



HAL
open science

CODEUR CELP À COMPLEXITÉ RÉDUITE

M. Mauc, G. Baudoin

► **To cite this version:**

M. Mauc, G. Baudoin. CODEUR CELP À COMPLEXITÉ RÉDUITE. Journal de Physique IV Proceedings, 1992, 02 (C1), pp.C1-327-C1-330. 10.1051/jp4:1992170 . jpa-00251242

HAL Id: jpa-00251242

<https://hal.science/jpa-00251242>

Submitted on 4 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CODEUR CELP À COMPLEXITÉ RÉDUITE

M. MAUC et G. BAUDOIN

ESIEE Département Signaux et Télécommunications, BP. 99, F-93162 Noisy le Grand cedex, France

ABSTRACT

CELP coder enables speech coding with good quality at bit rates as low as 4 kps. Several methods have been developed to lessen the computation task.

We present a new multistage method that preserves the qualities of the CELP coder with reduced complexity. The basic idea is to find a subset S_1 of candidate codewords by approximating the synthetic signal energy. This approximation is obtained by restricting the frequency range and sub-sampling. A full exact search is done on this subset. The procedure can be iterated. Typically, for an initial codebook of 1024 words and a single stage of sub-sampling by a factor equal to $q_1 = 5$, the size of the subset S_1 is 70 and the calculation gain is 9. This performance can be improved up to a calculation gain of 13 by introducing several stages of sub-sampling $q_1 = 5$ and $q_2 = 2$.

1. INTRODUCTION

A la suite des travaux de M.R Schroeder et B.S Atal [Schroeder 85] sur le codage CELP (Code Excited Linear Prediction) de nombreuses structures ont été proposées dans la littérature [kleijn 90][Trancoso 86] en vue de réduire la complexité du système d'analyse par synthèse (fig.1).

Le filtre $H_0(z) = B(z) A(z/\gamma)$ est le filtre de synthèse sans

conditions initiales. Le filtre $B(z) = 1 - b z^{-M}$ est le prédicteur long-terme dont le rôle est de reproduire la structure périodique du signal de parole pour les sons voisés. M représente en échantillons la période du signal de parole [Ramachandran 89]. Le filtre $A(z)$ est le filtre LPC classique dont le rôle est de reproduire l'enveloppe spectrale court-terme du signal de parole. Le facteur $\gamma = 0.9$ permet de réduire l'amplitude des résonances du filtre LPC. Le filtre

$W(z) = A(z)/A(z/\gamma)$ est appelé filtre perceptuel; son rôle est de masquer le bruit par le signal de parole [Atal 82]. Le calcul des filtres $A(z)$ et $B(z)$ est classique. La recherche du mot de code optimal au sens du critère des moindres carrés dans le dictionnaire nécessite le filtrage de l'ensemble des séquences. Cette étape est la plus coûteuse en calculs.

Après soustraction de la mémoire des filtres \hat{p}^0 , une séquence du dictionnaire C_{n_j} sera choisie si le signal synthétique $\hat{p}^1_{n_j}$ maximise le rapport :

$$\frac{\langle p^1, \hat{p}^1_{n_j} \rangle^2}{\langle \hat{p}^1_{n_j}, \hat{p}^1_{n_j} \rangle} = \frac{\beta_j^2}{\alpha_j^2} \quad (1) \quad \text{où} \quad \langle u, v \rangle = \sum_{n=0}^{N-1} u_n v_n$$

$p^1 = p^0 - \hat{p}^0$ est le signal perceptuel sans mémoire.

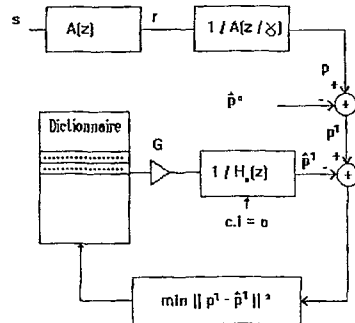


Figure 1 Codeur CELP

N est la durée de la trame.

On peut montrer que, par propriété du produit scalaire, le terme β_j peut se mettre sous la forme $\beta_j = \langle c_j, d \rangle$ (2) où d est calculée une fois pour toute et ne dépend pas des mots de codes du dictionnaire [Adoul 87].

Il reste toutefois à évaluer les normes α_j qui dépendent des mots de code du dictionnaire.

Nous présentons dans ce papier une nouvelle méthode pour le calcul des α_j cf(1). Cette méthode opère en plusieurs étapes de sous-échantillonnage et de tris successifs. Elle permet de réduire la complexité d'un facteur légèrement supérieur à 9 tout en n'augmentant que raisonnablement la taille de la mémoire nécessaire. Dans une première partie, nous décrivons notre méthode, dans une seconde partie, nous donnons les résultats obtenus.

2. NOUVELLE METHODE DE REDUCTION DE LA COMPLEXITE PAR SOUS-ECHANTILLONNAGE ET TRIS SUCCESSIFS.

2.1 Description de la méthode.

On supposera que le dictionnaire original est de taille Δ . Il ne sera fait aucune hypothèse quant au contenu du dictionnaire.

Dans une première phase, on recherche un sous-ensemble de M mots (de longueur N) ayant une très bonne probabilité de contenir la meilleure séquence CELP; cette première phase devant se faire de façon peu coûteuse en calculs.

Dans une deuxième phase, on explore le sous-ensemble M pour y trouver la meilleure séquence CELP, et ceci en effectuant tous les calculs du schéma CELP classique (fig. 2).

2.1.1 Structure à un étage

Première étape : Le calcul du numérateur β_j n'est pas modifié. La séquence d est calculée une fois pour toute en utilisant la technique du backward filtering [Adoul 87]. Le terme α_j est évalué différemment. Tous les mots de code ainsi que la réponse impulsionnelle du filtre de synthèse sont sous-échantillonnés par un facteur q (fe devient fe/q). Chaque sous-trame de taille N sous-échantillonnée par un facteur q sera donc de longueur N/q .

Le calcul de α_j nécessite le calcul d'une convolution d'un mot de code par la réponse impulsionnelle du filtre de synthèse. Dans ce cas, le nombre de multiplications est de $N(N+1)/2$. Pour le terme approximé, qui sera noté par la suite $\alpha_{j,q}$, le nombre d'opérations sera réduit à $N/q (N/q + 1)/2$ multiplications. Le nombre d'opérations est donc divisé par q^2 .

On cherche alors les M_q mots du dictionnaire de taille Δ qui maximisent le rapport $\beta_n^2 / \alpha_{j,q}^2$.

Deuxième étape : Parmi ces M_q séquences d'excitation, on cherche ensuite la meilleure séquence CELP de l'algorithme classique.

A l'étape de tri près, la complexité pour la méthode à un seul étage est approximativement égale à :

$$\text{complexité du CELP de base} \times \left(\frac{1}{q^2} + \frac{M_q}{\Delta} \right)$$

2.2.2 Structure à plusieurs étages

La méthode peut être cascadée de la façon suivante :

Sous-échantillonnage des différents signaux par un facteur q_1 (après filtrage anti-repliement du spectre)

Recherche des M_{q_1} meilleurs rapports $\beta_n^2 / \alpha_{j,1}^2$ dans le dictionnaire de taille Δ .

Sous-échantillonnage des M_{q_1} différents signaux par un facteur $q_2 < q_1$ (après filtrage anti-repliement du spectre).

Recherche des M_{q_2} meilleurs mots de code candidats parmi les M_{q_1} mots retenus précédemment

$$(M_{q_2} < M_{q_1} < \Delta)$$

Recherche de la meilleure séquence CELP sans sous-échantillonnage parmi les M_{q_2} mots retenus à l'étape précédente.

Si la méthode est utilisée en cascade, on peut schématiquement caractériser sa complexité aux étapes de tri près par :

$$CELP \times \left(\frac{1}{q_1^2} + \frac{M_{q_1}}{\Delta} \times \frac{1}{q_2^2} + \frac{M_{q_2}}{\Delta} \times \frac{1}{q_3^2} + \dots + \frac{M_{q_p}}{\Delta} \right)$$

et ceci pour une structure à P étages ayant comme facteur de sous-échantillonnage q_1, q_2, \dots, q_p successivement.

2.2.3 Taille mémoire

La taille mémoire nécessaire est augmentée modestement par rapport au CELP de base. En effet, en plus du dictionnaire original, il faut stocker le dictionnaire préfiltré et sous - échantillonné par un facteur q dans le cas d'un seul étage, ainsi que le dictionnaire préfiltré et sous - échantillonné pour différentes valeurs de q_i ($i = 1, \dots, p$) dans le cas de P étages. Pour P étages, avec les facteurs q_i la taille mémoire nécessaire est :

$$Taille \text{ du CELP original } \times \left(1 + \frac{1}{q_1} + \frac{1}{q_2} + \dots + \frac{1}{q_p} \right)$$

Ainsi, pour un étage avec $q=5$, la taille est multipliée par 1,2 et pour deux étages avec $q_1=5$ et $q_2=2$, la taille est multipliée par 1,7.

2.2.4 Remarque

Nous avons essayé d'appliquer le sous - échantillonnage globalement pour β_k et α_k de (1) dans l'espoir de sélectionner dans une première étape les mots du dictionnaire proche de l'optimal dans la plage $[0, f_s/2q]$. Mais les résultats sont moins convaincants que précédemment. Le nombre de séquences à conserver étant au moins deux fois plus grand qu'avec la méthode décrite. On l'explique par le fait que la fonction de transfert étant globalement décroissante, l'importance des fréquences supérieures à $f_s/2q$ sera généralement faible. Pour les trames où ces fréquences sont importantes, la vraie séquence CELP devra approximer au mieux ces zones de fréquences. Elle ne sera pas nécessairement optimale pour la bande de base.

2.2 Justification théorique

Le dénominateur α_q^2 correspond à l'énergie du signal synthétique. Avec la méthode appliquée ici, on commence par calculer l'énergie dans la plage de fréquence $[0, f_s/2q]$, ce qui constitue en général une bonne approximation de l'énergie totale, car le filtre de synthèse a une fonction de transfert décroissante avec la fréquence. Cette remarque n'est vraie que si q n'est pas trop grand (i.e $q > 5, f_s/2q > 800$ Hz). Pour calculer l'énergie dans la plage $[0, f_s/2q]$, il suffit de travailler à la fréquence d'échantillonnage f_s/q . Si la méthode est utilisée en cascade, l'étape suivante améliore l'approximation de α_q , en élargissant la plage de fréquence $[0, f_s/2q]$.

L'étape finale recherche parmi les séquences précédemment sélectionnées, la séquence qui optimise le rapport (1) en prenant en compte toute la plage de fréquence pour le calcul de α_q (recherche classique).

2.3 Description des conditions expérimentales.

La structure utilisée est celle de la fig.2 avec un filtre LPC d'ordre 10 (covariance stabilisée), un filtre LTP à un pas, un dictionnaire gaussien linéaire de taille $\Delta = 1024$ trames de 40 échantillons. La méthode a été expérimentée sur des signaux de parole de voix féminine et masculine échantillonnés à 8 kHz. Les différents tests ont été effectués sur plus de 2000 trames de 40 échantillons choisies dans 8 fichiers de voix différentes (bruits, son voisés, son non voisés, plosives).

2.4 Résultats obtenus : performances et complexité.

2.4.1 Evaluation de M_q

Trouver la séquence optimale dans le dictionnaire nécessite de calculer le terme (1) pour toutes les séquences du dictionnaire. Soit m l'index de la séquence optimale, le terme β_m^2/α_m^2 est donc le plus grand. Par contre, $\beta_m^2/\alpha_{m,q}^2$ ne sera pas nécessairement le plus grand. Après avoir calculé et trié tous les termes $\beta_l^2/\alpha_{l,q}^2$ pour chacune des séquences du dictionnaire, nous avons pu déterminer la taille du sous-ensemble M_q en regardant pour chaque trame la place occupée par $\beta_m^2/\alpha_{m,q}^2$. Dans le tableau suivant, nous donnons la taille de M_q pour que la probabilité que M_q contienne la séquence CELP

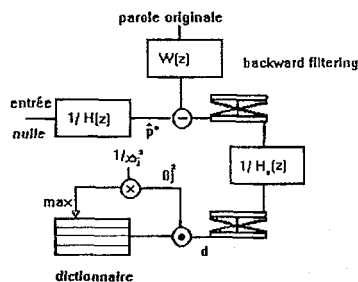


Figure 2 structure utilisé

optimale soit égale à 99%. (pour $q = 2,4,5,6$ (resp.) M_q sera noté M_2, M_4, M_5, M_6 (resp.)). Nous donnons également la taille de M_q pour être sûr à 99% de trouver l'une des deux meilleures séquences CELP au sens de (1). (dans ce cas M_q sera noté $M_q^{(2)}$). La figure 3 (4 resp.) donne la fonction de distribution de la probabilité de trouver la séquence CELP en fonction de $M_2, M_2^{(2)}$ ($M_5, M_5^{(2)}$ resp.)

| q | 2 | 4 | 5 | 10 |
|-------------|----|----|----|-----|
| M_q | 20 | 58 | 70 | 180 |
| $M_q^{(2)}$ | 13 | 22 | 28 | 104 |

2.4.2 complexité

La complexité dépend du facteur de sous-échantillonnage choisi : La taille de M_q augmente avec la valeur de q . La Figure 5 illustre la complexité globale pour différentes valeurs de q et en conservant à la fin de la première étape un nombre de séquences M_q tel qu'on ait une probabilité très grande (99%) d'y trouver la meilleure séquence CELP. (C'est le cas pour les 2000 trames testées).

Nous avons trouvé que $q = 5$ était la valeur la plus intéressante pour une structure à un étage. Dans ce cas, le gain en complexité est à peu près égal à 9 ($M_q = 70$, $\Delta = 1024$). Pour deux étages avec $q_1 = 5$ et $q_2 = 2$, nous avons obtenu un gain en complexité à peu près égal à 13 pour la recherche de la séquence d'excitation, aux étapes de tri près.

3. CONCLUSION

Nous avons présenté une nouvelle méthode de réduction de la complexité du codeur CELP. Elle consiste à effectuer la recherche de la meilleure séquence CELP en plusieurs étapes (au moins deux). Dans une première phase, on sélectionne un sous-ensemble dans le dictionnaire contenant avec une très grande probabilité la meilleure séquence CELP. Dans cette phase, on diminue la complexité du calcul de α_1^2 (l'énergie du signal synthétique) en l'approximant par sa valeur sur la bande basse des fréquences, ce qui permet de réduire la fréquence d'échantillonnage. La deuxième étape consiste à chercher la meilleure séquence dans le dernier sous-ensemble de mots du dictionnaire retenus. Dans cette phase, α_1 est calculé plus précisément en prenant en compte toute la plage de fréquence. On ne peut donc plus sous-échantillonner. Cette méthode permet de conserver les performances du codeur CELP original, tout en diminuant sa complexité par un facteur inférieur aux méthodes comme la décomposition en valeurs singulières. La taille mémoire est par ailleurs augmentée (moins que doublée). On peut de plus diminuer encore la complexité si l'on tolère un faible pourcentage de trames pour lesquelles le mot du dictionnaire sélectionné n'est pas l'optimum CELP, et ceci en limitant plus sévèrement la taille des séquences retenues.

Bibliographie

- [Adoul 87] J.P. Adoul, P. Mabileau, M. Delprat, et S. Morissette, "Fast CELP coding based on algebraic codes", Proc. int. Conf. Acoust., Speech, Signal processing (Dallas, Texas), 1987, pp. 1957-1960.
- [Atal 82] B.S. Atal, "Predictive Coding of Speech at low bit rates", IEEE Trans. On Com, vol. Com. 30, N° 4, April 1982, pp. 600-614
- [Kleijn 90] W.B. Kleijn, D.J. Krasinski, et R.H. Ketchum, "Fast methods for the CELP speech coding Algorithm", IEEE trans ASSP, vol. 38, No 8, August 1990, pp. 1330-1342.
- [Ramachandran 89] R.P. Ramachandran, P. Kabal, "Pitch Prediction Filters In Speech Coding", IEEE Trans. On Acous., Sp., and Sig. Proc., vol 37, No. 4, pp. 467-478, April 1989
- [Schroeder 85] M.R. Schroeder, B.S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates", Tampa (FL). Proc. ICASSP, vol. 3, pp. 937-940 1985
- [Trancoso 86] I.M. Trancoso, B.S. Atal, "Efficient procedures for finding the optimum innovation in stochastic coders", Proc. ICASSP (Tokyo, Japan), 1986, pp. 2379-2382.

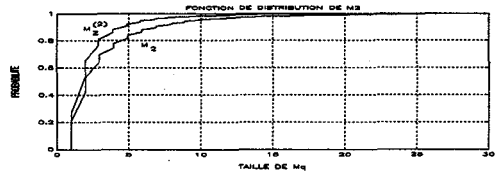


Figure 3

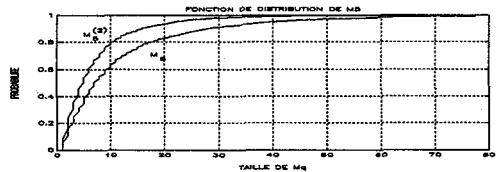


Figure 4

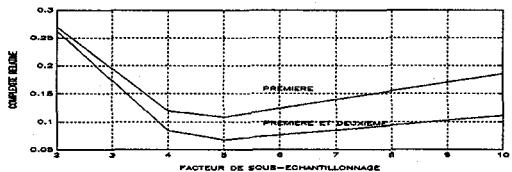


Figure 5