

# Contribution à l'amélioration des signaux parole synthétisés par LPC en utilisant l'information de phase

Ali Khelfa<sup>2</sup>, Djamel Chikouche<sup>1</sup>, Khaled Rouabah<sup>2</sup>, Nouredine Amardjia<sup>2</sup>

<sup>1</sup>Département d'Electronique, Université de M'sila, Algérie

Laboratoire LIS, Département d'Electronique, Université de Sétif, Algérie

<sup>2</sup>Laboratoire LIS, Département d'Electronique, Université de Sétif, Algérie

E-mail : [khalfaali2000@yahoo.fr](mailto:khalfaali2000@yahoo.fr), [dj\\_chikou@yahoo.fr](mailto:dj_chikou@yahoo.fr), [rouabah\\_khaled@yahoo.fr](mailto:rouabah_khaled@yahoo.fr)

**Résumé** - Cet article présente une contribution à l'amélioration des signaux parole synthétisés par la technique de codage à prédiction linéaire LPC en exploitant l'information de phase issue de la transformée de Fourier à court terme lors de la phase d'analyse. Les performances de la technique proposée en termes de la qualité des signaux synthétisés sont analysées sur la base des tests effectués sur deux phonèmes prononcés par des locuteurs masculins. Les résultats obtenus démontrent la validité de cette approche.

## I. INTRODUCTION

Les méthodes sophistiquées qui réduisent des redondances dans le signal parole ont permis aux codeurs de la parole d'accomplir une qualité élevée à de faibles débits d'information. Certains algorithmes de compression de la parole de très faibles débits binaires emploient le codage à prédiction linéaire (LPC). Cette technique de codage a été largement utilisée par les chercheurs dans le domaine du traitement de la parole depuis 1970. Elle tient sa bonne réputation à sa rapidité de calcul et son faible débit binaire [1-3]. Toutefois, la parole synthétisée par LPC reste de mauvaise qualité relativement à celle obtenue par les célèbres techniques conventionnelles de codage numérique comme la MIC ou la MIC différentielle. Ces derniers demandent un haut débit binaire de l'ordre 32-64 kilobits/s et une faible complexité du codeur [1-15].

L'objectif de cet article est de contribuer à l'amélioration des signaux parole synthétisés par la LPC en exploitant l'information de phase issue de la transformée de Fourier à court terme. Les performances de la technique proposée en termes de la qualité des signaux synthétisés seront analysées sur la base des tests effectués sur deux phonèmes prononcés par des locuteurs masculins.

## II. ALGORITHME D'ANALYSE LPC

L'analyse de la parole est une étape indispensable à toute application de synthèse, et de codage, qui repose en général sur un modèle. Celui-ci possède un ensemble de paramètres numériques, dont les plages de variation définissent l'ensemble des signaux couverts par le modèle. Pour un

signal et un modèle donné, l'analyse consiste en l'estimation des paramètres du modèle en vue de lui faire correspondre le signal analysé. Pour ce faire, on met en oeuvre un algorithme d'analyse, qui cherche généralement à minimiser la différence, appelée erreur de modélisation, entre le signal original et celui qui serait produit par le modèle s'il était utilisé en tant que synthétiseur. Le modèle AR est plus important d'un point de vue pratique car il indique comment un signal à l'instant  $k$  dépend d'une manière linéaire d'un passé fini, plus un terme entièrement nouveau, non corrélé avec le passé, appelé innovation. Il est en relation directe avec la prédiction linéaire d'un signal. On peut estimer facilement les paramètres à l'aide des algorithmes d'estimation. Dans la section suivante, nous allons étudier la LPC et ses limitations.

### A. LE MODELE AUTOREGRESSIF (AR)

Fant a proposé en 1960 un modèle de production dont nous résumons ici la version numérique [1,2]. Un signal voisé peut être modélisé par le passage d'un train d'impulsions  $u(n)$  à travers un filtre numérique récuratif de type tout pôles.

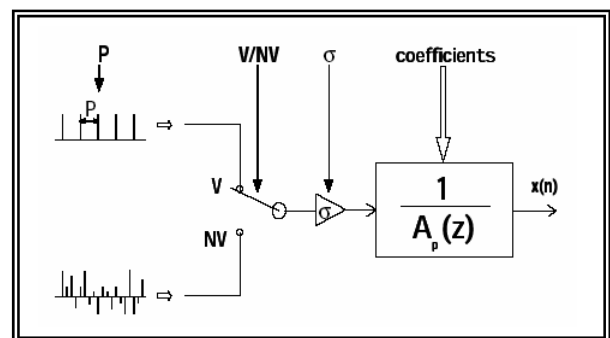


Fig 1. Le modèle LPC.

On montre que cette modélisation reste valable dans le cas de sons non voisés, à condition que  $u(n)$  soit cette fois un bruit blanc.

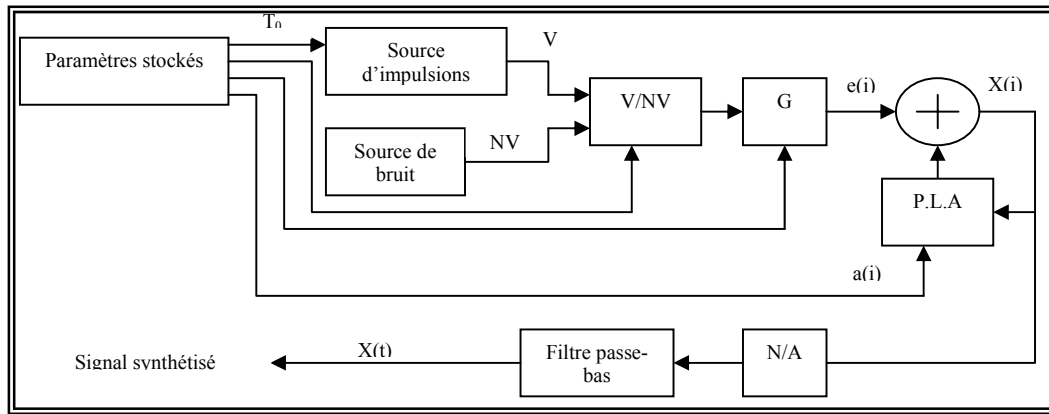


Fig 2. Schéma bloc d'un système de synthèse par LPC  
(PLA: Prédicteur Linéaire Adaptatif).

Le modèle autorégressif, illustré à la figure 1, correspond dans le domaine temporel à une régression linéaire de la forme :

$$x(n) = \sigma \cdot u(n) + \sum_{i=1}^p -a_i x(n-i). \quad (1)$$

(où  $u(n)$  est le signal d'excitation); ce qui exprime que chaque échantillon est obtenu en ajoutant un terme d'excitation à une prédiction obtenue par combinaison linéaire de  $p$  échantillons précédents [1-3].

Les paramètres du modèle LPC sont: la période du train d'impulsions (sons voisés uniquement), la décision Voisé/non Voisé ( $V/NV$ ), le gain  $\sigma$ , et les coefficients du filtre de synthèse  $1/A(z)$  (modèle AR). Ils sont estimés durant l'étape d'analyse.

#### B. ESTIMATION DU MODELE AR PAR LA METHODE DU TREILLIS

La structure en treillis est une conséquence directe de l'algorithme de Levinson-Durbin et s'applique uniquement au cas stationnaire. Le signal vocal est considéré comme quasi-stationnaire sur des intervalles de temps limités. Donc, en considérant des tranches successives, on peut utiliser cette méthode et estimer un modèle AR pour chacune des tranches de quelques dizaines de ms avec décalage d'une tranche à la suivante. En fait, une tranche n'est jamais isolée.

La méthode du treillis présente plusieurs avantages : La matrice d'autocorrélation  $R_x^{(p)}$  ne doit pas être calculée. Il n'est pas nécessaire d'appliquer une fonction de pondération aux échantillons de la tranche considérée. La stabilité du modèle AR est assurée,  $|K_m| < 1$ . Il est possible de quantifier les paramètres  $K_m$  au fur et à mesure qu'ils sont obtenus sans pour cela compromettre la stabilité et on peut déterminer les coefficients de prédiction  $a_p(i)$  ( $i = 1, 2, \dots, p$ ) et l'énergie résiduelle  $\alpha_p$ . Par contre, il est clair que cette méthode est coûteuse en temps de calcul. Les  $m$  premières cellules du treillis étant supposées connues, on détermine le coefficient  $K_{m+1}$  associé à la cellule suivante. La valeur choisie doit rendre minimale l'énergie de prédiction avant  $\alpha_{m+1}$  ou arrière

$\beta_{m+1}$ . Pour toutes ces raisons, la méthode du treillis est devenue une approche importante et fiable pour la mise en œuvre de l'analyse par prédiction linéaire.

#### C. LA SYNTHÈSE PAR LPC

Dans l'opération de synthèse, on procède par l'inverse de l'opération d'analyse [1,4]. Le signal de parole sera donc synthétisé autour des mêmes paramètres extraits de l'opération d'analyse. Les paramètres de contrôle fournis au synthétiseur sont : la période du pitch  $T_0$ , les  $p$  coefficients de prédiction  $a(i)$  ou de réflexion  $K(i)$ , les paramètres binaires de voisement ou non voisement ( $V/NV$ ), et le gain  $G$ .

Le générateur d'impulsions produit une impulsion d'amplitude unité à chaque période du pitch. Le générateur de bruit blanc produit une séquence d'échantillons aléatoires non corrélés de variance unité à chaque période d'échantillonnage. La sélection entre ces deux générateurs est faite par un commutateur de voisement/non voisement. L'amplitude de l'excitation est ajustée par un amplificateur de gain  $G$ . Pour former l'échantillon  $X(n)$  d'un signal de parole synthétisée, la valeur prédite linéairement  $X(n)$  du signal de parole est combinée avec le signal d'excitation  $U(n)$ . Les échantillons de parole synthétisée sont finalement convertis en analogique après passage par un filtre passe-bas pour obtenir, ainsi, le signal synthétique continu  $X_s(t)$ . Pour une parole voisée, les paramètres de contrôle du synthétiseur sont remis à leurs nouvelles valeurs au début de chaque période du pitch. Le problème majeur dans la synthèse de la parole par la méthode LPC est celui des transitions entre deux trames de synthèse. Pour palier ce problème, deux solutions sont à envisager : soit la synthèse asynchrone, soit l'interpolation linéaire.

#### D. TEST DE LA TECHNIQUE LPC SUR LES SIGNAUX PAROLE

L'analyse et la synthèse par LPC sont effectuées sur deux phonèmes prononcés par un locuteur masculin: la lettre "a" et le mot arabe "kataba". La fréquence d'échantillonnage est

fixée à 16000 Hz, la durée des segments de parole à 0.03 s et l'ordre du modèle AR à 12.

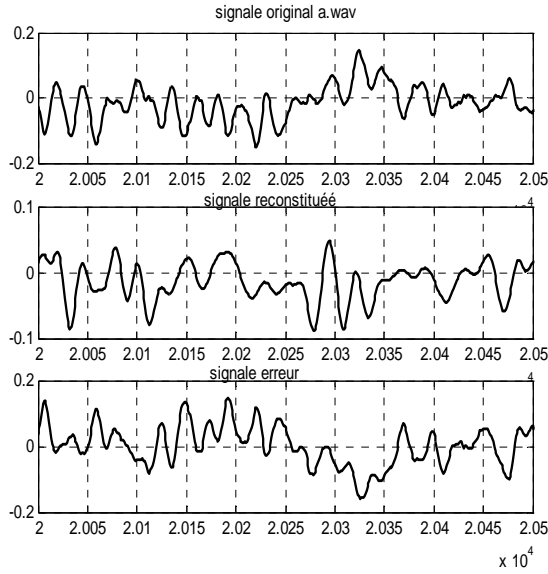


Fig 3. Analyse/synthèse LPC du phonème a.  
Erreur quadratique = 0.0715.

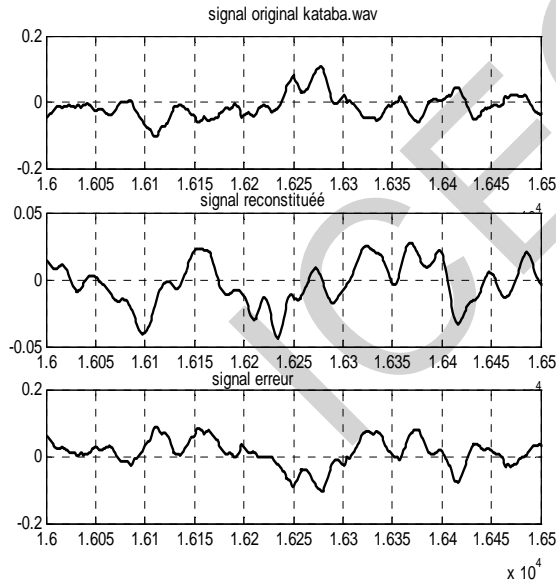


Fig 4. Analyse/synthèse LPC du phonème (kataba).  
Erreur quadratique = 0.0742.

Les tests de la technique LPC, effectués sur les différents signaux parole considérés montrent que la LPC permet de concentrer toute l'information contenue dans le signal à analyser sur un nombre restreint de paramètres LPC (coefficients de prédiction, coefficient de corrélation et fréquence fondamentale). Toutefois, les signaux synthétisés par LPC sont de faible qualité. Dans la section suivante, nous

allons présenter une version modifiée de la technique LPC qui associe l'information de phase en vue d'améliorer la qualité des signaux synthétisés.

### III. LA VERSION LPC AMELIOREE PAR L'ASSOCIATION DE L'INFORMATION DE PHASE.

Dans cette section, nous utilisons un algorithme pour la reconstruction du signal en se basant sur l'association de la phase du signal avec le spectre d'amplitude du modèle autorégressif AR obtenu lors de la phase d'analyse. La motivation de ce codage est basée sur le travail réalisé par Curtis et Atal [5, 6]. L'information du signe de la transformée de Fourier est essentiellement codée sur 1-bit.

#### A. ALGORITHME

Soit  $X(n)$ ,  $0 \leq n \leq N-1$ , une séquence de durée finie de longueur  $N$  et  $X(k)$  sa DFT à  $M$  points déterminée à partir de  $M$  fréquences distinctes telle que :

$$X(k) = |X(k)| \exp(j\theta_x(k)), \quad 0 \leq k \leq N-1$$

Où  $|X(k)|$  est l'amplitude de  $X(k)$  et  $\theta_x(k)$  sa phase.

Première étape:

Former la transformée de Fourier  $X_1(k)$  telle que

$$X_1(k) = |X'(k)| \exp(j\theta'_x(k))$$

Où  $|X'(k)|$  est l'estimation du spectre d'amplitude obtenu par l'analyse LPC, et

$$\theta'_x(k) = \begin{cases} \alpha + \pi/2 & \text{si } \alpha \leq \theta_x(k) \leq \alpha + \pi \\ \alpha - \pi/2 & \text{si } \alpha - \pi \leq \theta_x(k) \leq \alpha \end{cases}$$

$\alpha$  est choisi arbitrairement. Dans notre cas  $\alpha = \pi/4$ .

Deuxième étape:

Calculer la DFT inverse à  $M$  points de  $X_1(k)$ . Pour obtenir une estimation  $X_1(n)$ , on impose la contrainte suivante :

$$\begin{cases} X_2(n) = X_1(n) & \text{pour } 0 \leq n \leq N-1 \\ = 0 & \text{pour } N \leq n \leq M-1 \end{cases}$$

Troisième étape:

Calculer la transformée de Fourier  $X_2(k)$  de  $X_2(n)$  pour former  $X_3(n)$  comme suit :

$$X_3(k) = |X'(k)| \exp(j\theta'_x(k))$$

où  $\theta'_x$  est la phase de  $X_2(k)$ .

Quatrième étape:

En utilisant  $X_3(k)$  de la deuxième étape, une nouvelle estimation de  $X_3(n)$  est obtenue.

Les étapes 2, 3 et 4 sont répétées pour un certain nombre d'itérations fixé d'avance ou jusqu'à ce que un niveau spécifié de reconstruction soit atteint.

#### B. RESULTATS ET COMMENTAIRES

L'analyse et la synthèse sont effectuées sur les mêmes phonèmes.

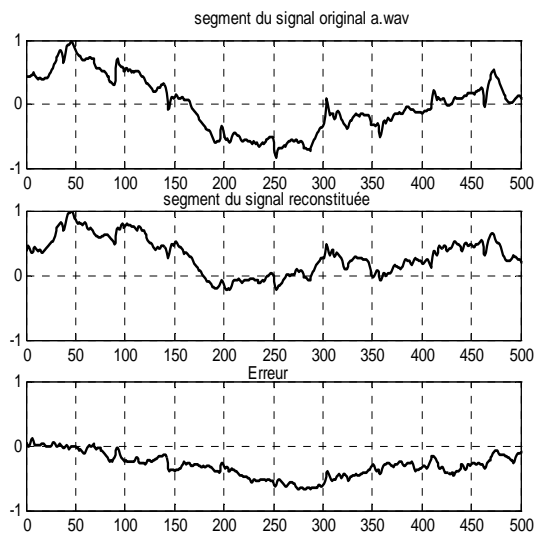


Fig. 5. Reconstruction du phonème (a).

Erreur quadratique = 0.0654.

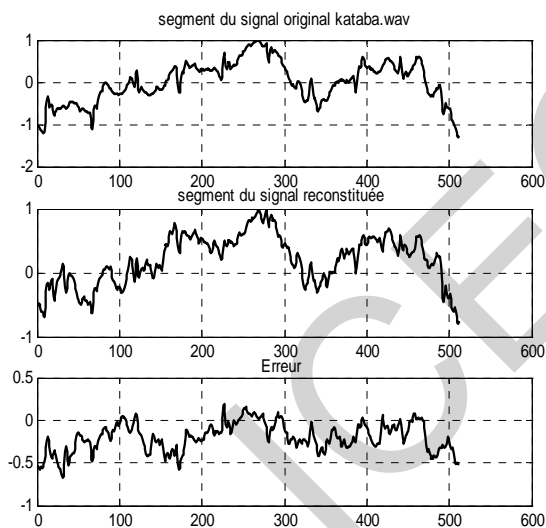


Fig. 6. Reconstruction du phonème (kataba).

Erreur quadratique = 0.0333.

Les tests de la technique LPC améliorée sur les phonèmes considérés montrent clairement l'amélioration considérable de la qualité des signaux synthétisés lorsqu'on associe l'information de phase.

#### IV. CONCLUSION

Dans cet article, nous avons étudié l'analyse /synthèse du signal parole au moyen de la technique de codage par prédiction linéaire LPC. Les tests de la LPC sur deux phonèmes prononcés par des locuteurs masculins montrent que la parole synthétisée par LPC est de très faible qualité.

Toutefois, l'association de l'information de phase avec le modèle LPC participe pleinement à améliorer la phase de synthèse de la parole par LPC donnant naissance à une parole de très bonne qualité.

#### REFERENCES

- [1] J. Wiley, "Speech coding algorithms foundation and evolution of standardized coders", published in Canada, 2003.
- [2] R. Kumaresan and D. W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise", IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP, vol. 30, N°6, Décembre 1982, pp.833-840.
- [3] I. Lim, and B. G. Lee, "Lossless Pole-Zero Modeling of Speech Signals", IEEE Transactions on Speech and Audio Processing, Vol. 1, No. 3, July 1993, pp. 269-276.
- [4] J. L. Flanagan, "Speech Analysis Synthesis and Perception", PhD Thesis, New York, Springer, 1972.
- [5] B. S. Atal, and J. R. Remde, "A new Method of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates", IEEE ICASSP, 1982, pp. 614-617.
- [6] S. R. Curtis, A. V. Oppenheim, and J. S. Lim, "Signal reconstruction from Fourier transform sign information", IEEE Transactions Acoustics, Speech and Signal Processing, vol. 33, June 1985, pp.643-657.
- [7] B. Yegnanarayana, S. Tanveer Fathima, H. A. Murthy, "Reconstruction from Fourier transform phase with applications to speech analysis", Proceedings IEEE International Conference on Acoustics, and Speech Signal Processing, Dallas, Texas, 1987, pp.301-304.
- [8] L. D. Alsteris, K. K. Paliwal, "Importance of window shape for phase only reconstruction of speech", Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004, pp. 573-576.
- [9] J. Samuelsson, and P. Hedelin, "Recursive Coding of Spectrum Parameters", IEEE Transactions on Speech and Audio Processing, vol. 9, No. 5, July 2001, pp. 492-503.
- [10] D. W. Griffin and J. S. Lim, "Signal estimation from modified short time Fourier transform", IEEE Transactions on Acoustics, Speech and Signal Processing, April 1984, vol. 32, pp. 236-243.
- [11] L. D. Alsteris, K. K. Paliwal, "ASR on speech reconstructed from short time Fourier phase spectra", Proceedings IEEE International Conference on Spoken Language Processing, October 2004.
- [12] A. Ramalingam, and S. Krishnan, "Gaussian mixture modelling using short time Fourier transform features for audio fingerprinting", Proceedings International Conference on Multimedia, Amsterdam, July 2005, pp.1146-1149.
- [13] K. Umapathy, S. Krishnan, J. Bejar and S. Jimaa, "Time-frequency modeling and analysis of wideband speech and audio", Proceedings Micronet Annual Workshop, Aylmer, Quebec, April 2002.
- [14] K. Umapathy and S. Krishnan, "Efficient compression of wideband audio signals using time-frequency functions", Proceedings. Annual Micronet Workshop, Aylmer, Quebec, April 2001.
- [15] L. D. Alsteris, K. K. Paliwal, "Iterative reconstruction of speech from short-time Fourier transform phase and magnitude spectra", Elsevier School of Microelectronic Engineering, Griffith University, Brisbane, Qld, Australia, May 2005, pp.1-13.