TRANSMISSION DE LA PAROLE A TRES FAIBLE DEBIT : REALISATION D'UN VOCODEUR 800 BITS/S

Denis ROCHETTE
Alain ALBARELLO

THOMSON-CSF DTC, Laboratoire Traitement du Signal
B.P. 156, 92231 GENNEVILLIERS CEDEX

RESUME

Nous présentons, dans cet article, la simulation et la réalisation temps-réel d'un vocodeur 800 bits/s fondé sur la prédiciton linéaire et la quantification vectorielle du spectre. Après un résumé des algorithmes mis au point en simulation (algorithme à seuil de construction du dictionnaire de spectres et algorithme de recherche du plus proche voisin dans ce dictionnaire), nous décrivons l'architecture matérielle de la maquette temps-réel réalisée.

Des tests d'intelligibilité et des auditions subjectives montrent que la qualité de la parole synthétisée à 800 bits/s est proche de celle obtenue avec les vocodeurs LPC-10 d'un débit trois fois supérieur.

ABSTRACT

This paper describes both simulation and real time implementation of an 800 bits/s linear predictive vocoder based upon vector quantization. A summary of the software algorithms is presented including the codebook generation by a threshold algorithm and the fast nearest neighbour search algorithm.

The hardware configuration is described. It consists of a signal processor board and a memory board. Preliminary informal listening tests and Diagnostic Rhyme Tests indicate that the speech quality at 800 bits/s is close to that achieved at 2400 bits/s with the standard LPC-10 algorithm.

1. INTRODUCTION

Depuis une dizaine d'années, la technique désormais classique de la prédiction linéaire du signal de parole (1) a permis de réaliser des vocodeurs qui transmettent à faible débit (2 à 5 kbits/s) une parole d'intelligibilité et de confort d'écoute parfaitement acceptables (et acceptés) pour des transmissions militaires à bande limitée, et ce pour un coût que l'évolution technologique rend de plus en plus raisonnable. Rappelons que la prédiction linéaire consiste à modéliser le processus de production de la parole en extrayant périodiquement du signal (typiquement toutes les 20 ms) trois facteurs de commande : un filtre numérique auto-régressif (modélisation de la fonction de transfert du conduit vocal), les paramètres d'excitation de ce filtre (modélisation de la source vocale) et l'énergie.

Les paramètres d'excitation comprennent un indicateur binaire caractérisant la nature voisée ou non voisée du signal et, dans le cas voisé, la période de vibration des cordes vocales ou pitch : ainsi le signal d'excitation se limite à deux cas extrêmes, train d'impulsions périodiques à la période pitch pour les sons voisés ou bruit blanc pour les sons non voisés.

Afin de garantir l'interopérabilité des matériels, une norme de codage par prédiction linéaire a été définie par les Etats-Unis et reprise par l'OTAN : le standard LPC-10 2400 bits/s (2). La figure 1 présente un vocodeur compatible de cette norme.

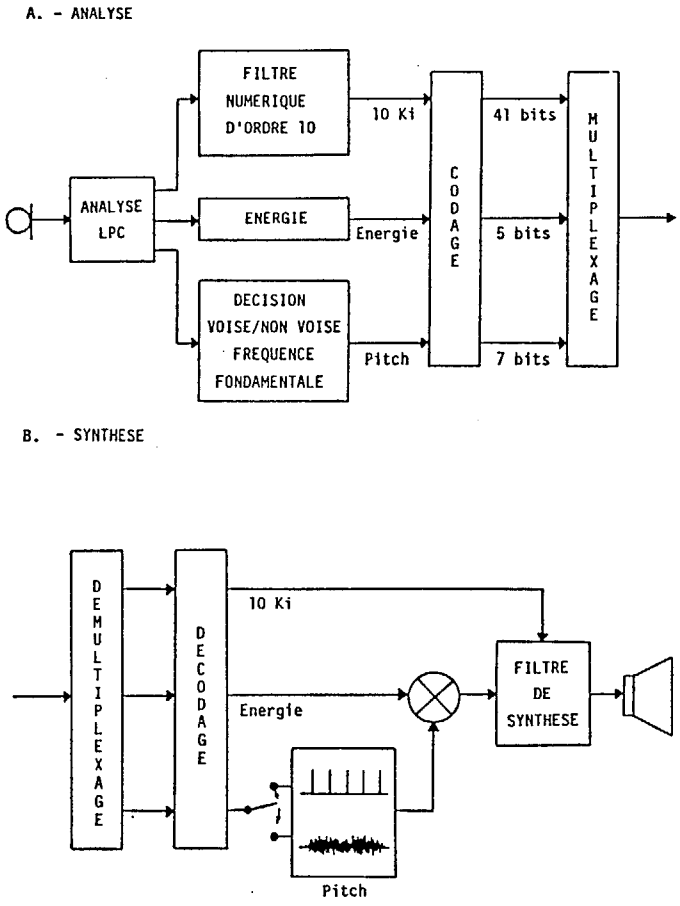

FIGURE 1 - Schéma d'un vocodeur LPC-10 2400 bits/s

Il est important de noter que 2400 bits/s correspond à l'ordre de grandeur du débit alloué aux communications téléphoniques longues distances utilisant la gamme HF (2-30 MHz). Or, des liaisons expérimentales ont permis d'établir que, pour des taux d'erreurs en ligne supérieurs à quelques 10^{-3} (taux fréquemment observés en transmission HF), le vocodeur LPC-10 n'assure plus une qualité opérationnelle suffisante. Il est donc important de savoir coder le signal vocal à un débit encore plus faible (typiquement moins de 1 200 bits/s) sans dégradation significative de la qualité, de manière à utiliser le débit économisé pour assurer aux données numériques transmises un puissant code correcteur d'erreurs.

Depuis quelques années, la quantification vectorielle du spectre (associée au codage par prédiction linéaire) est la technique la plus utilisée pour numériser le signal vocal à très faible débit (3). Elle repose sur le principe suivant : chaque filtre obtenu par prédiction linéaire modélise une certaine configuration du conduit vocal. Supposons alors que l'on dispose d'un dictionnaire de quelques centaines de filtres de référence (ou représentants) susceptibles de refléter statistiquement l'ensemble des configurations possibles. Dans la partie émission du vocodeur, le codage consistera à chercher, dans le dictionnaire, le filtre de référence le plus proche du filtre à traiter. Le numéro de ce représentant sera transmis en lieu et place des coefficients du filtre initial et permettra, à la réception, d'utiliser comme filtre de synthèse le filtre de référence correspondant (fig. 2).

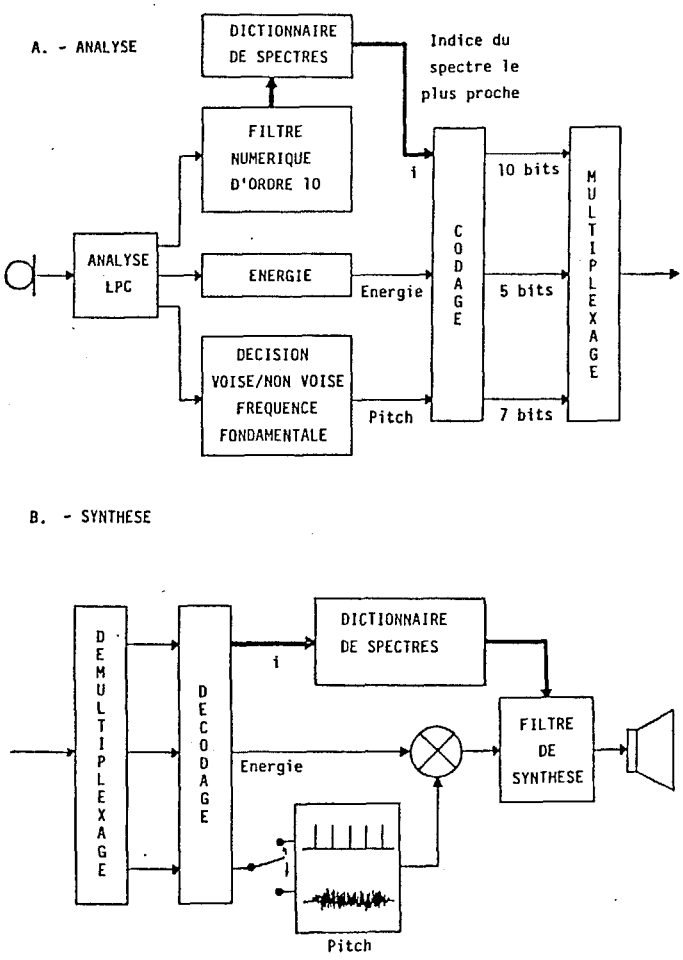

FIGURE 2 - Schéma d'un vocodeur à dictionnaire LPC .

Ce choix de l'indice de numérotation du dictionnaire comme paramètre de transmission de l'information spectrale permet une forte réduction du nombre de bits de codage et par conséquent, du débit de transmission. Par exemple, le standard LPC-10 2400 bits/s utilise 41 bits pour coder scalairement les 10 coefficients de chaque filtre, ce qui correspond en théorie à 241 configurations possibles. En réalité, la distribution non uniforme des vecteurs de dimension 10 que sont les filtres considérés peut être bien représentée par 1024 vecteurs de référence, soit un indice à 10 bits.

Deux problèmes doivent par conséquent être résolus :

- *construire un dictionnaire pertinent (phase d'apprentissage),*
- *disposer d'un algorithme d'accès rapide à ce dictionnaire, afin de pouvoir effectuer la recherche du plus proche voisin d'un spectre à coder en temps réel (phase de codage).*

2. ALGORITHMES

Les algorithmes mis au point ayant déjà été largement décrits (4) (5), nous n'en présentons ici qu'un bref résumé. Rappelons, par ailleurs, que le standard LPC-10 2400 bits/s est utilisé ici comme étage de prétraitements.

L'algorithme de construction du dictionnaire reprend, en l'adaptant, le principe des algorithmes à seuil de distance (6) : un vecteur de la séquence d'apprentissage n'est retenu comme nouveau représentant que si sa distance aux représentants déjà créés est supérieure à un certain seuil. Cet algorithme a été appliqué à une base de données d'environ 15000 vecteurs obtenus par analyse LPC de phrases phonétiquement équilibrées prononcées par 10 locuteurs (5 hommes et 5 femmes). La distance spectrale choisie est le rapport de vraisemblance (7), non pas tant pour sa signification "physique" que pour la simplicité des calculs qu'il requiert. Le dictionnaire obtenu comporte 1024 représentants.

La méthode triviale de recherche du plus proche voisin d'un filtre à coder consiste à calculer sa distance aux 1024 représentants du dictionnaire et à choisir le représentant qui minimise cette distance. Cependant, le coût élevé en temps de calcul d'une telle recherche exhaustive impose d'utiliser un algorithme d'accès rapide au dictionnaire qui réalise un bon compromis entre l'efficacité du codage et le nombre de distances à calculer. Dans cette optique, nous avons conféré au dictionnaire une structure d'arbre binaire équilibré à 5 niveaux. Chaque noeud de l'arbre est caractérisé par un hyperplan, les deux branches reliées au noeud étant associées au signe de l'équation de l'hyperplan. Le niveau 5 de l'arbre se compose de 32 sous-niveaux comportant chacun 32 représentants parmi lesquels on a sélectionné 8 "super-représentants" ou centroïdes. Les 24 autres représentants de chaque sous-niveau sont remplacés par des centroïdes des sous-niveaux voisins. Enfin, on associe à chaque super représentant les numéros de ses 40 plus proches voisins (dans tout le dictionnaire).

L'algorithme d'accès rapide au dictionnaire se déroule en 3 étapes :

- *progression jusqu'au niveau 5 de l'arbre binaire équilibré,*
- *première recherche parmi les 32 (8 + 24) super-représentants dont les numéros appartiennent au sous-niveau atteint.*
- *deuxième recherche parmi les 40 plus proches voisins du meilleur centroïde obtenu lors de l'étape précédente*

Cet algorithme ne requiert que 77 calculs de distances et fournit un représentant qui est à plus de 70 % le plus proche voisin du spectre à coder et à plus de 90 % l'un des 3 plus proches.

Le codage vectoriel du spectre (sur 10 bits) associé à un codage différentiel sur 3 trames de 22,5 ms de l'énergie et du pitch permet d'atteindre le débit de 800 bits/s (tableau 1).

	2400 bits/s	800 bits/s
Spectre	41 bits	10 bits par trame
Pitch	7 bits	12 bits pour 3 trames
Énergie	5 bits	11 bits pour 3 trames
Synchro	1 bit	1 bit pour 3 trames
TOTAL	54 bits/trame	54 bits pour 3 trames (18 bits/trame)

TABLEAU 1 - Comparaison des codages à 2400 bits/s et à 800 bits/s

3. ARCHITECTURE MATERIELLE

L'architecture de la maquette temps-réel d'un vocodeur à dictionnaire 800 bits/s (figure 3) se compose de 3 cartes standards au format 1/2 ATR développées par THOMSON-CSF/DTC pour les applications en traitement du signal (8) :

- *une carte processeur TMS 32010,*
- *une carte conversion (filtrage dans la bande téléphonique, conversions A/N et N/A pour une fréquence d'échantillonnage de 8 kHz),*
- *une carte mémoire (32 K mots d'EPROM disponibles, le stockage du dictionnaire en utilisant 23 K).*

Le même processeur de traitement du signal effectue en temps réel les traitements d'analyse (partie émission du vocodeur) et de synthèse (partie réception) du signal de parole : deux maquettes peuvent ainsi fonctionner en liaison duplex.

La durée totale de traitement de chaque trame de 22,5ms de signal est égale à 20 ms, la décomposition étant la suivante (tableau 2) :

Prétraitements d'analyse	*= 4 ms*
Codage vectoriel du spectre	*= 2*
Calcul de l'excitation et codage sur 3 trames	*= 7,5*
Synthèse	*= 6*
Gestion et contrôle	*= 0,5*
TOTAL	*= 20 ms*

TABLEAU 2 - Temps de traitement d'une trame de 22,5 ms.

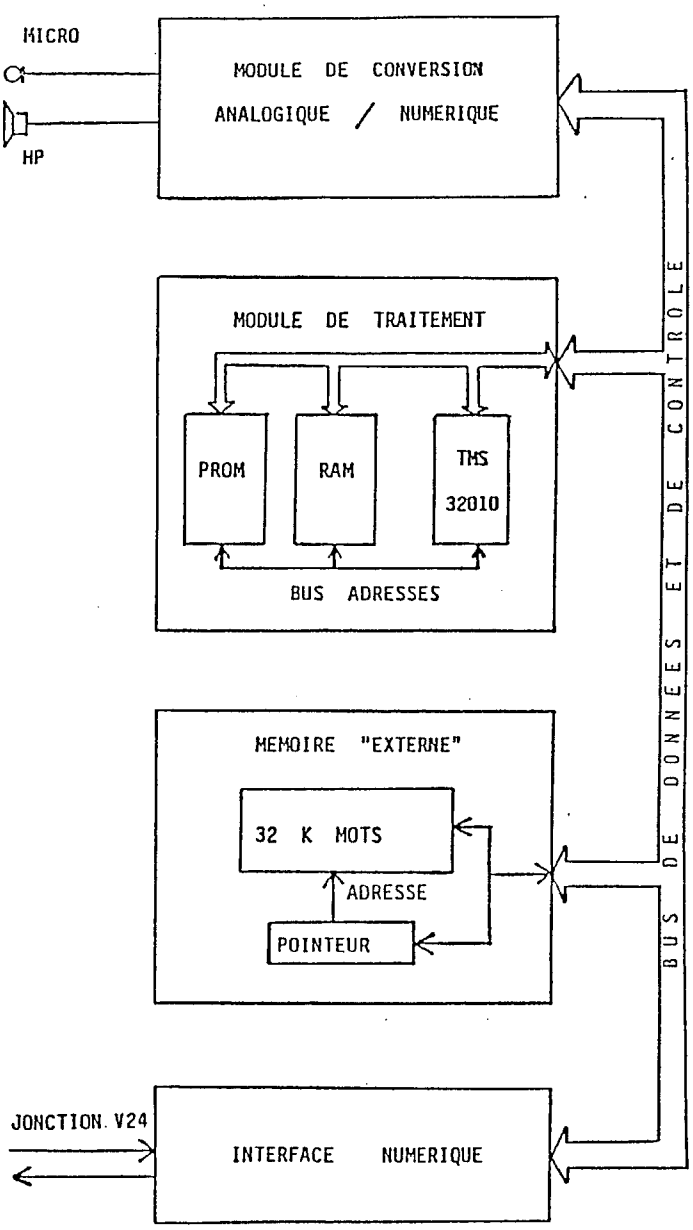

FIGURE 3 - Schéma de l'architecture matérielle du vocodeur 800 bits/s

4. RESULTATS ET CONCLUSIONS

Selon le test de rimes, le vocodeur à dictionnaire 800 bits/s de THOMSON-CSF/DTC a une intelligibilité supérieure à 90 %, ce qui est tout à fait satisfaisant; la perte d'intelligibilité par rapport au standard LPC-10 2400 bits/s (94 % selon le même test) est minime, compte-tenu de la division du débit de transmission par un facteur 3.

Le confort d'écoute, apprécié grâce à de nombreuses comparaisons subjectives, est acceptable bien que légèrement altéré par rapport au standard LPC-10 2400 bits/s.

Cette maquette temps-réel d'un vocodeur à dictionnaire 800 bits/s démontre, tant au niveau des résultats obtenus que de la faible complexité de l'architecture mise en oeuvre, qu'il est désormais possible de répondre aux besoins opérationnels concernant la transmission de phonie numérique protégée sur canal HF.

Cette étude a été partiellement supportée par la Direction des Recherches, Etudes et Techniques (DRET) du Ministère de la Défense.

REFERENCES

(1) MARKEL J.D. & GRAY A.H (1976)
Linear Prediction of Speech
Springer Verlag, Berlin

(2) TREMAIN T.E. (1982)
The Government Standard Linear Predictive Coding Algorithm : LPC-10
Speech Technology, April 1982, 40-49

(3) BUZO A., GRAY A.H., GRAY R.M. & MARKEL J.D. (1979)
Speech Coding based upon Vector Quantization
I.E.E.E. Trans. ASSP-28, 562-574

(4) ROCHETTE D. (1986)
Etude et réalisation d'un vocodeur à dictionnaire LPC 800 bits/s
Thèse Docteur-Ingénieur, I.N.P., GRENOBLE

(5) POTAGE J., ROCHETTE D. & MATHEVON G. (1986)
Les techniques de numérisation de la parole à bas débit applicables aux liaisons navales
Revue Technique THOMSON-CSF vol. 18, N° 1, 171-205

(6) DABOUZ M. & MICLET L. (1983)
Expériences en transmission de la parole à faible débit par vocodeur à classification
Séminaire GALF/GRECO, PARIS, 78 - 89

(7) GRAY A.H. & MARKEL J.D. (1976)
Distances Measures for Speech Processing
I.E.E.E. Trans. ASSP-24, 380-391

(8) LENORMAND E. & ALBARELLO A. (1987)
Famille de cartes standards pour applications multiprocesseurs à base de TMS 32010 ou de TMS 32020
11ème colloque GRESTSI, NICE.