux mêmes quand elles apparaissent comme sonnantes. Pour les six consonnes, nous avons pu écrire, jusqu'à présent, 34 règles de production. Nous avons testé les algorithmes de segmentation du système sur un corpus constitué pour l'occasion et qui contient un grand nombre de consonnes vélaires et uvulaires. Les résultats obtenus pour trois locuteurs et calculés par comparaison avec la segmentation manuelle de ce corpus, sont donnés par la table 4.16.

Phon	Nb	Plo	Voy	Fri	Aut	FriVoy	PloFri	Omis	Taux
/k/	310	291	0	2	1	0	1	5	94 %
/q/	367	341	0	0	7	0	2	17	93 %
/χ/	65	0	1	60	1	0	0	3	92 %
/γ/	135	2	5	33	85	3	0	7	63 %

TABLE 4.16 – Résultat de la segmentation des consonnes vélaires et uvulaires

Les plosives non voisées /k/ et /q/ sont très bien segmentées. Les rares omissions se produisent lorsque deux plosives se suivent, dans ce cas le système rend un seul segment.

Le /χ/ est une fricative, elle est bien segmentée même si parfois, elle se présente comme ayant une structure formantique.

Bien que définie comme fricative, /γ/ apparaît plutôt comme sonnante avec des structures formantiques. Elle est rarement vue comme une fricative.

Pour l'identification, les trois phonèmes qui totalisent le meilleur score sont pris

pour étiquettes du segment. Le pourcentage de reconnaissance est résumé par la table 4.17.

Phonèmes	Présents	Reconnus	Insérés
/k/	310	272 (88%)	73 (19%)
/q/	367	261 (71%)	70 (16%)
/χ/	65	47 (73%)	10 (13%)
/γ/	135	74 (55%)	13 (9%)

TABLE 4.17 – Résultat de la reconnaissance des consonnes vélaires et uvulaires

Les résultats obtenus pour /k/ et /χ/ sont encourageants par rapport à ceux de /q/ et /γ/ et dans l'ensemble, les taux de reconnaissance sont satisfaisants.

7 Conclusion

Nous avons présenté dans ce chapitre les différentes phases du décodage acoustico-phonétique de l'Arabe et les résultats obtenues jusqu'à présent. L'amélioration du taux de reconnaissance passe nécessairement par une étude plus approfondie ou expertise de la phonétique. La prochaine étape du travail sera l'intégration du décodeur phonétique dans un système de reconnaissance et/ou de compréhension de phrases en langage naturel. Nous développerons les idées directrices dans le chapitre suivant.

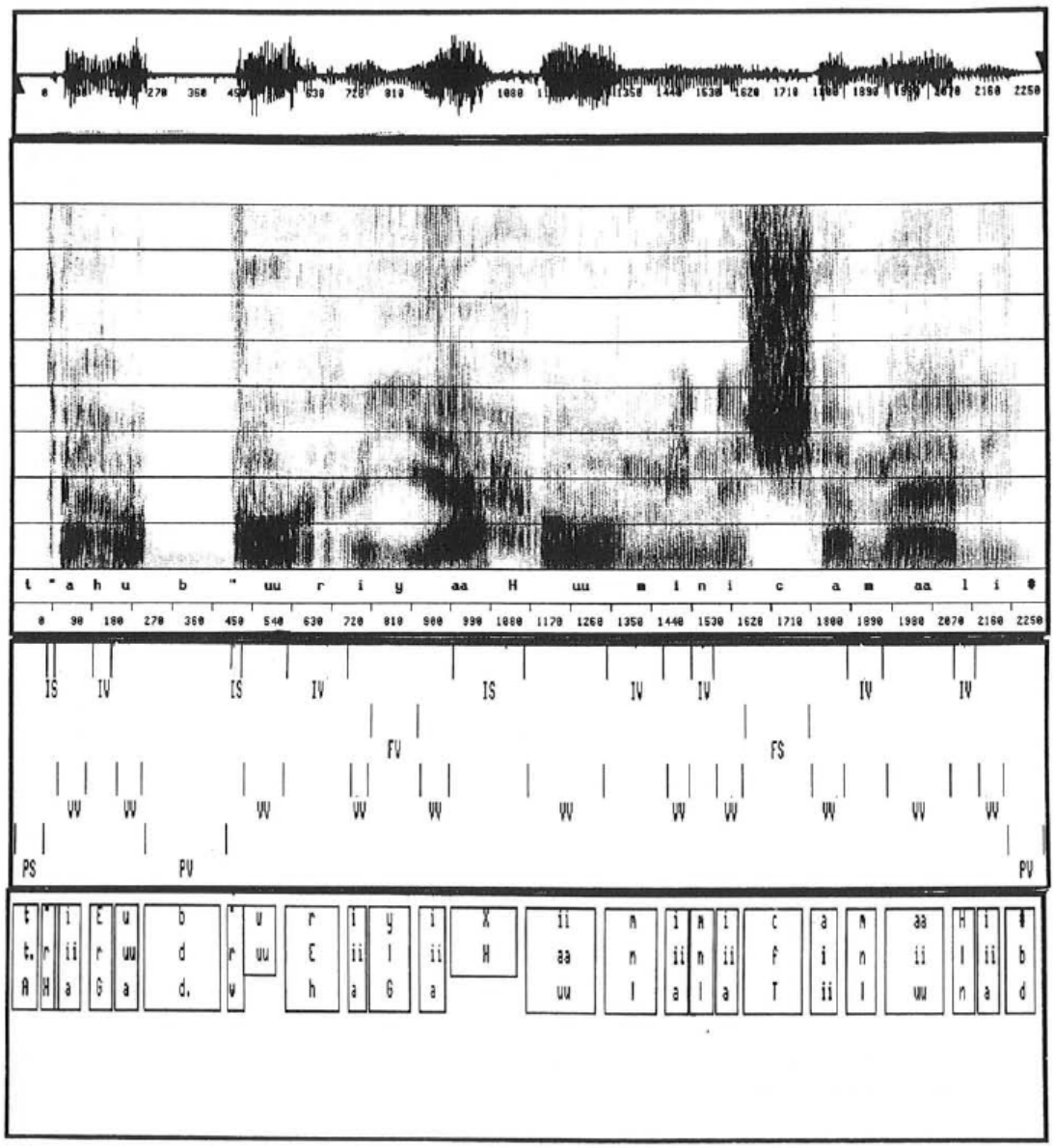


FIGURE 4.7 – exemple d'étiquetage

Chapitre 5

La reconnaissance automatique

1 Introduction

Dans ce chapitre, nous présentons une démarche d'intégration du décodeur acoustico-phonétique dans un système de reconnaissance de phrases utilisant divers sources de connaissances : phonétique, phonologique, morphologique, syntaxique, sémantique et prosodique. Avant de décrire le système de reconnaissance, nous présentons une brève introduction à une étude morphologique et grammaticale de l'Arabe standard indispensable pour la suite de l'exposé. Le lecteur trouvera des études détaillées dans [Fleisch 61] et [Benhamouda 83]

2 La morphologie l'Arabe

La morphologie a pour tache d'étudier les mots considérés isolément sous le double aspect de leur nature propre et des variations qu'ils peuvent subir. Les parties les plus importantes de la morphologie concernent la formation et la conjugaison des verbes, la formation des noms et leurs flexions, l'étude des pronoms et des particules.

2.1 Les éléments morphologiques

L'Arabe possède un système morphologique régulier qui n'a pas beaucoup changé depuis plusieurs siècles. La langue accorde aux consonnes une importance capitale dans la constitution des mots. Elle offre un système complet basé sur la notion de la racine qui constitue ainsi le pilier du dictionnaire arabe. A cette racine est attachée une idée générale qui se propage dans les formes qui en dérivent.

La racine

Le système morphologique de l'Arabe utilise la notion de racine. Une racine est un mot primitif composé de 2, 3, 4 ou 5 consonnes radicales.

— Un mot composé de 2 consonnes est dit bilitère :
exemple /fam/ : (bouche).

— Un mot composé de 3 consonnes est dit trilitère :
exemple /kataba/ : (écrire).

— Un mot composé de 4 consonnes est dit quadrilitère :
exemple /ɛaqrab/ : (scorpion).

— Un mot composé de 5 consonnes est dit quinquilitère :
exemple /safarzal/ : (coing).

Selon une étude récente faite par Mrayati [Mrayati 87] sur 5 grands dictionnaires, la répartition des racines dans la langue est la donné par la table 5.1.

Type de racine	Nombre	Pourcentage
Bilitère	115	1.3%
Trilitère	7198	63.4%
Quadrilitère	3739	33.0%
Quintilitère	295	2.3%
Total	11347	100%

TABLE 5.1 – Répartition des racines dans la langue

Les racines sont donc purement consonantiques et sont remarquables par leur fixité. A chaque racine est attachée une idée générale plus au moins précise. La réalisation de cette idée en mots autonomes se fait par le jeu des voyelles à l'intérieur de cette racine (alternance vocalique). Il n'y a d'exception que pour les pronoms et quelques particules.

La formation des mots

La formation du mot ou dérivation se fait par modification de voyelles ou par annexion de lettres additionnelles qui ajoutent une idée accessoire à l'idée primitive exprimée par la racine. Les lettres additionnelles servent à former les temps, les

personnes, les nombres, les genres, les noms de lieu et de temps etc. Ces lettres au nombre de 10 sont /s/, /?/, /l/, /t/, /m/, /n/, /w/, /y/, /h/ et le /?alif/. L'augmentation se fait par l'adjonction de préfixes, suffixes ou infixes.

Les schèmes

Tous les mots arabes, à l'exception des particules peuvent se rapporter à un certain nombre de schèmes, c'est à dire de types en formes caractérisant toute une classe de noms ou de verbes. Pour former le schème d'un mot, il suffit de remplacer les consonnes radicales de ce mot par les trois consonnes du groupe /f/, /ε/ et /l/ ou les quatre consonnes du groupe /f/, /ε/, /l/ et /l/ dans lesquels /f/ représente la première radicale, /ε/ la deuxième, /l/ la troisième et /l/ éventuellement la quatrième, et d'y ajouter les même voyelles et les mêmes lettres additionnelles. Par mesure de commodité, on préfère utiliser la représentation 123 pour schématiser une racine trilitère et 1234 une racine quadrilitère. Ces chiffres représentent donc les 3 ou les 4 consonnes d'une racine réelle. On parle alors de schèmes 1a2a3, 1a2a3a4, 1aa2i3, 1u22aa3, etc et on dit que le mot /kaatib/ a pour schème 1aa2i3.

La flexion

La flexion est la modification que subit un mot dans sa terminaison suivant le rôle qu'il joue dans la proposition. C'est donc une information grammaticale indiquant la fonction des mots dans la phrase qui peut être obtenue par l'analyse morphologique. Elle est purement casuelle sans relation avec le genre et le nombre. Les mots en arabe sont soit flexibles ou à flexion, c'est à dire que la fin du mot varie selon le contexte, soit inflexibles ou construits, dans ce cas, la fin du mot reste inchangée quelque soit son contexte.
Sont flexibles les verbes à l'inaccompli, les adjectifs et la grande partie des noms. Sont construits les verbes à l'accompli, les pronoms et conjonctions. L'Arabe pratique la flexion à trois cas, pour la conjugaison du verbe et pour la déclinaison du nom. Pour les verbes, ce sont l'indicatif, le subjonctif et l'apocopé auquels on associe respectivement les voyelles /u/, /a/ et le /sukun/ ou absence de voyelle. Pour les noms, ce sont le nominatif pour le cas du sujet, l'accusatif pour le cas du direct et le génitif pour le cas indirect auxquels on associe respectivement les voyelles /u/, /a/ et /i/.
Pour la déclinaison, il y a lieu de considérer d'une part la notion de nombre : singulier, duel et pluriel et d'autre part la notion de détermination et indétermination.

2.2 La morphologie verbale

Le verbe et la racine

Selon le nombre des consonnes radicales, le verbe peut être trilitère (ex. /kataba/ : il a écrit) ou quadrilitère (ex. /tardjama/ : il a traduit). Les verbes trilitères forment la grande majorité des verbes, les quadrilitères sont relativement peut nombreux.

Verbe fort et verbe faible

Suivant la nature des consonnes radicales, le verbe est fort ou faible.

— Le verbe fort est celui dont la racine ne renferme que des consonnes fortes non susceptibles de disparition. Il peut être
  — sain : /saalim/ s'il ne contient ni /ʔ/, ni consonnes géminés ni de consonnes faibles (ex /kataba/ : il a écrit),
  — médial géminé ou doublé (/muḍaaɛif/ dont les deux derniers phonèmes de la racine sont semblables (ex /ʃadda/ : il a tenu)
  — ou hamzé : /mahmuuz/ dont l'une des trois radicales peut être une hamza /ʔ/ (ex /ʔakala/).
— Le verbe faible, au contraire, est celui dont la racine renferme une ou plusieurs lettres faibles susceptibles de disparaître entièrement dans la conjugaison. On distingue 4 sortes de verbe faible : l'assimilé, le concave, le défectueux et le lafif.
  — l'assimilé : /miθaal/ dont la première radicale est une lettre faible (ex /wait zada : il a trouvé),
  — le concave : /ʔaẓwaf/ dont la deuxième radicale est une lettre faible (ex /qaala/ : il a dit),
  — le défectueux : /naaqis/ dont la dernière radicale est une lettre faible (ex /maʃaa : il a marché),
  — le /lafif/ dont deux des trois radicales sont des lettres faibles.

Verbe primitif et verbe augmenté

Un verbe primitif ou à forme simple est celui qui ne renferme que les consonnes constitutives de la racine. Un verbe augmenté ou à forme augmenté est caractérisé par le redoublement d'une consonne ou par l'adjonction de lettres additionnelle ou enfin par l'allongement de certaines voyelles. L'Arabe présente 14 formes dérivées pour le verbe trilitère et 3 formes pour le quadrilitère en considérant le type primitif.

Verbe transitif et verbe intransitif

Le verbe transitif est celui qui fait passer directement l'action du sujet au complément. Le verbe intransitif est celui qui exprime une action qui s'applique au sujet et ne passe pas sur un objet.

Les voix

A l'exception de quelques formes de verbes qui n'ont pas de forme passive, tous les verbes possèdent deux voix, la voix active et la voix passive, marquées seulement par des différences de vocalisation.
Exemple :

— /kataba ?ad-darsa/ "il a écrit la lec on",
— /kutiba ?ad-darsu/ "la lec on est écrite".

Les temps ou aspects

Les aspects sont les formes particulières que prend le verbe pour indiquer à quelle époque se rapporte l'action ou l'état exprimés par le verbe. Le verbe arabe a deux aspects : l'accompli et l'inaccompli.

— L'accompli /maadhi/ indique l'achèvement d'une action ou la pleine réalisation d'un état.
— L'inaccompli indique l'inachèvement d'une action ou l'incomplète réalisation d'un état. L'inaccompli possède trois modes : l'indicatif présent et futur (/marfuuɛ)/, le subjonctif pour les subordinations (/manṣuub/) et l'apocopé pour les phrases conditionnelles ou négatives (/ma*zz*uum/).

Le verbe a aussi un impératif / ?amr/ qui exprime l'ordre ou la demande ou la prière et ne s'emploie qu'à la deuxième personne.
Un verbe est désigné par la troisième personne du masculin singulier à l'accompli.

Les nombres

Dans le verbe, il y a trois nombres.

— Le singulier ou /mufrad/ qui s'applique à une seule personne.
— Le duel ou /muthana/ qui s'applique à deux personnes.
— Le pluriel ou /djamɛ/ qui s'applique à plus de deux personnes.

Les genres

Dans le verbe, il y a deux genres : le masculin et le féminin.

Les personnes

Dans le verbe, il y a trois personnes :
— la première, celle qui parle,
— la deuxième, celle à qui on parle,
— la troisième, celle dont on parle.

2.3 La morphologie nominale

Le nom arabe est en réalité un terme générique sous lequel sont compris le substantif, l'adjectif, les pronoms, etc.

Le nom et la racine

Les noms primitifs peuvent avoir de 2 à 5 lettres radicales. Selon le nombre des ses consonnes, le nom est dit bilitère, trilitère, quadrilitère ou quinquilitère.

Nom primitif et nom dérivé

Le nom peut être primitif ou dérivé. Il est primitif quand il ne dérive d'aucun verbe ou nom. Il est dérivé quand il est formé d'un autre nom ou verbe.
Un nom dérivé d'un nom est dit dénominatif et il est dit déverbatif lorsqu'il est dérivé d'un verbe.
Certains noms dérivent soit d'un prénom, soit d'une particule mais cette dérivation appartient à une époque tardive.
Les noms dérivés se forment à partir de leur primitif soit par modification des voyelles, soit par l'addition d'une ou plusieurs des 10 lettres formatives.
Les noms dénominatifs sont le collectif, le nom d'unité, le nom d'abondance, le nom abstrait de qualité, l'adjectif de relation, le diminutif, le nom de nombre, etc.
Les noms déverbatifs comprennent le nom de manière, le nom de lieu et de temps, le nom d'instrument, le nom d'agent, le nom de patient, l'intensif, le qualificatif assimilé, etc.

De même, le nom peut être commun ou propre, concret ou abstrait, déterminé ou indéterminé.

Le genre

Le nom peut être masculin ou féminin, mais le genre grammatical peut ne pas correspondre avec le genre naturel. De même, il existe parfois un seul nom qui désigne l'individu de l'espèce, abstraction faite du sexe.

Le nombre

Les noms ont comme dans la conjugaison des verbes, 3 nombres : le singulier, le duel et le pluriel. Il y a en Arabe deux sortes de pluriels :

— le pluriel externe appelé sain parce qu'il conserve intactes les consonnes et les voyelles du singulier,

— le pluriel interne appelé brisé parce que son singulier est brisé par un ou plusieurs des cas suivants : préfixation, infixation, suffixation, modification de voyelles.

2.4 Les pronoms et les particules

Les pronoms

Le pronom est un mot qui tient la place du nom dont il prend le genre et le nombre. Le pronom personnel, en Arabe, est isolé ou affixe. Isolé, il correspond au franc ais, moi, toi, lui, etc. Affixe, il se joint à un verbe pour en marquer le complément direct, soit à un nom pour rendre le possessif, soit à une préposition. Les pronoms personnels isolés (sujet ou attribut de sujet) sont :

Les pronoms personnels affixes (compléments) sont :

Personne	Singulier	Duel	Pluriel
1	/ʔana/	/naħnu/	/naħnu/
2	/ʔanta/ fem. /ʔanti/	/ʔantuma/	/ʔantum/ fem. /ʔantunna/
3	/huwa/ fem. /hiya/	/huma/	/hum/ fem. /hunna/

TABLE 5.2 – Les pronoms personnels isolés

Les pronoms personnels compléments des noms, des verbes et des particules deviennent des enclitiques, c'est à dire qu'ils doivent être unis au mot dont ils dépendent de fac on à ne former graphiquement et phonétiquement avec lui qu'un tout

Personne	Singulier	Duel	Pluriel
1	/ii/	/naa/	/naa/
2	/ka/ fem. /ki/	/kumaa/	/kum/ fem. /kunna/
3	/hu/ fem. /haa/	/huma/	/hum/ fem. /hunna/

TABLE 5.3 – Les pronoms personnels affixes

homogène. Certaines prépositions et conjonctions formées d'une seule consonne sont proclitiques. Il existe aussi les démonstratifs, et les relatifs dits noms de conjoints et des pronoms interrogatifs.

Les particules

La particule est un mot invariable qui accompagne un nom ou un verbe. Les particules unilitères se lient graphiquement avec les mots qui les accompagnent ; les particules bilitères et trilitères constituent des mots indépendants. Les particules servent à marquer l'adjonction, la négation, l'affirmation, l'interrogation, la condition, le vocatif, la détermination, l'exception, etc.

3 La syntaxe de l'Arabe

La langue arabe est caractérisée par une grande simplicité de la syntaxe qui se contente le plus souvent de faire suivre une phrase d'une autre, sans les rendre dépendantes ou subordonnées. Elle a recours à la coordination plutôt qu'à la subordination.
Une phrase simple de l'Arabe peut être soit nominale ou verbale. Cette classification étant faite en fonction de l'absence ou la présence du verbe dans la phrase.

3.1 La phrase nominale

La phrase nominale simple

La phrase nominale est formée par le rapprochement de deux termes : le sujet (dit /mubtad ?a/ : inchoactif) et l'attribut (dit /χabar : énonciatif). Ce type de phrases sert à exprimer une définition ou énoncer un jugement. La structure de la phrase nominale peut être normale (le sujet précède l'attribut) ou inversée (l'attribut précède le sujet). Le sujet est dans la plupart des cas un nom déterminé, mais il peut être aussi un pronom personnel démonstratif ou autre.

L'attribut est simple (formé d'une seule expression) ou composé (formé d'une proposition ou d'une similiproposition composée d'une préposition avec son complément). L'attribut simple peut être un nom, un pronom personnel ou démonstratif.

La phrase nominale et les particules

La flexion commune du premier cas du sujet et de l'attribut de la phrase nominale simple peut être modifiée par des verbes ou des particules qu'on leur prépose. et qu'on appelle abrogatifs. Il existe deux catégories d'abrogatifs :

— /kaana/ et ses analogues : précédé de /kaana/, le sujet devient nom de /kaana/ et demeure au premier cas, l'attribut devient attribut de /kaana/ et se met au deuxième cas.
Les verbes de la catégorie de /kaana/ sont dits verbes incomplets.

— /ʔinna/ et ses analogues : précédé de /ʔinna/ le sujet devient nom de /ʔinna/ et se met au deuxième cas, l'attribut devient attribut de /ʔinna/ et reste au premier cas. Les particules de la catégorie de /ʔinna/ sont considérées par les grammairiens arabes comme des assimilées au verbe.

3.2 La phrase verbale

On appelle phrase verbale, toute phrase contenant au moins deux éléments : le sujet (dit /faaɛil/ : agent) et le verbe (dit /fiɛl : procès). Cette phrase exprime une action attribuée à un certain sujet, rapportée à un certain temps et dirigée s'il y a lieu, vers un certain objet. Les structures syntaxiques des phrases verbales les plus fréquentes sont Verbe + Sujet, ou bien Verbe + Sujet + Complément. Cependant, il se peut que le sujet précède le verbe.

3.3 Les règles d'accord

L'accord du verbe

L'accord du verbe avec son sujet obéit à un certain nombre de règles :

— Si le sujet est un nom désignant des êtres humains, deux cas :
  — Le verbe précède son sujet (cas général) ; l'accord se fait seulement en genre, jamais en nombre, le verbe restant au singulier.

  — Le verbe suit son sujet : il s'accorde en genre et en nombre.

— Si le sujet est un nom désignant des animés non humains ou des inanimés, deux cas :
  — Le sujet est un singulier : le verbe s'accorde en genre.

  — Le sujet est pluriel : le verbe s'accorde au féminin singulier, qu'il précède ou non son sujet.

L'accord de l'adjectif

Les règles d'accord de l'adjectif sont fonction du nom auquel il se rapporte :
— Si le nom désigne des êtres humains, l'adjectif s'accorde en genre et en nombre.
— Si le nom désigne des animés non humains ou des inanimés, l'adjectif s'accorde :
  — Avec un nom singulier, en genre.
  — Avec un nom pluriel au féminin singulier.

  L'adjectif épithète s'accorde toujours en cas et détermination avec le nom auquel il se rapporte.

4 Le système MARS

L'Arabe moderne, dit aussi littéraire ou standard a fait l'objet de plusieurs travaux anciens ou récents, ayant trait soit à l'aspect phonétique [Ani 70] [Bonnot 77] [Giannini 82] [Djoudi 89a] soit à la composante linguistique de la langue [Sibawayh 89] [Jinni 54] [Sacy 10] [Ghazali 87a]. Toutefois, le problème de reconnaissance automatique n'a été que très peu abordé jusqu'à présent [Djoudi 90b].

La reconnaissance automatique de la parole continue fait appel a divers sources de connaissances. Cette démarche conduit à la conception de systèmes complexes ayant de fortes intéractions entre les différents modules qui les composent. Le système que nous proposons a pour but la compréhension de phrases en arabe moderne, dans un contexte multilocuteur. Sa mise en œuvre tient compte bien sûr des particularités de la langue.

La structure générale du système fait apparaître deux grands sous systèmes. D'une part, un système de décodage acoustico-phonétique, en l'occurrence SAPHA et d'autre part, le décodeur linguistique SALAM. Chaque sous système utilise diverses sources de connaissances. Le décodeur phonétique fournit au système linguistique, en mode proposition, un treillis phonétique. Il peut être réactivé, en mode vérification, par le décodeur linguistique pour savoir si un phonème donné est l'éti-

quette d'un morceau de signal donné. Le décodeur linguistique part des données fournies par le système phonétique et celle provenant de la définition du langage, détermine la signification du message oral (voir figure 5.1).

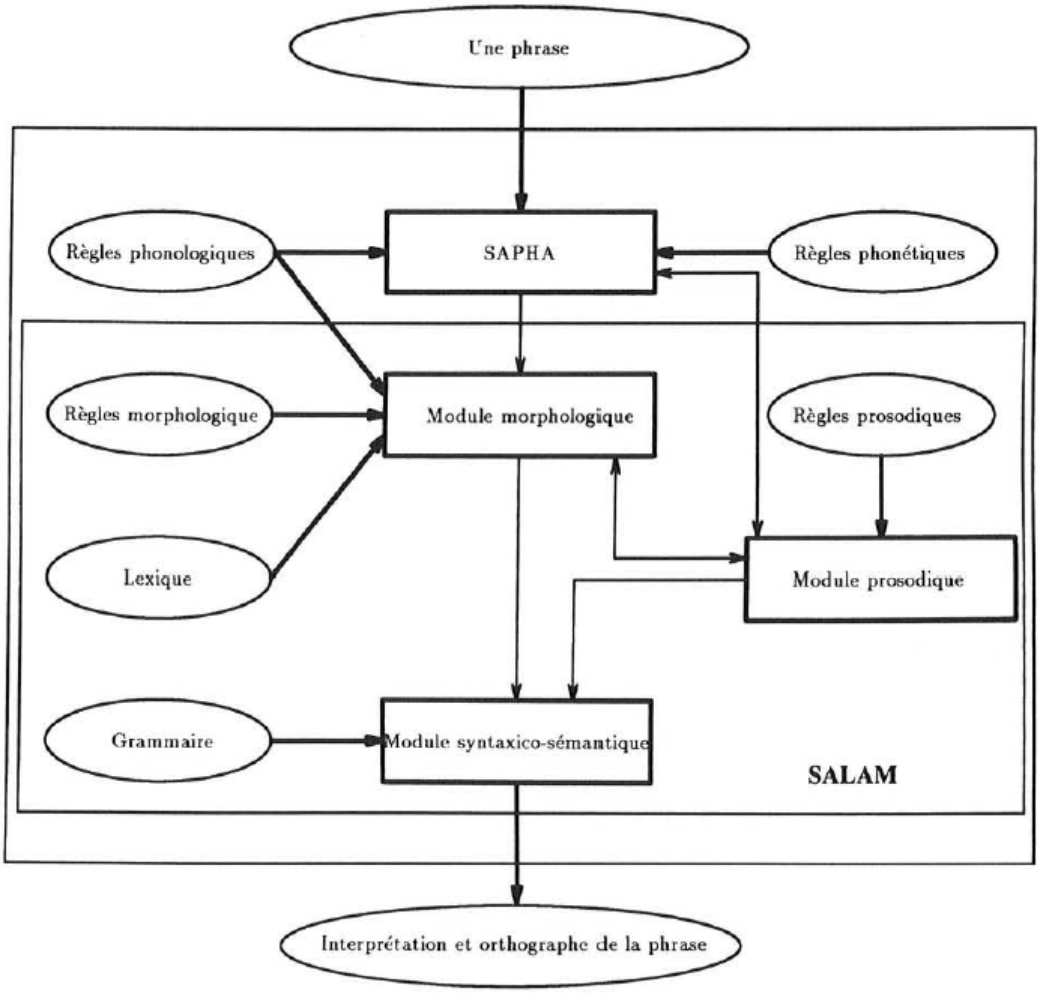


FIGURE 5.1 – Architecture de MARS

5 Le décodeur acoustico-phonétique

Ce système est considéré comme le premier maillon du processus de compréhension. Il construit la description phonétique d'un énoncé, à partir du signal acoustique numérisé. Nous rappelons que les principales étapes du système sont :

— la segmentation du signal en grandes classes phonétiques,

— le calcul des indices phonétiques pertinents pour la reconnaissance,

— et l'identification phonétique des segments.

Le résultat du système est un treillis de phonèmes, représenté par une suite de segments. A chaque segment, est associée une liste des phonèmes les plus probablement prononcés par le locuteur (voir chapitre 3).

6 Le décodeur linguistique

SALAM comme Système Approprié pour le décodage Linguistique de l'Arabe Moderne rec oit en entrée une suite de phonèmes sous forme de treillis phonétique (résultats fournis par le décodeur SAPHA) et effectue la reconnaissance de la phrase traitée. Son résultat consiste pour l'instant en l'interprétation sémantique et l'écriture de la phrase prononcée. Le décodeur linguistique SALAM comporte plusieurs modules :

6.1 Le module morphologique

L'objectif de ce module est de détecter des mots dans le treillis de phonèmes fourni en entrée, de vérifier leur appartenance à la langue et de fournir les valeurs grammaticales fixes (catégorie syntaxique, genre, nombre, temps, mode, cas, ...) Le processus peut être déclenché de différentes manières :

— **la vérification de mots :** cela se produit lorsque le module syntaxico-sémantique émet une hypothèse concernant la présence d'un mot particulier à un instant donné de la phrase. Cette hypothèse, qui constitue une donnée du module morphologique, est un couple composé :

  — d'une forme de référence représentant une transcription phonétique du mot-hypothèse

  — et d'un repère temporel indiquant le début du mot dans le treillis de phonèmes.

  Le traitement consiste à apparier la forme de référence et le sous-treillis correspondant au mot à vérifier. Le résultat est une validation ou un rejet de l'hypothèse traitée.

— **la recherche de mots :** lorsque le module ne dispose pas d'information temporelle sur un mot à valider, il procède à une recherche itérative en effectuant un balayage du treillis de phonèmes. A chaque itération, un appariement, entre le schème de référence et la partie sélectionnée dans le treillis, est tenté.

Cette opération séquentielle rend le processus très coûteux en temps de traitement. Pour réduire ce dernier, on peut s'appuyer sur les hypothèses prosodiques qui vont permettre d'éliminer les appariements avec un sous-treillis contenant un marqueur prosodique.

Les difficultés de l'analyse morphologique

Les problèmes rencontrés pendant l'analyse morphologique peuvent être de deux types, les uns sont liés à la structure radicale de la langue et les autre aux erreurs introduites par le décodage acoustico-phonétique.

Difficultés liées à la langue

- N'importe quelle racine ne peut pas être instanciée à n'importe quel schème.
- Deux racines analogues ne donnent pas lieu à des dérivations analogues.
- Cas de la racine faible.
- Cas des mots homographes, ou une même chaîne recouvre deux notions différentes suivant le contexte, exemple /fii/ verbe impératif de /wafa/ et /fii/ préposition.

Incertitudes du décodage phonétique En l'état actuel du décodeur phonétique, le treillis phonétique contient différents types d'erreurs :

- Une sur-segmentation peut provoquer la duplication de certains phonèmes.
- Une sous-segmentation risque de faire disparaître d'autres phonèmes.
- D'autre part, pour un phonème prononcé, le système SAPHA fournit plusieurs solutions possibles.
- De plus, il arrive que le phonème réellement prononcé n'apparaisse pas dans le choix proposé.

Pour résoudre les problèmes des sur-segmentation et de sous-segmentation, l'algorithme doit admettre des élisions et des insertions de phonèmes. Le résultat n'est plus un booléen indiquant la validation ou le rejet d'une hypothèse, mais un score de reconnaissance qui constitue une mesure de certitude d'apparition du mot de référence. Ce score est calculé en appliquant une pénalité à la suite de chaque élision ou insertion. Dans le cas ou' le phonème recherché n'apparaît pas parmi les choix possibles, la solution à mettre en œuvre consiste à pénaliser la reconnaissance en fonction du degré de ressemblance entre le phonème recherché et les phonèmes proposés. Le résultat de l'analyse morphologique est un ensemble de mots ; à chaque mot sont associés un score de validité et un repère sur l'axe temporel de l'énoncé. Cette structure est représentée à l'aide d'un *treillis de mots*.

Organisation du lexique

Le système morphologique rigoureux a sûrement joué un rôle important dans l'organisation originale du lexique arabe. La conception du lexique peut se faire de différentes manières et sa réalisation est déterminée par l'architecture globale du système. Deux attitudes se dégagent selon l'objectif qu'on voudrait atteindre.

— L'une consiste à construire explicitement le lexique, celui-ci doit inclure toutes les formes fléchies des mots avec leurs informations relatives. La construction d'un tel dictionnaire peut être obtenue à l'aide de deux générateurs : un conjugueur de verbes et un dérivateur de noms. La génération se fait à partir d'une liste de racines auxquelles sont associées les informations relatives à la génération. Les changements concernant l'assimilation et la dissimilation seront résolus à l'aide de règles de réécriture [Debili 85]. Le problème se pose pour les applications faisant référence à un large vocabulaire ou dans le cas de la construction d'un lexique indépendant des applications. Un autre problème se pose au niveau de la construction du lexique, il s'agit de définir les informations à associer aux unités lexicales et les méthodes de les exploiter par les niveaux morphologique, syntaxique, sémantique et phonétique.

— La seconde approche consiste à créer une partition sur l'ensemble des mots et de n'enregistrer dans les lexiques que les informations permettant de vérifier l'appartenance d'un mot à une classe d'équivalence. Chaque entrée du lexique constitue une classe d'équivalence à laquelle est associée une liste de valeurs grammaticales qui correspond à l'union des valeurs grammaticales des mots appartenant à cette même classe. Dans ce cas, l'analyse d'un mot se traduit par la recherche de préfixes, suffixes et schèmes vocaliques et puis la vérification de l'affinité entre les différents éléments. Cette idée doit, d'une part supposer la formulation des règles émises par les grammairiens est facile à mettre en œuvre et d'autre part savoir utiliser les règles et pouvoir juger leurs limites. [Andreewsky 73] et [Flhur 77] proposent d'obtenir ces lexiques par apprentissage.

Les attributs sémantiques Ce système de dérivation n'est pas seulement formel, mais il propage aussi des valeurs sémantiques.

Les objets qu'on cherche à représenter sont les entités linguistiques : mots, phrases, schèmes, racines etc... et leur description relative. La caractérisation de ces objets nous permet de définir des subdivisions dans cet ensemble d'objets. Chaque subdivision définit un univers qui sera appelé par la suite concept. Les entités linguistiques seront appelées individus. Les concepts sont définis soit en extension, en énumérant

ses individus avec leur description, c'est notamment le cas des mot-outils et des pronoms qui sortent du système de la racine ; soit en intention à partir de règles d'inférence.

La transcription phonétique-orthographique Le lexique est utilisé pour extraire la (les) représentation(s) phonétique(s) d'un mot à partir de sa forme orthographique. Le passage d'une forme à une autre doit prendre en compte certaines caractéristiques de la langue en particulier :

- La quantité de la voyelle.
- La gémination.
- La nounation.
- L'assimilation.
- La prononciation du /ta/ /marbuta/ (liée) à la fin des noms et des adjectives.
- Les mots irréguliers.

L'absence d'un clavier bilingue nous contraint de négliger cette phase et de travailler sur la structure phonétique interne du mot. Il arrive que le module syntaxico-sémantique demande la validation d'un ensemble de mots ayant la même fonction syntaxique ou possédant des traits sémantiques communs. Afin de permettre une sélection selon ces critères, il est indispensable d'affecter aux éléments du lexique des attributs syntaxiques et sémantiques.

6.2 Le module syntaxico-sémantique

La structure grammaticale de l'Arabe standard est rigide, et obéit à des règles qu'on peut formaliser. Cette rigidité n'est pas évidente lorsqu'il s'agit du langage parlé. Comme restriction, nous prenons, dans un premier temps, une grammaire artificielle qui engendre la plus grande partie des phrases. Le rôle de module syntaxico-sémantique est d'étudier les règles à suivre pour la construction des phrases, c'est à dire l'ordre dans lequel les mot doivent être disposés et les rapports grammaticaux qui doivent exister entre eux. La finalité du module est de produire une ou plusieurs structures syntaxico-sémantiques correspondant à un énoncé en langage pseudo naturel. Ce modèle a la particularité de proposer une représentation sémantique de chaque énoncé selon une structure composée d'un prédicat -dérivé d'une primitive- et d'un certain nombre d'arguments de cette primitive. Le traitement syntaxique s'appuie sur une base de connaissances sous forme de règles de production. Ce traitement construit une structure syntaxique intermédiaire. Le passage à une représentation conforme au modèle syntaxico-sémantique., est assuré par l'application des règles de correspondance.

6.3 Le module prosodique

Ce module détient un double rôle dans le système :

— il contribue, d'une part, à déterminer la signification de l'énoncé par l'émission des hypothèses concernant la forme de la phrase : Cette information vient en complément des données destinées au processeur syntaxico-sémantique qui s'en sert pour lever certaines ambigui"tés.

— il s'efforce, d'autre part, de localiser des marqueurs prosodiques qui indiquent, sur le plan morphologique, une frontière de mot et, sur le plan syntaxique, une frontière de syntagme. Ainsi, les hypothèses morphologiques ou syntaxiques, qui entrent en contradiction avec les informations prosodiques, peuvent être abandonnées par le processeur correspondant. Par conséquent, le module prosodique doit retenir uniquement les marqueurs non-ambigus, c'est à dire ceux qui sont détectés avec un taux de certitude suffisamment élevé. Le travail revient à détecter la position de l'accent dans le mot et utiliser les règles qui régissent la présence de l'accent dans les mots arabes. Les paramètres acoustiques utilisés sont la durée, l'énergie et surtout l'évolution de la fréquence fondamentale.

En Arabe, il existe cinq types de phrases :

Les phrases déclaratives : ce type de phrases est le plus fréquent et pour qui la prononciation est un modèle pour le pitch.

Les commandes : il s'agit d'un ordre, ou‘ la phrase est centrée autour du verbe conjugué comme à l'impératif. Le pitch est élevé au niveau du verbe même.

Les questions : le pitch est plus élevé au niveau du mot de la question. Ce dernier se trouve généralement au début de la phrase.

Les appels : ils sont constitués de l'article d'appel (/ya/ ou /Aya/) suivi de l'appelé. Les phrases sont généralement courtes et le pitch est pratiquement le même que pour les phrases déclaratives.

Les exclamations : le pitch est plus élevé à la fin du mot principal de la phrase.

Dans une étude sur la prosodie de l'Arabe en synthèse de la parole faite sur des phrases affirmative, il se dégage les effets intonatifs suivants [Esskalli 87] :

— L'attaque intonative d'une phrase se situe sur la dynamique de base (120 Hz).

— Chaque mot de la phrase conserve son accent de l'état isolé.

— Il existe deux classes de schémas intonatifs qui se réalisent en fonction de la structure syntaxique de la phrase.

— Toutes les phrases ont un schéma intonatif final descendant.

7 Conclusion

Dans ce chapitre, nous avons présenté comment s'effectue l'intégration du système de décodage acoustico-phonétique dans un système de compréhension de phrases. Les problèmes liés à la particularité de la langue ont été soulevés. La prochaine étape consiste à rendre opérationnels les modules du traitement linguistique et établir une communication entre les niveaux inférieur et supérieur du système de reconnaissance.

Conclusion et perspectives

Ce travail constitue une contribution à la reconnaissance automatique de la parole continue en Arabe standard dans un contexte multilocuteur.
Nous avons, dans un premier temps effectué une étude phonétique et phonologique de la langue en se basant particulièrement sur l'examen de spectrogrammes de mots et de phrases et en tenant compte des différents contextes de production des phonèmes. Cette étude nous a permis de dégager les caractéristiques propres à la langue, en particulier :

- l'existence d'une opposition temporelle brève-longue des voyelles. La durée de la voyelle a une importance capitale et elle dépend du contexte de production.
- la présence des consonnes pharyngales, glottales, emphatiques et uvulaires. Ces consonnes possèdent chacune un lieu d'articulation à l'arrière de l'appareil phonatoire. Cette caractéristique se manifeste par une influence particulière sur les voyelles adjacentes.
- l'existence de phénomènes phonologiques très importants telle que la nounation, l'assimilation et la gémination. Ces phénomènes ont une incidence sur la phonétique, la morphologie et la syntaxe de la langue.
- la détermination de la structure syllabique de langue et la place de l'accent par rapport à cette structure.

Cette étude nous a permis de cerner les difficultés dûes à la langue lors du décodage phonétique. Ensuite, nous avons présenté l'architecture générale du système de décodage acoustico-phonétique SAPHA, les outils d'analyse. Nous étions amené dans un premier temps à constituer le corpus DJOUMA de 50 phrases phonétiquement équilibrées prononcées par 11 locuteurs. Ce corpus nous a servi à effectuer une analyse phonétique et ensuite à faire une évaluation des performances du système de décodage phonétique. L'étiquetage manuel et le stockage d'une partie du corpus sur disque est donc nécessaire. Pour connaître la répartition des phonèmes, nous avons effectué une analyse statistique sur le corpus DJOUMA. Le système a été développé autour du logiciel de traitement de la parole Snorri dont nous avons présenté les fonctions d'acquisition et d'affichage de spectrogramme et des courbes.

L'étape de décodage phonétique proprement dite comporte les phases suivantes :

— la segmentation du signal de parole en grandes classes phonétiques (voyelles, plosives, fricatives et sonnantes) en utilisant des algorithmes non contextuels,

— l'extraction des indices phonétiques pertinents utilisés en reconnaissance,

— l'étiquetage des segments utilisant un système à base de connaissances adapté du système expert APHODEX développé au sein de notre équipe pour le décodage acoustico-phonétique du Frančais. Le formalisme adopté pour la représentation des connaissances étant les règles de production.

En l'absence d'un expert phonéticien, nous étions amené à construire nous-même la base de connaissances en s'appuyant sur l'expérience acquise lors de l'étude phonétique. L'évaluation des performances du système a été effectuée à partir de l'étiquetage manuel des phrases phonétiquement équilibrées du corpus DJOUMA pour trois locuteurs masculins. Nous avons évalué les algorithmes de segmentation du signal en classes phonétique et le décodage phonétique par le système SAPHA. Cette évaluation nous permis de connaître le taux global de reconnaissance en monolocuteur comme en multilocuteur et de déceler les insuffisances du système.

Comme nous l'avons mentionné, une grande particularité de la langue réside dans la présence des consonnes arrières. Nous avons travaillé sur ces consonnes avec trois stagiaires venus d'Algérie pour effectuer des stages de Magister.

Avec le premier nous avons étudié les consonnes pharyngales (/ħ/ et /ε/) et glottales (/?/ et /h/). Nous avons pu remarquer la forte dépendance de la structure acoustique de ces phonèmes en fonction des contextes. Nous étions ainsi amené à utiliser dans la même règle les deux contextes gauche et droit.

Avec le second nous avons touché au problème délicat de l'emphase. Nous avons analysé la structure des consonnes emphatiques (/ṭ/, /ḍ/, /ṣ/ et /ð̣/) et leur influence sur les phonèmes adjacents. En comparant les consonnes emphatiques à leurs homologues non emphatiques, nous pouvons affirmer que l'emphase modifie la structure acoustique de la consonne et influe considérablement sur les voyelles précédentes et suivantes et sur les consonnes suivantes du mot. Cette influence se manifeste par le changement de la position de la fréquence du burst des plosives, le déplacement de la limite inférieure du bruit des fricatives ou bien des formants des sonnantes.

Nous avons aussi examiner de plus près les consonnes vélaires (/k/ et /χ/) et uvulaires (/q/ et /γ/). A l'exception du /q/, ces consonnes agissent de la même fačon sur les voyelles adjacentes. Le /q/ influe sur les voyelles adjacentes exactement comme les consonnes emphatiques mais cette influence ne s'étend pas au delà des phonèmes voisins.

Les résultats de ces travaux nous ont permis d'élaborer de nouvelles règles phoné-

tiques et d'enrichir donc notre base de connaissances.

La question "Comment intégrer le décodeur phonétique dans un système de reconnaissance de phrases" nous a amené à faire une étude morpho-syntaxique de la langue, de proposer une architecture d'un système de reconnaissance et de relever les difficultés à surpasser sur le plan linguistique et prosodique.

Perspectives

L'analyse des résultats du décodage phonétique de l'Arabe nous a permis d'expliquer certaines erreurs. Les principales causes sont :

— Les erreurs de segmentation, en particulier lorsque deux phonèmes appartenant à une même classe, se suivent dans la phrase, le système rend un seul segment. Ce cas est relativement fréquent dans la classe des sonnantes. Pour pallier à cette insuffisance, nous proposons d'utiliser la segmentation par dendogrammes. Cette méthode consiste en une segmentation multiple du signal de parole [Hajislam 90].

— Les procédures de calcul des indices phonétiques ne fournissent pas toujours les valeurs correctes. Un effort supplémentaire doit être consenti dans cette direction pour proposer des nouveaux algorithmes plus efficaces. Il fallait penser aussi à utiliser d'autres indices phonétiques comme l'énergie du bruit du friction ou la visibilité des formants des sonnantes.

— La base de règles est insuffisante. Il reste un certain nombre de règles à écrire ou à modifier. Une expérience dans le domaine de la lecture des spectrogrammes s'impose pour vérifier la cohérence de la base. Il est nécessaire d'étudier avec plus de détail certains phonèmes pour mieux les reconnaître.

Afin d'améliorer le taux global de reconnaissance, nous pensons mêler une autre approche à base de modèles markoviens et des réseaux de neurones. Par la suite, il faudra réaliser les modules du décodeur linguistique et intégrer le système de décodage dans un système de reconnaissance et/ou de compréhension de phrases en Arabe standard.

ANNEXE : Le corpus DJOUMA

SERIE A

تهب الرياح من الشمال Les vents soufflent du nord	1
لقد إحتفظوا بعادات إفطارهم Ils ont garde leurs traditions de manger	2
غضب الأب على إبنه الصغير Le pere s'est mis en colere contre son petit enfant	3
كثرة الكلام تميت القلب Trop parler tue le cœur	4
أصبحت الفتاة من أبرز السيدات La fille est devenue l'une des plus celebres dames	5
تحول الموقع إلى محطة للقطار Le lieu s'est transforme en gare ferroviere	6
وجدت نفسها في عالم الإقتصاد Elle s'est retrouvee dans le domaine de l'economie	7
قرأت في المجلة خبرا سارا J'ai lu dans le magazine une bonne nouvelle	8
إن الشاب في حالة جيدة Le jeune est en bonne forme	9
كان يود دراسة الطب Il voulait poursuivre des etudes de medecine	10

SERIE B

حـاول دائـمًا إستغـلال الوقـت Essaye toujours de profiter du temps	11
تـمـكـن فـريـقنـا من الـفـوز بالمـقـابلةِ Notre equipe a pu gagner le match	12
كـنـت أسيـر بجـوار قـصر عـظيـم Je marchais a proximite d'un fabuleux chateau	13
تـسـلـلـتْ إلي عالـمِ الفن فتـاة جـمـيـلـة Une jolie fille s'est infiltree dans le monde artistique	14
لـقـد أدرك أنـه غـلـب عـلى أمـره Il s'est rendu compte qu'il n'avait plus le choix	15
أجـرى الطـبـيـب عـمـلـيـة جـراحـيـة Le medecin a effectue une operation chirugicale	16
خـطـرتْ لي أفـكـار كـثـيـرة Il m'est venu a l'esprit plusieurs idees	17
غـدا سيـكـون الطـقـس بـاردًا Demain, il fera froid	18
الـقـط تـحـت المـنـضـدة Le chat est sous la table	19
إبـتـعـدتُ عـن ذكـريـات شـبـابي Je me suis eloigne de mes souvenirs d'enfance	20

SERIE C

سقطت الأمطار بالمناطق الوسطى Il a plu dans les regions du centre	21
تقوم الحياة العائلية على المودة La vie familliale est basee sur l'affection	22
كن متفائلا في تحقيق مشاريعك Sois optimiste dans la realisation des tes projets	23
لقد عاشت في منزل صغير Elle a vecu dans une petite maison	24
عثر الشرطي على السيارة المسروقة Le policier retrouve la voiture volee	25
الشمس كانت جد ساخنة Le soleil etait trop chaud	26
حضر الوالد بعد إنتهاء الحفل Le pere est arrive apres la fin de la ceremonie	27
العلم نور والجهل ظلام La science est lumiere, l'ignorance une obscurite	28
لماذا ناديتني البارحة Pourquoi m'as tu appele hier ?	29
رأى الصبي طائرة تحلق في السماء Le gamin a vu un avion qui vole dans le ciel	30

SERIE D

علم الصياد بوجود الغزال Le chasseur a pris connaissance de la presence de gazelles	31
كان يحلم بمكان بعيد Il revait d'un endroit lointain	32
تمكن السائق من شق طريقه Le chauffeur a pu suivre son chemin	33
لدى الحيوانات حاسة غامضة Les animaux sont dotes d'un sens misterieux	34
أجرى المعلم إمتحانا لتلاميذه L'instituteur a passe un examen a ses eleves	35
كان الجد نائمًا في الفراش Le grand pere dormait dans son lit	36
لم أكتب حرفًا واحدًا في الرواية Je n'ai pas ecris un seul mot dans le roman	37
إتصلت بدائرة شؤون الموظفين J'ai contacte le departement du service du personnel	38
الإحساس بضيق الوقت يرهق الحياة La sensation d'avoir peu de temps est fatigante	39
جثم الصياد في وضعه المختار Le pecheur est reste dans sa position preferee	40

SERIE E

باتت السماء صافية زرقاء Le ciel etait clair	41
أحاط بالمكان رعبٌ كبير Une grande terreur regnait dans le site	42
إحتضنت المدينة معرضا دوليا La ville a abrite une foire internationale	43
إكتشف الحيوان قرب حدوث الخطر L'animal a pressente l'approche du danger	44
كثر الضجيح في شوارع المدينة Le bruit s'est accru dans les rues de la ville	45
كان يحفظ حكاية مسلية Il apprenait une drole d'histoire	46
لايزال المغامر يجول في العالم L'aventurier continue a voyager a travers le monde	47
ها قد وصل الطبيب الشرعي! Voila, le medecin legiste qui arrive!	48
أعاد الولد السنة الدراسية L'enfant a redouble son annee scolaire	49
لا تتردد في تنفيذ أعمالكَ N'hesite pas a realiser tes projets	50

Bibliographie

[Adem 83] S. A. Adem. *Are The Emphatic Consonnats of Egyptian Arabic Rounded ?* Rapport technique 11, Reports from uppsala University, Dept. of Linguistics (RUUL), 1983.

[Andreewsky 73] A. Andreewsky. Apprentissage, analyse automatique du langage, application à la documentation. Document de linguistique quantitative, Paris, 1973.

[Ani 70] S. H. Al. Ani. Arabic Phonology. An Acoustical and Physiological Investigation. Mouton & Co N.V., 1970.

[Ani 83] S. H. Al. Ani & M.S. EL Dalees. *Tafkhim in Arabic : The Acoustic and Psychological Parameters.* Rapport technique IIB, Abstracts of the Tenth Intrenational Congress of Phonetic Sciences, 1983.

[Baker 75] Baker. *Stochastic Modelling for Automatic Understanding.* In Speech Recognition, pages 51–58. R. Reddy editor, New York, Academic Press, 1975.

[Belkaid 84] Y. Belkaid. *Les Voyelles de l'Arabe littéraire moderne. Analyse spectrographique.* Rapport technique 16, Travaux de l'institut de phonétique de Strasbourg, 1984.

[Benhamouda 83] A. Benhamouda. Morphologie et syntaxe de la langue arabe. Société Nationale d'édition et de Diffusion, 1983.

[Benkirane 87] T. Benkirane & C. Cavé. *Hié rarchie de sonorité et segmentation syllabique dans le parler arabe marocain.* In Actes des $16^{ème}$ Journées d'Etudes sur la Parole, pages 274–277, Hammamet, Tunisie, Octobre 1987.

[Bonnot 77] J. F. Bonnot. *Recherche expérimentale sur la nature des consonnes emphatiques de l'Arabe classique.* Rapport technique 9, Travaux de l'institut de phonétique de Strasbourg, 1977.

[Bonnot 79] J. F. Bonnot. *Recherche expérimentale de certains axpects de la gémination et de l'emphase en Arabe.* Rapport technique 11, Travaux de l'institut de phonétique de Strasbourg, 1979.

[Carbonell 86] N. Carbonell, J. P. Haton, D. Fohr, F. Lonchamp & J. M. Pierrel. *APHODEX, Design and Implementation of an Acoustic-Phonetic Decoding Expert System.* IEEE International Conference on Acoustics, Speech and Signal Processing, 1986.

[Cohen 69] D. Cohen. *Sur le statut phonologique de l'emphase en Arabe.* Rapport technique 3, Word, 1969.

[Colmerauer 77] A. Colmerauer. *Programmation en logique du premier ordre.* In Actes journée de la compréhension, I.N.R.A., 1977.

[Combescure 81] P. Combescure. *Vingt listes de dix phrases phonétiquement équilibrées.* Revue d'Acoustique, vol. 14, no. 56, 1981.

[Datta 90] S. Datta & M. Al Zabibi. *Discrimination of Words in Large Vocabulary Speech Recognition System.* In IEEE International Conference on Acoustics, Speech and Signal Processing, pages 249–251, Mexico, April 1990.

[Debili 85] F. Debili & L. Zouari. *Analyse morphologique de l'Arabe écrit voyellé ou non fondée sur la construction automatisée d'un dictionnaire arabe.* In proceedings of Cognitiva'85, pages 87–93, Paris, 1985.

[Djoudi 86] M. Djoudi. *Détection et localisation de la barre d'explosion en parole continue et dans un contexte multilocuteur.* Rapport de D.E.A, Centre de Recherche en Iinformatique de Nancy, 1986.

[Djoudi 89a] M. Djoudi. *Etude phonétique de l'Arabe standard.* Rapport technique 89-R-057, Centre de Recherche en Informatique de Nancy, 1989.

[Djoudi 89b] M. Djoudi, D. Fohr & J. P. Haton. *Phonetic Study for Automatic Recognition of Arabic.* In Proceedings of European Conference on

Speech and Technology, volume 2, pages 268–271, Paris, September 1989.

[Djoudi 89c] M. Djoudi, D. Fohr & J. P. Haton. *SAPHA : Un Système expert pour le décodage acoustico-phonétique de l'Arabe standard.* In Première Conférence Maghrébine sur le Génie logiciel et l'Intelligence Artificielle, Constantine, Septembre 1989.

[Djoudi 90a] M. Djoudi, H. Aouizerat & J. P. Haton. *Phonetic Study and Recognition of Standard Arabic Emphatic Consonants.* In 1990 International Conference on Spoken Language Processing, Kobe, Japan, 18-22 November, 1990.

[Djoudi 90b] M. Djoudi, D. Fohr & J. P. Haton. *MARS : Un Système de Reconnaissance de l'Arabe Moderne.* In Actes des $18^{ème}$ Journées d'Etudes sur la Parole, pages 217–221, Montréal, Mai, 1990.

[Djoudi 90c] M. Djoudi & J. P. Haton. *The SAPHA Acoustic Phonetic Decoder System for Standard Arabic.* In 1990 International Conference on Spoken Language Processing, Kobe, Japan, 18-22 November, 1990.

[Djoudi 91] M. Djoudi. *Utilisation des techniques d'intelligence artificielle pour le décodage acoustico-phonétique de l'Arabe standard.* In First Maghrebin Symposium on Programming and Ssystems, Alger, Octobre 1991.

[Duda 82] K. O. Duda. *The PROSPECTOR Consultation System.* Rapport technique Final Report, SRI 8172, 1982.

[Esskalli 87] L. Esskalli, M. Rajouani, M. Najim, M. Zyoute & D. Chiadmi. *Eléments d'un modèle intonatif de la phrase affirmative en Arabe.* In Actes des $16^{ème}$ Journées d'Etudes sur la Parole, pages 282–285, Hammamet, Tunisie, Octobre 1987.

[Fleisch 61] H. Fleisch. Traité de philologie arabe : Préliminaires, phonétique, morphologie nominale, volume 1. Imprimerie Catholique Beyrouth, Beyrouth, 1961.

[Flhur 77] C. Flhur. *Algorithme à apprentissage et traitement automatique des langues.* Thèse d'Etat, Université de Paris sud, 1977.

[Fohr 86] D. Fohr. *APHODEX : Un système expert en décodage acoustico-phonétique de la parole continue.* Thèse de Doct. Univ. de NANCY 1, 1986.

[Ghazali 87a] S. Ghazali. *Elements of Arabic Phonetics.* In Applied Arabic Linguistics and Signal & Information Processing, pages 51–58. Hemisphere publishing corporation, 1987.

[Ghazali 87b] S. Ghazali. *Etude EMG préliminaire sur les consonnes arrière de l'Arabe.* In Actes des $16^{ème}$ Journées d'Etudes sur la Parole, pages 286–289, Hammamet, Tunisie, Octobre 1987.

[Giannini 82] A. Giannini & M. Pettorino. The Emphatic Consonants in Arabic. Giardini editori e stampatori, 1982.

[Gillet 84] Gillet & et al. *SERAC : Un système expert en reconnaissance acoustico-phonétique.* Actes du $4^{ème}$ congrès AFCET Reconnaissance des Formes et Intelligence Artificielle, 1984.

[Gong 83] Y. Gong. *Conception et Réalisation d'un Système de Transformée de Fourier Rapide (FFT).* Rapport du dea génie electrique et instrumentation, Département d'Electronique, Université de Pierre et Marie CURIE (Paris VI), Oct, 1983.

[Gong 85] Y. Gong. *Introduction au Traitement du Signal pour la Représentation Paramétrique en Reconnaissance Automatique de la Parole.* Cours dea informatique, Département Mathématique Appliquée et Informatique, Université de Nancy I, 1985.

[Guerti 87] M. Guerti. *Contribution a la synthese de la parole en Arabe standard.* In Actes des $16^{ème}$ Journées d'Etudes sur la Parole, pages 290–293, Hammamet, Tunisie, Octobre 1987.

[Hajislam 90] R. Hajislam. *Amélioration de la segmentation dans APHODEX.* Rapport de D.E.A, Centre de Recherche en Iinformatique de Nancy, 1990.

[Jakobson 72] R. Jakobson. Preliminaries to speech analysis, the distinctives features and their correlates. MIT Press, 1972.

[Jinni 54] Ibn Jinni. Sirr sinaaeat al ieraab. Mustapha Al Halabi, 1954.

[Kayser 84] D. Kayser. *Représentation les connaissances : Pourquoi ? Comment ?* In Actes du séminaire "Dialogue homme-machine à composante orale", pages 155–174, 1984.

[Korichane 87] D. Korichane & F. Wioland. *De quelques aspects rythmiques de l'Arabe dialectal tunisien.* In Actes des 16ème Journées d'Etudes sur la Parole, pages 294–295, Hammamet, Tunisie, Octobre 1987.

[Laprie 88] Y. Laprie. *Notice d'utilisation de Snorri.* Rapport technique, Centre de Recherche en Informatique de Nancy, 1988.

[Laurière 82a] J. L. Laurière. *Représentation et utilisation des connaissances.* TSI, vol. 1, no. 2, pages 109–139, 1982.

[Laurière 82b] J. L. Laurière. *Représentation et utilisation des connaissances. Les systèmes experts.* TSI, vol. 1, no. 1, pages 25–42, 1982.

[Marçais 48] P. Marçais. *L'articulation de l'emphase dans un parler arabe maghrébin.* Annales de l'institut d'études orientales, 1948.

[Markov 54] A. Markov. The Theory of Algorithm. US Dpt of Commerce, 1954.

[Meloni 82] H. Meloni. *Contribution à la recherche sur la recoonaissance automatique de la parole continue.* Thèse de Doctorat d'état es sciences, université d'Aix-Marseille, 1982.

[Meyer 78] B. Meyer & R. Baudois. Méthodes de programmation. Eyrolles, 1978.

[Minsky 75] M. Minsky. A Framework for Representing Knowledge, pages 211–277. Mc Graw Hill, 1975.

[Mouradi 87] A. Mouradi. *Validité et limites du diphone en tant qu'unité de synthèse pour la langue arabe standard.* In Actes des 16ème Journées d'Etudes sur la Parole, pages 296–297, Hammamet, Tunisie, Octobre 1987.

[Mrayati 87] M. Mrayati. *Statistical Studies of Arabic Language Roots.* In Applied Arabic Linguistics and Signal & Information Processing, pages 97–103. Hemisphere publishing corporation, 1987.

[Obrecht 68] D. H. Obrecht. Effects of the second formant on the perception of Velarization Consonants in Arabic. Mouton, 1968.

[Pierrel 87] J.M. Pierrel. Le dialogue oral homme-machine : Connaisssance linguistiques, stratégies et architectures des systèmes. Collection Hermès, Paris, 1987.

[Prade 87] H. Prade. *Problématiques et Méthodes en Raisonnement Approché (conférence invitée).* In Actes du $6^{ème}$ congrès AFCET Reconnaissance des Formes et Intelligence Artificielle, volume I, Antibes, France, Nov. 1987. AFCET, INRIA.

[Puech 87] G. Puech, N. Louali & R. Hamdi. *La pharyngalisation des consonnes labiales.* In Actes des $16^{ème}$ Journées d'Etudes sur la Parole, pages 298–301, Hammamet, Tunisie, Octobre 1987.

[Rabiner 76] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg & C. A. McGonegal. *A Comparative Study of Several Pitch Detection Algorithms.* IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pages 399–418, Oct. 1976.

[Rajouani 87] D. Rajouani, M. Chiadmi, M. Najim & M. Ouadou. *Synthèse et perception de l'accent lexical en Arabe.* In Actes des $16^{ème}$ Journées d'Etudes sur la Parole, pages 302–305, Hammamet, Tunisie, Octobre 1987.

[Sacy 10] S. De Sacy. Grammaire arabe. De Sacy, 1810.

[Sakoe 8] H. Sakoe & S. Chiba. *Dynamic Programming Algorithm Optimisation for Spoken Word Recognition.* IEEE Trans. Acoust., Speech, Signal Processing, February 1978.

[Schank 77] R. Schank & R. Abelson. Scripts, Plans, Goals and Understanding. Laurence Erlbaum, 1977.

[Shafer 76] G. Shafer. A mathematical theory of evidence. Princeton University Press, 1976.

[Shortliffe 76] E. H. Shortliffe. Computer-based medical consulation : MYCIN. American Elsevier, New York, 1976.

[Sibawayh 89] Sibawayh. EL KITAB, traité de grammaire arabe. H. Derembourg, 1889.

[Troubetzky 70] N. S. Troubetzky. Principes de phonologie. Klincksiek, Paris, 1970.

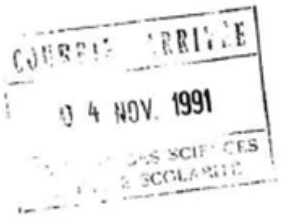
[Zadeh 65] Lotfi A. Zadeh. *Fuzzy Set.* Information and Control, vol. 8, pages 338–353, 1965.

[Zadeh 78] L. A. Zadeh. *Fuzzy sets as a basis for a theory of possibility.* Fuzzy sets and systems, vol. 1, pages 3–28, 1978.

UNIVERSITE DE NANCY I

NOM DE L'ETUDIANT : Monsieur DJOUDI Mahieddine

NATURE DE LA THESE : DOCTORAT DE L'UNIVERSITE DE NANCY I
en INFORMATIQUE



VU, APPROUVE ET PERMIS D'IMPRIMER

NANCY, le 31 OCT. 1991 n°478

LE PRESIDENT DE L'UNIVERSITE DE NANCY I

M. BOULANGÉ

.

Résumé

La reconnaissance automatique de la parole continue pose un certain nombre de problèmes. Les caractéristiques phonétiques et linguistiques de la langue sont largement impliquées dans le processus. Ce travail consiste à présenter une contribution à la reconnaissance automatique de l' Arabe standard. Nous avons d'abord effectué une étude phonétique et phonologique de la langue basée essentiellement sur l'examen de spctrogrammes de mots et de phrases en tenant compte des différents contextes de production des phonèmes. Cette étude nous a permis de définir les caractéristiques acoustiques des phonèmes nécessaires au système de reconnaissance.

Ensuite, nous avons réalisé un système de décodage acoustico-phonétique, baptisé SAPHA (Système Acoustico-PHonétique de l'Arabe) qui reçoit en entrée le signal de parole d'une phrase et retourne comme résultat un treillis de phonèmes. Les principales étapes du système sont :

— la segmentation du signal de parole en grandes classes phonétiques (voyelles, plosives, fricatives et sonnantes),

— l'extraction des indices phonétiques utilisés en reconnaissance,

— l'étiquetage des segments utilisant un système à base de règles de production. Les méthodes utilisées ont été adaptées du système APHODEX développé dans notre équipe pour le décodage phonétique du francˆais.

L'évaluation des performances du système a été effectuée à partir de l'étiquetage manuel des phrases phonétiquement équilibrées du corpus DJOUMA que nous avons constitué.

Enfin, nous avons développé quelques idées pour la conception d'un système de reconnaissance de phrases en Arabe (MARS) intégrant le décodeur phonétique et nous avons soulevé les problèmes d'ordre morphologique, phonologique, syntaxique et prosodique qu'il faut résoudre.

Mots clés :

- Arabe standard - Reconnaissance de la parole

- Décodage phonétique - Système expert