ANALYSE ET SYNTHESE DE LA PAROLE ARABE

Z. HAMAIZIA[(1)], M. BEDDA[(2)]

[(1)] Université Mohamed Khidder de Biskra Faculté des Sciences et de la technologie Département de génie électrique, B.P 145 RP, Biskra 07000 ALGERIE

[(2)] Université Badji Mokhtar de Annaba Faculté des Sciences de l'Ingénieur département d'électronique B.P12 Sidi Amar, Annaba 23000 ALGERIE

RESUME

Le signal de parole n'est pas un signal ordinaire, il est le vecteur d'un phénomène extrêmement complexe : communication parlée. Il est difficile de modéliser le signal de parole car ses propriétés statistiques varient au cours du temps. Sa redondance lui confère une robustesse à certains types de bruit. Il possède aussi une très grande variabilité. Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur, située au croisement du traitement du signal numérique et du traitement du langage. Nous nous sommes intéressés au traitement de la parole arabe par ordinateur car cette dernière n'a pas été l'objectif de beaucoup de travaux contrairement à l'anglais, le français, …etc.

Dans le cadre de ce travail, nous nous sommes consacrés à étudier l'analyse du signal vocal arabe par deux méthodes : l'autocorrélation et le cepstre, pour extraire quelques paramètres caractérisant ce signal tel que le pitch, puis exploitant ces paramètres pour réaliser la synthèse par LPC et comparer les résultats.

MOTS-CLES : Parole, analyse, synthèse, fondamental, formant, cepstre, autocorrélation, LPC.

1 INTRODUCTION

Le signal de parole n'est pas un signal ordinaire, il est le vecteur d'un phénomène extrêmement complexe : communication parlée. Il est difficile de modéliser le signal de parole car ses propriétés statistiques varient au cours du temps. Sa redondance lui confère une robustesse à certains types de bruit. Il possède aussi une très grande variabilité. Une même personne ne prononce jamais un mot deux fois de façon identique. La vitesse d'élocution peut varier, la durée du signal est alors modifiée. Toute altération de l'appareil phonatoire peut modifier la qualité de l'émission (par exemple rhume, fatigue). La variabilité interlocuteur est encore plus évidente, la hauteur de la voix, l'intonation, l'accent différent selon le sexe, l'origine social, régionale ou nationale [1, 2, 3, 4].

2 EXTRACTION DES PARAMETRES

L'analyse du signal vocal constitue dans tout système de reconnaissance, de synthèse ou de compréhension de la parole, l'étape préliminaire et primordiale. D'autre part l'analyse et la synthèse sont deux activités duales.

L'objectif principal d'analyse du signal vocal est d'extraire certains paramètres pertinents tels que : le voisement, le pitch et les formants. Les méthodes de traitement du signal vocal se sont de plus en plus affinés et sont des plus simples comme l'énergie ou le taux de passage par zéro vers les plus complexe comme le spectre court-terme, codage prédictif linéaire (LPC) et les modèles homomorphiques. On peut classer grossièrement les méthodes d'analyse en deux groupes : méthodes temporelles et méthodes spectrales.

2.1 Méthodes temporelles

Il existe différentes méthodes temporelles pour la détection du fondamental, citons : l'autocorrélation, l'autocorrélation modifiée, taux de passage par zéro, LPC et l'analyse mixte SIFT (Simplified Inverse Filter Tracking algorithm) [2, 5].

2.1.1 Analyse par autocorrélation

Cette méthode est basée sur la détection des maximas de la fonction d'autocorrélation d'un signal. Les positions de ces maximas nous informe sur l'existence du fondamental d'un signal. On calcule la fonction d'autocorrélation sur une tranche de N échantillons qui recouvre plusieurs périodes du fondamental. La détection de pitch par autocorrélation reste l'un des détecteurs robustes.

La procédure à suivre pour cette méthode peut être résumée comme suit (Figure.1) :

- Acquisition du signal y(n)=y (n.T_e)
- Pré-accentuation éventuelle par passage du signal dans un filtre de transmission : $(1-\mu.z^{-1})$; μ=0.95 cette opération vise à accentuer la partie haute fréquence.
- 3Segmentation en tranche dont la durée varie de 27 à 32ms.
- Application de la pondération consiste à pondérer la tranche par une fenêtre Hamming dont l'expression est la suivante : $w(n) = 0.54 - 0.46.\cos(\frac{2\pi n}{N})$ (1.1)
- Calcul de la fonction d'autocorrélation par la formule suivante : $r(k) = \sum_{n=1}^{n-1+k} x(n).x(n+k)$ (1.2)

L'autocorrélation travaille directement sur le signal temporel, elle est largement insensible à la variation de phase.

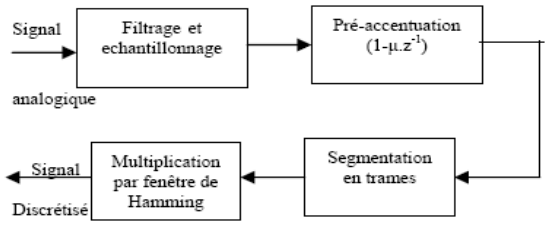

Figure 1: Mise en forme du signal vocal.

2.2 Méthodes spectrales

2.2.1 Analyse par la méthode cepstrale

L'opération « cepstre » est une méthode numérique permettant d'étudier séparément la source et le conduit vocal. Une façon simple de décrire le cepstre est de dire qu'il tend à séparer un composant harmonique fort du reste du cepstre [3, 4, 6]. Il existe deux méthodes de calcul des coefficients cepstraux:

Analyse spectrale.

Analyse paramétrique.

Analyse spectrale

Le signal de parole est modélisé en utilisant le spectre. Il est possible de calculer directement les coefficients cepstraux en suivant le processus décrit à la Figure.2 :

Il suffit pour obtenir les coefficients cepstraux de prendre le spectre (c.à.d. transformée de Fourier, puis passer dans le domaine logarithmique et enfin de prendre la fonction inverse de FFT (IFFT).

Le cepstre est basée sur une reconnaissance du mécanisme de la production de la parole. On part de l'hypothèse que la suite constituant le signal vocal est les résultats de la convolution du signal de la source par le filtre correspondant au conduit.

$$s(t)=u(t)*b(t) \quad (2.1)$$

avec s(t) : signal temporel, u(t) : le signal excitateur, b(t) : la contribution du conduit.

Le but du cepstre est de séparer ces deux contributions par déconvolution. Une transformation de Fourier permet de transformer la convolution en produit :

$$S(f)=U(f).B(f) \quad (2.2)$$

$$Log|S(f)| = Log|U(f)| + Log|B(f)| \quad (2.3)$$

par transformation inverse, nous obtenons le cepstre, enfin nous aurons une relation dans le domaine temps donnée par :

$$TF^{-1}(Log|S(f)|) = TF^{-1}(Log|U(f)|) + TF^{-1}(Log|B(f)|) \quad (2.4)$$

Les résultats du calcul cepstral est une séquence temporelle comme le signal d'entrée lui-même. Si le signal d'entrée possède une période de hauteur fondamental forte, elle apparaît dans le cepstre sous forme de pic, en mesurant la distance entre le temps zéro et le temps pic on trouve la période fondamentale de cette hauteur.

Engendré par un filtre dont il faut trouver les coefficients a(i). Ce modèle de production du signal de parole est appelé AR (autorégressif).

3 METHODES DE SYNTHESE

La synthèse vocale est une technologie qui permet d'automatiser la production d'une parole artificielle par une machine. L'objectif de ce travail est la synthèse vocale à partir d'une expression paramétrique. Les méthodes de synthèse sont aussi variées: la synthèse par LPC, la synthèse par formant et synthèse par vocoder.

4 SYNTHESE PAR LPC

La synthèse basée sur la prédiction linéaire est connue sous la domination LPC. L'excitation du modèle LPC est soit un train d'impulsion, soit un bruit blanc selon le voisement du signal vocal [2, 5, 8].

La modélisation du signal vocal prend en compte non seulement de la transmittance du conduit vocal mais aussi la forme réelle de l'excitation.

5 RESULTATS ET DISCUSSION

L'étude pratique est consacrée à la détermination des paramètres de commande d'un synthétiseur à prédiction linéaire, grâce à une analyse automatique de la parole. Pour justifier plus l'utilité de la précision de la fréquence fondamentale, on étudiera la variation de cette dernière comme caractéristique intrinsèque acoustique pour des

phonèmes arabes. Nous avons déterminé la fréquence fondamentale par une méthode temporelle (autocorrélation) et une méthode spectrale (cepstre) des voyelles arabes Figure.4, 5). Chaque phonème est prononcé par cinq hommes et cinq femmes, chacun a fait deux locutions. Dans le tableau n°1, on représente la moyenne du pitch. On se contenter de prendre deux phonèmes : un court (ِ) et le deuxième longue (ِي) qui est spécifique à l'arabe.

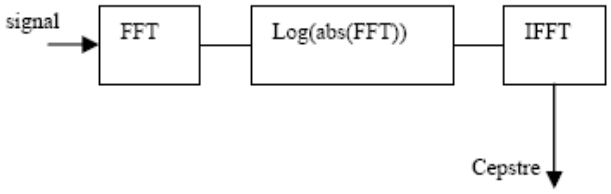

Figure 2: Calcul des coefficients cepstraux par analyse spectrale.

L'information prosodique est déterminée par la variation de F_0 fréquence fondamentale, ce paramètre est donc important pour la synthèse. L'estimation de F_0 et la décision de voisement/non voisement est un problème qui tient principalement aux raisons suivantes :

- L'excitation glottale n'est pas rigoureusement périodique.
- Il y a une interaction entre l'excitation et le conduit vocal.
- La segmentation des débuts et fin du voisement est difficile.
- La longueur de l'intervalle de mesure de F0 est difficile à choisir.

L'utilisation des impulsions du Dirac comme une excitation du filtre de synthèse donne des résultats plus accessibles que celle si on utilise une onde glottique pour raison que les impulsions de Dirac ont une périodicité plus forte par rapport à l'onde glottique.

On constate que la méthode d'autocorrélation donne des bons résultats pour la détection du pitch par rapport à la méthode cepstrale Cette dernière représente des inconvénients sachant que l'importance place mémoire qu'elle nécessite et le temps de calcul relativement long.

Ce qui nous conduit à conclure que les résultats de la synthèse utilisant le pitch par l'autocorrélation sont meilleurs (Figure 6).

Enfin, nous avons réalisé la synthèse de quelques mots arabes (ضرب, غرق) par la méthode de concaténation d'unité acoustique (Figure 8). Lorsqu'on veut comparer ces signaux on remarque qu'il est impossible d'obtenir un signal de synthèse égal échantillon par échantillon au signal original, c'est pourquoi nous sommes obligés d'utiliser certains critères de comparaison : comparaison au niveau spectral et comparaison au niveau perceptif. Ce dernier critère nous permet de décider si notre système est acceptable ou non.

Tableau 1: Les valeurs du fondamental Fo.

Phonème	Sexe	Autocorrélation	Cepstre
ِ	Homme	160	157
	Femme	286	276
ِي	Homme	167	163
	Femme	333	320

Les synthétiseurs à prédiction linéaire présentent le gros avantage, grâce à la représentation globale du signal de parole qu'il opère d'être commandés par des paramètres très facilement obtenus par l'analyse et de manière automatique. Les signaux synthétisés ont été entendus par cinq personnes différentes, ces tests de perception ont montrés que la parole est intelligible même le signal (signal/bruit) est faible.

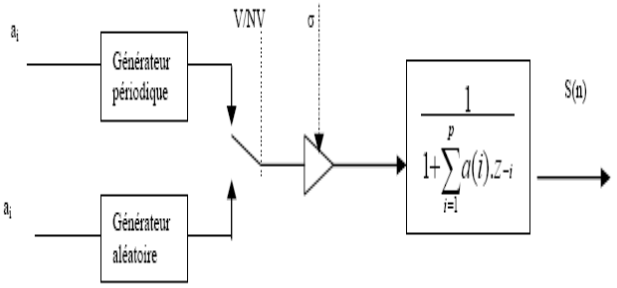

Figure 3: Modèle autorégressif de la production de la parole.

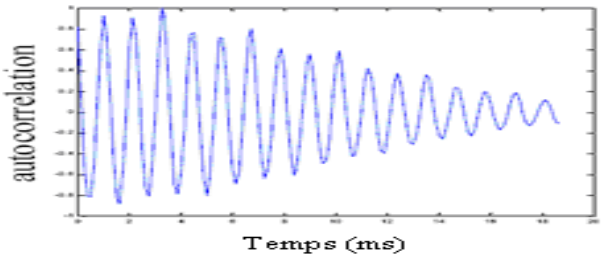

Figure 4: La fonction d'autocorrelation du phonème (ِ).

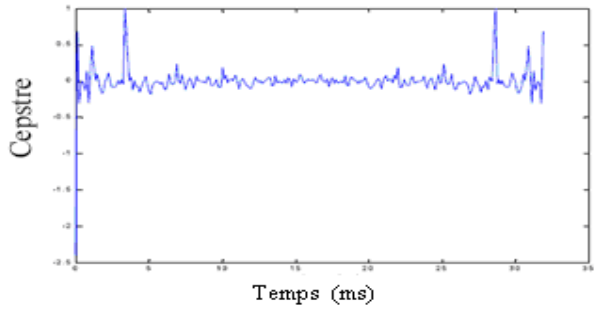

Figure 5: Le cepstre du phonème (ِ).

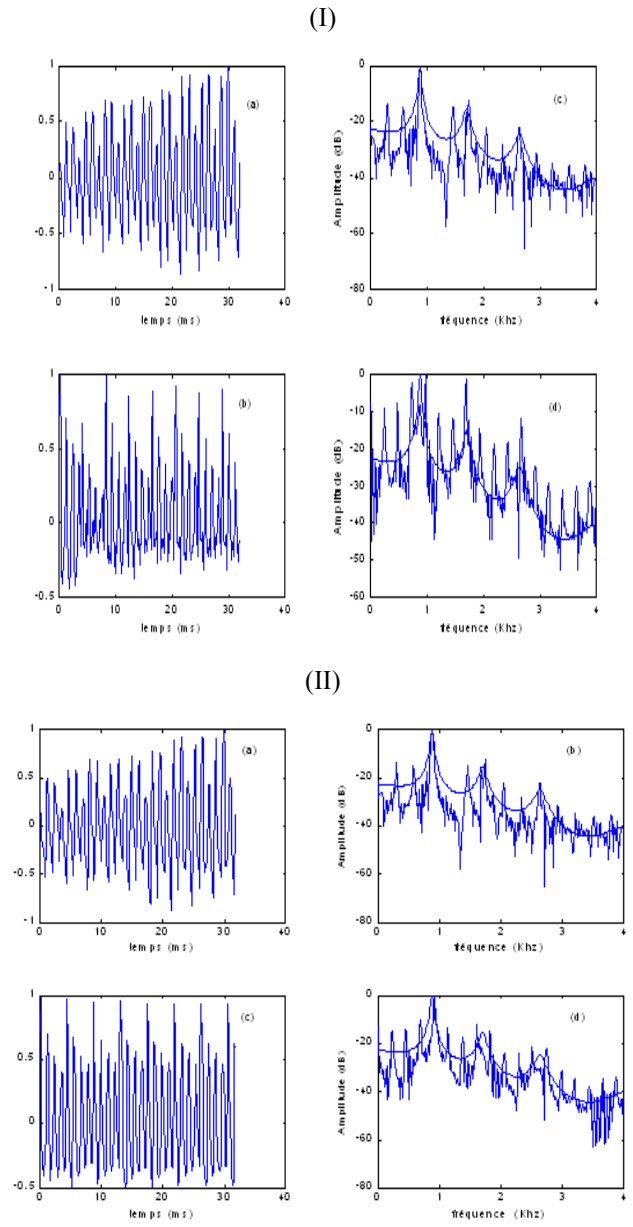

Figure 6: Synthèse de la voyelle 'ا' : (I) détection du pitch par l'autocorrelation, (II) détection du pitch par le cepstre, ((a), (d) : signal original et signal synthétisé respectivement ; (c), (b) le spectre et l'enveloppe du signal original et synthétisé respectivement ;

6 CONCLUSION

L'analyse du signal vocal n'est pas complète tant qu'on n'a pas mesuré l'évolution de la fréquence fondamentale, c'est un paramètre très important pour la synthèse de la parole.

L'estimation du pitch est bien sûr liée à la localisation des tranches voisées. Cette tache est difficile pour des nombreuses raisons telles que la non-stationarité du signal de parole, l'existence de certains irrégularités dans l'excitation glottique encore une interaction avec les formants.

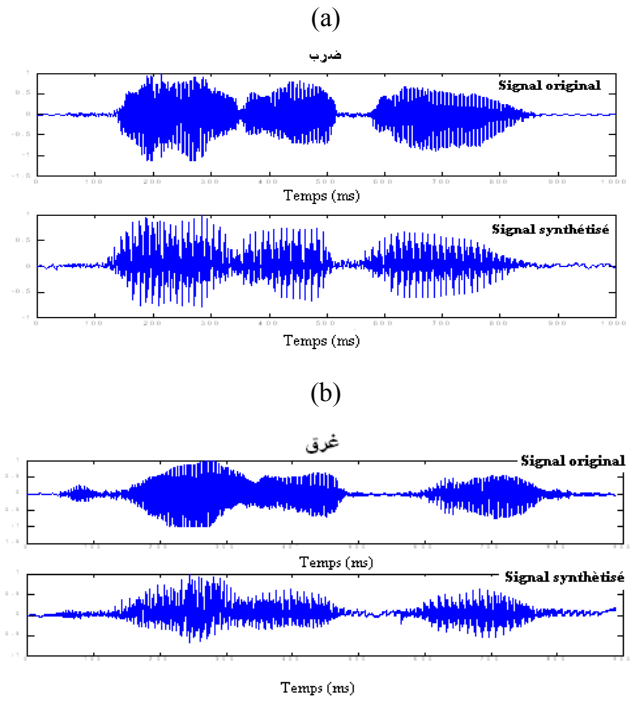

Figure 7: (a) Synthèse du mot arabe « ضرب »; (b) Synthèse du mot arabe « غرق ».

BIBLIOGRAPHIE

[1] C. Barras, "Reconnaissance de la parole continue adaptation du locuteur et contrôle temporel dans les modèles de Markov cachés", université de Paris, 1998.

[2] R. Boite, M. Kunt, "Traitement de la parole", Presse polytechnique romandes, 1987.

[3] Calliope, " La parole et son traitement automatique ", Paris, Masson, 1989.

[4] O. Deroo, " Modèles Dépendants du contexte et Méthodes de Fusion de données Appliquées à la reconnaissance de la Parole par Modèles Hybrides HMM/MPL ", Faculté Polytechnique de Mons, 1998.

[5] M. Taba, "Analyse-Synthèse de la parole par vocoder numérique à canaux", Université de Annaba, 1992.

[6] J. D. Reydellet, "L'audionumérique", 1998.

[7] M. Bacchiani, k. Aikawa," Optimatization of Time Frequency Masking Filters using the minum Classfication error criterion", IEEE, 1994, Japon.

[8] F.Bendjedi, Y. Meroud, "Analyse et Synthèse de la Parole par Prédiction Linéaire", université de Sétif, 1998.

[9] E.A. Echmants, L. Y. Pan & S. M O'Brien, "Automatic Feature Extraction from Spectrogram for Acoustic Phonitic

[10] Analysis", E .A. Edmands, L.Y Pan & S. M O'Brien, Lutchi Research Center.

[11] M .J. F Gales, S. J Young, "Cepstral parameter compensation of HMM recognition", cambridge Department.