

Synthèse de la parole en Arabe Standard

- Cas des huruf el medd -

K. FERRAT * & M. GUERTI **

* Laboratoire de la Communication Parlée
Centre de Recherche Scientifique et Technique pour
Le Développement de la Langue Arabe
CRSTDLA, BP 225, Rostomia-Alger, 16011 Alger
Tel : 213 21 94 10 88, Fax : 213 21 94 10 88
Email : kamelferrat@yahoo.fr

** Ecole Nationale Polytechnique BP 182-El Harrach
16200 Alger, Algérie.
Fax : 213 21 52 29 73
E-mail : mhaniag@yahoo.com

Abstract

In Standard Arabic language, the long vowels called "huruf el medd" represents a very important phenomenon. Our present work will contribute to modeling long vowels in order to synthesize speech using Mbrola method.

Indeed, we are going to study vowels temporal durations and formants' values, as well as the modifications undergone by long vowels energy graphs in relation to short vowels' ones. For it, we are going to use the Computing Speech Laboratory CSL of Kay Elemetrics together with the Praat and Matlab Windows Programs. Concerning the application on speech synthesis in Standard Arabic, we will use the Mbrola method (Multi Band Re-synthesis Overlap Add) which is considered as one of the most efficient methods in terms of natural speech.

Keywords: arabic speech synthesis, Mbrola, vowels, huruf el medd, formants, energy.

1. Introduction

Les « huruf el medd », confondus à tort ou à raison avec les voyelles longues en langue française, représente un phénomène très important en Arabe Standard. Ce phénomène qui ne présente qu'un simple moyen d'expression dans beaucoup de langues comme le français, est très pertinent en arabe. Ce travail vise à contribuer à la modélisation des phénomènes propres à l'Arabe Standard, dans le but d'une synthèse de qualité, proche du naturel.

Les Anciens Grammairiens Arabes définissaient le système phonologique de l'Arabe comme étant basé sur l'étude des catégories grammaticales associées à des notions physiologiques et physiques de la parole qui sont la [haraka] et le [sukūn]. Ces catégories sont respectivement :

- la [fatha] [a] qui se prononce en ouvrant largement la bouche et en conservant la langue dans une position horizontale ;
- le [damma] [u] en contractant la langue en arrière et arrondissant les lèvres jusqu'à ce qu'elles se joignent ;
- la [kasra] [i] en portant l'apex en avant et en l'étalant largement tandis que l'arrière frôle presque le palais et que les commissures des lèvres s'étirent.

Aux trois voyelles brèves [a, i, u], s'opposent trois voyelles longues [ā, ī, ū]. Le système vocalique de l'Arabe se présente comme suit (figure 1).

2. Les notions de [haraka] et de [sukūn] en Arabe Standard

La [haraka] est définie comme un mouvement aérien et organique dont a besoin un [harf] pour se produire dans un continuum sonore [2].

Lorsque le mouvement se fait d'un [harf] à un autre, on dit qu'il y a [haraka], si le mouvement se fait d'un [harf] à une "position vide", on dit qu'il y a [sukūn].

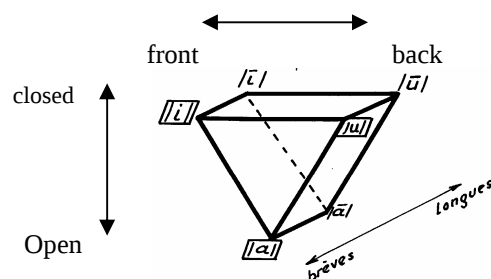


Figure 1: Système vocalique de l'Arabe Standard [1].

3. Etude acoustique des hurūf el-medd en Arabe Standard

En étudiant la composante phonétique de l'Arabe Standard, nous pouvons dégager les caractéristiques relatives aux différents phonèmes et cerner ainsi l'ensemble des paramètres acoustico-phonétiques nécessaires à l'élaboration de notre système de synthèse.

Le choix d'une étude préalable des voyelles, prises dans un contexte isolé, est guidé par le souci de les prendre comme références, afin d'étudier les variations qu'elles subissent en présence des différentes consonnes de l'Arabe Standard.

En faisant le relevé des valeurs moyennes en utilisant le sonographe Kay 5500, avec une précision de ± 20 Hz, fréquence d'échantillonnage fixée à 10 KHz, nous obtenons le tableau suivant :

Table 1 : Valeurs moyennes des formants des voyelles de l'Arabe Standard.

voyelles	F1	B1	F2	B2	F3	B3	d (ms)
[a]	600	360	1520	280	2580	340	75
[i]	280	360	2360	280	2780	360	65
[u]	300	300	760	320	2320	300	70
[ā]	560	360	1520	280	2520	340	160
[ī]	260	360	2360	280	2760	360	140
[ū]	320	300	660	320	2340	300	150

D'après le tableau1, nous remarquons que :

- [i], [ī], [u] et [ū] ont respectivement les F_1 les plus bas ;
- [a] et [ā] ont respectivement les F_1 les plus élevés ;
- [u] et [ū] ont respectivement les niveaux des formants les plus bas, en particulier pour le second formant F_2 .
- Le niveau du formant F_2 de [i] et [ī] est assez important par rapport aux F_2 des autres voyelles. De même, l'écart entre F_1 et F_2 dans le cas de [i] et [ī] est très important, comparé aux autres voyelles (figure 3).
- Les niveaux des formants F_2 et F_3 de [u] sont à des niveaux assez bas, comparés à ceux de F_2 et F_3 des autres voyelles.

Ces résultats confirment ce qui a été déjà démontré expérimentalement et qui est représenté par un triangle vocalique.

En comparant les niveaux des formants entre chaque voyelle brève et son opposée longue, nous remarquons qu'ils sont assez rapprochés. Cela nous permet de dire qu'en Arabe Standard, l'allongement ou le [medd] n'influe pas sensiblement sur les niveaux formantiques des voyelles. La partie stable de la voyelle longue se trouve beaucoup plus allongée. De même, la durée de la voyelle reste tributaire du contexte de prononciation (influence de l'entourage).

- Résultats et Interprétation

L'ouverture de la cavité buccale correspond à un niveau haut du formant F_1 (cas de la voyelle [a]). Sa fermeture correspond à un niveau bas de F_1 (cas des voyelles [i] et [u]).

Lors de l'ouverture de la cavité buccale, nous assistons à un agrandissement de la cavité buccale et un rétrécissement de la cavité pharyngale. Comme F_1 prend naissance dans cette dernière cavité, nous pourrions donc dire qu'un rétrécissement de cette cavité de résonance correspond à une montée de la fréquence du niveau formantique de F_1 . Par opposition à ce rétrécissement, un agrandissement de cette cavité correspond à une chute de fréquence du niveau formantique de F_1 .

Une articulation antérieure correspond à un niveau haut du formant F_2 (cas de la voyelle [i]). L'apex se trouve porté vers l'avant. La langue se trouve étalée largement tandis que l'arrière frôle le palais. Ce mouvement lingual provoque un rétrécissement de la cavité buccale, donc à une montée de F_2 . Par contre, la voyelle [u] se prononce en contractant la langue au fond de la bouche. Cette contraction de la langue fait qu'il y ait élargissement de la cavité buccale, donc diminution de F_2 .

L'arrondissement des lèvres se traduit par une baisse du formant F_3 . L'ouverture des lèvres lors de la prononciation de la voyelle [a], ainsi qu'une articulation antérieure lors de la prononciation de la voyelle [i], font monter le niveau de ce formant.

De même, suivant les valeurs des formants, nous pouvons dire que le [medd] n'influe pas sensiblement sur les niveaux formantiques des voyelles. Le [harf el-medd] est en fait une extension quantitative de la voyelle brève. La durée temporelle de la zone stable se trouve donc allongée.

Une étude réalisée à partir d'un enregistrement au moyen d'une carte d'acquisition sur micro-ordinateur PC (carte OROS AU21), et dont l'analyse a été faite à partir du logiciel Matlab, a donné, approximativement pour l'amplitude d'énergie, les résultats illustrés par la figure 2.

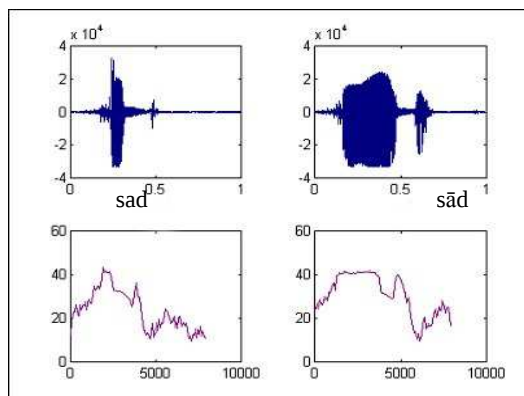


Figure 2 : Comparaison des courbes de l'amplitude d'énergie des mots [sad] et [sād]

Nous constatons un niveau supérieur de l'amplitude d'énergie plus marqué et plus étendu pour le cas de présence d'un [harf el medd] dans [sād], contrairement à son opposée voyelle brève dans [sad]. L'effort articulatoire pour prononcer le [harf medd] est donc plus important. C'est également le cas, en présence de [i].

4. Synthèse de la parole en Arabe Standard

Nous avons déjà entamé l'étude acoustique de la gémiation en Arabe Standard et son application pour la synthèse de la parole, en appliquant la méthode MBROLA [4]. Ce travail va dans le même sens pour inclure l'ensemble des phénomènes caractéristiques de la langue arabe.

4.1. La technique MBROLA

MBROLA (**M**ulti **B**and **R**esynthesis **O**ver**L**ap **A**dd) est une technique de synthèse vocale fondée sur la concaténation des sons [5]. L'avantage principal d'une synthèse de parole basée sur cette technique est qu'elle permet :

- le lissage des discontinuités spectrales apparaissant de part et d'autre des points de concaténation (problème que l'on rencontre dans la lecture automatique de textes à partir du système Arphon du laboratoire de la Communication Parlée du CRSTDLA) [6];
- une parole de synthèse très fluide ;
- une capacité mémoire du dictionnaire de diphtongues sensiblement réduite.

Le module MBROLA reçoit les informations exposées ci-dessus en entrée, sous forme d'un fichier de format « fichier.pho ». Il génère ensuite en sortie, un fichier audio de format « fichier.wav ». Le principe de MBROLA est d'écrire une phrase en symboles phonétiques et de fournir les informations suivantes :

- les pauses et la ponctuation ;
- la durée de chaque phonème ;
- le mouvement du pitch F_0 (début du pitch par rapport à la durée du phonème et sa valeur).

4.2. Application pour le système Arphon

Le système ARPHON, mis au point au laboratoire de la Communication Parlée du CRSTDLA, utilise un dictionnaire de données sonores de 3760 unités de tailles variables (polysons)[6]. La taille du dictionnaire est de 18 Mo. Ce qui est comme même assez important, notamment pour une éventuelle mise en place d'une démonstration du lecteur sur le site Web du centre. Il est donc plus que nécessaire de compresser cette base de données pour son exploitation à partir du serveur de l'application. La modélisation des huruf el-medd répond à ce besoin en diminuant considérablement cette base de données et surtout contribue à une meilleure sécurité des données puisque les règles seront directement intégrées dans le programme du lecteur et par conséquent son exécutable.

Ces règles une fois mises au point dans les différents contextes sont appliquées au lecteur automatique de textes et nombres Arphon. Les différentes étapes suivies pour générer notre synthèse de la parole sont :

- la génération des mots à partir du lecteur automatique Arphon ;

- leur resynthèse (passage à travers le module MBROLA) permettant de modifier les mots par une bonne amélioration du naturel, modifiant les paramètres pertinents que sont la durée et le pitch (en respectant les modèles obtenus lors de l'analyse acoustique).

La qualité de la parole obtenue est assez bonne. Les discontinuités perceptibles à travers le lecteur automatique Arphon sont éliminées grâce à l'apport de ces paramètres. Quelques exemples suivants montrent les résultats obtenus après resynthèse de la parole par la technique MBROLA (figure 3).

```
; .pho file created with MBROLIGN v1.0
; Ferrat Kamel, kamelferrat@yahoo.fr,
; resynthesis of the word "sin"
; generated by the Arphon reader
```

```
_ 74
s 113
i 65 31 106 50 106
n 167 0 105 26 103 30 101
_ 123
```

Pour les huruf al medd, nous augmentons les valeurs des durées temporelles des voyelles brèves ou harakate.

```
; .pho file created with MBROLIGN v1.0
; Ferrat Kamel, kamelferrat@yahoo.fr
; resynthesis of the word "sī n"
; generated by the Arphon reader
```

```
_ 74
s 110
i 140 31 106 50 106
n 137 0 105 26 103 30 101
_ 123
```

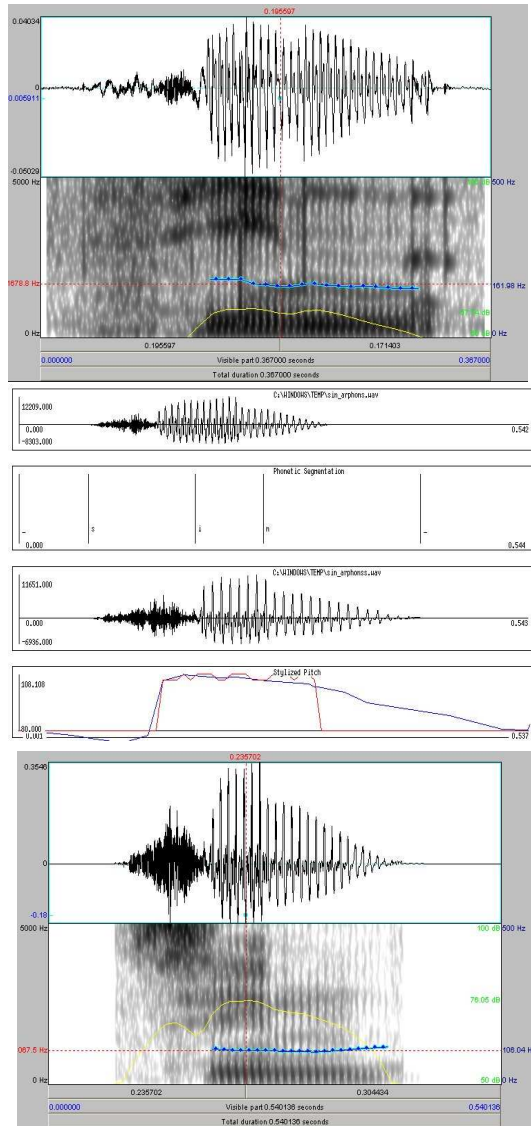


Figure 3 : Resynthèse par MBROLIGN du mot [sin] générés par Arphon.

Nous procédons de la même façon pour les cas des deux mots [sad] et [sād] avec un prolongement de la voyelle ou haraka dans le second.

5. Conclusion générale

Un travail dans ce sens a été déjà mis au point, concernant la gémation (tachdid et idgham) en Arabe Standard [3]. Dans cette présente étude, nous avons extrait les paramètres acoustiques des huruf al-medd appelés encore voyelles longues par opposition aux voyelles brèves. Nous avons ensuite appliqué les résultats obtenus pour la

synthèse de la parole en Arabe Standard. Cette étude vise une étude approfondie des phénomènes propres à la langue Arabe, en vue d'une synthèse vocale de bonne qualité.

Cette expérience nous a montré l'apport de la technique MBROLA dans la synthèse de la parole en Arabe Standard. L'insertion du module MBROLA, dans l'organigramme du lecteur automatique de texte Arphon, s'avère très fructueux vue la qualité de la parole resynthétisée obtenue. De plus ce travail nous a montré que par une modélisation adéquate des huruf el medd dans les différents contextes, nous pouvons obtenir une parole synthétique sans avoir à recourir à la conception d'unités sonores contenant les voyelles longues dans le dictionnaire du lecteur Arphon. Ce qui nous permettra de réduire considérablement la taille de ce dictionnaire. Cette réduction est d'autant importante que nous projetons de mettre une démonstration du lecteur automatique de texte Arphon sur le site Web du centre de recherche du CRSTDLA (Alger).

BIBLIOGRAPHIE

- [1] M. Guerti " *Contribution à la synthèse de la parole par diphtongues en Arabe Standard* " (Synthèse par diphtongues et technique de prédiction linéaire). Thèse de magister en Electronique Acoustique et Physiologique de la parole. ILP- Université d'Alger (Algérie), 1983.
- [2] A. Hadj-Salah " *La notion de syllabe et la théorie cinétique-impulsionnelle des phonéticiens arabes* ". Al-Lissaniyat. Revue Algérienne de Linguistique. Vol.1, pp. 63-83, ILP Alger (Algérie), 1971.
- [3] K. Ferrat " *La technique MBROLA dans la synthèse de la parole en Arabe Standard. Application pour le cas du Tachdid et du Idgham* ", 4^{ème} Séminaire National en Informatique SNIB04, Biskra (Algérie), 4-6 mai 2004.
- [4] T. Dutoit, V. Pagel, N. Pierret, F. Bataille, O. van der Vrecken " *The MBROLA Project : Towards a set of high quality speech synthesizers free of use for non commercial purposes* ". Faculté Polytechnique de Mons, TCTS Lab, 31 Bvd Dolez, B-7000 Mons, Belgium, mail: MBROLA@tcts.fpms.ac.be, 1998.
- [5] K. Benbellil, K. Ferrat, G. Droua-Hamdani, M. Abbas " *Conception et réalisation d'un lecteur automatique de textes en Arabe Standard* ", Séminaire International sur l'Automatisation de la Langue Arabe, organisé par l'Académie de langue arabe de Aman (Jordanie) en collaboration avec les centres de recherche CRSTDLA (Algérie) et CERIST (Algérie), 16 septembre 2002.