

Codage par Prédiction Linéaire (LPC) à Bas et très Bas Débit

Mohamed Saidi, Leila Falek, Bachir Boudraa

المُلخَص

يعرض هذا العمل بإيجاز للمرمّزين : CELP و MELP؛ بداية تمّ توضيح المفاهيم العامّة لترميز الكلام بنوعيه : ذي التدفّق المنخفض، وذي التدفّق المنخفض جدًّا، تاليا، جرى التّحرّي الواعي بتركيب الكلام، بواسطة المرمّزين المذكورين سالفًا، لغرض مقارنة تركيب الكلام ذي التدفّق المنخفض بنظيره الأدنى منه تدفّقًا. لذلك استعملنا المرمّزين المشتغلين تباعا على : 4.8 كيلو بايت/ الثانية بالنّسبة لـ CELP و 2.4 كيلو بايت/ الثانية بالنّسبة لـ MELP. **الكلمات المفتاح** : إشارة الكلام، ترميز الكلام، MELP، CELP.

Codage par Prédiction Linéaire (LPC) à Bas et très Bas Débit

Mohamed Saidi^{1,2}, Leila Falek², Bachir Boudraa²

¹Centre de Recherche Scientifique et Technique pour le Développement de la Langue Arabe (CRSTDLA), Alger, Algérie.

²Laboratoire de Communication Parlée et de Traitement du Signal (LCPTS), USTHB, Alger, Algérie.

mosaidi@usthb.dz, lfalek@hotmail.fr, b.boudraa@yahoo.fr

Résumé

Dans ce travail, nous avons donné une présentation brièvement sur ces deux codeurs CELP et MELP. En premier lieu, notre objectif était d'éclaircir d'abord, les notions générales du codage de la parole à Bas et très Bas Débit. En second lieu, synthétiser la parole avec une bonne perception avec ces deux codeurs afin de comparer la synthèse de la parole à bas et à très bas débit. Pour cela, nous avons utilisé les deux codeurs qui fonctionnent à 4.8 kbps pour le CELP et 2.4 kbps le MELP.

Mots clés: Signal de parole, Codage LPC, CELP, MELP.

1. Introduction

Tout système de codage de parole réalise un compromis parmi plusieurs contraintes. Idéalement, on voudrait un système capable de représenter le signal de parole avec un débit très faible, produisant un signal synthétisé d'une qualité transparente. Pour les applications de télécommunications, il faudrait aussi maintenir le délai du codage très court et assurer une parfaite robustesse contre les erreurs de transmission.

La réalisation d'une synthèse de la parole de haute qualité fonctionnant à bas débit et très bas débit exige un traitement par blocs qui sont codés comme une unité. Ces blocs sont appelés des trames et leur longueur varie habituellement entre 80 et 320 échantillons pour la fréquence d'échantillonnage de 8 kHz.

L'accumulation des échantillons nécessaires pour le traitement par trame augmente en général le délai de codage, la complexité de l'algorithme et aussi la demande en mémoire. La plupart des codeurs de parole modernes réalisent une modélisation paramétrique du signal sous la forme d'un signal d'excitation passant au travers d'un filtre, en exploitant d'une certaine manière les propriétés de la perception humaine. Le filtre, appelé filtre de synthèse, est généralement modélisé par la prédiction linéaire (LP). Le plus souvent, il s'agit d'un filtre autorégressif pur. Dans ce travail, concerne la compression du signal de parole numérique de bande téléphonique (200-3400 Hz). On va s'intéresser en particulier aux codeurs avec un débit de transmission entre 4.8 et 2.4 kbps et avec un objectif de qualité téléphonique du signal synthétisé (qualité de parole analogique en bande téléphonique). Ce débit étant trop bas pour une représenta-

tion fidèle de la forme d'onde temporelle, les techniques de codage cherchent à bien modéliser l'évolution du contenu spectral du signal de parole.

2. Model de synthèse par LPC

La parole peut être considérée comme étant un signal pseudo-stationnaire, c.-à-d. stationnaire sur de courtes durées allant en général de 5 jusqu' à 30 ms. Sur cette priode, il est possible de caractériser le spectre du signal par deux attributs :

- L'enveloppe spectrale ;
- La structure fine du spectre.

Le codage par prédiction linéaire se fonde sur un modèle fortement simplifié pour la production de la parole [1].

3. Prédiction Linéaire Excitation par Code (CELP)

Ce codage est une variété des codeurs MP-LPC. Il permet d'obtenir des débits encore plus faibles (<5 kbps). Le codage CELP (Code Excited Linear Prediction) a été introduit par Atal et Schroeder en 1985 [2]. Il fait parti des codeurs hybrides. Depuis, il n'a pas cessé d'être modifié.

Aujourd'hui, la majorité des systèmes de codage bas et moyen débit utilisent ce type de codage. Il est très efficace pour les débits intermédiaires de 4.8 kbps à 16 kbps, comme en témoignent les nombreuses normes qui l'utilisent en téléphonie. La figure 1 représente le principe du codage CELP [3]. Globalement les codeurs de type CELP modélisent le système de production de la parole en trois étages :

- Un étage d'excitation ;
- Un étage modélisant l'effet des cordes vocales ;
- Un étage modélisant la fonction de transfert du conduit vocal.

Le standard FS1016 travaille à 4.8 kbps avec des trames de 30 ms de longueur donc l'encodeur doit transmettre au décodeur 144 bits pour $c(n)$ chaque trame.

La distribution de ces 144 bits est illustrée dans la table 1.

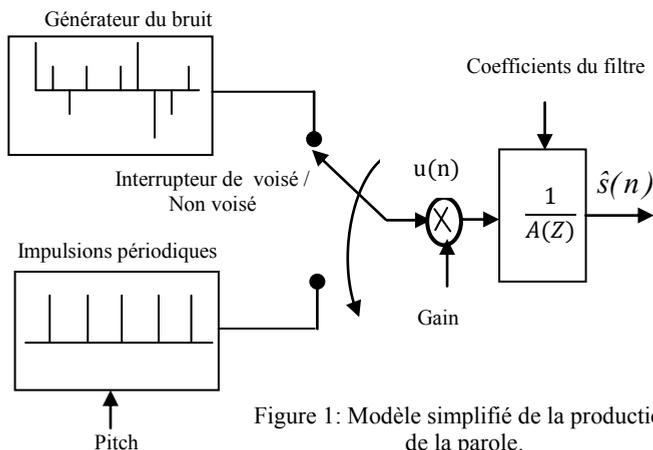


Figure 1: Modèle simplifié de la production de la parole.

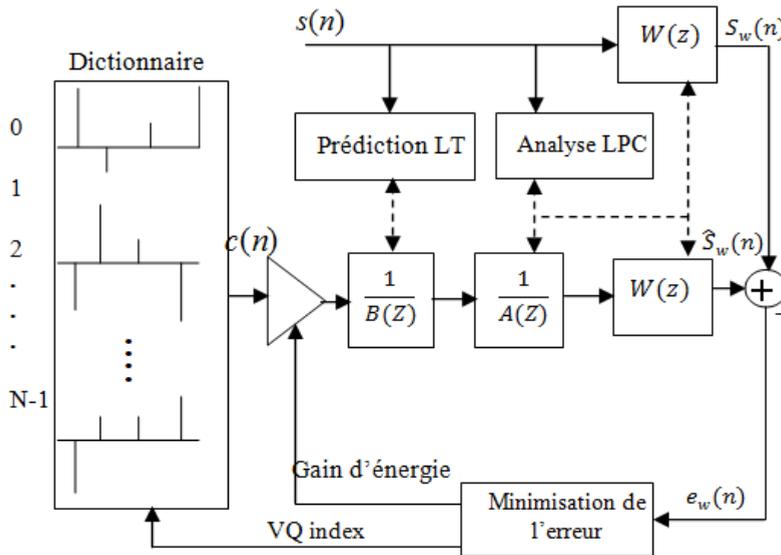


Figure 2: Schéma de principe du codeur CELP

Paramètre	Nombre de bit par trame
Fréq. d'échantillonnage	8khz
Taille de la trame	240 échantillons (30 ms)
Débit en trame	33,33trame/seconde
LPC	34
Pitch	28
Gain adaptative	20
Indice d'excitation	36
Gain stochastique	20
Synchronisation	1
Protection	4
Expansion	1
Total bits par trame	144 bits
Débit total	4800bps

Fréq : fréquence

Table 1. Le standard FS1016.

4. Vocoder MELP

Le codeur LPC à Excitation Mixte (MELP) est basé sur un modèle paramétrique, qui inclut cinq fonctionnalités améliorées comparativement aux codeurs LPC [4]. Celles-ci sont :

- Une excitation mixte ;
- Une impulsion aperiodique ;
- Amélioration spectrale adaptative ;
- Un filtre de dispersion d'impulsions ;
- Une Modélisation par les amplitudes de Fourier.

L'excitation mixte est formée de la somme d'une composante impulsionnelle et d'une composante de bruit en utilisant un mixage de filtres adaptatifs multi-bandes en vue de réduire le bruit introduit par le vocodeur LPC classique. Lorsque le signal de parole est voisé, le codeur MELP synthétise ce signal en utilisant soit un train d'impulsions périodique ou des impulsions aperiodiques. Ces dernières sont souvent utilisées dans les zones des transitions c'est-à-dire situées entre les segments voisés et les segments non-voisés du signal de parole. Ceci per-

met d'améliorer, lors du décodage, la reproduction des impulsions glottiques [5].

Paramètres	MELP à 2.4 kbps	
Fréq d'échantillonnage	8kHz	
Taille de la trame	180 échantillons (22.5 ms)	
Débit en trame	44,44 trames/seconde	
Mode de voisement	V	N/V
10 LSFs	25	25
Pitch	7	7
10 Amplitudes de Fourier	8	-
5 Bandes de voisement	4	-
2 Gains	8	8
Flag	1	-
Protection	-	13
Synchronisation	1	1
Total de bits par trame	54 bits	
Débit total	54*44,44= 2400 bps	

Table 2. Allocation des bits des codeurs MELP de 2.4 kbps.

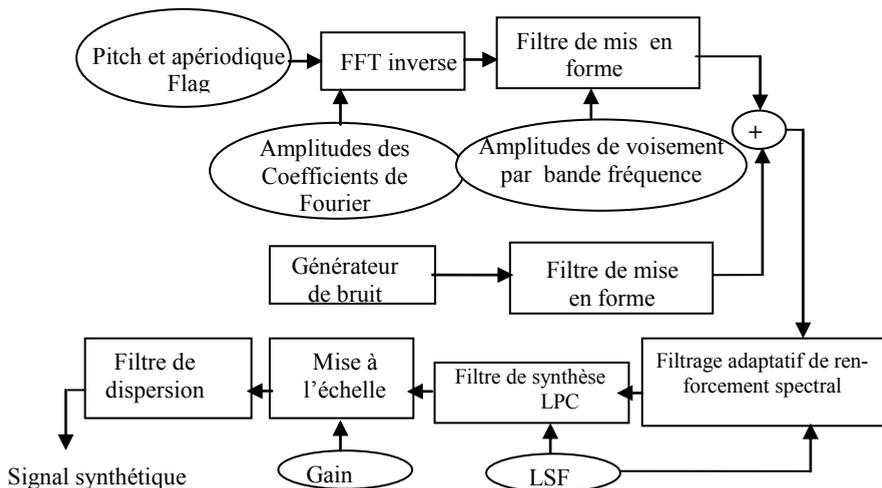


Figure 3: Schéma synoptique synthèse par le codeur MELP.

Cette excitation est une excitation multi-bande avec une intensité de voisement déterminée pour chaque bande de fréquence. Les paramètres transmis par le codeur MELP pour reconstituer la parole synthétique sont: la fréquence fondamentale (pitch), le flag (drapeau d'apériodicité), les cinq intensités de voisement, les deux gains (correspondant aux énergies de demi trames), les dix coefficients LPC transformés en LSF et les dix amplitudes de Fourier du pitch codées par une quantification vectorielle. Le tableau suivant précise l'allocation de bits de tous les paramètres, les 54 bits pour une trame codée à 2.4 kbps.

5. Evaluation perceptuelle de la qualité vocale (PESQ)

Pour tester la qualité de la parole obtenue par la méthode proposée, nous avons utilisé la méthode d'évaluation objective normalisée dite PESQ (perceptual evaluation of speech quality). Celle-ci a été décrite dans la recommandation P.862 de l'ITU-T (union internationale des télécommunications), [6]. La méthode PESQ est capable de modéliser les distorsions non linéaires engendrées par le codage ou par les procédures de masquage d'erreur. Très schématiquement, il est considéré comme un algorithme qui calcule la distance spectrale perceptuelle suivie d'un modèle « cognitif » qui permet de prendre en compte le fait qu'une dégradation n'a pas le même impact selon qu'elle est additive, soustractive, ou selon son contexte (segment de parole ou non) et sa distribution (localisée ou non).

6. Description de corpus de parole utilisés

Pour tester et valider notre méthode, nous avons utilisé un corpus de langue Arabe. Ce dernier est composé de phrases arabes

phonétiquement équilibrées (PAPE) [7] conçu au niveau de laboratoire LSPTS à l'USTHB. Ce corpus contient un total de 60 phrases, 10 phrases prononcées par 3 locuteurs féminins et 3 locuteurs masculins. La fréquence d'échantillonnage du signal parole des fichiers était de 10 kHz, nous avons dues effectuer un sous-échantillonnage de toute la base de données à 8 kHz, pour mettre dans les conditions de la téléphonie.

7. Evaluation des Résultats

Plusieurs simulations, ont été réalisées pour évaluer les performances de ces deux vocodeurs CELP et le MELP fonctionnant respectivement à 4.8 kbps et 2.4kbps. Des évaluations comparatives entre ces deux codeur CELP et MELP afin d'évaluer leur performances. Le but est de chiffrer la qualité perceptuelle de nos codeurs. Le tableau suivant le score du PESQ pour la base PAPE.

PESQ FS1016		PESQ MELP	
Féminin	Masculin	Féminin	Masculin
3.209	3.362	2.936	3.10

Table 3. Résultats des tests objectifs.

D'après les résultats obtenus, nous constatons que :

- Le score PESQ obtenu est meilleur pour les locuteurs masculins que les locutrices féminines pour les deux codeurs CELP et MELP ;
- La qualité synthétique du codeur CELP (FS1016) est meilleure que celle du MELP ;
- Dans les deux cas, l'intelligibilité synthétique est assurée par ces deux codeurs avec un niveau assez bon.

Nous donnons deux exemples de résultats où l'on voit le signal reconstitué avec le CELP (FS 1016) et le MELP (figures 4 et 5).

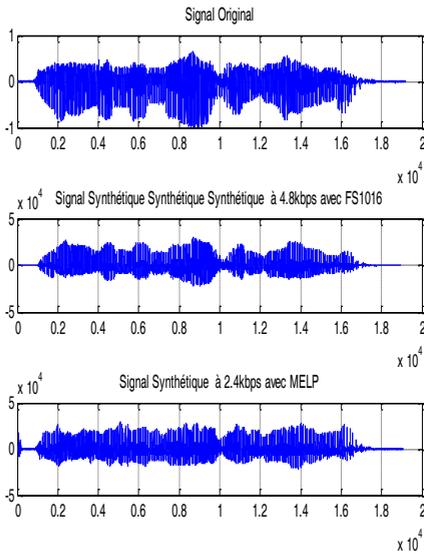


Figure 4: Phrase prononcée par un locuteur
’’ نعم ماء اليوم‘‘

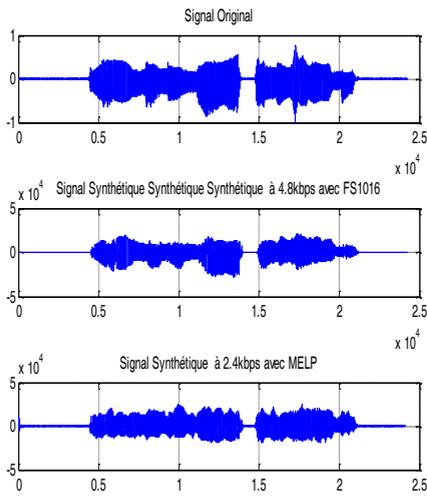


Figure 5: Phrase prononcée par une locutrice
’’ نعم ماء اليوم‘‘

9. Conclusion

Dans ce travail nous avons présenté une implémentation des deux codeurs MELP et FS1016. Par la suite nous avons montré du point de vue théorique la technique du codeur CELP appliqué au codeur FS1016. Les codeurs de type CELP sont des systèmes de codage mixtes qui utilisent à la fois des représentations paramétriques et temporelles du signal de parole.

Dans la seconde partie, notre objectif était focalisé envers l'étude du codeur MELP.

Le codeur MELP permet de surpasser les problèmes rencontrés dans le codeur vocal LPC classique. En plus, le MELP utilise un filtre d'amélioration spectral qui permet d'avoir une bonne concordance entre le son traité et le son original dans la région formantique. Ce qui contribue d'une manière significative dans la synthèse d'un son avec une bonne qualité.

Pour valider réellement les performances du codeur MELP pour la langue arabe, nous avons effectué une implémentation software de deux types de codeurs hybride et paramétrique nommés FS1016 à 4.8 kbps et le MELP 2.4 kbps que nous avons testé avec la base de données PAPE.

Pour comparer deux codeurs de différentes familles, le codeur FS1016 appartient aux codeurs hybrides qui préservent la forme d'onde et l'autre appartient aux codeurs paramétriques qui ne préservent pas la forme d'onde, mais il synthétise un signal perceptiblement similaire au signal d'origine. Pour cela on se réfère sur un fort critère d'évaluation qui est le PESQ, car il prend en compte plusieurs facteurs. Dans la phase finale, Les différents résultats d'évaluation objective avec le PESQ ont montré que les deux codeurs appartiennent à la bonne catégorie.

D'après le score PESQ obtenu, le codeur FS1016 présente une qualité d'écoute meilleure que le codeur MELP.

10. Références

- [1] Roy, G. and Eng, B., “Low-rate analysis-by-synthesis wideband speech coding”, Department of Electrical Engineering McGill université de Montreal, Canada, 1990.
- [2] Moreau, N., ”Codage prédictif du signal de parole à débit réduit : une présentation unifiée”, Annales Télécom, 46:223-239, 1991.
- [3] Rowe, D.G.,”Techniques for Harmonic Sinusoidal Coding”, July 1997.
- [4] Chu, W.C., “Speech Coding Algorithms Foundation and Evolution of Standardized Coders”, Wiley & Sons Inc., 2004.
- [5] Boudraa, B., “Analyse et Synthèse Multi-Impulsionnelle”, Thèse de doctorat d'état, USTHB, 2006.
- [6] ITU-T, Recommendation P.862, “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs”, 2001.
- [7] Boudraa, M., Boudraa, B. and Guérin, B., “Twenty lists of Ten Arabic Sentences for Assessment”, Acustica, 86: 870-882, 2000.