

CODAGE DE LA PAROLE A EXCITATION ALGEBRIQUE APPORT D'UN CODE CORRECTEUR DE HAMMING

M. Bouraoui^{(*)(**)}, G.Feng^(*) & F.B Druilhe^(**)

(*) Institut de la Communication Parlée - URA CNRS 368
BP 25 - 38040 GRENOBLE CEDEX - FRANCE
e-mail : feng@icp.grenet.fr

(**) SGS-THOMSON Microelectronics - Telecom Division
BP 217 - 38019 GRENOBLE - FRANCE
e-mail : mustapha.bouraoui@st.com

RESUME

La complexité d'un codeur CELP est principalement concentrée sur la procédure de recherche dans le dictionnaire. Plusieurs études ont montré que les structures algébriques pour les dictionnaires permettaient de réduire considérablement cette charge de calcul entraînant une très légère dégradation de la parole reconstruite. Nous proposons un nouvel algorithme de recherche basé sur un code de Hamming 1-correcteur. Le meilleur vecteur-code est directement circonscrit en utilisant la procédure de correction d'erreurs qu'offrent de tels codes. L'algorithme présenté ne requiert qu'une charge de calcul de 0.7 MIPS, ce qui représente une réduction de complexité d'un facteur de 3 à 4 par rapport aux algorithmes de recherche fréquemment utilisés.

1. INTRODUCTION

Un codeur CELP utilisé pour le stockage de la parole cohabite souvent avec d'autres fonctions (Décodage DTMF, Annulation d'écho etc ...), fonctions qui doivent fournir des réponses rapides quand elles sont sollicitées. Cela implique la subdivision de l'algorithme de codage en tâches multiples à faible charge de calcul, permettant ainsi l'insertion des fonctions précitées. Etant donné que la complexité d'un codeur CELP est principalement concentrée sur la procédure de recherche dans le dictionnaire, un algorithme rapide est nécessaire pour cette tâche. Une manière très répandue de réduire cette charge de calcul consiste en l'utilisation de dictionnaires algébriques qui permettent de définir des excitations ternaires spécifiques. Salami [1] a proposé une définition directement dérivée de la valeur de l'indice de l'excitation dans le dictionnaire. Cette technique connue sous le nom de BCELP (Binary-CELP) apporte une grande réduction de complexité et une bonne robustesse contre les erreurs dues aux canaux bruités. Lamblin & al. a proposé une autre approche basée sur les réseaux sphériques [2][3] dont la théorie est basée sur celle des codes correcteurs d'erreurs. Cette technique très prometteuse a été appliquée dans la définition du standard de l'ITU-T G729 correspondant au codage de la parole à bande étroite à un débit de 8 kbps [4][5]. Notre approche est une solution intermédiaire. En effet, le meilleur vecteur-code est directement circonscrit en utilisant la procédure de correction d'erreurs d'un code de Hamming. Une description détaillée de la définition du dictionnaire sera donnée au paragraphe 2. Adapté à la structure proposée, un algorithme de recherche sous-optimale appelé HCELP est défini. Il sera détaillé au paragraphe 3. Une évaluation des

ABSTRACT

The complexity of a CELP algorithm is mainly concentrated in the codebook search procedure. Many studies show that the algebraic approach for codebook structures considerably lessens the computational load with a slight degradation compared to stochastic or trained codebook performances. We propose a novel search algorithm based on a Hamming single-error-correcting -code (ECC) structure. The best codevector is directly circumscribed using the error correction ability of the ECC. This algorithm requires only 0.7 MIPS which represents a complexity reduction by a factor of 3 or 4 compared to search algorithms based on current algebraic codebooks.

performances de cet algorithme a été effectuée en comparant ses résultats avec ceux que procure une recherche exhaustive et en présentant la charge de calcul qu'il requiert. Enfin, l'algorithme proposé a été inséré dans un codeur complet dont les performances sont rapportées dans le paragraphe 4.

2. DEFINITION DU DICTIONNAIRE

2.1 Généralités

D'un point de vue géométrique une excitation de dictionnaire est un vecteur de R^L . Chacune des excitations est ternaire $\{-1,0,+1\}$. Si on considère de plus que le nombre des impulsions non-nulles est constant (N) et que leur positions sont fixes, un vecteur-code est totalement défini par la combinaison des amplitudes non-nulles. Ainsi un mot de N bits peut complètement définir l'excitation en remplaçant les symboles $+1$ et 0 par $+1$ et -1 . Ce mot binaire peut tout aussi bien être l'indice I dans le dictionnaire comme dans la technique BCELP ou $T(I)$ où T représente une transformation algébrique. T peut être par exemple, la procédure de codage pour un code correcteur de Hamming. Le code $H(15,11,3)$ a été choisi sachant que l'on traite des excitations de 60 échantillons contenant 15 impulsions non-nulles.

2.2 Description du code $H(15,11,3)$

Le code binaire $H(15,11,3)$ peut facilement être défini en termes de bits de parité. Chacun de ces bits est une combinaison linéaire en OU-exclusif de certains des 11 bits de données. La figure 1 décrit les bits impliqués dans chacun des



4 bits de parité. L'ensemble des 2^{11} indices de 11-bit I ($0 \leq I \leq 2^{11} - 1$) va générer un ensemble Θ de 2^{11} codes binaires C_I sur 15 bits représentant les formes d'ondes de l'excitation. Si on définit la distance d_H entre deux codes comme le nombre de positions par lesquels ils diffèrent, on a :

$$\forall (I, J) \in \{0..2^{11} - 1\}, I \neq J, \quad d_H(C_I, C_J) \geq 3 \quad (1)$$

Ainsi, tout mot de 15 bits se trouve à une distance d'un élément de Θ d'au plus 1 i.e en ne modifiant qu'un seul bit. On a de plus la symétrie :

$$\forall I \in \{1..1024\}, \quad C_I = \overline{C_{2048-I+1}}$$

ce qui entraîne :

$$\forall I \in \{1..1024\}, \quad E_I = -E_{2048-I+1} \quad (2)$$

où E_I est l'excitation déduite de C_I .

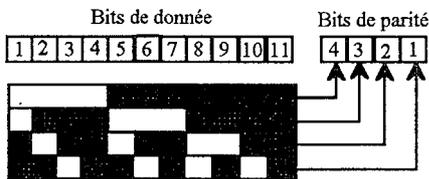


Figure 1. Code de Hamming $H(15,11,3)$: Calcul des bits de parité à partir des bits de donnée.

2.3 Structure du dictionnaire

Dans cette application, le signal de parole après décorrélation long-terme et court-terme est modélisé par l'excitation E qui maximise la fonction :

$$m(E) = \frac{\langle R, HE \rangle^2}{\|HE\|^2} \quad (3)$$

où H est la matrice triangulaire inférieure de la réponse impulsionnelle tronquée du filtre de prédiction linéaire. En utilisant la parité de la fonction m et l'expression (2), seulement la moitié du dictionnaire (1024 vecteurs) a besoin d'être testée. L'indice du dictionnaire est donc réduit à 10 bits. Chaque excitation contient 15 impulsions non-nulles représentées par un élément du sous-ensemble Θ' de Θ défini par

$$\Theta' = \{C_I \in \Theta \mid I \in \{0..2^{10} - 1\}\}$$

De plus, les impulsions non-nulles sont séparées de 3 "zéros" et 4 positions possibles (2 bits) sont réservées pour la première impulsion. Cela conduit à un dictionnaire sur 12 bits dont la construction de l'excitation est illustrée dans la figure 2.

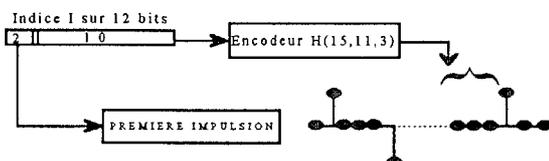


figure 2. Définition de l'excitation de Hamming intervenant dans le dictionnaire proposé

3. L'ALGORITHME HCELP PROPOSE

3.1 Description

Si T_{BF} est la cible issue du "backward filtering" [6], on a, en reprenant les notations de (3) :

$$m = \frac{\langle T_{BF}, E \rangle^2}{\|HE\|^2}$$

En faisant l'approximation que l'énergie de l'excitation filtrée est à peu près constante, la meilleure combinaison des amplitudes correspond aux signes de T_{BF} aux positions autorisées par le dictionnaire. Autrement dit, le meilleur vecteur d'excitation, au sens du critère simplifié, est construit en choisissant le signe de ses impulsions égal au signe de celle de T_{BF} aux positions non-nulles. Il existe 4 "meilleures combinaisons" puisqu'il y a 4 positions pour la première impulsion non-nulle.

Soient V_p , $p \in \{1..4\}$ les quatre mots de 15 bits correspondant à ces combinaisons. V sera le mot binaire correspondant à l'excitation qui maximise le critère complet m . On n'a pas nécessairement $V \in \Theta'$. Si $V \notin \Theta'$, V est alors dégradé en lui changeant un seul bit afin de former un ensemble de 16 mots V'_i , $i \in \{1..15\}$ plus V lui-même. V'_i sont les plus proches voisins en termes de distance de Hamming. V et chacun des V'_i sont par la suite corrigés pour former un ensemble C_i , $i \in \{1..16\}$ en utilisant la procédure classique de correction d'un code $H(15,11,3)$. On obtient alors un sous-ensemble du dictionnaire contenant 16 vecteurs, parmi lesquels on choisit celui qui maximise le critère complet m . Remarquons que si $V \in \Theta$ (ce qui arrive avec une probabilité 1/16) alors tous les V_i sont égaux et V est directement corrigé en C . C est finalement décodé en un mot de 10 bits qui sera transmis au décodeur. Il est à noter que même si le décodeur $H(15,11,3)$ produit un mot de 11 bits, la symétrie mentionnée en (2) permet de forcer la sortie à 10 bit par inversion de signe, celle-ci étant reportée sur le gain. L'algorithme est résumé dans l'organigramme de la figure 3.

3.2 Performances de l'algorithme

3.2.1 Degré d'identification de la cible

Une évaluation de l'algorithme a été effectuée en comparant ses résultats à ceux fournis à l'issue d'une recherche exhaustive dans le dictionnaire avec le critère complet m . S désignera la recherche par HCELP et S_{full} la recherche exhaustive. Une longue séquence de parole (3 min.) a été convertie en une longue séquence de cibles T par filtrage LP inverse. Les excitations E issues de S et E_{full} issues de S_{full} obtenues pour modéliser les cibles T sont pondérées par leur gain. Les enveloppes des histogrammes d'énergies en dB :

$$e_1 = \frac{\|T - E_{full}\|}{\|T\|} \quad \text{and} \quad e_2 = \frac{\|T - E\|}{\|T\|}$$

sont tracées sur la figure 4. Comme on peut le constater, le caractère sous-optimal de l'algorithme n'affecte pas de manière sensible le degré d'identification des cibles puisqu'on obtient pratiquement les mêmes courbes.

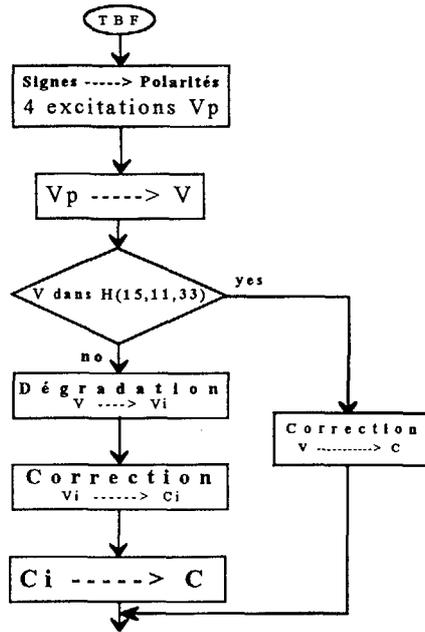


Figure 3. Organigramme de l'algorithme HCELP

amplitudes	positions
1	2,9,16,23,30,37,44,51
-1	4,11,18,25,32,39,46,53
1	6,13,20,27,34,41,48,55
-1	8,15,22,29,36,43,50,57

Tableau I. Amplitudes et positions potentielles des impulsions dans la version modifiées du dictionnaire du G729.

On obtient un dictionnaire sur 12 bits dont l'algorithme de recherche est effectué dans quatre boucles imbriquées. La complexité d'un tel algorithme est fixée par le taux de pénétration dans les boucles. La version adoptée en référence pour l'évaluation de la qualité et de la charge de calculs est un codeur à 4800 bps utilisant le dictionnaire modifié du G729. Son algorithme de recherche exige 2.6 MIPS de calculs. La réduction de complexité due à la structure HCELP est d'un facteur 3.7.

4. PERFORMANCES DU CODEUR

4.1 Description générale du codeur

L'algorithme proposé a été inséré dans un codeur CELP et a été simulé avec les caractéristiques suivantes. Le signal de parole échantillonné à 8 kHz, est découpé en trames de 30 ms (240 échantillons). Une analyse LPC d'ordre 10 est effectuée par la méthode de l'autocorrélation avec une extension de bande de 50 Hz. L'analyse du pitch et la recherche de l'excitation sont réalisées quatre fois par trame sur des sous-trames de 7,5 ms (60 échantillon). Une estimation du pitch est obtenue par la méthode décrite par Salami & al. [4], ce qui permet de réduire considérablement la fourchette de recherche dans le dictionnaire adaptatif contenant les excitations précédemment sélectionnées. On évite ainsi le phénomène de doublement du pitch et on peut coder une fois sur deux l'indice du dictionnaire adaptatif de manière différentielle. Les caractéristiques du codeur HCELP sont résumées dans le tableau ci-dessous :

3.2.2 Charge de calculs

La répartition de la charge de calculs est détaillée ci-dessous :

Calcul de V et V_i	:	195 cycles
identification des C_i	:	1328 cycles
Boucle fermée sur les C_i	:	3632 cycles
TOTAL	:	5155 cycles

Etant donné que la recherche est effectuée toutes les 7,5 ms, la procédure ne demande que 0.7 MIPS. Nous avons comparé cet algorithme avec celui utilisé dans le codeur G729 [5] muni de quelques modifications concernant le débit et la taille du dictionnaire. Dans ce dernier cas, le vecteur d'excitation contient 4 impulsions non-nulles. Leurs amplitudes sont fixées à +1, -1, +1 et -1, respectivement, et leur positions sont données dans le tableau I.

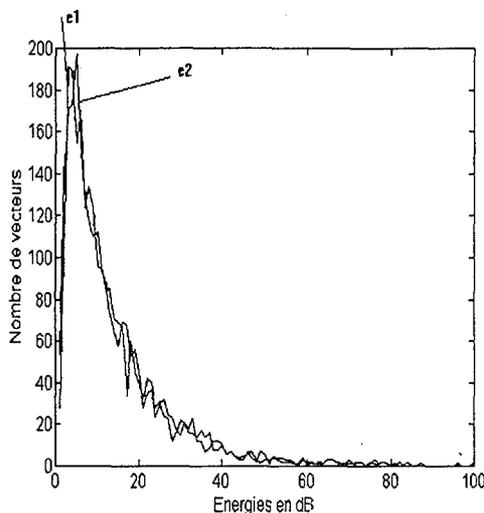


Figure 4. Enveloppes des histogrammes d'énergie d'erreur

	LPC	Pitch	Dictionnaire
Fenêtre	30 ms	7.5 ms	7.5 ms
Analyse	Autocorr. 50 Hz B.E	Demi boucle fermée	Boucle fermée sur 16 vecteurs présélectionnés
Bits/Trame	34 LSP	Indice : 7+5+7+5 Gain : 4x4	Indice : 12x4 Gain : 5x4
Débit (4800 bps) (2 bits/trame "non utilisés")	1133.3 bps	1333.3 bps	2266.7 bps

Tableau II. Description du codeur HCELP et allocation binaire



4.2 Postfiltrage

Nous avons remarqué une légère dégradation à l'écoute, due probablement au caractère sous-optimal de l'algorithme de recherche. Un module de postfiltrage adaptatif a donc été ajouté au décodeur dans le but de rehausser la qualité du signal de parole reconstruite. Nous avons implémenté un schéma classique utilisé dans la plupart des standards CELP actuels [7]. Il consiste en la cascade de trois filtres adaptatifs (figure 5): un postfiltre long-terme, un postfiltre court-terme et un compensateur de pente. Les coefficients sont recalculés à la fréquence sous-trame.

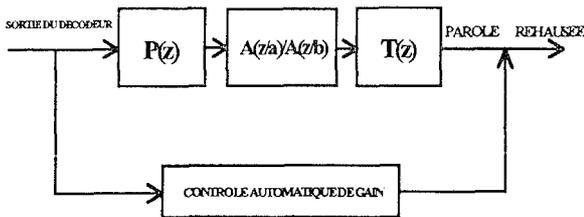


figure 5. Structure du Postfiltrage

Nous avons opté pour les valeurs $a = 0.5$ et $b = 0.8$. Nous avons aussi: $P(z) = 1 + \gamma z^{-T}$ où T est le pitch et γ le gain du pitch. $A(z)$ est le filtre LPC inverse. Le compensateur de pente est un filtre passe-haut du premier ordre: $T(z) = 1 - 0.5z^{-1}$.

4.3 Evaluation de la qualité

La qualité du codeur a été évaluée en effectuant des tests d'écoute informels sur un corpus multilocuteur (4 hommes + 4 femmes + 3 enfants). Le codeur de référence est celui décrit plus haut. Nous n'avons pas constaté de différences pour les voix d'hommes. De plus le codeur HCELP nous a semblé meilleur pour les voix de femmes et d'enfants (figure 6). Une évaluation "objective" a donné 9.1 dB en termes de rapport signal sur bruit segmentaire.

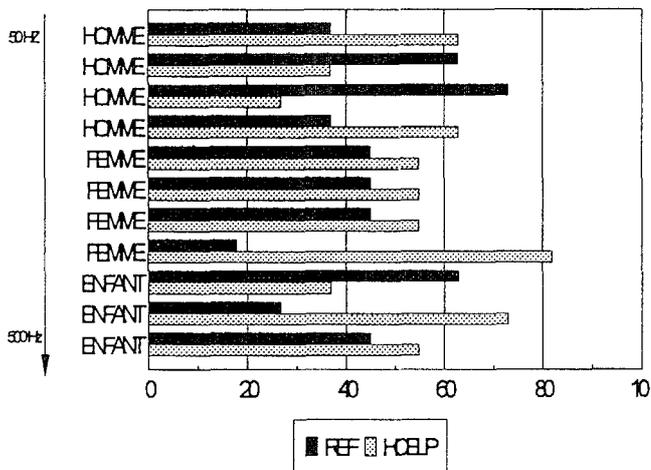


figure 6. Résultats des tests d'écoutes. REF représente un codeur ACELP à 4800 bps utilisant un dictionnaire dérivé de celui du G729.

5. CONCLUSION

Un codeur HCELP a été proposé dans cette communication. Un dictionnaire structuré par un code correcteur d'erreurs de Hamming est utilisé et permet l'implantation d'un algorithme rapide basé sur la capacité de correction de $H(15,11,3)$. Une réduction notable de la complexité a été observée et la légère dégradation due au caractère sous-optimal de l'algorithme est masqué par le module de postfiltrage au décodage.

BIBLIOGRAPHIE

- [1] Salami R. A "Binary code excited linear prediction (BCELP): new approach to CELP coding of speech without codebooks" *Electron. lett.*, vol.25 no. 6, pp. 401-403, March 1989.
- [2] Lamblin C. "Quantification vectorielle algébrique sphérique par le réseau de Barnes-Wall : application à la parole." *Ph.D, University of Sherbrooke, P.Q., Canada, April 1988.*
- [3] Ireton M.A., Xydeas C.S. "On improving vector excitation coders through the use of spherical lattice codebooks (SLC's)" *Proc. ICASSP 1989.* pp. 57-60.
- [4] Salami R.A., Laflamme C., Adoul J.P., and Massaloux D. "A Toll quality 8 kbps Speech codec for PCS systems" *IEEE Transactions on Vehicular Technology Vol. 43 no 3 pp. 808-816 - August 1994*
- [5] Salami & al. "Description of the ITU-T 8kbps Speech Coding Standard" *Proc. IEEE Speech Coding Workshop 1995.*
- [6] Adoul & al. "Fast CELP coding based on algebraic codes" pp.1957-1960. *Proc. ICASSP 1987.*
- [7] Chen J.H., Gersho A. "Adaptive postfiltering for quality enhancement of coded speech" pp.59-71 *IEEE Transactions on Speech and Audio Processing Vol.3 no 1 - January 1995*