

IMPLANTATION ET EVALUATION D'UN ALGORITHME DE CODAGE DE PAROLE A
16 KBIT/S POUR UN EQUIPEMENT DE RADIOTELEPHONE NUMERIQUE

T. LANGLAIS et Z. PICEL

T.R.T. Télécommunications Radioélectriques et Téléphoniques
BP. 21,92352 LE PLESSIS ROBINSON CEDEX - FRANCE

Les récents développements des technologies VLSI ont permis l'apparition de processeurs de traitement de signal capables de réaliser en temps réel des algorithmes d'une complexité croissante. Dans le domaine du codage de la parole, ces progrès se traduisent par une recherche de réduction du débit numérique sans altération de la qualité.

Nous présentons dans cet article un codeur de signaux de parole fonctionnant à un débit de 16 kbit/s utilisant les techniques du codage en sous-bandes avec une attribution adaptative des débits, et de la quantification vectorielle sphérique pour le codage de l'information auxiliaire.

Des maquettes opérationnelles ont été développées et intégrées à un équipement de radiocommunications numériques avec les mobiles, le SFH 900.

Elles utilisent le processeur de traitement du signal TMS 32020.

1 - INTRODUCTION

Nous présentons dans cet article un codeur de signaux de parole fonctionnant à un débit de 16 kbit/s, utilisant une technique de codage en sous-bandes. Des maquettes opérationnelles ont été développées et intégrées dans un équipement de radiocommunications avec les mobiles, SFH 900, et des essais de fonctionnement du système sur le terrain ont été effectués.

L'algorithme proposé utilise un banc de 12 filtres miroirs en quadrature implantés sous forme d'un réseau polyphasé [1], afin de minimiser la complexité de calcul et les capacités taille de mémorisation. Cette étape est discutée dans la première partie.

Le codage de l'information auxiliaire, constituée par l'énergie de blocs dans les sous-bandes, est effectué en utilisant les principes de la quantification vectorielle sphérique [2].

Cette information auxiliaire quantifiée est utilisée à la fois pour normaliser les blocs d'échantillons dans les sous-bandes, et pour commander une procédure d'attribution adaptative des débits dans les sous-bandes retenues [3]. En fonction de cette attribution des débits, les signaux normalisés sont quantifiés à l'aide de quantificateurs scalaires fixes de 1 à 5 bits. Codage et quantification des informations font l'objet de la seconde partie.

Dans une troisième partie nous décrivons l'implantation du codeur et son insertion dans le système de radiocommunications numériques SFH 900, en développant notamment les points concernant la détection d'activité vocale et l'élaboration d'une stratégie de protection contre les erreurs de transmission.

Recent developments in VLSI technologies have made available the use of dedicated signal processing chips, allowing real time implementation of ever more complex algorithms. This enables an easy evaluation of new concepts in medium and low bit rate speech algorithms.

This article deals with the implementation of a real time 16 kbit/s sub-band speech coder with adaptive bit allocation and vector quantization of short time power levels.

Prototypes have been developed and integrated into a digital radiocommunication network called SFH 900. Field tests have been performed with this experimental system.

The prototype coder uses commercially available DSP chips, the TMS 32020.

Enfin, les performances du codeur sont évaluées en présentant les résultats de tests objectifs et de tests d'écoute.

2 - FILTRAGE EN SOUS-BANDE UTILISANT UN RESEAU POLYPHASE

L'implantation du banc de filtres est basée sur les travaux de Rothweiler [4], généralisés par Masson et Picel [1]. Un filtre prototype passe-bas à réponse implusionnelle finie de N_c coefficients, où N_c est un multiple entier du nombre N de sous-bandes désirées, est d'abord calculé. Ses coefficients peuvent être déterminés en utilisant l'algorithme de Hooke et Javeas [5].

Si f_e est la fréquence d'échantillonnage, des modulations successives du filtre prototype aux fréquences multiples impaires de $f_e/4N$, permettent d'engendrer le banc de filtres.

Les contraintes du filtre prototype doivent permettre d'assurer une quasi parfaite reconstitution du signal (voir figure 2.1), ce qui impose une minimisation des repliements spectraux entre canaux non adjacents.

Elle est réalisée grâce à un choix adéquat des termes de phase [1].

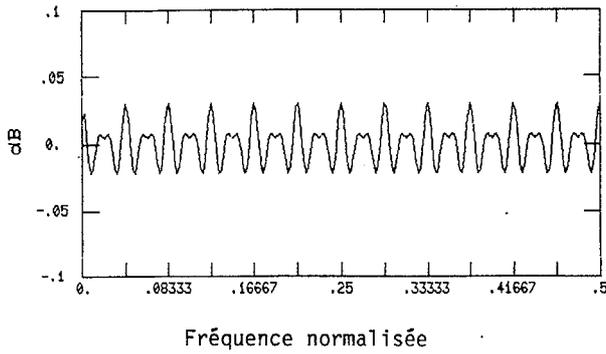


Figure 2.1 - REPOSE COMPOSITE EN FREQUENCE

Faisant appel aux définitions de N , N_c et f_e données ci-dessous, le même coefficient du même filtre du banc de N filtres s'exprime par

$$h_k(n) = 2 \cos \left[2 \pi \frac{2k+1}{4N} \left(n - \frac{N_c-1}{2} \right) + (2k+1) \frac{\pi}{4} \right] h(n)$$

où $h(n)$ est le même coefficient du filtre prototype.

Quant au banc de filtres de synthèse le même coefficient du même filtre s'écrit :

$$h'_k(n) = 2 \cos \left[2 \pi \frac{2k+1}{4N} \left(n - \frac{N_c-1}{2} \right) - (2k+1) \frac{\pi}{4} \right] h(n)$$

L'implantation sous forme de réseau polyphasé est basée sur la décomposition du filtrage en deux blocs. Après décimation d'un facteur N , les signaux sont filtrés par le filtre prototype lui-même sous-échantillonné. Cette opération est suivie d'une modulation utilisant la transformée en cosinus discrète doublement impaire. Par rapport à l'implantation classique utilisant une structure en parallèle, un gain important en complexité de calculs et en nombre de valeurs à mémoriser peut être obtenu (voir tableau 2.2).

ANALYSE + SYNTHÈSE	RESEAU POLYPHASE	IMPLANTATION PARALLELE
Nb de multiplications par bloc ($N \cdot f_e$)	$\frac{1}{2} (2\lambda + \log_2 N) N^*$ ou $(2\lambda + N) N$	λN^2
Nb de coefficients à mémoriser	$(\lambda + 1) N^*$ ou $(\lambda + N) N$	$2\lambda N^2$

* Avec une DCT rapide

$\lambda = \frac{N_c}{N}$ est le nombre de coefficients du filtre prototype utilisé dans chaque sous-bande.

Tableau 2.2

COMPARAISON ENTRE L'IMPLANTATION PARALLELE ET L'IMPLANTATION SOUS FORME DE RESEAU POLYPHASE

Le filtre prototype retenu a 108 coefficients ($\lambda = 9$) et sa fréquence de coupure est 167.7 Hz. Le banc de filtre (analyse ou synthèse) nécessite 500 mots de programme et 90% de la RAM interne du TMS 32020, principalement pour le stockage des blocs d'échantillons traités. Parallèlement à l'analyse,

un calcul de l'information auxiliaire est effectué.

Celle-ci est constituée par les valeurs des carrés des énergies calculées sur des blocs de huit échantillons par sous-bande, représentant 12 ms de signal de parole.

Les échantillons des 10 premières sous-bandes et l'information auxiliaire correspondante sont transmis au module de codage à la fréquence de 83.3 Hz, par blocs de 80 échantillons.

30% de l'activité du processeur TMS 32020 est imputable au filtrage et à la compression ou l'expansion des données selon la loi de codage MIC-A, dans le cas de données échantillonnées à 8 kHz.

3 - CODAGE ET QUANTIFICATION DES SIGNAUX DE SOUS-BANDES

Le sous-système de codage peut être décomposé en trois modules (voir figure 3.1), que nous examinerons séparément :

- 1 - Quantification du vecteur dont la même composante représente le carré de l'énergie du bloc de huit échantillons de la même sous-bande. Ce vecteur constitue l'information auxiliaire.
- 2 - Normalisation des blocs d'échantillons.
- 3 - Quantification des signaux de sous-bandes normalisés.

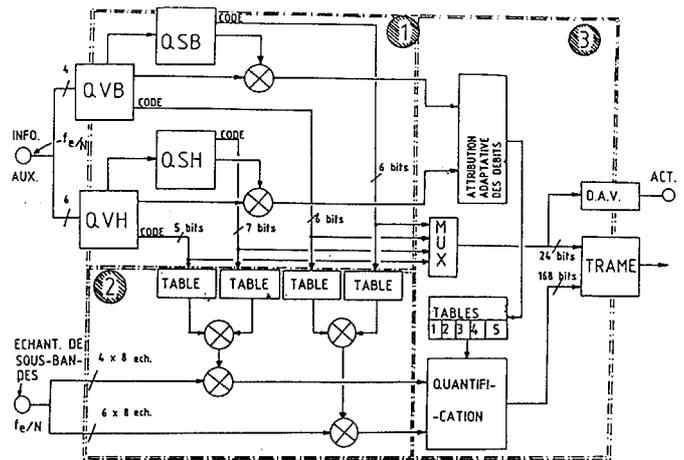


Figure 3.1

SYNOPTIQUE DU CODEUR

La fréquence de calcul de l'information auxiliaire est de 83.3 Hz et accompagne 8 échantillons dans chaque sous-bande.

Pour chaque bloc de 12 ms, 24 bits sont affectés au codage de cette information, et 168 restent disponibles au codage de l'information résiduelle constituée par les signaux de sous-bandes normalisés, soit 14 kbit/s.

Le débit affecté à l'information auxiliaire 2 kbit/s interdit l'utilisation de la quantification scalaire.

Nous utiliserons le concept de la quantification vectorielle sphérique décrit dans [2] et [7]. La direction du vecteur est d'abord quantifiée par comparaison à une table de vecteurs de références obtenus à partir de séquences "d'apprentissage" et de l'algorithme de Linde, Buzo et Gray [6]. Le vecteur choisi est celui qui minimise la distance euclidienne, i.e. qui maximise le produit scalaire.

Pour éviter le calcul de la norme du vecteur à quantifier, on quantifie ensuite la norme du vecteur "projection du vecteur à quantifier" sur la direction choisie précédemment.

Cette norme n'est autre que le produit scalaire calculé à l'étape précédente. Cette dernière quantification est cette fois scalaire ; les coefficients de la table de référence sont eux aussi déterminés par l'algorithme de Linde, Buzo et Gray (voir figure 3.2).

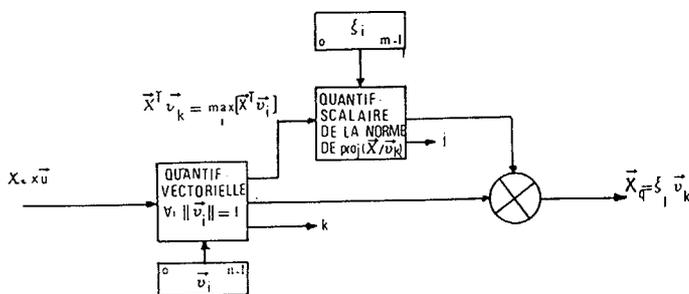
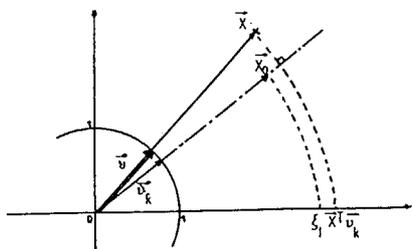


Figure 3.2
PRINCIPE DE LA QUANTIFICATION VECTORIELLE

Pour minimiser la taille des dictionnaires lors de l'implantation, on sépare le vecteur initial de dimension 10 en deux vecteurs de dimensions respectives 4 et 6 ayant chacun leurs propres dictionnaires de formes et représentant, l'un la partie basse du spectre [0,1333 Hz], l'autre la partie haute du spectre [1333 Hz, 3333 Hz] à traiter.

Les quantificateurs vectoriels des parties basse et haute, soient QVB et QVH, contiennent respectivement 32 vecteurs de dimension 4 et 64 vecteurs de dimension 6, soient 512 valeurs, à comparer aux 1024 vecteurs de dimension 10 qui seraient nécessaires dans le cas d'un seul vecteur.

Les quantificateurs scalaires des parties basse et haute, soient QSB et QSH, contiennent respectivement 128 et 64 valeurs.

Tout vecteur de dimension 10 peut donc être quantifié à l'aide de 24 bits.

3.2 - Normalisation des blocs d'échantillons

La grande dynamique des signaux de parole rend nécessaire leur normalisation avant quantification.

Dans le cas du traitement par bloc, on doit donc diviser chaque échantillon de sous-bande par l'énergie du bloc considéré, c'est-à-dire la racine carrée de la même composante du vecteur de l'information auxiliaire quantifiée.

Pour éviter les calculs de racines carrées et les opérations de divisions coûteuses en calculs, on mémorise l'inverse de la racine carrée de chaque composante des dictionnaires QVB, QVH, QSB et QSH. Deux multiplications (une pour obtenir l'inverse de l'énergie quantifiée, l'autre pour normaliser l'échantillon) suffisent pour traiter un échantillon.

3.3 - Quantification des signaux normalisés

Pour chaque bloc de 80 échantillons, 168 bits sont à attribuer. Chaque échantillon se voit attribuer au moins un bit. La répartition des bits restants est effectuée de façon adaptative en se basant sur les valeurs des énergies de blocs dans chaque sous-bande, selon un principe décrit dans [3]. Le nombre maximum de bits attribués à un échantillon est 5. 5 quantificateurs scalaires centrés, obtenus à l'aide de l'algorithme décrit dans [6], sont donc nécessaires.

3.4 - Implantation du système de codage/décodage

L'implantation du sous-système de codage sur le TMS 32020 requiert 2800 mots de programme et utilise 40% de la RAM interne. L'exécution du codage d'un bloc nécessite 2 ms. Le décodage utilise 1800 mots de programme, utilise 40% de la RAM interne et l'exécution du décodage d'un bloc nécessite 1,5 ms.

Les informations nécessaires au codeur sont aussi utilisées par le décodeur, notamment les dictionnaires de forme. L'implantation sur un même processeur en est facilitée.

4 - IMPLANTATION DU PROTOTYPE

Un prototype expérimental d'un codeur de parole à 16 kbit/s basé sur l'algorithme décrit a été développé. Des interfaces analogiques utilisant un COFIDEC pour la conversion en échantillons MIC-loi A, permettent de raccorder un combiné téléphonique standard ou tout équipement audio. Une interface numérique (jonction J64) permet un raccordement au réseau.

La carte ordinateur est basée sur trois processeurs TMS 32020, réalisant respectivement les parties filtrage d'analyse et calcul des carrés des énergies de blocs (P1), codage et décodage (P2) et filtrage de synthèse et conversion en données MIC loi A (P3). La faisabilité à l'aide d'un processeur de type TMS 320C25 a été démontrée.

Le motif de base de la trame à 16 kbit/s comporte 24 octets, soient 192 bits. La fréquence de répétition est 83.3 Hz. La protection contre les erreurs, dans le cas d'une transmission en radiotéléphonie mobile, doit être soigneusement étudiée. Trois codes de Reed-Solomon sont utilisés, apportant une redondance moyenne de deux et un débit de 32 kbit/s après protection.

Les trois premiers octets contiennent l'information auxiliaire et sont protégés par un code de Reed-Solomon (3,5). Les seize octets suivants sont protégés par un code de Reed-Solomon (4,4) et les cinq derniers par un code de Reed-Solomon (5,3).

La trame est constituée de façon à limiter au maximum la durée de l'information contenue dans chaque octet : 1,5 ms pour les octets 3 à 18 ; 3 ms pour les octets 19 à 22 et 12 ms pour le dernier octet. Ceci permet de limiter la durée de propagation d'une erreur de transmission.

La connaissance de l'ordre d'attribution des bits dans la procédure d'attribution adaptative des débits permet de ranger en tête de trame les bits les plus significatifs, et d'assurer une meilleure protection.

Ce schéma de protection est associé à une procédure de détection d'activité vocale pour maximiser la charge du système de radiotéléphonie. Le module de détection d'activité vocale est basé sur les valeurs choisies pour quantifier les normes des vecteurs de l'information auxiliaire c'est-à-dire sur une mesure d'énergie sur toute la bande téléphonique. Dans le



cas d'une liaison half duplex d'une station fixe (environnement non bruité) vers une station mobile 35% de silences peuvent être détectés. Les motifs correspondants ne sont alors pas transmis. La figure 4.1 montre le synoptique de la carte ordinateur du codeur prototype.

Le retard total (émission et réception) inférieur à 50 ms, ne nécessite pas de mise en oeuvre d'un procédé d'annulation d'écho.

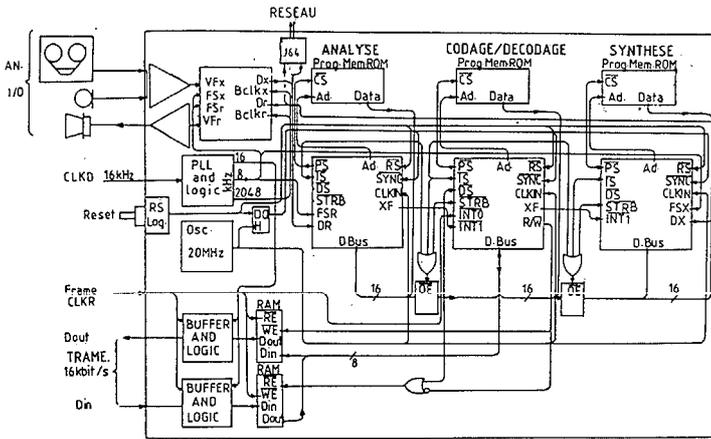


Figure 4.1
SYNOPTIQUE DE LA CARTE CALCULATEUR DU
CODEUR/DECODEUR

5 - RESULTATS DES TESTS OBJECTIFS

Les performances du codeur à 16 kbit/s ont été comparées aux spécifications données dans l'Avis G.712 du CCITT, ainsi qu'au bouclage MIC (64 kbit/s) présent sur le prototype, et nommé "bouclage analogique".

Les courbes désignées par le terme "bouclage numérique" correspondent aux résultats obtenus après codage à 16 kbit/s, transmission sur un canal parfait, non bruité, et décodage.

Elles illustrent la robustesse et la bonne qualité de l'algorithme présenté.

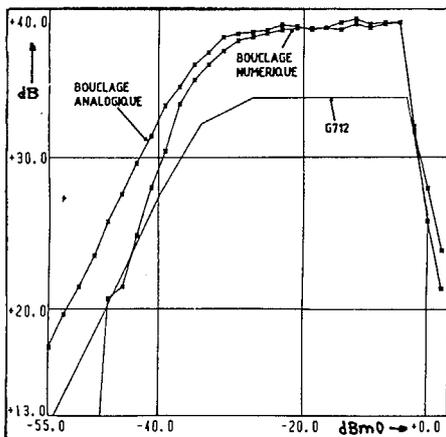


Figure 5.1
VARIATION DE LA DISTORSION EN FONCTION DU NIVEAU
D'ENTREE POUR UN BRUIT COMPRIS ENTRE 350 HZ ET
550 HZ - REFERENCE A - 10 dBm0

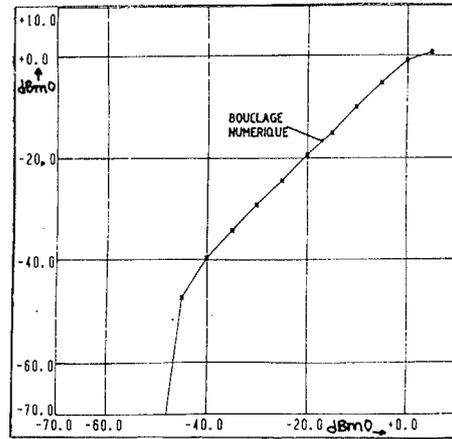


Figure 5.2
POURSUITE DE NIVEAU BRUIT D'ENTREE ENTRE 350 ET
550 HZ - REFERENCE A - 10 dBm0

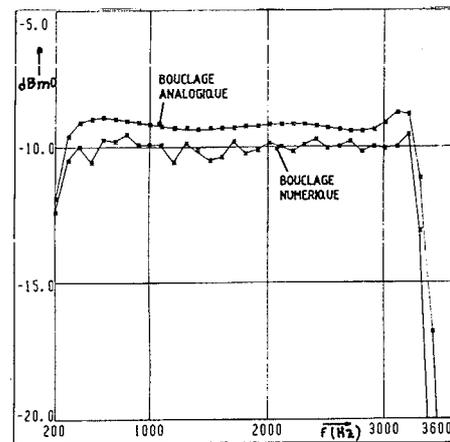


Figure 5.3
VARIATION DU NIVEAU EN FONCTION DE LA FREQUENCE A
- 10 dBm0 (référence : seuil de saturation du
COFIDEC a 813 Hz)

6 - CONCLUSION

Nous avons présenté un prototype de codeur de parole à 16 kbit/s utilisant les techniques du codage en sous-bandes avec attribution adaptative des débits et de la quantification vectorielle de l'énergie. L'implantation du calculateur est basée sur l'utilisation du processeur de traitement de signal TMS 32020.

Codage et décodage peuvent être effectués à l'aide de deux processeurs TMS 32020, ou d'un seul TMS 320C25.

Le codeur réalisé associe une faible complexité d'implantation, une compatibilité vis-à-vis des standards MIC, un faible retard induit (moins de 50 ms) et une bonne résistance aux erreurs de transmission, tout en conservant une restitution de qualité téléphonique du signal.

REFERENCES

- [1] MASSON J., PICEL Z., Proc. ICASSP'85, pp 541-544
- [2] SABIN M.J., GRAY R.M., IEEE Trans. ASSP (1984), pp 474-488
- [3] RAMSTAD J., Proc. ICASSP'82, pp 203-207
- [4] ROTHWEILER J.H., Proc. ICASSP'83, pp 1980-1983
- [5] HOOKE R, JEAVER T. Direct Search Solution of Numerical and Statistical Problems, J. Ass. Comp. Mach. Vol. 8 (1961) pp 212-229,
- [6] LINDE Y., BUZO A., GRAY R.M., IEEE Trans. on Comm. (1980), pp 89-95
- [7] LANGLAIS T., PICEL Z., proc. EUSIPCO 86, pp 419-423