



CODEUR A EXCITATION PAR CODE MULTI-IMPULSIONNEL:
UN ALGORITHME, UN DICTIONNAIRE D'EXCITATION

J.-G. FRITSCH

TELIC-ALCATEL 1, route du Dr Albert Schweitzer 67408 ILLKIRCH CEDEX

RESUME

Le codage de la parole est une discipline qui a subi une forte évolution depuis le début des années 1982 par l'avènement d'un nouveau concept de codage du signal d'excitation pour les codeurs APC (Adaptive Predictive Coding). Ce nouveau concept appelé modélisation par analyse-synthèse a vu l'introduction successive du codeur à excitation multi-impulsionnelle /4/ (MPLPC) puis du codeur à excitation par code /1/ (CELPC). Ce dernier permet d'atteindre des débits inférieurs à 7 kbits/s, mais au prix d'une complexité qui dépasse les 20 millions de multiplications et additions par seconde ($x, +/s$).

On présente ici un codeur, appelé codeur à excitation par code multi-impulsionnel (MPCELPC) qui combine les avantages des deux codeurs cités précédemment afin d'obtenir un codeur de qualité sub-téléphonique permettant de réduire le débit global à moins de 8 kbits/s tout en limitant la complexité à environ 3 millions de $x, +/s$.

En outre, la souplesse d'utilisation de ce codeur permet de satisfaire de nombreuses applications exigeant d'une part une qualité sub-téléphonique, d'autre part un débit inférieur à 10 kbits/s.

On propose également une technique d'extraction statistique du dictionnaire d'excitation, qui fait intervenir la même métrique que celle définie par le critère de minimisation dans la procédure de codage du signal excitation.

Finalement les performances du codeur MPCELPC sont comparées à celle des codeurs MPLPC et CELPC.

SUMMARY

Analysis by Synthesis modelisation of the excitation signal for APC (Adaptive Predictive Coding) coders, has introduced successively the multi-pulse (MPLPC) and the code exciting (CELPC) coders. The CELPC coder has very interesting characteristics because it allows the production of sub-telephonic speech quality with a bit rate less than 7 kbits/s. But the draw back of this coder is its complexity wich is greater than 20 million of multiplications and additions per seconde ($x, +/s$).

In this paper, we describe a new coder, called multi-pulse code exciting coder (MPCELPC). This coder combines a low bit rate, less than 8 kbits/s with a complexity less than 3 million $x, +/s$.

In addition, a procedure for the extraction of the exciting code book is proposed. This procedure is based on the mean-square error minimisation criteria used for coding.

A comparative study of the MPLPC, CELPC and MPCELPC coders is also presented.



1. INTRODUCTION:

Le choix d'un codeur pour une application donnée est toujours le résultat d'un compromis entre plusieurs paramètres antagonistes qui sont la qualité de la parole décodée, le débit intermédiaire après codage et la complexité. La complexité peut d'ailleurs être décomposée en trois autres paramètres qui sont, la complexité opératoire qu'on évalue généralement en nombre de $x, +/s$, l'effort de mémorisation RAM engendré par le stockage des signaux intermédiaires et l'effort de mémorisation ROM lié plutôt aux dictionnaires scalaires ou vectoriels de codage-décodage des données constituant le débit intermédiaire. Ajoutons que dans le cas d'une implémentation sur circuit VLSI spécifique (ASIC) un autre paramètre, dont l'appréciation est plutôt subjective, intervient également. Il s'agit de la régularité et du caractère systématique du traitement, qui sont le reflet de la simplicité du séquencement de l'algorithme.

Dans le cadre d'applications sub-téléphoniques, il apparaît que les codeurs associant une prédiction adaptative (à court et à long terme) à une modélisation par analyse-synthèse de l'excitation /3/ offrent le meilleur compromis entre les différents paramètres décrits précédemment.

La modélisation par analyse-synthèse met en oeuvre des méthodes globales qui supposent disponibles tous les échantillons du bloc de signal à modéliser. Son principe, dont le bloc diagramme général est présenté figure 1, peut se formuler de la manière suivante.

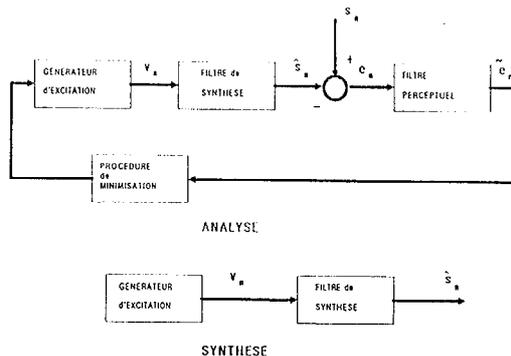


Figure 1: Principe des codeurs à modélisation de l'excitation par analyse-synthèse

Connaissant le filtre de synthèse, il faut trouver le signal d'excitation optimal, qui au sens des moindres carrés minimise l'erreur pondérée e_n par le filtre perceptuel, entre le signal original s_n et le signal synthétique \hat{s}_n . Plusieurs alternatives, donnant lieu aux différents codeurs, sont possibles selon que le signal d'excitation est défini comme:

- un train d'impulsions indépendantes, qui donne lieu au codeur à excitation multi-impulsionnelle.
- des séquences d'excitation plus élaborées, de longueur identique à celle du bloc au quel on applique la minimisation et qui sont choisies dans un dictionnaire de référence. On obtient dans ce cas le codeur à excitation par code.
- un signal multi-impulsionnel, dont les amplitudes des impulsions sont considérées de manière globale dans le bloc de signal à modéliser. Les séquences d'amplitudes sont choisies dans un dictionnaire de référence. Il en résulte le codeur à excitation par code multi-impulsionnel.

Comme cela a été précisé précédemment le codeur à excitation multi-impulsionnelle et le codeur à excitation par code sont complémentaires. En effet, leurs principales caractéristiques peuvent se résumer de la manière suivante.

La modélisation multi-impulsionnelle ne permet pas de restituer de la parole de qualité sub-téléphonique pour des débits, au niveau de l'excitation, inférieurs à 8 kbits/s, mais il a l'avantage d'être d'une complexité qui n'exède pas 0.5 millions de $x, +/s$. En revanche, la modélisation par code permet de réduire ce débit à moins de 3 kbits/s, mais au prix d'une complexité de quelques 19 millions de $x, +/s$.

2. MODELISATION PAR CODE MULTI-IMPULSIONNEL:

Cette procédure combine les modes d'excitation multi-impulsionnel et par code. Elle modélise le signal d'excitation en deux temps.

Premier temps, les positions m_k des impulsions sont déterminées de manière itérative à partir de la procédure de modélisation multi-impulsionnelle /4/. Les positions sont transmises séparément.

La combinaison des vecteurs v^i du dictionnaire des amplitudes avec les positions m_k des impulsions permet de constituer les vecteurs d'excitation u^i :

$$v^i = \sum_{k=1}^{M_i} A^i \cdot u^i \cdot \delta_{m_k} \quad (1)$$

où A^i représente le facteur de gain global

Deuxième temps, les amplitudes des impulsions regroupées en séquences sont encodées par une procédure de quantification vectorielle sphérique, qui prend en compte la fonction de pondération perceptuelle. Cette procédure de quantification n'est autre que la procédure de modélisation par code /1/, qui dans ce cas ne s'applique pas à tous les échantillons du bloc de signal à modéliser, mais uniquement aux quelques impulsions dont les positions ont été déterminées précédemment. Compte-tenu de la forme particulière des séquences multi-impulsionnelles, la procédure de modélisation par code se simplifie. Minimiser au sens des moindres carrés l'erreur perceptuelle revient à maximiser G^i par rapport aux vecteurs d'excitation v^i . On obtient:

$$G^i = \frac{(T e \cdot Q \cdot u^i)^2}{T u^i \cdot C \cdot u^i} \quad (2)$$

La valeur optimale du facteur de gain A^i est donnée par:

$$A^i = \frac{T e \cdot Q \cdot u^i}{T u^i \cdot C \cdot u^i} \quad (3)$$

La matrice C , de dimension $K \times K$ (K étant le nombre d'impulsions placées dans un bloc de signal de longueur N) correspond à la matrice d'autocorrélation symétrique constituée par les termes $c_{m_l - m_j}$. m_l et m_j sont les positions des impulsions. Les indices l et j varient 1 à K .

La matrice Q , de dimension $K \times N$ est également une matrice d'autocorrélation dont les termes sont définis par $c_{n - m_j}$ où n varie de 0 à $N-1$ et j varie de 1 à K .

Les termes c_j représente l'autocorrélation normalisée à l'ordre j de la réponse impulsionnelle f du filtre de pondération $1/A(z/g)$; g étant le facteur perceptuel.

c_j est défini par:

$$c_j = \sum_{n \in N} f_n \cdot f_{n-j} \quad (4)$$



On passe ainsi d'un système de dimension N , dans le cas du codeur CELPC, à un système de dimension K . Le rapport de réduction de la complexité, par vecteur d'excitation, est alors au moins égal à $(N \times N + 2N) / (K \times K + 2K)$. Ceci correspond pour des valeurs standard de $N=40$ et $K=6$ une réduction dans un rapport 35 de la complexité opératoire. En réalité, ce rapport est encore plus important car les vecteurs du dictionnaire passent d'une dimension N à K . C'est pourquoi, à qualité égale, le nombre de vecteurs du dictionnaire peut être également réduit. Il en résulte une réduction supplémentaire de la complexité qui ne dépasse pas 3 millions de $x, +/s$.

L'introduction de la diagonalisation de la matrice d'autocorrélation C [5] apporte, comme pour le codeur à excitation par code, une réduction supplémentaire du coût opératoire.

Contrairement à une approche qui consiste à appliquer à posteriori une quantification vectorielle aux amplitudes des impulsions, la procédure de modélisation par code multi-impulsionnel prend en compte l'effet de masquage que procure le filtre perceptuel.

On peut observer que cette modélisation est équivalente à une modélisation par code lorsque le nombre d'impulsions placées dans les blocs successifs de signal est égal au nombre d'échantillons des blocs.

3. EXTRACTION STATISTIQUE DU DICTIONNAIRE D'EXCITATION:

L'extraction statistique du dictionnaire contenant les séquences des amplitudes de l'excitation multi-impulsionnelle fait appel à un algorithme de classification à seuil [2]. Notre choix s'est porté sur cette méthode, car elle présente un bon compromis en terme de complexité et de vitesse de convergence. Pour une telle méthode, les éléments déterminants sont d'une part la mesure de distance (ou métrique) retenue d'autre part la valeur du seuil de classification.

3.1. CHOIX DE LA METRIQUE:

La métrique retenue est celle définie par la relation 2. La classification se faisant à partir d'un seuil fixe, il est nécessaire de normaliser cette métrique, qui est homogène à une intercorrélation. Elle varie donc entre 0 et 1. La relation définissant la métrique s'écrit alors:

$$L^2 = \frac{(\sum e_n \cdot Q \cdot u_n^2)}{(\sum u_n^2 \cdot C \cdot u_n^2) \cdot (\sum e_n \cdot Q \cdot e_n)} \quad (5)$$

On remarque que $L^2 = 1$, lorsque le bloc de signal e_n à modéliser et la séquence d'excitation v_n^i sont colinéaires. En revanche, $L^2 = 0$ lorsque e_n et v_n^i sont orthogonaux. On peut donc prévoir, que plus la valeur du seuil de classification est proche de 1, plus le nombre de classes ou le nombre de vecteurs du dictionnaire est élevé.

3.2. CHOIX DU DICTIONNAIRE:

La démarche retenue pour déterminer le contenu du dictionnaire des amplitudes consiste à faire varier le seuil de classification par petits pas (par valeurs croissantes). A chaque expérimentation, des mesures du rapport signal sur bruit segmental, ainsi que des tests d'écoute permettent d'évaluer les performances du dictionnaire constitué.

Ceci est illustré par la figure 2, qui représente l'évolution de la taille du dictionnaire en fonction du nombre de blocs de signal modélisés, avec pour paramètres le seuil de classification et le facteur perceptuel.

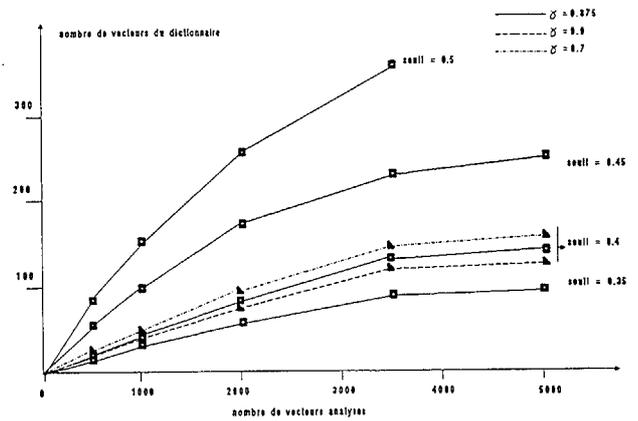


Figure 2: Evolution de la taille du dictionnaire en fonction du nombre de blocs analysés avec pour paramètres, le seuil de classification et le facteur perceptuel.

Ainsi, passer d'un seuil de classification de 0.35 à 0.45 multiplie approximativement par un facteur 2.5 la taille du dictionnaire. On peut noter également, que la taille du dictionnaire, pour une valeur donnée du seuil de classification, dépend du facteur perceptuel. Diminuer ce dernier augmente de manière significative la taille du dictionnaire. Ceci s'explique du fait que plus le facteur perceptuel est faible, plus la matrice d'autocorrélation C se rapproche d'une matrice unité, de sorte que l'information de phase ou de périodicité contenue dans les blocs de signal à modéliser devient prépondérante.

Une autre représentation intéressante est illustrée par la figure 3. Elle met en évidence que le rapport signal sur bruit segmental tend asymptotiquement vers la valeur qui correspond à une modélisation multi-impulsionnelle sans quantification de l'amplitude des impulsions.

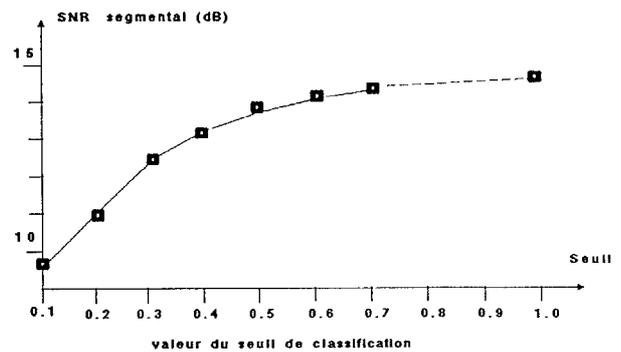


Figure 3: Evolution du rapport signal sur bruit segmental en fonction de la valeur du seuil de classification.

4. PERFORMANCES ET RESULTATS EXPERIMENTAUX:

Un codeur n'est pas constitué uniquement de procédures de modélisation mais également de procédures de quantification des paramètres résultant des modélisations. Ainsi, l'introduction de ces modules de quantification dans la chaîne d'analyse n'est pas sans incidence sur les performances globales du codeur. Pour les codeurs de qualité sub-téléphonique, le placement de ces modules nécessite un soin tout particulier de façon à ne pas anéantir les efforts consentis au niveau des différentes modélisations. En effet, selon l'endroit où les modules de quantification sont placés, la dégradation inhérente au codage peut être partiellement compensée.

