

# SEPTIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 28 MAI au 2 JUIN 1979

---

CODAGE DIFFERENTIEL ADAPTATIF DE LA PAROLE POUR  
LE RESEAU TELEPHONIQUE

Alain LE GUYADER

André PISSARD

Centre National d'Etudes des Télécommunications LANNION

---

## RESUME

L'extension de la numérisation des réseaux de Télécommunications donne un intérêt croissant à l'étude de procédés de codage du signal téléphonique offrant une réduction de débit par rapport au procédé MIC actuellement utilisé (64 kbit/s).

On présente dans une première partie trois codeurs différentiels destinés au codage à 32kbit/s. Ces codeurs, étudiés par simulation, utilisent une quantification adaptative du signal de différence basée sur une estimation locale de la variance ; ils se différencient par la nature du prédicteur, qui est soit figé par ajustement sur le spectre moyen de la parole, soit rendu adaptatif pour suivre l'évolution des caractéristiques spectrales à court terme du signal. Les techniques d'adaptation du prédicteur mettent en oeuvre l'algorithme bien connu du gradient, et une procédure que nous avons dérivée de l'algorithme du filtre en treillis ; elles ne nécessitent pas la transmission des coefficients.

Dans une seconde partie nous donnons les résultats objectifs (rapports signal sur bruit) et subjectifs (test d'écoute) sur des signaux de parole pour les trois codeurs, dont nous avons simulé le fonctionnement aux débits de 32, 24 et 16 kbit/s. Les procédures de test appliquées (isopréférence et jugement de dégradation par catégories) indiquent pour les deux codeurs à prédicteur adaptatif fonctionnant à 32 kbit/s une qualité supérieure à celle du MIC 7 bit.

## SUMMARY

The increasing digitization in Telecommunications networks gives more interest to speech coding processes with reduced transmission rates in comparison with a standard PCM (64 kbit/s).

We describe first three different types of DPCM coders intended to operate at the rate of 32kbit/s. Those coders, which have been simulated on a computer, are equipped with the same adaptive quantizer using a local estimate of the variance of the difference signal. However, each of them uses a different predictor : (1) a stationary one i.e. fitted on the average speech spectrum, and (2) adaptive ones which follow the varying short-term spectral characteristics of signals. The two predictor adaptation techniques implemented are : (a) the well-known gradient algorithm, and (b) a process that we have derived from a lattice filter algorithm. The transmission of the coefficients is not required in both techniques.

Second, we describe objective results (signal to noise ratios) and subjective results (listening tests) for the three coders operating on speech signals at the rates of 32, 24 and 16 kbit/s. Listening tests (isopreference and goodness rating scale) showed that the quality of the two coders with adaptive predictors is better, at the rate of 32kbit/s, than that of PCM with 7 bits.



CODAGE DIFFERENTIEL ADAPTATIF DE LA PAROLE POUR  
LE RESEAU TELEPHONIQUE

**1. - INTRODUCTION : LE CODAGE DIFFERENTIEL DU SIGNAL TELEPHONIQUE ET LA REDUCTION DU DEBIT NUMERIQUE**

La numérisation du réseau des télécommunications permet la transmission de messages de nature diverse dont la plus grosse part est encore, à l'heure actuelle, constituée de signaux téléphoniques. Le codage numérique de ces derniers est obtenu classiquement par le procédé M.I.C. (Modulation par Impulsions et Codage) fournissant un débit numérique de 64 kbit/s. De nouveaux procédés semblent permettre un codage de qualité comparable en exigeant un débit moindre (par exemple 32 kbit/s) ce qui améliorerait l'efficacité de la transmission. L'intérêt économique de tels procédés est manifeste à condition que la complexité des codeurs ne soit pas trop grande.

Nous exposons dans cet article une étude sur le codage différentiel, procédé bien connu dans son principe, mais dont certains aspects n'ont été que peu envisagés pour le réseau téléphonique.

Le codage différentiel est susceptible de répondre à certaines contraintes existant dans le réseau (en particulier l'introduction d'un retard inférieur à 4 ms). Son insertion dans le réseau reste cependant conditionnée par sa capacité à prendre en compte d'autres contraintes comme la présence d'erreurs de transmission et la mise en cascade (codages et décodages successifs dus à la coexistence de tronçons analogiques et numériques), ainsi que la présence sur les lignes de signaux autres que la parole (données de modems).

Dans cette étude nous nous sommes attachés à déterminer l'influence de la prédiction adaptative sur la qualité du codage différentiel. Deux algorithmes de prédiction adaptative dont l'un est basé sur une structure de filtre en treillis sont décrits dans une première partie. Les performances objectives et subjectives correspondantes sont présentées ensuite pour des débits de 32 kbit/s, 24 kbit/s et 16 kbit/s.

**2. - DESCRIPTION DES CODEURS DIFFERENTIELS - ALGORITHMES DE PREDICTION ADAPTATIVE**

**2.1 - Principe du codage différentiel**

Le codage différentiel (voir figure 1) utilise la redondance contenue dans le signal de parole pour calculer une prédiction  $\hat{s}(n)$  du signal  $s(n)$  à chaque instant d'échantillonnage de telle sorte que le signal de différence  $e(n)$  ait une énergie inférieure à celle du signal d'origine  $s(n)$ . On peut alors quantifier  $e(n)$  avec un nombre réduit de niveaux tout en conservant une bonne qualité au signal codé ce qui permet une réduction du débit de transmission par rapport à celui requis par un codage direct (M.I.C).

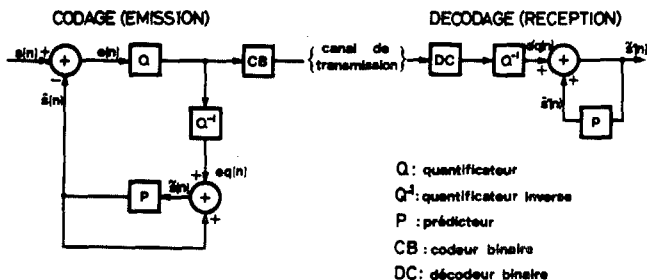


Figure 1 : Codage différentiel

Le quantificateur Q compare le signal de différence  $e(n)$  à des seuils de décision délimitant des plages adjacentes et fait correspondre à chaque valeur  $e(n)$  le numéro de la plage où cette valeur apparaît. Le quantificateur inverse  $Q^{-1}$  transforme le numéro de la plage en la valeur médiane de celle-ci. Le prédicteur P calcule la prédiction  $\hat{s}(n)$  du signal d'entrée à partir des échantillons reconstitués précédents  $\hat{s}(n-1), \hat{s}(n-2)...$ . L'introduction du quantificateur dans la boucle formée par le codeur permet de réinjecter le bruit de quantification ; le signal reconstitué  $\hat{s}(n)$ , à l'émission, et à la réception en l'absence d'erreurs de transmission, est alors égal au signal d'entrée  $s(n)$  au bruit de quantification près :

$$\begin{aligned}
 e_q(n) &= e(n) + q(n) & (1) \\
 \hat{s}'(n) &= \hat{s}(n) + e_q(n) & (2) \\
 e(n) &= s(n) - \hat{s}(n) & (3) \\
 \hat{s}'(n) &= s(n) + q(n) & (4)
 \end{aligned}$$

d'où

La prédiction  $\hat{s}(n)$  du signal à l'instant n est calculée à partir d'une combinaison linéaire des échantillons précédents du signal reconstitué :

$$\hat{s}(n) = \sum_{k=1}^N a_k \hat{s}'(n-k) \quad (5)$$

Les coefficients  $a_k$  caractérisent la corrélation existant entre les échantillons du signal  $s(n)$ .

Le signal de parole est non-stationnaire : il présente des variations de niveau et de spectre à court terme auxquelles un système de codage différentiel efficace doit pouvoir s'adapter. On peut envisager d'adapter le quantificateur seul ou le quantificateur et le prédicteur simultanément.

**2.2 - Adaptation du quantificateur**

L'adaptation du quantificateur consiste à rendre les seuils de décision et les valeurs médianes des plages dépendants du niveau présent estimé du signal de différence. Comme l'adaptation de  $Q^{-1}$  doit se faire également côté réception, l'estimation à l'instant n du niveau actuel du signal de différence (signal transmis) est faite à partir du code binaire transmis, c'est à dire à partir des numéros des plages de quantification utilisés aux instants  $n-1, n-2...$ . Cette estimation peut être faite de différentes façons (GIBSON, JONES et MELSA, 1974 ; CASTELLINO, MODENA, NEBBIA et SCAGLIOLA, 1974 ; GOODMAN et WILKINSON, 1975).

La stratégie d'adaptation du quantificateur que nous avons choisie est celle proposée par GOODMAN et WILKINSON qui revient à faire une estimation séquentielle de la variance du signal différence à l'instant n à partir de l'estimation à l'instant  $n-1$ , compte tenu de la sortie  $\eta(n-1)$  du quantificateur. Dans les faits l'opération de quantification est réalisée comme l'indique la figure 2.

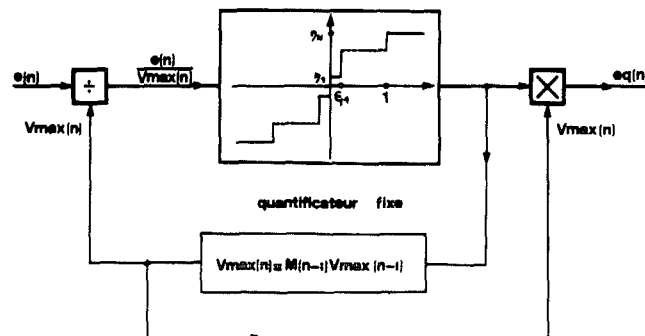


Figure 2 : Adaptation du quantificateur

Le facteur multiplicatif M(n) est donné par (GOODMAN et WILKINSON, 1975) :

$$M(n) = [\hat{y}^2 (1 - \gamma) \eta_i^2(n) + \gamma]^{1/2} \quad (6)$$

Le paramètre  $\gamma$  agit sur la vitesse d'adaptation, qui est d'autant plus faible que  $\gamma$  est voisin de 1.  $\hat{y}$  est un facteur de charge égal au rapport entre le seuil de saturation du quantificateur et l'écart-type de la loi à laquelle est adaptée le quantificateur. Parmi les lois de quantification testées la loi non linéaire optimisée au sens des moindres carrés pour des signaux à répartition exponentielle donne les meilleurs résultats en rapport signal sur bruit ; elle sera seule utilisée dans la suite.

2.3 - Adaptation du prédicteur

Les coefficients  $a_k$  du prédicteur définis par la formule (5) sont calculés de telle sorte que la variance du signal de différence soit minimum :

$$E\{e^2(n)\} = E\left\{s(n) - \sum_{k=1}^N a_k \tilde{s}(n-k)\right\}^2 \quad (7)$$

Compte tenu de l'équation (4), en admettant que le bruit de quantification  $q(n)$  n'est pas corrélé au signal d'origine  $s(n)$ , cela revient à déterminer le prédicteur optimal pour ce signal. Dans un prédicteur fixe les coefficients  $a_k$  sont calculés à partir des caractéristiques moyennes de corrélation de la parole, alors que dans un prédicteur adaptatif ils sont modifiés dans le temps pour suivre les évolutions du spectre à court terme. Un prédicteur fixe n'est pas très efficace car il est constamment désadapté par rapport aux caractéristiques spectrales à court terme de la parole. Nous vérifierons plus loin que les performances d'un codeur équipé d'un tel prédicteur sont nettement inférieures à celles que permet la prédiction adaptative.

2.3.1 - Calcul d'un prédicteur fixe

Nous avons calculé un prédicteur à 4 coefficients à partir de données expérimentales sur le spectre moyen de la parole (LORAND, MINIER et SALIOU, 1975). Les 4 premières valeurs de la fonction d'autocorrélation du signal  $s(n)$  sont calculées à partir des points du spectre moyen pris entre 0 et 4 000 Hz à intervalles réguliers de fréquence (100 Hz) ; les équations classiques d'autocorrélation provenant de la minimisation de (7) (MARKEL et GRAY, 1973) sont ensuite inversées. On obtient alors les coefficients du filtre prédicteur ainsi qu'une approximation du spectre moyen expérimental du signal de parole (fig. 3).

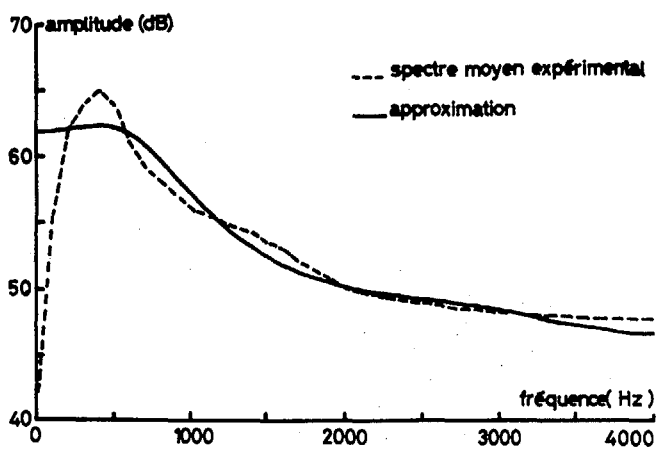


Figure 3 : Spectre moyen de la parole et son approximation.

2.3.2 - Adaptation par l'algorithme du gradient

Principe de la méthode. Les coefficients  $a_k$  optimaux minimisent l'erreur quadratique moyenne (7) et le gradient  $\partial e^2(n)/\partial a_k, k=1, \dots, N$  est alors nul. Afin de se rapprocher de l'optimum les estimations séquentielles des coefficients  $a_k$  (au nombre de 4 dans notre étude) à l'instant  $n$  sont corrigées par un terme proportionnel à la composante du gradient instantané de  $e^2(n)$  suivant la direction  $a_k$  (GIBSON, JONES et MELSA, 1974).

$$a_{k(n+1)} = a_k(n) + \alpha(n) e(n) \tilde{s}(n-k) \quad (8)$$

Le gain  $\alpha(n)$  détermine la vitesse de convergence vers l'optimum :

$$\alpha(n) = \frac{G}{C + \sum_{k=1}^N \tilde{s}^2(n-k)} \quad (9)$$

Les constantes G et C contrôlent la stabilité de l'algorithme : G réduit l'amplitude du gain et C évite que celui-ci ne devienne trop grand lorsque les échantillons du signal reconstitué  $\tilde{s}(n)$  sont très petits (par exemple dans les périodes de silence).

Calculs en virgule fixe. Les signaux d'entrée et de sortie sont représentés par des entiers de valeurs extrêmes  $\pm 2^{11}$  (12 bits), certains calculs intermédiaires pouvant nécessiter 16 bits. Les coefficients du prédicteur sont codés sur 12 bits. Le calcul du gain  $\alpha(n)$  est simplifié en posant  $G = 2^{-5}$  et en mettant le dénominateur de (9) sous la forme d'une puissance de 2 ne pouvant pas prendre de valeurs inférieures à  $2^{-6}$ .

2.3.3 - Filtre prédicteur à structure en treillis.

La méthode du filtre en treillis est un algorithme de déconvolution du signal de parole équivalent au filtrage inverse utilisant le filtre linéaire défini par l'équation (5). Cet algorithme a été utilisé pour la première fois par ITAKURA et SAITO (cités par MARKEL et GRAY, 1976, page 273) en vue d'un système d'analyse-synthèse de la parole à bas débit. Le filtre en treillis comporte une suite de cellules caractérisées par des coefficients de corrélation partielle  $k_i$  liés de façon biunivoque aux coefficients  $a_i$ .

Dans notre étude les coefficients  $k_i$  sont calculés à partir des signaux d'entrée de chaque cellule d'un premier filtre en treillis, dit d'analyse, de sorte que l'énergie des signaux à la sortie de chaque cellule soit minimum. Les coefficients  $k_j$  sont alors utilisés pour calculer la prédiction  $\hat{s}(n)$  à la sortie d'un second ensemble de cellules (fig. 4) constituant un deuxième filtre en treillis (filtre de synthèse) excité par le signal de différence  $e(n)$  (GOLDBERG et ARCESE, 1975). Comme le montre le schéma, l'analyse du signal  $\tilde{s}(n)$  se traduit par des équations récurrentes sur l'ordre  $m$  des  $N$  cellules du filtre :

$$s_{m+1}^+(n) = s_m^+(n) + k_{m+1}(n) s_m^-(n) \quad (10)$$

$$s_{m+1}^-(n+1) = s_m^-(n) + k_{m+1}(n) s_m^+(n) \quad (11)$$

ceci pour  $m = 0, 1, \dots, N-2$

et :

$$s_0^+(n) = \tilde{s}(n), \quad s_0^-(n) = \tilde{s}(n-1) \quad (12)$$

La cellule  $n^\circ m+1$  du filtre d'analyse effectue ainsi une décorrélation partielle des signaux  $s_m^+$  et  $s_m^-$  présents à son entrée.



CODAGE DIFFERENTIEL ADAPTATIF DE LA PAROLE POUR  
LE RESEAU TELEPHONIQUE

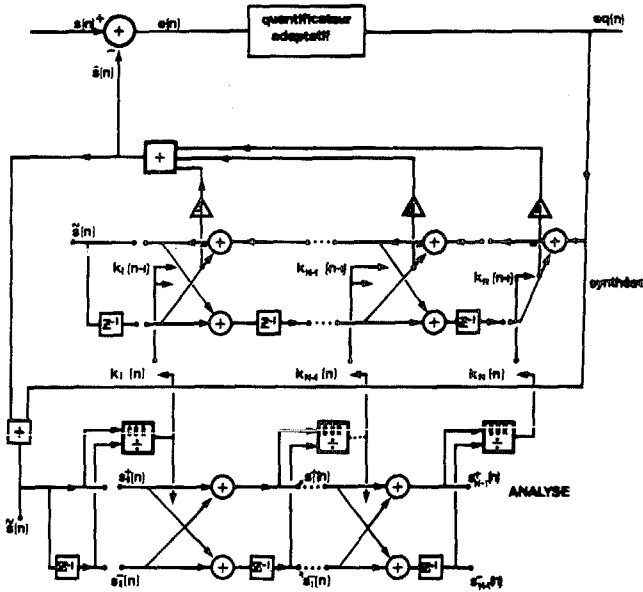


Figure 4 : Codeur différentiel adapté par l'algorithme du filtre en treillis.

L'énergie des signaux à la sortie de la cellule  $m + 1$  est diminuée par rapport à celle des signaux à l'entrée, et s'écrit dans le cas d'un signal analysé stationnaire :

$$E_{m+1}^{\pm} = E_m^{\pm} ( 1 - k_{m+1}^2 ) \quad (13)$$

Dans le filtre de synthèse, le signal codé est progressivement reconstitué, chaque cellule introduisant un terme de corrélation supplémentaire. La relation (10) retournée permet le calcul des signaux  $s_m^+(n)$  de la voie "aller" :

$$s_m^+(n) = s_{m+1}(n) - k_{m+1}(n-1) s_m^-(n) \quad (14)$$

La prédiction  $\hat{s}(n)$  est obtenue par la sommation des signaux  $s_m^-(n)$  de la voie "retour" pondérés par les coefficients  $k_m(n-1)$ : en effet, l'équation (14) itérée donne :

$$s_0^+(n) = - \sum_{m=0}^{N-1} k_{m+1}(n-1) s_m^-(n) + s_N^+(n) \quad (15)$$

qui est identifiée immédiatement à (voir figure 4) :

$$\tilde{s}(n) = \hat{s}(n) + eq(n) \quad (16)$$

Les coefficients  $k_i$  à l'instant  $n$  sont obtenus par la méthode d'ITAKURA et SAITO citée plus haut : chaque  $k_i(n)$  est défini par une approximation du rapport de l'intercorrélacion entre les signaux  $s_m(n)$  et  $s_m^-(n)$  à l'entrée de chaque cellule du filtre d'analyse. Les estimations des intercorrélacions  $C_m(n)$  et des énergies  $D_m(n)$  sont obtenues par filtrage passe-bas du premier ordre défini par la constante  $\beta$ :

$$C_m(n) = \beta C_m(n-1) - 2(1-\beta) s_m^+(n) s_m^-(n) \quad (17)$$

$$D_m(n) = (\beta D_m(n-1) + (1-\beta) (s_m^{+2}(n) + s_m^{-2}(n)) ) \quad (18)$$

$$k_{m+1}(n) = \frac{C_m(n)}{D_m(n)} \quad (19)$$

En fait nous avons utilisé le coefficient calculé à l'instant  $n$  pour la synthèse à l'instant  $n+1$  des signaux en sortie de chaque cellule du filtre d'analyse : cela équivaut à remplacer  $k_{m+1}(n)$  par  $k_{m+1}(n-1)$  dans les équations (10) et (11). On peut montrer (MAKHOUL et VISWANATHAN, 1978) que cette procédure de calcul des coefficients est équivalente à l'adaptation des coefficients  $k_i$  par la méthode du gradient où le gain de l'algorithme est calculé à partir de l'énergie à l'entrée de chaque cellule du filtre d'analyse.

Le nombre de coefficients est de 4, ce qui assure un bon compromis entre l'efficacité de la prédiction et la complexité (nombre d'opérations). Un filtre à 8 coefficients a également été testé dans les codeurs à 24 kbit/s et 16 kbit/s, en vue de compenser la perte de qualité due à la quantification plus grossière.

Calculs en virgule fixe. Le choix d'un filtre passe-bas pour les estimations conduit à des coefficients  $k_i$  dont les valeurs absolues sont inférieures à l'unité ce qui garantit la stabilité du filtre de synthèse ; c'est un avantage par rapport à la méthode du gradient appliquée aux coefficients  $a_k$ , où il n'existe pas de moyen simple de contrôle de la stabilité. De plus, le procédé d'estimation défini par (17) et (18) facilite les calculs en virgule fixe : les estimations sont définies sur 12 bits.

Dans le filtre de synthèse, les coefficients  $k_i$  sont définis sur 12 bits et les calculs intermédiaires sur 16 bits, ce qui s'est avéré largement suffisant.

Un inconvénient de la méthode est qu'elle requiert une quantité importante de calculs : 26 multiplications et 4 divisions par échantillon (avec  $N = 4$  coefficients) dans sa forme actuelle contre 8 multiplications pour l'algorithme du gradient simplifié.

2.34 - Comparaison entre treillis et gradient sur un signal stationnaire (32 kbit/s)

On a représenté sur la figure 5 le spectre original de la voyelle "O" prononcée de façon soutenue par un locuteur masculin, ainsi que les spectres modèles représentant le module de la fonction de transfert des filtres prédicteurs gradient et treillis. On observe que le treillis donne un meilleur ajustement global sur les maxima du spectre, ainsi qu'une meilleure approximation aux hautes fréquences, ce dernier point s'expliquant par l'influence plus grande du bruit de quantification dans l'algorithme du gradient. La prédiction fournie par le filtre en treillis est donc plus efficace, ce qui est confirmé par la valeur efficace trois fois plus faible de l'erreur de prédiction.

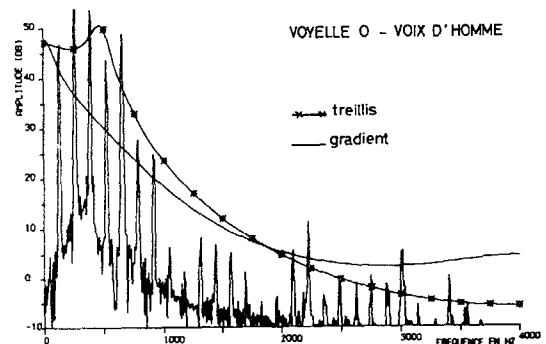


Figure 5 : Spectre original et modèles - voyelle "O"



CODAGE DIFFERENTIEL ADAPTATIF DE LA PAROLE POUR  
LE RESEAU TELEPHONIQUE

3. - EVALUATION ET COMPARAISON DES PERFORMANCES

L'insertion du codage différentiel dans le réseau téléphonique est conditionnée par le respect de contraintes de qualité. En l'absence de normalisation sur la question à l'heure actuelle, il semble raisonnable de rechercher une qualité au moins égale à celle du codage MIC à 7 bits théorique dont il est facile de déterminer le comportement dans tous les cas de figure que l'on peut rencontrer sur le réseau : affaiblissements, nombre de codages et décodages, présence d'erreurs, etc... L'étude décrite ici est relative à un seul codage - décodage et à un fonctionnement des codeurs à leur niveau nominal en l'absence d'erreurs de transmission. La part la plus importante des résultats concerne le codage à 32 kbit/s ; nous présentons également des résultats à 24 kbit/s et 16 kbit/s, débits pour lesquels la qualité du codage différentiel est insuffisante pour l'utilisation dans le réseau ; l'étude à ces derniers débits avait pour but de montrer que les algorithmes de prédiction adaptative fonctionnent correctement en présence d'un signal fortement bruité par une quantification grossière.

3.1 - Méthodes d'évaluation des performances

Les performances des codeurs sont évaluées au moyen de critères objectifs et subjectifs. Les critères objectifs correspondent à des indices physiques sur le signal codé : puissance du bruit de codage ; par ailleurs, la nature particulière de ce bruit est telle que son influence sur la qualité du signal perçu n'est à l'heure actuelle déterminée avec sûreté que par des tests subjectifs. Les signaux utilisés sont des phrases enregistrées en chambre sourde à travers un microphone linéaire et filtrées dans la bande 200 Hz à 3400 Hz par un filtre Butterworth 8<sup>ème</sup> ordre ; la numérisation en vue du traitement par simulation est faite sur 12 bits en linéaire, à la fréquence d'échantillonnage de 8 kHz.

3.1.1.- Indices objectifs de qualité

Le programme de simulation calcule le rapport signal à bruit global RSB et le rapport signal à bruit par segments RSB seg.

Le rapport signal à bruit global est défini par :

$$RSB = 10 \log_{10} \frac{\sum s^2(n)}{\sum (g(n) - s(n))^2} \quad (20)$$

la sommation est effectuée sur l'ensemble des échantillons de chaque phrase.

Le rapport signal à bruit par segments est défini par :

$$RSB_{seg} = \frac{1}{N} \sum_{i=1}^N 10 \log_{10} \frac{\sum s^2(n)}{\sum (s(n) - \hat{s}(n))^2} \quad (21)$$

Les sommations partielles sur n sont effectuées sur des blocs de 128 échantillons (16 ms) ; N est le nombre de blocs constituant une phrase. Le rapport signal à bruit par segments défini par (21) est présenté par certains auteurs (MC DERMOTT, SCAGLIOLA et GOODMAN, 1978) comme un des indices qui sont de bons indicateurs des caractéristiques subjectives du codage ; nous nuancerons cette affirmation au vu des résultats présentés ici.

Le gain de prédiction Gp a été également calculé ; c'est une mesure de l'efficacité de la prédiction sur l'ensemble des échantillons de la phrase :

$$G_p = 10 \log_{10} \frac{\sum s^2(n)}{\sum e^2(n)} \quad (22)$$

3.1.2 - Tests subjectifs

Deux types de tests subjectifs ont été mis en oeuvre : l'isopréférence et le jugement de dégradation par catégories.

La méthode d'isopréférence consiste à dégrader le signal de référence en lui ajoutant au moyen d'un appareil adéquat (Mesureur d'Agrément des Lignes Téléphoniques, MALT) un bruit multiplicatif (c'est-à-dire dont l'amplitude est à chaque instant proportionnelle à celle du signal) d'un niveau tel que l'impression subjective soit la plus proche possible de celle produite par le bruit de codage : on obtient ainsi le rapport signal à bruit d'isopréférence  $RSB_{iso}$ . L'ajustement est effectué automatiquement à partir de la suite des réponses du sujet.

Le jugement de dégradation par catégories consiste à classer, par rapport à la référence, un ensemble de dégradations apportées par différents types de codages suivant une échelle à 5 notes couvrant une large plage de dégradations. L'échelle choisie est celle recommandée par le C.C.I.R., la note 5 traduisant une dégradation imperceptible et la note 1 une dégradation très gênante ; il était suggéré aux auditeurs d'utiliser l'ensemble de l'échelle. Ce test a été mis en oeuvre ici dans le but d'évaluer le codage à 32 kbit/s.

Le corpus utilisé est constitué comme suit :

Isopréférence : deux phrases, prononcées chacune par trois locuteurs (deux femmes et un homme) :

P1 : "le crédit est consenti par la banque"

P2 : "le sculpteur taille la pierre".

Les indices objectifs mentionnés plus haut ont été calculés sur ces phrases.

Jugement par catégories : aux deux phrases précédentes on a ajouté :

P3 : "le docteur a ordonné un médicament"

P4 : "le charretier fouette son cheval".

Quatre locuteurs (les trois précédents plus un deuxième locuteur masculin) ont prononcé chacun une phrase différente.

Nous considérons que les résultats obtenus à partir de ce corpus réduit sont significatifs, compte tenu d'autres résultats (non présentés ici) où il intervenait au sein d'un corpus plus étendu.

Six à dix auditeurs, dont quelques uns entraînés, ont participé aux tests ; l'écoute a été faite en salle acoustiquement isolée, sur casque binaural de qualité Hi Fi, à un niveau "confortable".

3.2 - Présentation et comparaison des résultats

Dans la suite on désignera par les lettres F, G et T les codeurs munis respectivement de prédicteurs fixes et adaptés par les méthodes du gradient et du filtre en treillis.

3.2.1 - Résultats à 32 kbit/s

Le tableau 1 donne les performances moyennes des 3 codeurs sur le corpus du test d'isopréférence (les valeurs entre parenthèses mesurent la demi largeur des intervalles de confiance à 99 % sur les moyennes).



CODAGE DIFFERENTIEL ADAPTATIF DE LA PAROLE POUR  
LE RESEAU TELEPHONIQUE

Indice Codeur (dB)	RSB	RSBseg	Gp	RSBiso
T	29.1	25.8	10.9	37.4 (1.5)
G	27.4	24.5	9.1	36.6 (1.4)
F	24.6	21.8	6.3	31.9 (1.16)

Tableau 1 : Performances moyennes à 32kbit/s.

Les résultats objectifs font apparaître une supériorité du codeur T sur le codeur G, de 1.7dB en RSB et 1.3 dB en RSB seg ; le codeur G est lui-même supérieur au codeur F de 2.8dB en RSB et 2.7 dB en RSBseg. Le gain de prédiction Gp présente des écarts correspondants voisins : 1.8dB entre les codeurs T et G et 2.7 dB entre les codeurs G et F ; l'amélioration des rapports signal à bruit objectifs est donc due au gain de prédiction, compte tenu du fait que l'adaptation du quantificateur est identique pour les 3 codeurs.

Les résultats d'isopréférence sont nettement supérieurs aux résultats objectifs : ce fait provient de la prépondérance du bruit de saturation de pente sur le bruit de quantification, plus gênant subjectivement à niveau égal (JAYANT, 1974). Les écarts observés vont cependant dans le même sens.

Nous avons appliqué le test du t pour déterminer le pourcentage de signification des écarts observés : ainsi, l'écart de 4.7 dB entre les codeurs G et F est significatif à plus de 99 % : ceci confirme ce que laissent présumer les résultats objectifs et fait apparaître le net avantage de la prédiction adaptative sur la prédiction fixe.

Le jugement de dégradation par catégories fait intervenir des dégradations s'étendant depuis le M.A.L.T réglé sur RSBiso = 15 dB jusqu'au meilleur codeur ; on a inclus entre autres un MIC 7 bits provenant d'une maquette de laboratoire. Les résultats portent sur 40 réponses par codeur. Les codeurs T et G ont des notes supérieures à 4.75, ce qui traduit une dégradation très peu audible.

Le codeur F a la note 4, pratiquement égale à celle du MIC 7 bit ; cela correspond à une dégradation audible, mais non gênante. Le test du t montre que l'écart entre ces deux notes est très significatif : l'avantage de la prédiction adaptative est ainsi confirmé.

On trouvera dans une autre publication (LE GUYADER et PISSARD, 1978) des résultats et des commentaires plus détaillés sur le codage à 32 kbit/s.

### 3.2.2 - Résultats à 24kbit/s et 16 kbit/s

Le tableau 2 donne les performances moyennes, dans les mêmes conditions que le tableau 1, des 3 codeurs précédents auxquels on a adjoint un codeur possédant un prédicteur à filtre en treillis à 8 coefficients (T8)\*.

\* Les paramètres G et  $\rho$  intervenant dans l'adaptation des prédicteurs ont été modifiés de telle sorte que celle-ci soit ralentie, afin de limiter l'effet des perturbations dues à la quantification plus grossière du signal de différence.

Indice (dB)		RSB	RSBseg	Gp	RSBiso
Codeur					
24 kbit / s	T8	23.7	19.1	11.3	32.1(1.5)
	T4	22.6	18.3	10	30.1(1.5)
	G	19.6	16.7	8.3	29.2(1.35)
	F	18.5	14.8	6	26.7 (0.4)
16 kbit / s	T8	16.4	12.1	9.4	24.1 (2)
	T4	15.3	11.4	8.3	22.6(2.15)
	G	13.9	10.6	6.9	21.1(1.9)
	F	12.7	8.7	5.5	17.5(0.43)

Tableau 2 : Performances moyennes à 24kbit/s et 16kbit/s.

#### Résultats objectifs :

- la perte de 6 à 7 dB en RSB et RSBseg lorsqu'on diminue le débit correspond à la disparition d'un bit de codage, ce qui en gros double l'amplitude du bruit de quantification. Cependant les codeurs à prédicteur adaptatif préservent leur avantage sur le codeur à prédicteur fixe.

- en dépit d'une certaine diminution, le gain de prédiction reste du même ordre de grandeur qu'à 32kbit/s ; le codeur T8 présente un gain de prédiction supérieur de plus de 1dB à celui du codeur T4.

On en déduit que l'adaptation séquentielle du prédicteur conserve son intérêt aux débits de 24kbit/s et 16kbit/s, bien que la quantification plus grossière perturbe les algorithmes plus fortement.

Les résultats subjectifs, encore très supérieurs aux résultats objectifs, montrent aussi un net avantage des codeurs à prédicteur adaptatif. Le test du t donne les indications suivantes :

- à 24kbit/s, l'écart de 2.5 dB entre les codeurs G et F est significatif à plus de 99 % ; l'écart de 2 dB entre T8 et T4 est significatif à plus de 95 %.

- à 16kbit/s, l'écart de 3.6 dB entre G et F est significatif à plus de 99 %. De même, l'écart de 3 dB entre T8 et G est significatif à plus de 99 %.

L'ensemble des résultats donnés dans les tableaux 1 et 2 est représenté figure 6.

CODAGE DIFFERENTIEL ADAPTATIF DE LA PAROLE POUR  
LE RESEAU TELEPHONIQUE

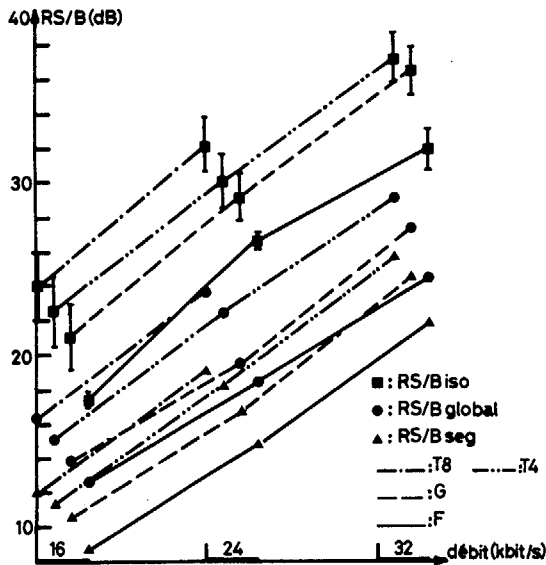


Figure 6. : Résultats moyens aux trois débits étudiés.

On observe :

- un écart entre RSBiso et RSB de 5 à 9 dB pour tous les codeurs et tous les débits : ceci provient, comme on l'a dit plus haut, de la nature du bruit de codage, moins gênant à l'écoute qu'un bruit multiplicatif de même puissance.

- RSBseg est toujours nettement inférieur à RSB et par conséquent à RSBiso. Le poids plus important donné au bruit sur les petits signaux et dans les silences produit par les codages étudiés ici semble intervenir de manière négligeable dans les résultats des test subjectifs.

Cela indique que le RSBseg défini par la formule (21) n'est pas un indice objectif adéquat pour caractériser la qualité subjective des codeurs tels que ceux étudiés ici.

#### 4. - CONCLUSION

Deux codeurs différentiels munis de prédicteurs à adaptation séquentielle ont été testés et comparés à un codeur muni d'un prédicteur fixe. Ces codeurs, destinés à la numérisation du signal téléphonique à 32 kbit/s, ont été également simulés à 24kbit/s et 16kbit/s, ce qui a fait apparaître le bon comportement des algorithmes en présence de perturbations dues à un fort bruit de quantification. De tous les résultats présentés il ressort un net avantage des codeurs à prédicteurs adaptatifs dont la qualité, supérieure à celle d'un MIC 7bits dans les conditions des simulations présentes, peut être considérée comme satisfaisante pour la transmission numérique de la parole.

Les auteurs remercient Patrick GRAILLOT, qui a conçu et mis en place la procédure automatisée d'isopréférence, et les personnes qui ont bien voulu participer aux tests d'écoute.

#### BIBLIOGRAPHIE

- CASTELLINO (P), MODENA (G), NEBBIA (L) et SCAGLIOLA(C), 1974, "Bit rate reduction by automatic adaptation of quantizer step-size in DPCM systems". Proc. Internat. Zurich Seminar on Digital Communications, communication B6.
- GIBSON (J.D), JONES (S.K) et MELSA (J.L), 1974, "Sequentially adaptive prediction of speech signals". IEEE Trans. on Communications 22, n°1, pp. 1789-1797.
- GOLDBERG (A.J) et ARCESE (A), 1975, "A quantitative comparison of residual encoding techniques". EASCON Record, communication 28.
- GOODMAN (D.J) et WILKINSON (R.M), 1975, "A robust adaptive quantizer". IEEE Trans. on Communications 23, n°11, pp.1362-1365.
- JAYANT (N.S), 1974, "Digital coding of speech waveforms : PCM, DPCM and DM quantizers". Proc. of IEEE 62, n°5, pp. 611-632.
- LE GUYADER (A) et PISSARD (A), 1978, "Codage différentiel de la parole à 32kbit/s". Recherches/Acoustique Vol. V (publication C.N.E.T.).
- LORAND (P), MINIER (G) et SALIOU (A), 1975, "Détermination du spectre moyen à long terme de la parole (langue française)". Recherches/Acoustique Vol.II (publication C.N.E.T.).
- Mc DERMOTT(B.J), SCAGLIOLA (C) et GOODMAN(D.J), 1978, "Perceptual and objective evaluation of speech processed by adaptive differential PCM". IEEE Internat. Conference on Acoustics, Speech and Signal Processing Record pp. 581-585.
- MAKHOUL (J) et VISWANATHAN (R), 1978, "Adaptive lattice methods for linear prediction". IEEE Internat. Conference on Acoustics, Speech and signal Processing Record pp.83-86.
- MARKEL (J.D) et GRAY (A.H), 1973, "On autocorrelation equations as applied to speech analysis". IEEE Trans. on Audio and Electroacoustics 21, n°2, pp.69-79.
- MARKEL (J.D) et GRAY (A.H), 1976, "Linear prediction of speech". Springer ed.