

Analyse acoustique pour la reconnaissance des phonèmes de l'Arabe moderne standard

Mahieddine DJOUDI

Laboratoire d'Informatique de l'Université de Poitiers
40, avenue du Recteur Pineau,
86022 Poitiers cedex

Mots clés : Arabe moderne standard, corpus, décodage phonétique, analyse acoustique, spectrogramme.

Résumé :

Nous proposons dans cet article une étude phonétique pour la reconnaissance automatique des phonèmes de l'Arabe. La langue de référence est celle utilisée dans l'enseignement, la littérature, l'administration et les médias de l'ensemble du monde arabe. Elle est dite moderne par opposition à l'Arabe classique littéraire, Elle est qualifiée de standard par opposition aux différents dialectes régionaux.

Notre étude est purment acoustique, elle est basée essentiellement sur des visions spectrographiques et nécessite un corpus représentatif. Elle consiste en l'extraction des indices acoustiques qui caractérisent un phonème ou une classe phonétique.

1 Introduction

La reconnaissance automatique des phonèmes de l'Arabe pose des problèmes supplémentaires du fait de la présence des consonnes emphatiques, pharyngales et glottales et l'existence d'une opposition brève-longue chez les voyelles.

L'approche utilisée en reconnaissance est basée sur l'extraction des indices phonétiques pertinents à partir d'un corpus représentatif. Dans un premier temps, nous avons réalisé un grand corpus de mots dans lequel chaque phonème apparaît dans les différents contextes. Comme notre travail s'insère dans le cadre de la reconnaissance automatique de la parole continue et dans un contexte multilocuteur, les mots isolés à eux seuls ne suffisent pas, ils ne répondent pas au critère de prise en compte réelle du contexte de production des phonèmes et du phénomène de coarticulation, il nous a été nécessaire

d'utiliser le corpus DJOUMA (DJOUmal MAqrou'a, qui veut dire en Arabe phrases lues) constituée de 50 phrases phonétiquement équilibrées prononcées par 11 locuteurs (7 hommes et 4 femmes) [5].

Pour effectuer notre étude phonétique de l'Arabe, nous avons utilisé le logiciel Snorri [9] qui 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.

2 Structure acoustique des voyelles

L'Arabe standard comporte trois voyelles /*ḥarakaat*/ qui s'opposent phonologiquement par le timbre. Ces voyelles sont : /a/ : /*fatḥa*/, /i/ : /*kasra*/ et /u/ : /*ḍamma*/. De même, on note le /*sukuun*/ pour désigner l'absence de voyelle après la consonne. 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*/.

2.1 La quantité vocalique

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 [2]. 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 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.

2.2 Les fréquences formantiques

Les voyelles se caractérisent principalement par la présence de zones de fréquences où 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.

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

L'extraction des formants peut se faire à partir des pics ou des coefficients 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 coefficients LPC (ordre 20) nous calculons les premiers pics significatifs qui seront des candidats à des positions de formants. Ensuite, nous calculons les trois premiers formants comme étant des pics visibles dans les bandes de fréquences [250-850], [750-2300] et [1800-3000] Hz respectivement pour F1, F2 et F3.

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

- labial : /b/, /f/, /m/ et /w/ ;
- dental : /t/, /d/, /θ/, /ð/, /n/, /s/, /z/, /r/ et /l/ ;
- palatal : /j/, /z/ et /y/ ;
- vélaire : /k/, /χ/ et /γ/ ;
- emphatique : /t̤/, /d̤/, /s̤/, /ð̤/ et /q/ ;
- pharyngal : /ħ/ 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 [7].

Nous donnons dans la table 2 les valeurs des formants des voyelles dans différents contextes. 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.

Contexte	Formant	a	i	u	aa	ii	uu
Labial	F1	630	430	420	590	280	370
	F2	1250	1720	1000	1200	2100	780
	F3	2420	2620	2170	2580	2700	2370
Dental	F1	490	420	420	500	280	390
	F2	1550	1970	1090	1330	2020	990
	F3	2500	2790	2290	2550	2880	2220
Palatal	F1	520	280	300	570	280	300
	F2	1530	2030	1100	1330	2050	1080
	F3	2490	2900	2650	2380	2750	2300
Vélaire	F1	530	370	310	590	370	380
	F2	1230	1570	1070	1510	1980	920
	F3	2170	2530	2270	2430	2780	2440
Emphatique	F1	650	440	440	650	380	440
	F2	1200	1680	900	1250	2050	890
	F3	2600	2300	2650	2740	2700	2700
Pharyngal	F1	620	300	400	580	331	340
	F2	1570	2000	880	1620	2100	810
	F3	2470	2500	2100	2610	2700	2300
Glottal	F1	660	370	440	670	250	380
	F2	1270	1860	1130	1130	2020	830
	F3	2390	2680	2340	2430	3070	2130

TABLE 2 – Valeurs des formants des voyelles en contexte

3 Structure acoustique consonnes

Le système consonantique de l'Arabe standard est généralement décrit comme étant composé de 28 consonnes /ħuruuf/ : ce sont les 28 lettres de l'alphabet [1]. Sur le plan acoustique, les consonnes forment une classe très hétérogène que l'on peut décomposer en sous classes ayant des caractéristiques distinctes.

3.1 Les consonnes selon leurs lieux d'articulation

Nous avons retenu les lieux suivants [11], [4] :

- labial : /b/, /m/ et /w/ ;
- labiodental : /f/ ;
- dental : /θ/, /ð/ et /ð/ ;
- alvéodental : /t/, /d/, /t̤/, /d̤/, /s/, /z/, /s̤/, /n/, r/ et /l/ ;
- palatal : /ʃ/, /z/ et /y/ ;
- vélaire : /k/, /χ/ et /γ/ ;
- uvulaire : /q/ ;
- pharyngal : /ħ/ et /ε/ ;
- glottal : /ʔ/ et /h/.

3.2 Les consonnes selon leurs modes d'articulation

Les modes articulatoires caractéristiques des consonnes sont :

- Le /zahr/ et son contraire le /hams/ qui correspond à l'opposition sourde/sonore.
- Le /ʔitbaaq/ (emphase) et son contraire le /ʔinfitaħ/. 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.
Les quatre articulations définies traditionnellement comme emphatiques sont /t/, /s̤/, /d̤/ et /ð/ mais leur nombre varie d'un auteur à un autre [3]. L'influence des dialectes sur la prononciation de l'arabe standard provoque l'emphatisation de certaines autres consonnes, particulièrement les labiales /b/, /m/ et /f/ [10].
- Les consonnes occlusives ou plosives : physiologiquement, une plosive est caractérisée par :
 1. la formation d'une fermeture à l'intérieur de la cavité vocale par un ou plusieurs articulateurs à l'endroit où le conduit de pression est bloqué et qui 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.
- Les consonnes fricatives : les fricatives 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.

- Les consonnes 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 le /m/ et le /n/
- Le vibrant : le /r/ est un vibrant apicoalvéolaire lingual voisé qui comporte une vibration accentuée de la pointe de la langue.
- Le latéral : le /l/ est un phonème lingual qui possède deux allophones : le plus commun de ces allophones est une latérale dentale voisée de durée L 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 :
/wallah/ : "jurer par Dieu " et /wallah/ : "il l'a nommé wali "
- Les semivoyelles : ce sont les consonnes /y/ et /w/.

4 Les indices acoustiques

4.1 La durée

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 ; 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 la table 3, les valeurs moyennes de la durée des phonèmes selon le contexte de production. A partir de ce tableau, nous pouvons faire les

consonnes	début	milieu	milieu géminé	fin	fin géminé
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 3 – Durée moyenne des consonnes

remarques suivantes :

- 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.

4.2 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 la table 4.

Le calcul du degré de voisement, nous amène à 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,

4.3 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 éventuellement vide de prélèvements correspondants au pics détectés.
- La localisation de la barre d’explosion en retournant le numéro de prélèvement qui contient le plus de pics sur l’ensemble des 8 bandes de fréquences et qui correspond bien à la visibilité 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 Français a donné un pourcentage de 87% de bonne localisation, avec un taux d’insertion égal à 13% [5]. Les tests effectués sur 293 plosives arabes du corpus DJOUMA ont donnés un pourcentage de 85%. Le taux d’insertion est d’environ 10%.

Phonème	Degré de voisement
t	0.1
k	0.0
ʔ	0.1
b	0.4
d	0.4
q	0.1
t̥	0.0
ɖ	0.7
z	0.5
f	0.2
θ	0.1
s	0.1
ʃ	0.1
χ	0.5
ħ	0.1
ð	0.7
z	1.0
γ	0.7
ε	1.0
h	0.9
ʂ	0.2
ʈ	0.6
m	1.0
n	0.9
r	0.7
l	0.8
w	0.9
j	0.8

TABLE 4 – Degré de voisement des consonnes

4.4 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 :

$$dc = (max/moy) - 1; \quad (1)$$

où 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. L'ensemble de ses indices nous permettent de connaître le lieu d'articulation de la plosive (labial, dental, vélaire, uvulaire ou glottal). Actuellement, seuls, la durée, 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 5 les valeurs du degré de concentration (dc) et la fréquence du burst (fb) des plosives dans les différents contextes vocaliques. D'une façon générale, le burst est diffus

Contexte gauche	<i>a/aa</i>		<i>i/ii</i>		<i>u/uu</i>	
	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
ʈ	3800	0.54	3000	0.66	2500	0.45

TABLE 5 – Degré de compacité et fréquence du burst

pour les consonnes labiales et dentales et compact pour les vélares et glottales.

4.5 Les structures formantiques

Les formants caractérisent aussi les consonnes à structure formantique ou sonnantes. Nous donnons dans la table 6 les valeurs des deux premiers formants de ces consonnes en contexte vocalique.

Consonne	<i>moyen</i>		<i>a/aa</i>		<i>i/iï</i>		<i>u/uu</i>	
	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
/ð/	300	1400	330	1450	300	1400	310	1380
/ð/, /d/	460	1320	450	1500	500	1400	450	1200
/ɣ/	480	1270	550	1240	470	1550	450	840
/h/	410	2050	670	1560	430	2000	550	1340
/ɛ/	600	1450	650	1300	470	1250	460	1030

TABLE 6 – Valeurs des formants des sonnantes

4.6 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. 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 coefficient 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 7 résume les transitions les plus remarquées des consonnes aux voisinage des classes des voyelles¹ Nous tenons à signaler, que pour certains phonèmes les transitions ne sont pas nettes dans 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 :

- pour les consonnes labiales F2 et F3 sont montants,
- pour les dentales F2 et F3 sont descendants,
- pour les vélares F2 est descendant et F3 montant,

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

Contexte gauche	<i>a/aa</i>			<i>i/ii</i>			<i>u/uu</i>		
	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
ḏ	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
ş	m	p	p	m	m	p	p	d	m
ḏ	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 7 – Les transitions formantiques

— pour les emphatiques F1 est descendant et F2 montant.

4.7 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 8. Les remarques que nous

Fricative	Limite du bruit
<i>z</i>	1750
<i>f</i>	1380
<i>θ</i>	3500
<i>s</i>	3300
<i>ʃ</i>	1800
<i>χ</i>	1000
<i>ħ</i>	900
<i>z</i>	3450
<i>ʒ</i>	3200

TABLE 8 – Limite inférieure du bruit des fricatives

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 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 */ʒ/* possède un bruit légèrement plus bas que son homologue non emphatique */s/*,
- le */ħ/* 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 est basse.

4.8 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 la table 9 :

Fricative	Centre de gravité
<i>z</i>	3160
<i>f</i>	4550
<i>θ</i>	4820
<i>s</i>	4760
<i>ʃ</i>	4570
<i>χ</i>	3580
<i>ħ</i>	3860
<i>ð</i>	2880
<i>z</i>	4900
<i>γ</i>	3420
<i>ε</i>	1570
<i>h</i>	2880
<i>ʂ</i>	4940
<i>ð</i>	2910

TABLE 9 – Centre de gravité énergétiques des fricatives

5 La reconnaissance automatique

En tenant compte de des particularités de la langue, nous avons conçu et réalisé le système SAPHA (Système Acoustico-PHonétique de l'Arabe). 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. Les principales étapes du système sont :

- la segmentation du signal en grandes classes phonétiques en utilisant des algorithmes non contextuels et reposant sur des critères simples. Les classes retenues sont les voyelles, les plosives, les fricatives et les sonnantes.
- le calcul des indices phonétiques pertinents à la reconnaissance. Ces indices sont : la durée, le degré de voisement, la barre d'explosion, le suivi de formants, les transitions formantiques, le centre de gravité et la limite du bruit de friction.
- l'identification phonétique des segments en utilisant un système expert à base de règles de production qui affecte des étiquettes phonétiques aux segments obtenus lors de la segmentation en utilisant les indices phonétiques et les connaissances se trouvant dans la base de règles.

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. L'évaluation du système, qui consiste en la comparaison du décodage automatique avec l'étiquetage manuel de 50 phrases phonétiquement équilibrées et prononcées par 3 locuteurs masculins donne actuellement un score global de 65% [6].

6 Conclusion

Nous avons présenté dans ce papier des résultats originaux qui permettent de caractériser les phonèmes de la langue arabe telle qu'elle est parlée actuellement. Ces résultats nous ont servi dans la reconnaissance automatique de la parole continue dans un contexte multilocuteur. Nous pouvons envisager l'utilisation de ces résultats dans un système de synthèse de la parole [8].

Références

- [1] S. H. Al. Ani. *Arabic Phonology. An Acoustical and Physiological Investigation*. Mouton & Co N.V., 1970.
- [2] Y. Belkaid. Les voyelles de l'Arabe littéraire moderne. analyse spectrographique. Technical Report 16, Travaux de l'institut de phonétique de Strasbourg, 1984.
- [3] J. F. Bonnot. Recherche expérimentale de certains aspects de la gémination et de l'emphase en Arabe. Technical Report 11, Travaux de l'institut de phonétique de Strasbourg, 1979.
- [4] J. Cantineau. *Cours de phonétique arabe*. Librairie Klincksieck, 1960.
- [5] M. Djoudi. Contribution à l'étude et à la reconnaissance automatique de la parole en arabe standard. Thèse de Doctorat de l'Université de Nancy 1, 1991.
- [6] M. Djoudi. Assessment of Acoustic Phonetic Decoder for Standard Arabic. In *13th National Computer Conference*, volume 2, pages 761–770, Riyadh Saudi Arabia, 28 November - 2 December, 1992.
- [7] M. Djoudi, H. Aouizerat, and 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.
- [8] M. Guerti. Contribution à la synthèse de la parole en Arabe standard. In *Actes des 16^{ème} Journées d'Etudes sur la Parole*, pages 290–293, Hammamet, Tunisie, Octobre 1987.
- [9] Y. Laprie. Un système d'étude interactif de la parole. *Actes des 17^{ème} Journées d'Etudes sur la Parole*, pages 71–76, Sep 1988.
- [10] G. Puech, N. Louali, and 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.
- [11] A. Hadj Salah. Arabic Linguistics and Phonetics. In *Applied Arabic Linguistics and Signal & Information Processing*, pages 3–22. Hemisphere publishing corporation, 1987.