QUATORZIEME COLLOQUE GRETSI - JUAN-LES-PINS - DU 13 AU 16 SEPTEMBRE 1993

IMPLEMENTATION EN TEMPS REEL DE CODEUR/DECODEUR (CELP) DE PAROLE A BAS DEBIT SUR PROCESSEUR DE SIGNAL MOTOROLA DSP 56001

A. GOALIC et S. SAOUDI

Département Mathématiques et Systèmes de Communications
ENST-Br BP 832 BREST FRANCE

RÉSUMÉ

La transmission de données numériques par onde acoustique dans le canal sous-marin est soumise à un certain nombre de lois physiques imposant le choix du codeur-décodeur de parole pour la définition d'un téléphone acoustique numérique sous-marin. Le choix d'un système de codage est un compromis entre la portée envisagée et le débit binaire à transmettre. Les Codeurs Prédictifs à Excitation Par Codes (CELP coder) permettent un codage à bas débit. Une bonne qualité de parole est obtenue par l'utilisation des Paires de Raies Spectrales (LSP : Line Spectrum Pairs). L'utilisation d'un dictionnaire de formes d'ondes ternaires (TCELP) facilite l'implémentation en temps réel. Le choix d'un processeur de signal à virgule fixe de 24 bits s'impose pour des contraintes temps réels, industrielles et économiques.

ABSTRACT

Data transmission in the acoustic underwater channel depends on physical rules, which imposes the choice of speech coding when designing an acoustic underwater phone. This is a compromize between the bit rate and the expected reachable distance. The Code Excited Linear Prediction Coder (CELP) has shown its ability to synthesize good quality speech at low bit rates. An improvement in quality is achieved by the use of Line Spectrum Pairs (LSP). The ternary codebook choice decreases computational time. Real time implementation, industrial and economic considerations lead us to use a 24 bit fixed point Digital Signal Processor (DSP MOTOROLA56001).

I. INTRODUCTION

Les codeurs/décodeurs CELP sont des systèmes de codage de type mixte. Les techniques paramétriques permettent d'extraire du signal de parole les redondances à court et long terme. Après décorrélation du signal original, le résidu est modélisé temporellement par une forme d'onde, cette dernière est sélectionnée dans un dictionnaire en utilisant le principe de l'analyse par synthèse. De nombreuses structures de dictionnaire ont, ainsi été définies : gaussiens, stochastiques, algébriques, combinaison linéaire de vecteurs indépendants. L'objet du présent article est de présenter les implémentations, dans une optique temps réel : au paragraphe 2, du codeur TCELP : au paragraphe 3, du décodeur TCELP. Dans le paragraphe 4, nous présentons les résultats obtenus du système en fonctionnement temps réel. Le dernier paragraphe présente les conclusions et les perspectives.

II. CODEUR TCELP

II.1 PRINCIPE DU CODEUR

Le codeur comprend trois parties principales (fig 1) :

- La prédiction court terme a pour but de modéliser le canal vocal, le filtrage inverse enlève donc les redondances court terme. La fonction de transfert du canal est modélisé par un filtre tous pôles :

$$H(z) = \frac{1}{A_p(z)}$$

$$\text{ou } A_p(z) = 1 + \sum_{i=1}^{p} a_i(z)$$

les paramètres $\{a_i\}_{i=1..p}$ sont les coefficients de prédiction, p étant l'ordre de prédiction. On définit alors les deux polynômes suivants :

$$P_{p+1}(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1})$$

$$P^*_{p+1}(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1})$$

Une autre forme de ces polynômes fait apparaître les paires de raies spectrales $\{w_i\}_{i=1,2,..p}$ pour p pair :

$$P_{p+1}(z) = (1+z^{-1}) \prod_{i \text{ impair}} (1-2\cos w_i z^{-1} + z^{-2})$$

$$P^*_{p+1}(z) = (1-z^{-1}) \prod_{i \text{ pair}} (1-2\cos w_i z^{-1} + z^{-2})$$

La version anti-symétrique de l'algorithme de LEVINSON éclaté nous fourni les coefficients de matrices dont les valeurs propres sont les paramètres LSP[2].

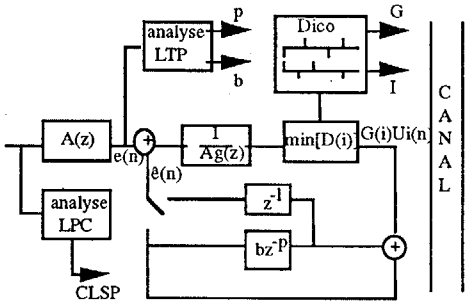

Fig 1 : Codeur TCELP

- La prédiction long terme a pour but de modéliser les redondances de la source vocale. Après filtrage inverse du signal original, nous obtenons le signal glottique e(n). Un échantillon de ce signal peut être prédit à partir d'une combinaison linéaire des échantillons précédents. Pour notre application, un prédicteur avec un seul coefficient est utilisé :

$$P(z) = 1 - bz^{-p}$$

Pour les signaux périodiques p correspond au pitch, c'est aussi le rapport entre la fréquence d'échantillonnage et la fréquence fondamentale. Le paramètre b est le coefficient de prédiction, sa valeur est d'autant plus près de 1 que le signal est voisé. Une méthode simplifiée appelée "Corrélation-peak-picking" constitue un bon compromis, précision-temps de traitement, pour déterminer ces paramètres. Elle consiste à maximiser la fonction d'autocorrélation du signal glottique.

- Le dictionnaire regroupe l'ensemble des formes d'ondes, l'une d'entre elles étant susceptible de modéliser le résidu de signal obtenu par décorrélation, court et long terme, du signal original. Basée sur la minimisation d'un critère énergétique, la sélection de la forme d'onde demande un très grand nombre de calculs. Ceci est du au fait qu'elle entraine de lourdes opérations de filtrage visant à masquer le bruit sous les formants du signal de synthèse et donc à en améliorer la qualité. Pour diminuer la charge de calcul on utilise un dictionnaire ternaire, chaque échantillon du dictionnaire est à valeur dans [+1,0,-1]. Ce qui entraine une diminution du nombre d'opérations, une simplification des opérations de filtrage et le non stockage du dictionnaire en mémoire[1].

II.2 IMPLEMENTATION DU CODEUR SUR DSP

Celle-ci est réalisé sur le processeur 24 bits à virgule fixe MOTOROLA 56001 (fréquence d'horloge : 27 Mhz). Elle reprend les 3 parties précédemment présentées :

1) *Analyse court terme*

Pour simplifier l'implémentation en temps réel, seuls les cosinus des LSP (CLSP) sont effectivement calculés, ce sont ces valeurs qui servent au décodeur[1]. Le signal est segmenté en trame de 20 ms (160 échantillons). Pour limiter les effets de bords l'analyse court terme est effectué sur 224 échantillons (recouvrement inter-trame). Le processus de calcul est le suivant :

- Acquisition des échantillons à 8 Khz sur un convertisseur analogique numérique de 16 bits par interruptions.
- Multiplication par fenêtre de hamming.
- Calcul des onze premiers termes de la fonction d'autocorrélation.
- Calcul des coefficients des matrices par l'algorithme de LEVINSON éclaté, l'extraction des valeurs propres est réalisée par la méthode de la bissection.
- Quantification scalaire non uniforme des coefficients CLSP, chacun est codé sur 3 bits (Tableau 1). Cette quantification assure également la relation d'ordre des coefficients, condition nécessaire et suffisante de stabilité du filtre de synthèse court terme au décodeur.

Paramètres	Trame(ms)	Bits/trame	Débit(bits/s)
CLSP(10)	20	30 x 1	1500
Glpc	20	7 x 1	350
B ltp	10	3 x 2	300
Pitch	10	7 x 2	700
Gq	5	3 x 4	600
I index	5	10 x 4	2000
	TOTAL	109	5450

Tableau 1 : Débit de codage

2)*Analyse long terme*

La première étape consiste à calculer les coefficients de prédiction $\{a_i\}_{i=1..p}$ à partir des valeurs quantifiées des CLSP. Une optimisation du calcul est obtenue en exprimant ces paramètres en fonction des $\{\cos w_i\}_{i=1..p}$:

$$a_i = F(\cos w_1,..,\cos w_p) \quad i = 1,..,p \quad a_0 = 1.$$

Après décorrélation court terme du signal original, le calcul des paramètres long terme (pitch p et gain de prédiction b) s'effectue 2 fois par trame (80 échantillons). La méthode consiste, dans sa version simplifiée, à chercher le maximum de la fonction d'autocorrélation :

$$\sum_{n=0}^{N-1} e(n)e(n-p) \quad N = 80, \quad p = 16..143.$$

La fonction d'autocorrélation est calculée pour les valeurs entières de p entre 16 et 143 (fréquence fondamentale entre 56 et 500 Hz, permettant donc de modéliser les voix enfantines, féminines et masculines).

Le coeur du processeur et son architecture parallèle [3] permettent une détermination précise et rapide des paramètres p et b qui sont respectivement codés sur 7 et 3 bits (tableau 1). Le gain syllabique est en même temps codé sur 7 bits.

3)*Sélection de la forme d'onde*

La mise à jour de la forme d'onde est effectuée 4 fois par trame, soit donc toutes les 5 ms. Pour diminuer la charge de calcul, l'algorithme est divisé en 2 parties (fig 2):

- La branche basse du schéma, qui est sans mémoire, est calculée une seule fois par trame.
- Les autres, haute et transversale, prennent en compte la mémoire du processus. Elles sont exécutées 4 fois par trame.

Le dictionnaire comporte 1024 formes d'ondes (codage sur 10 bits). Les caractéristiques du dictionnaire ternaire permettent également de limiter la recherche de la forme optimale à la moitié de ces formes, un simple test de signe permettant de la positionner dans l'une ou l'autre partie.

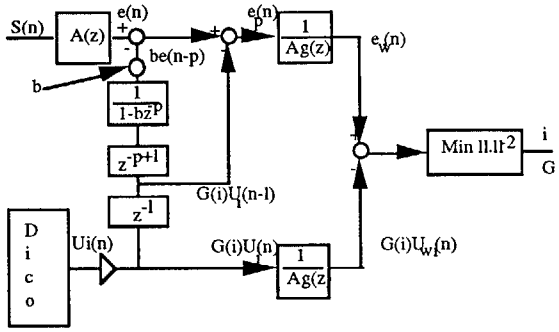

Fig 2 : Schéma d'implémentation

L'utilisation dans le critère d'erreur de la valeur quantifiée du gain d'excitation entraine de multiples accès mémoires. Une réduction significative du temps de traitement est obtenue en optimisant le stockage des données. D'autre part l'utilisation intensive de l'instruction division est évitée. Elle est efficacement remplacée par des opérations quantification-multiplication.

III. DECODEUR TCELP

La structure du décodeur (fig 3) est évidemment beaucoup plus simple et ne pose pas de problèmes particuliers au niveau de la synthèse de parole.

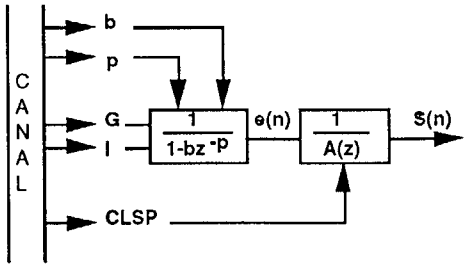

Fig 3 : Décodeur TCELP

Après décodage de la trame, les différents paramètres sont récupérés pour :

- sélectionner la forme d'onde et son gain.
- pour recréer les filtres court et long terme.

Le signal de synthèse s'obtient alors par filtrage long et court terme de la forme d'onde. Comme au codeur les coefficients de prédiction court terme sont calculés à partir des CLSP quantifiés.

IV. FONCTIONNEMENT TEMPS REEL

L'ensemble du dispositif fonctionne aujourd'hui en temps réel sur un banc d'essai permettant la réalisation de diverses mesures.

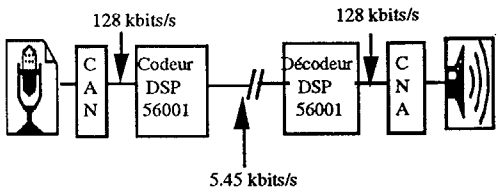

Fig 4 : Banc d'essai

Codeur et décodeur sont chacun implémentés sur un seul DSP MOTOROLA 56001. Ils communiquent à haut débit via leur port série. Fonctionnant en temps réel, l'ensemble du traitement au codeur est réalisé en moins de 20 ms (durée de trame).

1)*Temps de traitement*

Celui-ci n'est pas constant, il dépend principalement des niveaux de quantification atteints, les temps mesurés au codeur sont les suivants :

Analyse LPC : inférieur à 3ms (15%).

Analyse LTP : compris entre 3 et 4 ms (15 à 20%).

Forme d'onde : entre 9 et 13 ms (45 à 65%).

Les temps globaux mesurés varient de 15 ms (signal nul) à 18.5 ms (signaux de parole légèrement bruités de plongeurs). En ce qui concerne le décodeur de parole seul, le temps de traitement est relativement faible (1.5 ms).

Le temps maximal (codeur) étant proche de la valeur admissible, une prochaine implémentation est envisagée sur le DSP MOTOROLA 56002 (fréquence d'horloge 40 MHz).

2) *Réponse en fréquence*

Le système a été testé en temps réel dans la gamme des fréquences du signal de parole [300..3600Hz]. La figure 5 présente les résultats obtenus.

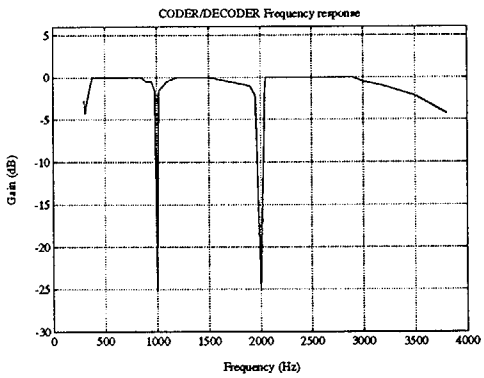

Fig 5 : Réponse en fréquence

Il apparaît clairement que le système ne peut pas synthétiser les 2 fréquences (1000..2000 Hz). Ceci est du à la structure du dictionnaire utilisé. Les formes d'ondes dont la période est un sous multiple pair (1/2,1/4) de 1ms ne sont pas synthétisables. Leurs échantillons significatifs correspondent à des zéros dans le dictionnaire. Une solution serait d'envisager 4 dictionnaires en décalant les bits significatifs, mais la charge de calcul deviendrait rapidement trop importante.

3)*Rapport signal à bruit*

Le système complet tourne également en langage C sur micro-ordinateur de type PC (en temps différé). Pour la même phrase les résultats sont les suivants :

Langage C : 6.53 dB

DSP 56001 : 6.22 dB

Une gestion efficace et "adaptative" des formats de données [4] permet d'effectuer un traitement dont les performances se rapprochent du traitement, en virgule flottante, du langage C. La qualité de parole synthétisée est bonne y compris pour les enregistrements effectués dans des masques de plongeurs.

4)*Compression*

Destiné à servir de codage source dans un téléphone acoustique numérique sous-marin, le système peut également être utilisé en compression de la parole :

Une heure de parole ----------> 2.5 Moctets.

V. CONCLUSIONS/PERSPECTIVES

Cet article présente le fonctionnement en temps réel d'un système unidirectionnel de codage de la parole à bas débit. Codeur et décodeur sont implémentés sur des processeurs de signaux en virgule fixe MOTOROLA 56001 (27 MHz). L'analyse LPC utilise la version anti-symétrique de l'algorithme de LEVINSON éclaté pour déterminer les paires de raies spectrales (LSP). Le dictionnaire de formes d'ondes est de type ternaire. La parole synthétisée est de bonne qualité.

La suite des travaux consiste :

- à mettre en place la transmission par voix acoustique (liaison codeur-modulateur, démodulateur-décodeur).
- à introduire uu codage canal pour corriger les erreurs de transmission.

REFERENCES

[1] A. Goalic, C Laot and S. Saoudi,"Real time implementation of a low bit rate coder for an acoustic underwater phone on a fixed point DSP MOTOROLA56001," ICSPAT, Boston, USA, pp. 921-927, November 1992.

[2] S. Saoudi, J.M. Boucher and A. Le Guyader,"A new efficient algorithm to compute the LSP parameters for speech coding", Signal Processing, Vol. 28, No 2, pp.201-212, August 1992.

[3] MOTOROLA semiconductors,"DSP56000/56001 User's manual", 1989.

[4] MOTOROLA semiconductors,"Fractionnal and Integer Arithmetic Using the DSP56000 Family of General-Purpose Digital Signal Processors", 1988.

Ce projet est soutenu financièrement par CBA (Collaboration BRETAGNE-ACOUSTIQUE) et par FRANCE TELECOM.