

Etude phonétique et identification des consonnes pharyngales et glottales de l'Arabe

Mahieddine DJOUDI¹

Mohamed DEBYECHE²

¹Département d'Informatique de l'Université de Poitiers
40 avenue du Recteur Pineau 86022 Poitiers cedex, France

²Institut d'Electronique
Laboratoire Traitement du Signal
USTHB BP 32 Bab Ezzouar, Algérie

Résumé :

Nous proposons dans cet article une étude phonétique des consonnes pharyngales /ħ/, /ε/ et glottales /ʔ/, /h/ de l'Arabe moderne standard. Cette étude nous a permis d'élaborer une base de connaissances phonétiques destinée à la reconnaissance automatique de ces consonnes en parole continue et dans un contexte multilocuteur.

A partir d'un corpus de mots dans lequel toutes les variantes contextuelles de ces consonnes apparaissent et à l'examen d'un nombre très important de spectrogrammes, nous avons relevé les caractéristiques acoustiques spécifiques à chaque consonne ainsi que les variations en fonction de l'environnement phonétique.

Les connaissances acquises sont destinées au système expert SAPHA, développé pour le décodage acoustico-phonétique de l'Arabe. Le formalisme adopté pour la représentation des connaissances est celui des règles de production [2].

Les performances du système sur les consonnes pharyngales et glottales ont été calculées sur un corpus de phrases pour trois locuteurs masculins.

Mots clés :

Reconnaissance de la parole, Décodage phonétique, Spectrogramme, Analyse acoustique, Consonnes pharyngales et glottales.

1 Introduction

La reconnaissance de l'Arabe pose un certain nombre de problèmes qui sont dues aux caractéristiques phonétiques particulières de la langue. L'originalité du système phonétique de l'Arabe se fonde pour une part très importante sur la présence des consonnes pharyngales et glottales. Notre étude est purement acoustique,

elle est effectuée par une analyse de spectrogrammes des consonnes dans différents contextes et est orientée vers la reconnaissance phonétique des consonnes pharyngales et glottales en parole continue et dans un contexte multilocuteur. La même approche a été utilisée pour la reconnaissance des consonnes emphatiques de l'Arabe [4].

2 Etude phonétique

2.1 L'outil d'analyse

Pour effectuer notre étude phonétique des consonnes pharyngales et glottales de l'Arabe, nous avons utilisé le logiciel Snorri 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. 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 postscript 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 [6].

2.2 Le corpus

L'étude des consonnes pharyngales et glottales dans le but de la reconnaissance nécessite un corpus représentatif. En prenant en considération la nature transitionnelle du phonème dans la chaîne verbale, nous avons réalisé un grand corpus de mots dans lequel chaque phonème apparaît dans les différents contextes. Chaque locuteur prononce les mots du corpus deux-à-deux séparés par un intervalle de temps allant de 1,5 à 2 secondes de telle sorte que lors de l'analyse, l'on obtienne toujours sur la console graphique les deux mots en même temps.

L'analyse spectrographique s'est portée sur les segments en question, ainsi que sur le contexte vocalique gauche et/ou droit dans lequel ils évoluent.

Comme notre travail s'insère dans le cadre de la reconnaissance continue de la parole, 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 constitué de 50 phrases phonétiquement équilibrées [5].

2.3 Les consonnes pharyngales

2.3.1 /ħ/

Le phonème /ħ/ possède une structure acoustique spécifique d'une fricative sourde, toutefois, il peut apparaître comme voisé en position intervocalique. Suivi de /a/ ou /aa/, le bruit de friction est concentré autour de 3500-4500 Hz et inexistant en dessous de 800-1000 Hz. Devant /i/ et /ii/, la concentration est dans la zone 3200-4500 Hz, et la limite inférieure vers 1800-2200 Hz. Avec /u/ et /uu/, la concentration du bruit est entre 3200 et 4200 Hz, sa limite inférieure dans la zone 600-1300 Hz. En position finale, la limite inférieure est autour de 1800-2000 Hz. Devant /ħ/, /a/ et /a/ possèdent des transitions plates du 2^{ème} formant (F2) et montantes de F3. Pour /i/ et /ii/ F2 est montant et F3 est plat et pour /u/ et /uu/, les transitions de F2 sont montantes et ceux de F3 descendantes.

2.3.2 /ε/

Le /ε/ est traditionnellement décrit comme fricative voisée. Al Ani dans son étude acoustique et physiologique des consonnes de l'Arabe [1] trouve que l'allophone le plus commun est une plosive sourde. Sur nos spectrogrammes, le /ε/ apparaît généralement comme une sonnante avec des pseudo formants dont la structure est très dépendante du contexte. Les mesures des formants de /ε/ sont avec /a/ et /aa/, F1 : 770, F2 : 1480 et F3 : 2440 Hz, avec /i/ et /ii/, F1 : 470, F2 : 1800 et F3 : 2650 Hz, avec /u/ et /uu/, F1 : 470, F2 : 1260 et F3 2190 Hz et dans un contexte non vocalique F1 : 650, F2 : 1640 et F3 2450 Hz. Devant /ε/, /a/ et /a/ possèdent des transitions plates de F1, montantes de F2 et plates de F3. Pour /i/ et /ii/ F1 est descendant et F2 et F3 sont montants et pour /u/ et /uu/, les transitions de F1 et F2 sont descendantes et ceux de F3 montantes.

2.4 Les consonnes glottales

2.4.1 /ʔ/

Le /ʔ/ est décrit comme une plosive glottale non voisée qui apparaît sur le spectrogramme comme un intervalle de silence suivi ou non d'un burst. Avec /a/ et /aa/, la concentration du burst est vers 3500-5000 Hz, avec /i/ et /ii, elle est dans la zone 4000-5200 Hz et avec /u/ et /uu/, le burst est le plus souvent absent. Les transitions formantique de l'ensemble des voyelles sont plates.

2.4.2 /h/

Le /h/ est décrit comme une fricative glottale non voisée de durée relative 100-150 ms qui apparaît comme un bruit ayant une structure de formants et qui devient voisé en milieu intervocalique. Devant /a/ et /aa/, la concentration du bruit est autour de 300-4000 Hz, devant /i/ et /ii/ le bruit est concentré dans la région 3200-4200 Hz. Quand le /h/ est devant /u/ et /uu/, le bruit est concentré en basse fréquence vers 1000 Hz. Les mesures des structures formantiques du /h/ avec /a/ et /aa/ sont F1 : 500-650 et F2 : 1450-1550 Hz, avec /i/ et /ii/, F1 : 280-350 et F2 : 2000-2200 Hz et avec /u/ et /uu/ F1 : 320-400 et F2 : 900-1100 Hz. Les transitions formantiques sont le plus souvent plates.

3 La reconnaissance automatique

3.1 Le système SAPHA

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. Le système de décodage proprement dit est un système expert à base de règles de production. L'acquisition des connaissances a nécessité le développement d'outils d'analyse et d'affichage graphique. Les principales étapes du système sont :

La segmentation

Elle 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. Le reste des phonèmes de l'Arabe standard sont mis dans la classe des sonnantes. Les phonèmes qui possèdent les caractéristiques de deux classes sont considérés comme appartenant à une classe ayant ces deux caractéristiques (exemple la classe Fricative/Voyelle pour quelques occurrences du /ε/).

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 chacun des indices phonétiques suivants :

- durée du segment ;
- degré de voisement ;
- position et caractéristiques de la barre d'explosion ;
- valeurs de formants ;
- transitions formantiques ;
- limite inférieure du bruit de friction.

L'étiquetage phonétique

Il consiste à identifier les phonèmes en utilisant un système expert à bases de règles de production.

Ce système consiste en une base de connaissances d'identification de phonèmes sous forme de règles de production et en un moteur d'inférence qui fonctionne en chaînage avant et en chaînage arrière en effectuant une analyse de gauche à droite, segment par segment.

3.2 Identification des consonnes pharyngales et glottales

Il s'agit pour nous de reconnaître les consonnes /?/, /ε/, /ħ/ et /h/ à partir des connaissances acquises de notre propre expérience dans la lecture des spectrogrammes de parole. Ces connaissances sont décrites sous forme de règles de production. Une règle prend la syntaxe suivante :

R (numéro règle)
CONTEXTE-DROIT (liste de phonèmes)
CONTEXTE-GAUCHE (liste de phonèmes)
SI (prémisses)
PHONEMES (listes de phonèmes pondérés)
FIN

Chaque prémisses correspond à un indice visuel, que nous avons déduit lors de la lecture des spectrogrammes caractérisant un segment donné. La partie contexte droit et/ou gauche est optionnelle dans une règle.

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 /ħ/ 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 Règle sur /h/ ctx 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. Le symbole & correspondant à l'opérateur de la logique floue " au milieu de ".

3.3 Résultats et commentaires

Les consonnes pharyngales et glottales se présentent en nombre réduit dans la langue par conséquent, nous étions amenés, pour l'évaluation la reconnaissance du système sur ces consonnes, à construire un corpus de phrases contenant un nombre très important de consonnes pharyngales et glottales. Les résultats que nous indiquons sont calculés par comparaison avec l'étiquetage manuel de ces phrases pour 3 locuteurs masculins.

3.3.1 Résultats de la segmentation

L'évaluation de la segmentation 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. L'évaluation est résumée par le table ??.

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 /h/ 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.

3.3.2 Résultats de la reconnaissance

L'évaluation de la reconnaissance 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 où 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.

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 ???. Le taux global du système étant de 64% [3], nous pouvons conclure que le /h/ est très bien reconnu et que le score des autres phonèmes reste encore insuffisant. Pour le /h/, la limite inférieure du bruit de friction est un indice pertinent et discriminatoire.

4 Conclusion

Nous avons présenté une méthode de reconnaissance des consonnes pharyngales et glottales de l'Arabe standard en utilisant un système expert à règles de productions. L'acquisition des connaissances est faite après une analyse acoustiques de spectrogrammes de ces consonnes produites dans des contextes variés. Notre travail a permis une amélioration sensible du taux global de reconnaissance du système SAPHA [3].

Références

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

- [2] 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.
- [3] 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.
- [4] 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.
- [5] M. Djoudi, D. Fohr, and 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.
- [6] 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.