Une Base de données Etiquetée Formantiquement en Langue Arabe Standard

Imen Jemaa [1, 2], *Oussama Rekhis* [1], *Kaïs Ouni* [1] *and Yves Laprie* [2]

[1] Unité de Recherche Traitement du Signal, Traitement de l'Image et Reconnaissance de Formes (99/UR/1119)
Ecole Nationale d'Ingénieurs de Tunis, BP.37, Le Belvédère 1002, Tunis, Tunisia
imen_jemaa@yahoo.fr, oussamarekhis@gmail.com et kais.ouni@enit.rnu.tn
[2]Equipe Parole, LORIA-CNRS – BP 239 – 54506 Vandœuvre-lès-Nancy, France
Yves.Laprie@loria.fr

ABSTRACT

There has been a lack of standard databases needed for the quantitative evaluation of the automatic formant extraction techniques especially in Arabic language. We report in this paper our recent effort to create a formant labeled database in standard Arabic language. The manually Formant labeling is carried out used the Winsnoori tool. Furthermore, we present in this paper an exploratory use of the database to quantitatively evaluate the automatic LPC method implemented in the open source Praat using the hand edited formant trajectories as reference. We noticed that in some cases the formant track F3 presents poor results probably due to the low energy of the high frequencies

1. INTRODUCTION

Pour décrire le conduit vocal, on mesure généralement les caractéristiques des formants qui correspondent à un renforcement spectral crée par une cavité du conduit vocal (telle que la bouche, le pharynx…) agissant comme un résonateur. Etant donné que les formants sont des porteurs fondamentaux de l'information, le suivi de formant peut jouer un rôle important dans certaines disciplines du traitement automatique de la parole. Il peut être très utile dans l'identification phonétique, en particulier celles des voyelles [Thi03] et autres sons vocaliques [Ahm02], le pilotage des synthétiseurs, la reconnaissance [Thi03] et le codage de la parole. Bien que le suivi automatique des formants soit un domaine d'applications très large, il est encore un problème ouvert en analyse de la parole. En particulier, lorsque les anti-formants sont présents comme dans la plupart des sons consonantiques où les fréquences de résonances, formants, sont souvent cachées.

En raison de l'importance des résonances de l'appareil vocal, de nombreux travaux ont été consacrés à élaborer des méthodes automatiques de suivi des formants dont la plupart sont basées sur la détection des racines de LPC [McC74] comme estimation initiale des fréquences des formants. Les résultats de plusieurs de ces méthodes ont été utilisés dans des applications de traitement de la parole. Cependant, il y a un manque manifeste de bases de données qui sont nécessaires pour l'évaluation quantitative de ces méthodes, en particulier pour la langue arabe.

D'où nous avons eu l'idée d'enregistrer et d'étiqueter en termes de formants un corpus en langue arabe standard.

La langue arabe standard a été bien étudiée par les phonéticiens arabes il ya bien des siècles. Son système phonétique est constitué essentiellement par 34 phonèmes qui se composent de 6 voyelles et 28 consonnes [Gha77][Bra97].

Le système vocalique de la langue arabe est composé de 6 voyelles qui sont 3 voyelles courtes (Haraka: t ou les signes diacritiques), qui sont opposées, par leur durée, par 3 voyelles longues (MADD). Les voyelles courtes sont à savoir: fathah ou / a /, kasrah ou / i / et dhammah ou / u /, et les voyelles longues sont: alif al-MADD ou / a: /, ya al-MADD ou / i: / et waw el-MADD ou / u: /.

À propos des consonnes, on peut classer ces phonèmes dans des classes différentes selon la manière de les prononcer: occlusive, fricative, trille, latérale, et approximante nasale ou semi-voyelle.

Le système phonétique de la langue arabe comprend cinq types de syllabes qui sont CV, CVV, CVC, CVCC et CVVC classées en fonction de traits ouverts et fermés [Bra97]. Une syllabe est ouverte (respectivement fermée), si elle se termine par une voyelle V (respectivement par une consonne C). On peut aussi classer les syllabes en fonction de leur longueur.

Nous présentons dans ce papier une élaboration d'un corpus Arabe étiqueté manuellement en termes de formants et une étude exploratoire de l'utilisation de ce corpus pour évaluer quantitativement la méthode automatique LPC mise en œuvre dans le logiciel Praat en utilisant les trajectoires formantiques éditées manuellement comme référence.

Ce papier est présenté comme suit. Dans la section 2, nous présentons le corpus, dans la section 3, l'étiquetage phonétique manuel, dans la section 4, les différentes étapes du processus d'étiquetage manuel des formants, dans la section 5, les résultats de l'étude et l'évaluation de la méthode automatique LPC (Praat) en prenant notre suivi de formants étiqueté manuellement comme référence. Enfin, nous donnons quelques perspectives dans la section 6.

2. DESCRIPTION DE CORPUS

Pour élaborer notre corpus, nous avons utilisé une liste de phrases arabes phonétiquement équilibrées proposée par Boodraa et al. [Bou00]. Ce corpus est également basé sur des études de Moussa [Mou73] à propos de la langue arabe standard à l'égard de la fréquence d'apparition de chaque phonème. Ainsi, il couvre l'ensemble de réalisations phonétiques et phonologiques de la langue arabe standard. La plupart des phrases de cette base de données sont extraites du Coran et Hadith. Elle est constituée de 20 listes chacune constituée de 10 phrases courtes. Chaque liste est constituée de 104 CV (C: consonne et V: voyelle), c'est-à-dire 208 phonèmes [Bou00].

Nous avons enregistré ce corpus dans une chambre sourde pour dix locuteurs tunisiens (cinq hommes et cinq femmes) dont leur âge varie entre 22 et 30 ans. Le signal est numérisé à une fréquence de 16 kHz. Ce corpus contient 2000 phrases (200 phrases prononcées par chaque locuteur) soit affirmative ou interrogative. De cette façon, la base de données présente une sélection équilibrée de locuteurs, de sexes et de phonèmes. Toutes les phrases du corpus sont riches en contextes phonétiques et sont ainsi une bonne collection de phénomènes phonétique-acoustique qui mettent en œuvre d'intéressantes trajectoires formantiques [Den06].

3. ETIQUETAGE PHONÉTIQUE MANUEL

Dans le but d'obtenir un étiquetage formantique précis, nous devons vérifier chaque fréquence formantique à l'égard du phonème prononcé au niveau de la phrase. Cela est nécessaire lorsque le signal d'entrée présente une faible énergie au niveau du spectrogramme. Ainsi, pour préparer notre base de données, nous avons phonétiquement annoté toutes les phrases du corpus à la main en utilisant de logiciel Winsnoori [Winsn]. Une capture d'écran de cet outil est présentée ci-dessous dans la Fig.1 montrant l'annotation phonétique de la phrase «هَلْ لذَعَتْهُ بقَوْلٍ؟» «Hal laðaʕathu biqawlin» prononcée par un locuteur. Nous avons utilisé pour cette annotation phonétique le code SAMPA [Sampa]. À ce stade de travail, nous avons rencontré plusieurs difficultés et surtout pour les consonnes. Une fois l'annotation terminée, le corpus a été révisé par des phonéticiens pour corriger toutes les erreurs qui ont été faites. Nous avons passé beaucoup de temps pour avoir une annotation phonétique précise. Cette étape est très importante pour obtenir un bon suivi de formants comme référence.

4. ETIQUETAGE FORMANTIQUE MANUEL

Pour faciliter le processus d'étiquetage formantique de notre corpus, nous avons d'abord obtenu un ensemble de fréquences formantiques candidates fournies par les racines de LPC [McC74] à l'aide de logiciel Winsnoori [Winsn]. Sur la base de ces valeurs candidates estimées, nous avons édité les trajectoires des formants à la main à l'aide du curseur de la souris. La Fig.1 montre un exemple d'une phrase prononcée par un locuteur illustrant le processus de l'étiquetage des formants et les résultats. Nous avons suivi et enregistré les trois premiers formants (F1, F2 et F3) toutes les 4 ms pour chaque phrase de la base de données. L'ordre de prédiction utilisé en LPC est 18 et la fenêtre d'analyse utilisée est de 16 ms. La durée temporelle de la fenêtre d'analyse spectrale est de 4 ms pour avoir un spectrogramme à large bande qui montre mieux l'évolution des trajectoires des formants. Des situations plus difficiles se posent dans certains cas quand il ya trop de racines candidates de LPC, qui sont proches les unes des autres pour deux formants. Mais la plupart des difficultés se manifestent pour les cas, où il ya une faible énergie au niveau du spectrogramme ou lorsque les protubérances spectrales ne coïncident pas avec les prévisions des fréquences des résonances et particulièrement pour les segments consonantiques. Pour surmonter ces difficultés, nous avons prévu des valeurs nominales spécifiques aux voyelles ainsi que les consonnes. [Gha77] [Bra97].

Enfin, afin de s'assurer de l'exactitude des trajectoires des formants de chaque phrase, nous avons synthétisé le son avec les trois premiers formants utilisant le synthétiseur Klatt mis en œuvre dans Winsnoori [Winsn] pour vérifier si le signal synthétisé correspond bien à l'original c'est-à-dire le signal d'entrée. L'évaluation a été tout d'abord subjective puisque les auteurs sont les seuls juges de la qualité des résultats lors de l'écoute du signal synthétisé ensuite on a passé à une évaluation objective en vérifiant la précision des trajectoires formantiques étiquetées avec des experts arabes en phonétique.

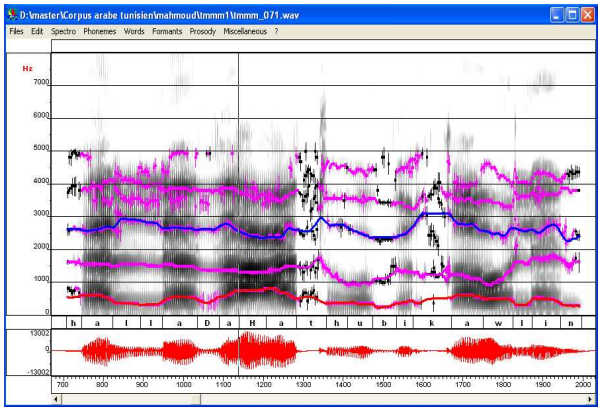

Figure 1 : *Etiquetage phonétique et formantique de l'enregistrement* «هَلْ لذَعَتْهُ بقَوْلٍ؟» (*« Hal laðaʕathu biqawlin »*) (*Lui a t – elle offensé par des paroles*) *prononcé par un locuteur.*

5. ETUDE ET ÉVALUATIONS

Nous avons tenté dans ce papier de faire une étude exploratoire en utilisant notre corpus étiqueté manuellement comme référence pour évaluer quantitativement la méthode LPC automatique mise en

œuvre dans le logiciel Praat [Praat]. Pour permettre d'abord une comparaison visuelle, les Fig.2 et 3 montrent le suivi automatique des formants (F1, F2 et F3) pour la phrase «آتُوْذِيهَا بآلامِهِمْ؟ » (« 3atu3Di:ha: bi3a:la: mihim») qui signifie *(Est-ce que tu souhaites la blesser avec leurs douleurs*) prononcée par une locutrice. La Fig.2 correspond à l'étiquetage référence obtenu à la main, et la Fig.3 correspond au suivi automatique des formants obtenu par la méthode LPC. L'ordre de prédiction de LPC utilisé par défaut au logiciel Praat est 16 et la fenêtre d'analyse utilisée est de 25 ms. La taille de la fenêtre d'analyse spectrale est de 5 ms pour avoir un spectrogramme à large bande. On peut voir que pour la plupart des segments vocaliques qui se manifestent par des parties denses en énergie au niveau du spectrogramme sont clairement identifiables, donc à ce niveau là on obtient bien un bon suivi pour la méthode automatique LPC. Les exceptions sont des erreurs occasionnelles en F2 et F3 et en particulier pour la voyelle longue / I /, contrairement à la voyelle courte / i / où on a un bon suivi pour les trois formants probablement à cause de la faible énergie de /I/ par rapport à /i/.

Pour évaluer quantitativement la méthode de suivi automatique LPC, nous avons utilisé notre corpus étiqueté comme référence en calculant la différence absolue moyenne (Eq.1) et l'écart-type normalisé par rapport aux valeurs de références (Eq.2) pour chaque trajectoire formantique (F1, F2 et F3). Nous avons donc étudié les résultats obtenus pour la voyelle courte / a / au sein de la syllabe CV. La (Table.1) montre les résultats obtenus sur la voyelle / a / précédée d'une consonne de chaque classe phonétique et pour les trois formants (F1, F2 et F3). La collection des différentes CV a été prise des deux phrases suivantes prononcées par un locuteur: «عَرَفَ وَالِيًا وَقَائِد» (« ɛarafa wa:liyan wa qa:3idan ») qui signifie (Il connaissait un gouverneur et un commandant) et « هِيَ هُنَا لَقَدْ آبَتْ» (« hiya Huna: laqad 3a:bat ») qui signifie (Elle est ici et elle était pieux).

$$Diff = \frac{1}{N} \times \sum_{p=1}^{N} \left| F_r(p) - F_c(p) \right| \ \text{Hz} \qquad (1)$$

Avec F_r est la fréquence de référence, F_c la fréquence calculée correspondant à la méthode LPC et N le nombre total de fréquences de chaque suivi de formant.

$$\sigma = \sqrt{\frac{1}{N} \sum_{p=1}^{N} \left(\frac{\left| F_r(p) - F_c(p) \right|}{F_r} \right)^2} \qquad (2)$$

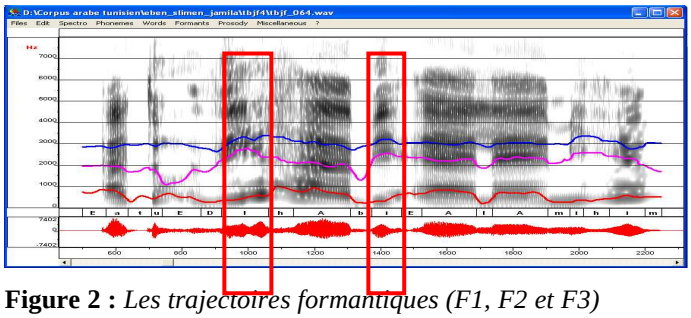

Figure 2 : *Les trajectoires formantiques (F1, F2 et F3) étiquetées manuellement de l'enregistrement «آتُوْذِيهَا بآلامِهِمْ؟" » (« 3atu3Di:ha: bi3a:la:mihim ») (Est-ce que tu souhaites la blesser avec leurs douleurs) prononcé par une locutrice.*

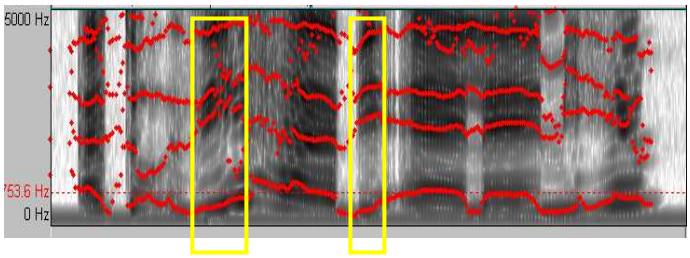

Figure 3 : *les trajectoires formantiques (F1, F2 et F3) par la méthode LPC (Praat) de l'enregistrement « آتُوْذِيهَا بآلامِهِمْ؟ » (« 3atu3Di:ha: bi3a:la:mihim ») (Est-ce que tu souhaites la blesser avec leurs douleurs) prononcé par une locutrice.*

Table 1: Les erreurs de suivi des trois premiers formants F1, F2 et F3 mesurées par le calcul de la différence absolue moyenne (Hz) et l'écart type normalisé par rapport aux valeurs de référence (%) pour la voyelle courte /a/ au sein de la syllabe CV avec différent type de consonne.

Suivi de		F1	F2	F3
Occlusive voisée :	Diff	119	93	212
/d /	σ	42	37	65
Occlusive non voisée:	Diff	71	58	67
/q/	σ	16	11	14
Fricative voisée:	Diff	49	48	36
/H/	σ	8	7	6
Fricative non voisée:	Diff	44	31	79
/f/	σ	14	8	24
Nasale:	Diff	97	30	101
/m/	σ	22	7	25
Latérale:	Diff	76	34	45
/l/	σ	24	8	11
Trille:	Diff	57	68	246
/r/	σ	13	16	59
Semi-voyelle:	Diff	98	91	140
/w/	σ	31	49	52

La comparaison des valeurs figurant dans la (Table.1) montre que pour certains cas il y a de grandes différences en termes d'erreurs entre la méthode de suivi automatique LPC et l'étiquetage de référence et cela lorsque /a/ est précédée par 1) une occlusive voisée en F1, F2 et F3, 2) une nasale en F1et F3, 3)une trille en F3 et 4) une semi-voyelle en F1, F2 et F3 et pour les autres cas, globalement, la méthode présente de bons résultats

Les (Tables 2 et 3) présentent les erreurs mesurées par la différence moyenne absolue et l'écart type normalisé par rapport aux valeurs de référence. La collection des différentes voyelles, c'est-à-dire les voyelles courtes (/ a /, / i / et / u /) et les voyelles longues (/ A /, / I / et / U /) a été prises depuis quatre phrases prononcées par deux différents locuteurs, respectivement deux différentes locutrices. Les phrases des tests effectués sont: « "هِيَ هُنَا لَقَدْ آبَتْ" , عَرَفَ وَالِيًا وَقَائِدً آتُوْنِيهَا بآلامِهِمْ؟ », citées ci-dessus et la dernière phrase est أسَرُونا بمُنْعَطف » (« ʒasaru:na: bimunɛatafin ») qui signifie (Ils nous ont capturés au niveau d'un virage).

Table 2: Les erreurs de suivi des trois premiers formants F1, F2 et F3 mesurées par le calcul de la différence absolue moyenne (Hz) et l'écart type normalisé par rapport aux valeurs de référence (%) pour chaque type de voyelle prononcées par deux différents locuteurs.

		Loc.Hom.1			Loc.Hom.2		
		F1	F2	F3	F1	F2	F3
a	Diff(Hz)	44	31	79	34	44	152
	σ	14	8	24	9	12	49
A	Diff(Hz)	52	91	89	58	114	63
	σ	14	45	29	13	23	14
i	Diff(Hz)	35	49	58	28	53	161
	σ	12	18	20	12	20	73
I	Diff(Hz)	42	67	64	64	47	83
	σ	13	19	21	20	14	25
u	Diff(Hz)	57	82	89	26	55	90
	σ	15	20	28	10	20	31
U	Diff(Hz)	65	83	265	86	191	215
	σ	19	22	64	40	125	103

Table 3: Les erreurs de suivi des trois premiers formants F1, F2 et F3 mesurées par le calcul de la différence absolue moyenne (Hz) et l'écart type normalisé par rapport aux valeurs de référence (%) pour chaque type de voyelle prononcées par deux différentes locutrices.

		Loc.Fem.1			Loc.Fem.2		
		F1	F2	F3	F1	F2	F3
a	Diff(Hz)	46	59	49	47	107	48
	σ	10	18	11	9	22	9
A	Diff(Hz)	93	115	100	44	51	113
	σ	14	21	14	6	7	14
i	Diff(Hz)	33	87	75	39	57	104
	σ	11	20	21	12	29	39
I	Diff(Hz)	32	110	185	38	445	289
	σ	11	41	66	10	137	78
u	Diff(Hz)	48	138	346	112	343	690
	σ	17	82	124	30	129	150
U	Diff(Hz)	36	96	182	43	90	158
	σ	10	29	58	13	24	46

La (Table.2) montre que les résultats de la méthode de suivi automatique LPC sont bons en particulier pour les voyelles (/ a /, / i / et / I /) contrairement aux voyelles (/ A /, / u / et / U /), probablement en raison de leur faible énergie. Nous avons également remarqué que la méthode présente de meilleurs résultats pour le locuteur (Hom.1) que pour le locuteur (Hom.2). La (Table.3) montre que les résultats sont bons pour les voyelles (/ a /, / i /) par rapport aux autres voyelles et particulièrement pour la locutrice (Fem.1) contrairement à la locutrice (Fem.2) où il ya des erreurs occasionnelles en F2 et F3. Cependant, les résultats ne sont pas très bons pour les autres voyelles prononcées par les deux locutrices. Enfin, on peut remarquer que pour certains cas le suivi en F3 présente des résultats médiocres probablement à cause de la faible énergie des hautes fréquences.

6. CONCLUSION

Dans ce papier, nous avons présenté l'élaboration d'un corpus étiqueté formantiquement en langue arabe standard. Les trajectoires des trois premiers formants de chaque phrase de ce corpus ont été étiquetées à la main et révisées après par des experts dans le domaine. Cette base de données est bien équilibrée en ce qui concerne la diversité des locuteurs, des sexes et des contextes phonétiques. En outre, nous avons exposé dans ce papier une étude exploratoire de l'utilisation de la base de données pour évaluer quantitativement la méthode de suivi automatique LPC en prenant notre corpus étiqueté comme référence. D'après les résultats trouvés nous avons constaté que pour certains cas le suivi en F3 présente des résultats médiocres probablement à cause de la faible énergie des hautes fréquences.

Nos futurs travaux viseront l'implémentation d'une nouvelle méthode automatique de suivi de formants.

6. REMERCIEMENTS

Ce travail est soutenu par le CMCU : Comité Mixte franco-tunisien de Coopération Universitaire (Projet de Recherche CMCU, code 07G 1112).

Nous sommes reconnaissants aux phonéticiens qui nous ont aidé dans ce travail : Selem Ghazeli, Abdelfeteh Brahem et Nedra Ben Slema de la faculté de lettres Menouba , Tunisie.

RÉFÉRENCES

[Thi03] Thibault F. (2003), "Formant trajectory Detection Using Hidden Markov Modals" *Rapport du projet NUMT609*, Sound Processing and Control Lab, McGill University, Montreal, Canada.

[Ahm02] Ahmed M. (2002) "Robust Auditory-based Processing using the Average localized Synchrony Detection", *In Proc. Of IEEE Trans. Speech and Audio Proc*, pp.279-292.

[McC74] McCandless S. (1974), "An Algorithm for automatic formant extrection using linear prediction spectra ", *In Proc. Of IEEE Trans.* Vol.22, pp. 135-141.

[Bou00] Boudraa M. (2000), "Twenty Lists of Ten Arabic sentences for Assessment ", *Act of Communication ACUSTICA*, Vol.86,pp.870-882.

[Mou73] Moussa A.H. (1973) "Statical study of Arabic roots on Moɛjam Al-Sehah", *livre*, Kuwait University.

[Den06] Deng L. (2006) "A Database of Vocal Tract Resonance Trajectories for Research in Speech Processing", *In Proc. Of ICASSP*.

[Gha77] Ghazeli S. (1977) "Back consonants and backing coarticulation in Arabic", *These*, University of Texas, Austin.

[Bra97] Braham A. (1997) "An Acoustic study of temporal organization in Arabic specific to Tunisian speakers", *These*, University of Manouba, Tunis.

[Winsn] http://www.loria.fr/~laprie/WinSnoori/.

[Sampa] http://www.phon.ucl.ac.uk/sampa/.

[Praat] http://www.praat.org/.