

CODAGE STOCHASTIQUE ADAPTATIF DE LA PAROLE

B.PUCCI, C.GALAND,
M.ROSSO

J.MENEZ

Centre d'Etudes et Recherches
IBM France
F-06610 La Gaude

LASSY équipe de l'URA CNRS 1376
41,Bd Napoléon III
F-06041 Nice cedex

RÉSUMÉ

L'apparition du codeur CELP (Code Excited Linear Predictive coder) introduit par B.S. Atal et M.R. Schroeder [1] a été le point de départ de nombreuses recherches dans le domaine de la compression de la parole. Cette technique qui fait appel à la quantification vectorielle et à la prédiction linéaire permet d'obtenir un signal de parole synthétique de bonne qualité pour un débit binaire compris entre 4 et 8 kbits/s. Dans cette communication on présente sous forme de synthèse, différents codeurs dérivés de la technique CELP de base: l'ACELP (Adaptive CELP), le VSELP (Vector Sum Excited Linear Prediction); et le codeur BACELP (Backward block adapted ACELP).

ABSTRACT

The introduction of CELP coder (Code Excited Linear Predictive coder) by B.S. Atal and M.R. Schroeder [1] has been the corner-stone of many researchs on speech compression. This technique uses vector quantization and linear prediction, and provides good quality with a bit rate ranging from 4 to 8 kbps. In this paper we present various coders derived from CELP technique: the ACELP (Adaptive CELP), the VSELP (Vector Sum Excited Linear Prediction) and the BACELP coder (Backward block adapted ACELP).

INTRODUCTION

Dès le début des années 80, la technique d'analyse par synthèse était utilisée dans le domaine de la compression de la parole avec les codeurs multi-impulsionnels. C'est en 1985 que la quantification vectorielle fut associée avec la technique précédente pour créer le codeur CELP (Code Excited Linear Predictive coder), qui pour un débit de l'ordre de 4 à 8 kbits/s, permet d'obtenir une bonne qualité. La figure 1 représente le schéma du codeur CELP de base. L'analyse par synthèse consiste à reconstruire à l'émetteur, un ensemble de séquences de parole obtenues à partir d'un jeu de paramètres (pour le codeur CELP : index et gain du vecteur du répertoire, coefficients LTP et LPC des filtres de prédiction à long terme et à court terme respectivement). C'est le jeu de paramètres qui engendre la séquence la plus proche, au sens d'un critère prédéfini, de celle du signal de parole, qui est transmis au récepteur.

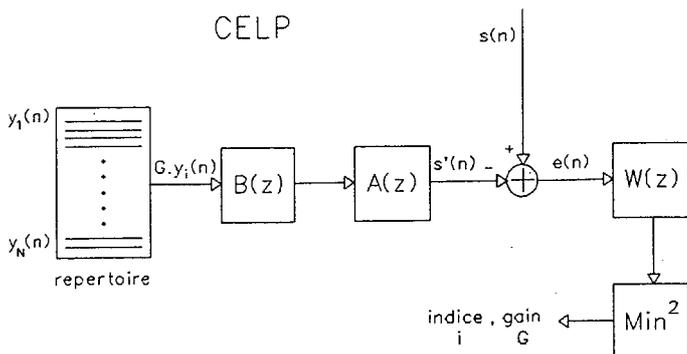


Figure 1. codeur CELP

Dans sa version originale, l'algorithme du codeur CELP était très complexe (40 MFlops env.) et nécessitait un volume de mémoire important (40 kOctets env.). De plus, il engendrait un délai de codage supérieur à 25 ms s'avérant contraire à la mise en oeuvre de ce codeur dans certains types d'applications.

Malgré ces problèmes, et du fait des bons rapports qualité/débit atteints, les recherches les plus récentes en compression numérique de la parole ont porté sur le codeur CELP avec pour objectif essentiel la réduction de la complexité sans détérioration de la qualité. Dans cette publication nous donnons trois approches différentes basées sur la technique CELP :

Le codeur ACELP 7 kbits/s (Adaptive CELP) qui utilise un dictionnaire linéaire adaptatif ALCB (Adaptive Linear Code Book) et applique une pondération de type globale qui évite le filtrage de chacun des vecteurs constituant le répertoire [2].

Le codeur VSELP 8 kbits/s (Vector Sum Excited Linear Prediction) sélectionné pour être le standard du radiotéléphone cellulaire Nord Américain [3]. Ce système utilise deux dictionnaires dont les éléments sont formés par combinaison linéaire de 7 vecteurs de base. Avec l'utilisation d'un code de Gray pour constituer les vecteurs candidats, et avec l'application d'un principe de quantification globale des gains, le système est robuste aux erreurs de transmission. Sur cette base de codeur, nous présenterons deux techniques différentes de masquage spectral permettant de réduire la complexité initiale.



- Le codeur BACELP (Backward block adapted ACELP) qui appartient à la catégorie des codeurs à faible délai, avec un retard de codage de 2ms [5]. Le principe est basé sur un calcul "à posteriori" des coefficients LTP et LPC. Ce codeur qui reste sensible aux erreurs de transmission, est de bonne qualité pour une complexité de l'ordre de 11 MIPS.

Après avoir présenté ces codeurs, nous donnons quelques résultats de simulations et nous comparons entre-elles ces différentes techniques.

LE CODEUR ACELP :

Le codeur ACELP se distingue du codeur CELP de base, d'une part avec l'usage d'une pondération de type globale, ou plus exactement par l'utilisation d'un filtre $W(z) = A(z)/A(z/\gamma)$ $0 < \gamma < 1$ sur le signal de parole original, de façon à obtenir un signal résiduel pondéré $r(n)$, et d'autre part avec le remplacement du répertoire stochastique bi-dimensionnel par un dictionnaire linéaire adaptatif (ALCB) qui permet d'accroître les performances du codeur initial. Un schéma de l'émetteur du codeur ACELP est représenté figure 2.

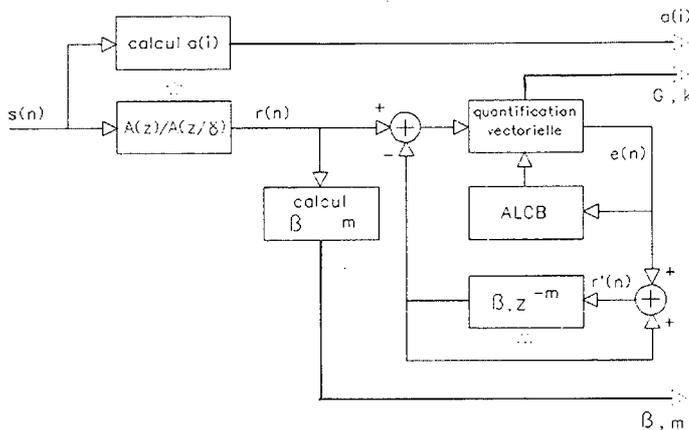


Figure 2. codeur ACELP

Le masquage spectral : Le but du masquage spectral est de distribuer la densité spectrale du bruit de codage d'une façon non uniforme, en tenant compte des propriétés de l'oreille humaine. L'application du filtre $W(z)$ sur le signal de parole $s(n)$ donne au signal $r(n)$ une enveloppe spectrale de même forme que la réponse harmonique du filtre $1/A(z/\gamma)$. La minimisation au sens des moindres carrés de l'erreur quadratique:

$$e(k) = \sum_{n=1}^N (r(n) - G_k \cdot v_k(n))^2$$

confère au vecteur $G_k \cdot v_k(n)$ une enveloppe spectrale de forme similaire au vecteur $r(n)$. Le signal de parole synthétique généré par filtrage du vecteur d'excitation au travers du filtre $1/W(z)$, suit l'allure de la réponse impulsionnelle du filtre modèle du conduit vocal $1/A(z)$. Le bruit de codage est pour sa part pondéré par le filtre de reconvolution $1/W(z)$ ce qui lui procure une densité spectrale non uniforme.

Contrairement au codeur CELP initial où il est nécessaire de filtrer chacun des vecteurs constituant le dictionnaire,

l'ACELP applique une pondération globale par l'intermédiaire du filtre $W(z)$ sur le signal de parole original, ce qui permet une réduction importante de la complexité.

Le répertoire linéaire adaptatif ALCB (Adaptive Linear CB):

Par son caractère linéaire ou mono-dimensionnel, l'espace mémoire requis pour le stockage de ce répertoire se trouve réduit par rapport à la structure CELP. Dans un répertoire ALCB, le vecteur candidat v_k s'écrit : $v_k(n) = CB(k+n)$, $1 < n < N$, $0 < k < K-1$. Le répertoire est renouvelé à chaque fenêtre en utilisant le vecteur sélectionné v_{k0} :

$$\begin{aligned} CB(n) &= CB(N+n) & n &= 1, \dots, K-1 \\ CB(K+N-1) &= v_{k0}(n) & n &= 1, \dots, N \end{aligned}$$

La nature initiale et la faculté de s'adapter en permanence à la partie non modélisée par les filtres de prédiction linéaire, procurent aux vecteurs candidats du dictionnaire les caractéristiques spectrales nécessaires pour modéliser au mieux le signal d'excitation (cf. paragraphe précédent). Afin de conserver au dictionnaire un caractère stochastique, seule une partie de ce dernier est rendue adaptative, l'autre restant fixe.

LE CODEUR VSELP

Le codeur VSELP (Vector Sum Excited Linear Prediction) récemment sélectionné pour être le standard du radiotéléphone cellulaire Nord Américain, utilise une nouvelle structure de dictionnaire pour modéliser le signal d'excitation: les vecteurs candidats sont formés par combinaison linéaire de 7 vecteurs de base (fig.3). En appelant $b_m(n)$ $n = 1, \dots, N$ le vecteur de base m , le vecteur candidat $v_k(n)$ s'écrit :

$$v_k(n) = \sum_{m=1}^7 \lambda_{k,m} \cdot b_m(n)$$

avec : $\lambda_{k,m} = 1$ si le bit m de la représentation binaire de k a pour valeur 1, dans le cas contraire, $\lambda_{k,m} = -1$. Par linéarité, les vecteurs candidats pondérés peuvent s'écrire :

$$v'_k(n) = \sum_{m=1}^7 \lambda_{k,m} \cdot b'_m(n)$$

avec : $b'_m(n)$ le vecteur $b_m(n)$ filtré par $H(z) = 1/A(z/\gamma)$. D'une manière plus précise, la structure VSELP possède une fonction d'orthogonalisation des vecteurs $b'_m(n)$ précédant la procédure de minimisation. En notant par $r(n)$ le vecteur sélectionné pour modéliser la composante périodique de $y(n)$ selon la méthode du codeur SELP [4], l'orthogonalisation des vecteurs de base pondérés s'effectue par rapport à $r(n)$ pour le premier étage de quantification vectorielle, et par rapport aux vecteurs $r(n)$ et $v''_{k0}(n)$ pour le second étage. $v''_{k0}(n)$ représentant le vecteur sélectionné du premier répertoire. Par ce principe, seuls les vecteurs de base sont pondérés. Utilisant par la suite une recherche de l'indice du vecteur suivant un code de Gray, la complexité algorithmique se trouve réduite par rapport à la méthode de sélection du vecteur d'excitation du codeur CELP.

L'emploi d'une combinaison linéaire permet à la structure d'être plus robuste aux erreurs de transmission. En effet, une erreur d'inversion d'un bit ne provoque le changement d'un seul signe dans la somme des vecteurs de base.

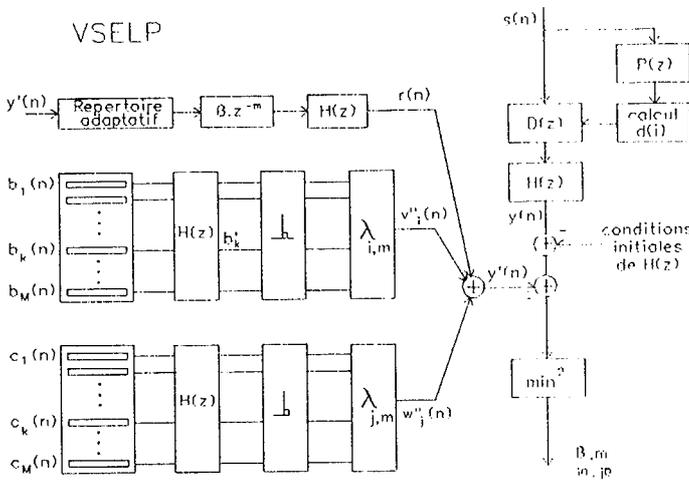


Figure 3. codeur VSELP.

VSELP initial (version A) :
 $P(z) = 1 \quad D(z) = A(z) \quad H(z) = 1/A(z/\gamma)$

VSELP étude 1 (version B) :
 $P(z) = 1 \quad D(z) = A(z) \quad H(z) = 1/(1 - \mu.z^{-1})$

VSELP étude 2 (version C) :
 $P(z) = (1 - \mu.z^{-1}) \quad D(z) = C(z) \quad H(z) = 1$

Sur cette structure, nous avons simulé une pondération en remplaçant $H(z)$ par le filtre $G(z) = 1/(1 - \mu.z^{-1})$, avec $\mu = -R(1)/R(0)$ où $R(k)$ représente la fonction d'autocorrélation du signal de parole. La réponse harmonique de ce filtre est représentée sur la figure 5. En appliquant cette méthode, nous retrouvons l'idée d'obtenir des vecteurs candidats possédant la même enveloppe spectrale que le vecteur $r(n)$. L'intérêt est de réduire le nombre d'opérations nécessaires au masquage spectral. Un test d'écoute par paire réalisé sur 8 phrases avec 6 auditeurs démontre l'efficacité de cette technique qui obtient un score de 37% contre 23% pour le codeur initial fig.6.

Une deuxième étude a consisté à mettre en oeuvre une technique de pré-emphase sur le signal de parole original et à éliminer par ce fait le filtre de pondération $H(z)$ de la structure initiale. Cette technique exposée en [2] réside sur un calcul des coefficients d'un filtre $C(z)$ (de forme similaire au filtre de prédiction linéaire à court-terme $A(z)$), sur un signal de parole filtré au préalable par $(1 - \mu.z^{-1})$:

$$v(n) = s(n) - \mu.s(n-1)$$

$$V(z) = (1 - \mu.z^{-1})S(z)$$

$$V(z) = S(z)/G(z)$$

Par application de cette méthode, le signal résiduel obtenu en filtrant le signal $S(z)$ par $C(z)$ possède une enveloppe spectrale de même allure que celle de la réponse

impulsionnelle du filtre $G(z)$. En conséquence, il n'est plus nécessaire de pondérer les vecteurs de base si l'on considère que leur caractéristiques spectrales sont similaires à celles du filtre $G(z)$. Au récepteur, le signal d'excitation formé par combinaison linéaire des vecteurs de base, est filtré par $1/C(z)$ pour reconstruire le signal de parole. Cette dernière possède donc par construction une enveloppe spectrale semblable à celle du signal original $s(n)$. Cette dernière méthode réduit la complexité de calcul, tout en maintenant au plan de la perception auditive la qualité du signal de parole reconstruit. Les résultats des tests d'écoute sont résumés à la fig. 7.

LE CODEUR BACELP

Le codeur BACELP (Backward block adapted ACELP) répond aux exigences de faible délai (<5ms) du CCITT concernant le nouveau standard à 16 kbps. La contrainte sur le retard de codage/décodage permet d'employer ces systèmes sans l'adjonction de techniques d'annulation d'écho.

Le codeur BACELP possède la même structure que le codeur ACELP, mais le calcul des coefficients des filtres LPC et LTP est à présent effectué "à posteriori". La figure 4 représente la structure de l'émetteur. Le signal de parole est divisé en blocs de 8 échantillons (1ms). Chaque bloc est filtré par $A(z)/A(z/\gamma)$ avec $\gamma = 0.8$. Les coefficients a_i du filtre de déconvolution $A(z)$ sont calculés à partir d'une fenêtre de signal reconstruit $s'(n)$. Le signal résiduel obtenu $r(n)$ est ensuite estimé par l'analyse de $r''(n)$ selon une boucle de prédiction à long terme. Les paramètres du filtre LTP, β et m , sont déterminés sur une fenêtre formée des blocs d'excitation reconstruite passés. Afin d'améliorer l'estimation de ces paramètres à partir d'un bloc de faible dimension, un filtrage passe-bas est appliqué au signal $y(n)$ pour augmenter son caractère périodique. Le signal d'erreur $e(n)$ non modélisé par les filtres de prédiction, est quantifié vectoriellement à partir d'un répertoire ALCB. Le gain associé à la séquence sélectionnée est codé sur 4 bits. Ce codage conduit à un débit de 13 kbps pour une structure possédant 512 vecteurs candidats.

Un test d'écoute réalisé pour évaluer la qualité du codeur BACELP 13 kbps par rapport à la norme G721 (ADPCM, 32 kbps, 0.125 ms de délai), montre que leurs qualités sont semblables étant donné les résultats obtenus de 33% de préférence pour le BACELP, 40% pour le codeur ADPCM 32kbps et 27% sans préférence.

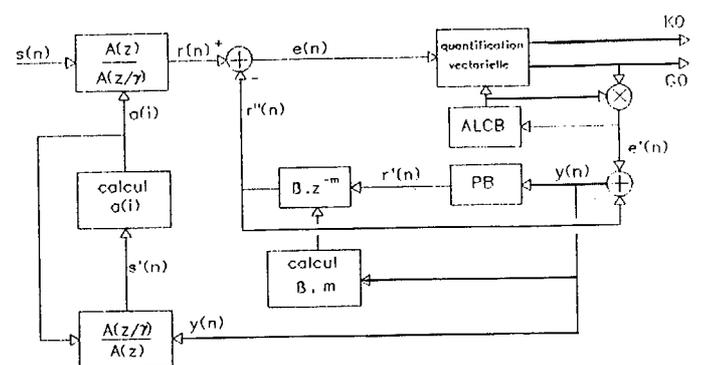


Figure 4. codeur BACELP



Malgré l'effort de calcul nécessaire à la détermination des coefficients des filtres de prédiction, cette technique reste d'une faible complexité de l'ordre de 11 MIPS, à comparer à 50 MFLOPS du codeur en cours de standardisation [6]. Le codeur BACELP est cependant sensible aux erreurs de transmission. L'algorithme BACELP peut être utilisé dans les systèmes de communication utilisant les nouvelles générations de réseaux large bande comme moyen de transmission de l'information.

CONCLUSION

Dans cet article nous avons présentés différents codeurs dérivés de la structure CELP originale, avec comme objectif essentiel une réduction de la complexité par rapport aux autres codeurs proposés, sans pour autant altérer la qualité.

C'est ainsi qu'à des débits binaire compris entre 4,8 et 8 kbps, les codeurs ACELP et VSELP permettent à la fois de réduire la complexité de calcul et l'espace mémoire.

Par rapport aux codeurs précédents, le codeur BACELP apporte également une solution au problème du retard de codage pouvant engendrer des phénomènes d'écho.

[1] M.R.SCHROEDER, B.S.ATAL: 'Code Excited Linear Prediction (CELP): High quality speech at very low bit rates' Proc. of Int. Conf. on ASSP-1985.

[2] J.MENEZ, C.GALAND, M.ROSSO, F.BOTTAU: 'Adaptive Code Excited Linear Predictive Coder (ACELPC)' Proc. of Int. Conf. on ASSP-GLASGOW 1989.

[3] I.A.GERSON, M.A.JASUIK: 'Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 KBPS' Proc. of Int. Conf. on ASSP-ALBUQUERQUE 1990.

[4] W.B.KLEINJ, D.J.KRASINSKI, R.H.KETCHUM: 'Improved speech quality and efficient vector quantization in SELP' Proc. of Int. Conf. on ASSP-TOKYO 1988.

[5] J.MENEZ, C.GALAND, M.ROSSO: 'A 2ms delay Adaptive Code Excited Linear Predictive Coder' Proc. of Int. Conf. on ASSP-ALBUQUERQUE 1990.

[6] J.H.CHEN: 'A robust low-delay CELP speech coder at 16 KBPS' GLOBECOM-DALLAS 1989.

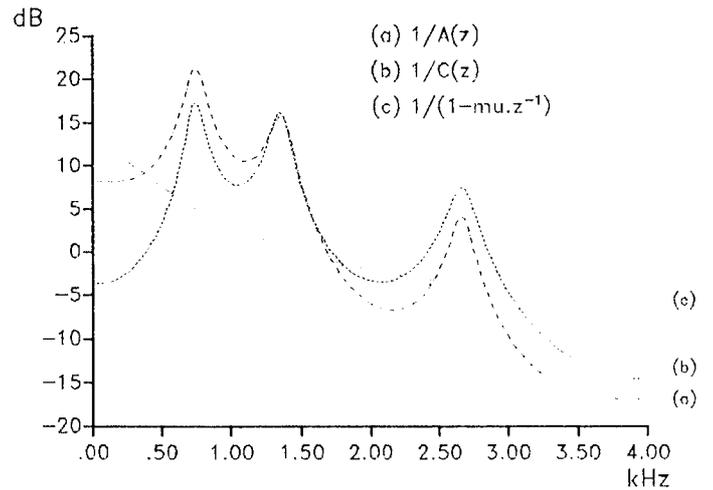


Figure 5. réponse impulsionnelle des différents filtres

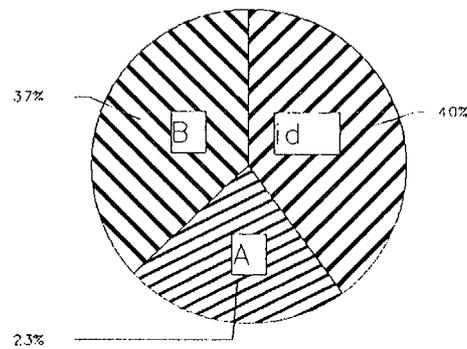


Figure 6. test d'écoute VSELP A / VSELP B.
id : sans préférence

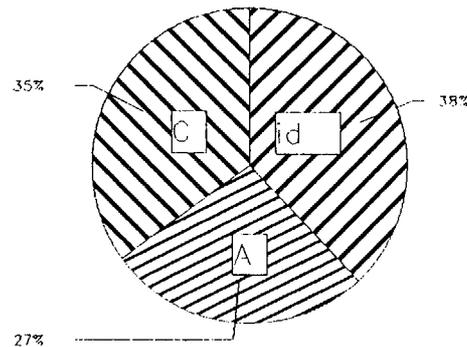


Figure 7. test d'écoute VSELP A / VSELP C.
id : sans préférence