

RECOMMANDATION UIT-R F.1112-1*

**TRANSMISSION DE SIGNAUX VOCAUX NUMÉRISÉS SUR DES SYSTÈMES
FONCTIONNANT À DES FRÉQUENCES INFÉRIEURES À 30 MHz ENVIRON**

(Question UIT-R 164/9)

(1994-1995)

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que les communications vocales en ondes décamétriques utilisent des canaux de 3 kHz;
- b) que la sécurité est essentielle pour un certain nombre de communications;
- c) que le chiffrement constitue la seule façon de parvenir à un niveau de sécurité suffisant;
- d) que le niveau de sécurité requis peut aisément être atteint en utilisant les techniques de numérisation de la parole;
- e) qu'il y a donc lieu de mettre en œuvre des codeurs de signaux vocaux (vocodeurs), associés à des modems HF;
- f) que pour des canaux HF de bonne qualité, le débit binaire admissible est au maximum de 4 800 bit/s;
- g) que le brouillage et les effets de la propagation tels que l'évanouissement entraînent pour les communications numériques une élévation du taux d'erreur binaire, nécessitant la mise en œuvre de procédés de correction (codes de correction d'erreur-entrelacement),

recommande

- 1** que pour les communications à courte portée (utilisant les ondes de sol), on emploie des vocodeurs à un débit de 2 400 ou 4 800 bit/s;
- 2** que pour les communications à longue portée (utilisant les ondes ionosphériques), on emploie, selon la qualité de la liaison, soit des vocodeurs à un débit de 2 400 bit/s, soit des vocodeurs à un débit de 600, 800 ou 1 200 bit/s avec codage de correction d'erreur;
- 3** que les systèmes de radiotéléphonie numérique utilisés soient conformes aux spécifications générales indiquées dans l'Annexe 1 et aux spécifications particulières données dans l'une des Annexes 2, 3 et 4 selon le type de vocodeur utilisé.

ANNEXE 1

**Généralités sur les systèmes de radiotéléphonie numérique
en ondes décamétriques**

Un système de radiotéléphonie numérique en ondes décamétriques se compose d'un circuit radiotéléphonique traditionnel, d'un codeur de signaux vocaux (vocodeur), d'un dispositif optionnel de chiffrement et d'un modem HF. Le schéma de principe de ce système est représenté à la Fig. 1.

Le signal vocal transmis arrive au codeur de voix où il est analysé et transformé en un flux numérique. Ce flux est éventuellement chiffré, puis arrive au modem où il est mis en forme pour être transmis dans la bande de fréquences de la voie téléphonique. Le flux numérique provenant du récepteur du modem est déchiffré si nécessaire, puis arrive au décodeur de voix, où le signal de parole est reconstitué.

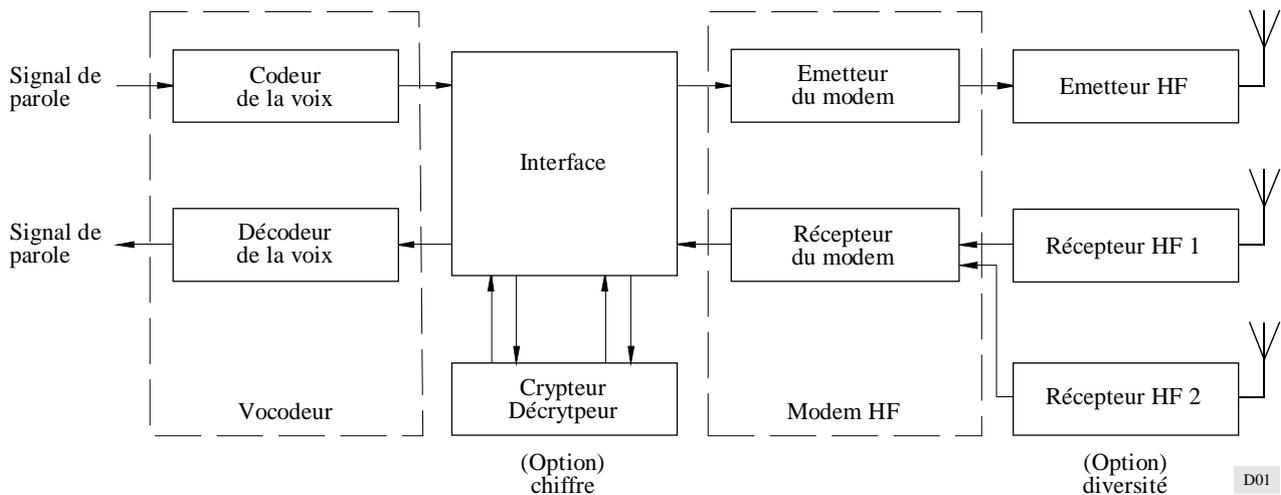
* Cette Recommandation doit être portée à l'attention de la Commission d'études 8 des radiocommunications.

Dans le cas de taux d'erreur élevés auxquels les vocodeurs à faible débit binaire (2 400 ou 4 800 bit/s) ne peuvent faire face, il faut utiliser des vocodeurs à *très* faible débit binaire (600, 800 ou 1 200 bit/s) avec un codage de correction d'erreur.

Il est d'autre part possible d'utiliser la réception en diversité; elle s'effectue à l'aide de deux récepteurs sur des antennes à diversité d'espace, avec traitement ultérieur dans la partie réceptrice du modem.

Il existe de nombreux types de vocodeurs à débit inférieur ou égal à 4 800 bit/s, en particulier les vocodeurs à canaux, les vocodeurs orthogonaux et les vocodeurs à prédiction linéaire; on trouvera dans les Annexes 2, 3 et 4 la description de systèmes utilisant ces types de vocodeurs.

FIGURE 1
Schéma de principe d'un radiotéléphone numérique
en ondes décamétriques



Les trois types de vocodeur ont tous un débit binaire utile de 2 400 bit/s et un taux d'intelligibilité du son supérieur à 90% lorsque les conditions de transmission sont favorables. Ils sont donc plus ou moins équivalents en ce qui concerne la qualité pour cette valeur du débit binaire.

Tous les vocodeurs sont présentés avec le modem associé qui, dans les trois cas, est du type modem parallèle (plusieurs sous-porteuses modulées indépendamment dans la bande audio).

Dans le cas du vocodeur à canaux et du vocodeur orthogonal, le symbole élémentaire transmis par le modem a une courte période (respectivement 13,3 ms et 8,33 ms). Ces deux vocodeurs résistent donc moins bien à des conditions de propagation très défavorables en ondes décamétriques, conditions caractérisées par la formation de trajets multiples introduisant un retard de quelques millisecondes et provoquant des brouillages importants entre symboles.

Pour le vocodeur LPC 10, lorsque le modem associé est le modem décrit par le STANAG 4197, la période des symboles est de 22,5 ms. Le fonctionnement de ce modem est donc moins affecté par les conditions indiquées ci-dessus.

De plus, dans le vocodeur LPC 10, certains bits importants de la trame sont protégés par un code de correction puissant (code de Golay). De ce fait, une augmentation du taux d'erreur binaire (TEB) entraîne une dégradation de qualité qui est plus progressive que dans les vocodeurs précédents: l'intelligibilité du son tombe à 80% pour un TEB de 2%, contre 1% pour les autres vocodeurs.

Une qualité de fonctionnement satisfaisante doit aussi être garantie dans des conditions médiocres. Cela est facile à réaliser avec le vocodeur LPC 10 en utilisant simultanément:

- une compression d'information pour réduire le débit binaire utile à 600/800/1 200 bit/s (au prix d'une diminution de la qualité intrinsèque du vocodeur);
- l'insertion d'une redondance qui augmente la résistance aux erreurs.

Si l'on procède ainsi, dans la version à 800 bit/s, le seuil d'intelligibilité du son de 80% n'est atteint que lorsque le TEB dépasse 5%.

ANNEXE 2

Système de radiotéléphonie numérique utilisant un vocodeur à canaux

1 Description du système

Ce système est composé d'un circuit radiotéléphonique sur ondes décamétriques traditionnel et d'un terminal de traitement numérique de la voix.

Les signaux vocaux provenant d'un microphone sont analysés et numérisés par le codeur du vocodeur à canaux. Le signal numérique est ensuite appliqué, à travers l'interface, au crypteur; à ce niveau il est crypté avec des signaux aléatoires. A la sortie du crypteur, le signal est ensuite dirigé à travers l'interface vers le modulateur et converti en signal à fréquences vocales par la méthode de modulation MDP4-D-MRF.

Le signal audio qui sort du récepteur radioélectrique est démodulé en codes numériques par le démodulateur MDP4-D-MRF. Le signal codé numériquement va à travers l'interface, au décrypteur où il est transposé dans les codes numériques d'origine. Ces signaux traversent ensuite l'interface jusqu'au vocodeur à canaux, où le signal de parole est reconstitué et appliqué à un récepteur téléphonique.

2 Vocodeur à canaux

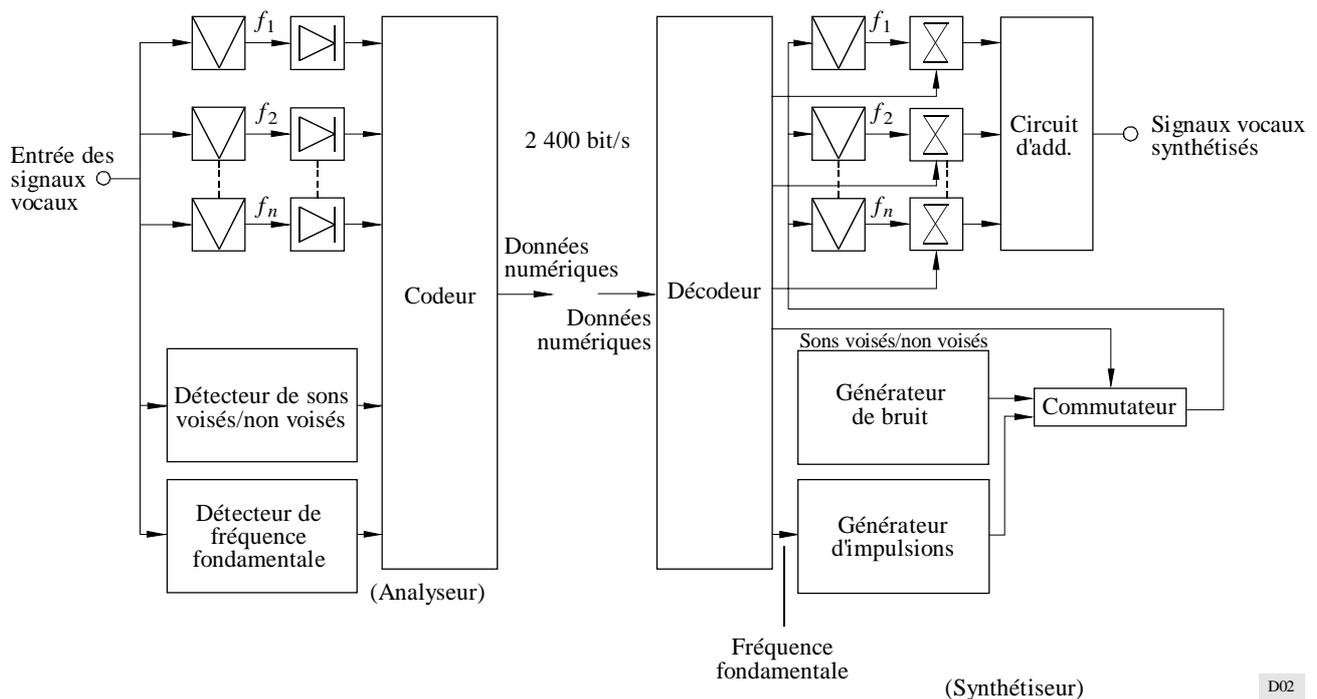
2.1 Principes théoriques

Le vocodeur à canaux divise le signal en bande vocale d'environ 300 à 3 000 Hz en un certain nombre de bandes spectrales contiguës et mesure l'intensité de chaque bande. Ces mesures sont codées et transmises.

On trouvera à la Fig. 2 le schéma de principe du vocodeur à canaux. Dans la partie analyseur, un certain nombre de filtres passe-bande (traitement équivalant à un filtre) sont utilisés pour séparer la largeur de bande des fréquences vocales et sélectionner la composante fréquence de chaque bande spectrale. Les sorties des filtres passe-bande sont mesurées afin de déterminer le niveau de chaque bande. En même temps, les fréquences voisées ou non voisées et la fréquence fondamentale («pitch») sont détectées. Ces paramètres sont échantillonnés et quantifiés au niveau de l'analyseur, qui formate le signal codé à 2 400 bit/s en vue de son émission.

FIGURE 2

Schéma de principe d'un vocodeur à canaux



Dans la partie synthétiseur, un générateur de bruit et un générateur d'impulsions excitent le synthétiseur de spectre, le générateur de bruit étant réservé aux sons non voisés et le générateur d'impulsions aux sons voisés. La fréquence de sortie du générateur d'impulsions est fixée de manière à être presque identique à la fréquence fondamentale. La sortie du générateur de bruit ou du générateur d'impulsions est analysée par une configuration de filtres passe-bande semblable à celle de la partie analyseur. Les niveaux des bandes spectrales analysées sont multipliés et ajoutés pour reconstituer le signal de parole.

2.2 Mise en œuvre

Le terminal de traitement numérique de la parole utilise la transformation de Fourier rapide (TFR) pour l'analyse du spectre de fréquences vocales. Le spectre calculé est divisé en un certain nombre de bandes spectrales dont la largeur est équivalente à celles des filtres passe-bande. On fait la moyenne des bandes spectrales ou des canaux pour déterminer le niveau de chaque bande spectrale. On détecte la fréquence en obtenant une autocorrélation maximale, tandis que la détection des signaux voisés et non voisés est basée sur le niveau des valeurs maximales de la fonction. Par ailleurs, on synthétise la parole en produisant la réponse en impulsion des filtres passe-bande avec un filtre numérique de réponse en impulsion finie. Le résultat est multiplié par le niveau de sortie de chaque bande spectrale. Enfin, les formes d'onde de toutes les bandes spectrales sont ajoutées pour obtenir un signal de parole.

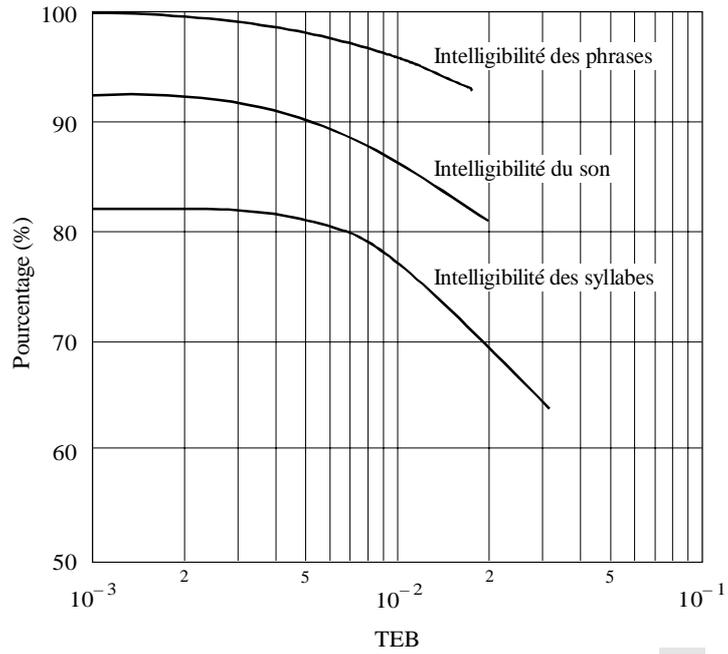
3 Modem MDP4-D-MRF

Le modem MDP4-D-MRF appliqué dans ce système est à peu près identique à celui que décrit l'Annexe 1 à la Recommandation UIT-R F.763. Ses principales caractéristiques sont les suivantes: le débit binaire est de 2 400 bit/s. 18 tonalités sont utilisées, parmi lesquelles 16 ont un espacement de 110 Hz (935 à 2 585 Hz) et sont modulées dans le mode MDP-4-CD à une rapidité de modulation de 75 Bd. La tonalité à 605 Hz sert à la correction des erreurs de fréquence de bout en bout; pour la synchronisation, on utilise la tonalité à 825 Hz, afin d'éviter une perte excessive à la limite de bande. Des temps de garde entre les trames sont introduits, pour lutter contre la propagation par trajets multiples et la distorsion de temps de propagation de groupe.

4 Résultats des essais

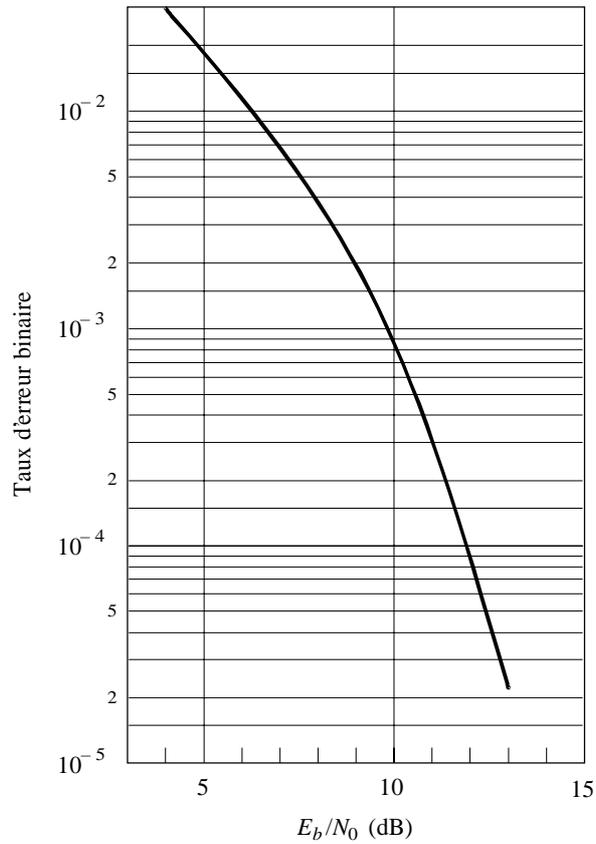
Les caractéristiques d'un vocodeur à canaux, c'est-à-dire l'intelligibilité des phrases, du son et des syllabes en fonction du TEB, sont présentées à la Fig. 3. Les Fig. 4 et 5 montrent les caractéristiques du modem MDP4-D-MRF, à savoir le TEB en fonction de E_b/N_0 . Les caractéristiques statiques sont indiquées à la Fig. 4 et les caractéristiques avec évanouissement à la Fig. 5.

FIGURE 3
Caractéristiques d'un vocodeur à canaux



D03

FIGURE 4
Caractéristiques statiques du modem MDP4-D-MRF

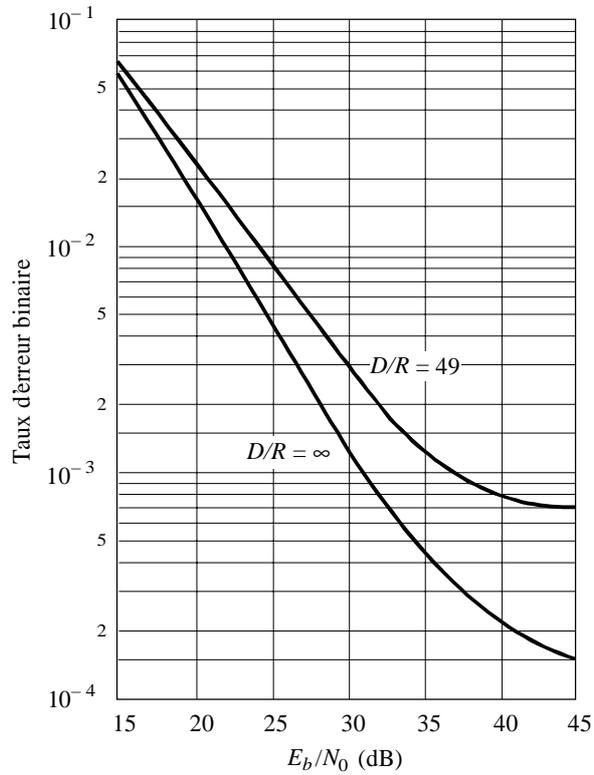


E_b : énergie par bit

N_0 : densité de puissance spectrale de bruit

D04

FIGURE 5
Caractéristiques du modem MDP4-D-MRF avec évanouissement



Fréquence Doppler: 0,5 Hz
Temps de propagation: 1 ms

D/R : rapport de puissance efficace entre l'onde directe et l'onde réfléchie

D05

ANNEXE 3

Système de radiotéléphonie numérique utilisant un vocodeur orthogonal

1 Description du système

La Fig. 1 représente le schéma de principe d'un type de système de radiotéléphonie numérique (RTN). Le circuit est équipé des moyens radioélectriques classiques d'une liaison principale sur ondes décimétriques: émetteurs, récepteurs pour la réception double en diversité d'espace, antennes et lignes de jonction.

L'équipement terminal de ligne se compose d'un vocodeur, qui transforme le signal vocal en un flux de symboles binaires à un débit de 2 400 à 4 800 bit/s et d'un modem qui en assure la transmission sur le circuit radio en ondes décimétriques.

2 Vocodeur

On a effectué des essais de fonctionnement du RTN avec deux types de vocodeur conçus pour un débit de 2 400 et de 4 800 bit/s. Il s'agit de deux vocodeurs orthogonaux dont les schémas de principe sont représentés aux Fig. 6 et 7. Le signal vocal initial est appliqué à un analyseur de spectre qui détermine la valeur Y_k de l'enveloppe du spectre du signal vocal sur plusieurs fréquences.

Le vocodeur 4 800 bit/s prélève 30 échantillons également distribués à l'intérieur de chacune des trois parties de la gamme de fréquences totale du signal vocal. Ces échantillons sont convertis en 16 coefficients γ_j de développement en série orthogonale de l'enveloppe du spectre; pour une fréquence de trame de 60 Hz, on constitue ainsi une séquence binaire de 3 840 bit/s. Simultanément à l'analyse du spectre, on extrait du signal vocal la valeur de la période du son fondamental (SF), ainsi que la caractéristique d'excitation son-bruit (S-B), qui sont transmis avec une fréquence double de la fréquence de trame (120 Hz) à l'aide d'un code à 8 chiffres, ce qui occupe 960 bit/s du flux total.

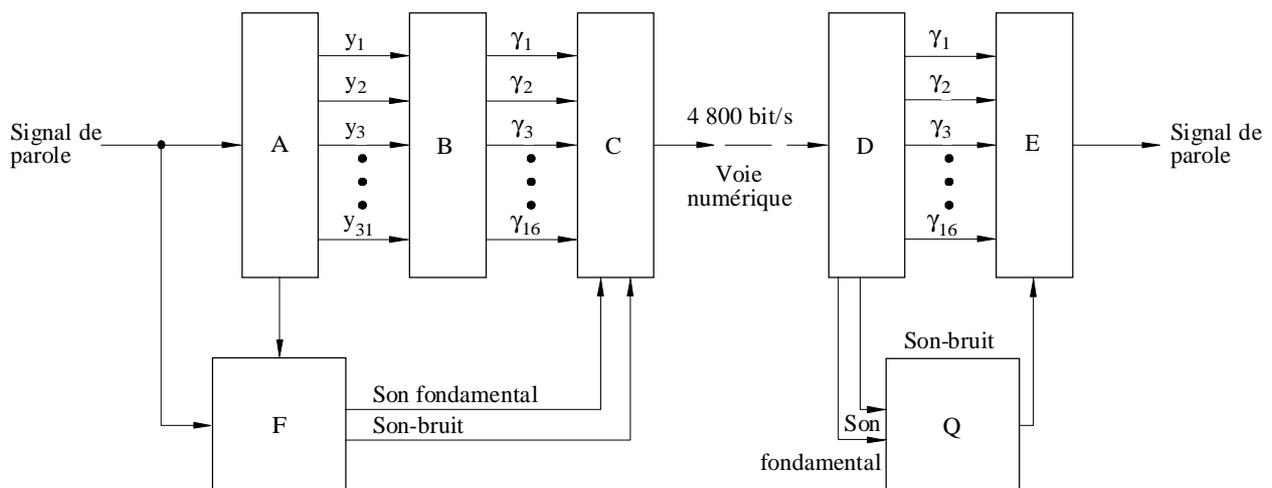
A la réception, on extrait du flux numérique les signaux γ_j qui commandent le synthétiseur orthogonal du spectre, et les signaux SF et S-B, qui commandent le générateur d'excitation du synthétiseur (GES). Le GES génère soit une séquence de groupes d'impulsions émises à la fréquence SF et ayant un spectre uniforme, soit une séquence d'impulsions pseudo-aléatoires. Le signal vocal est synthétisé à la sortie du synthétiseur.

Le vocodeur 2 400 bit/s (Fig. 7) est un vocodeur orthogonal à transformation non linéaire, où la racine carrée de l'enveloppe du spectre du signal vocal se présente comme une série orthogonale. Le nombre d'échantillons Y_k de l'enveloppe du spectre est ici de 21, et la distance entre eux croît progressivement avec la fréquence, selon une courbe d'égaux articulations. On extrait la racine carrée de ces échantillons, $\sqrt{Y_k}$, ce qui permet de déterminer 10 coefficients γ_j du développement en série orthogonale de la racine carrée de l'enveloppe du spectre. Les valeurs de ces coefficients sont émises toutes les 20 ms dans le flux total à l'aide de combinaisons codées à quatre chiffres, qui se succèdent également toutes les 20 ms.

A la réception, le flux à 2 400 bit/s est décomposé en signaux élémentaires, parmi lesquels les signaux FS et S-B commandent le générateur GES qui est analogue à celui du vocodeur 4 800 bit/s et, parallèlement, deux synthétiseurs de spectre. La sortie du premier synthétiseur est relié à la sortie d'excitation du deuxième, de sorte que le spectre à synthétiser est élevé au carré et qu'une relation linéaire s'établit ainsi entre le spectre de la parole d'origine et celui de la parole synthétisée.

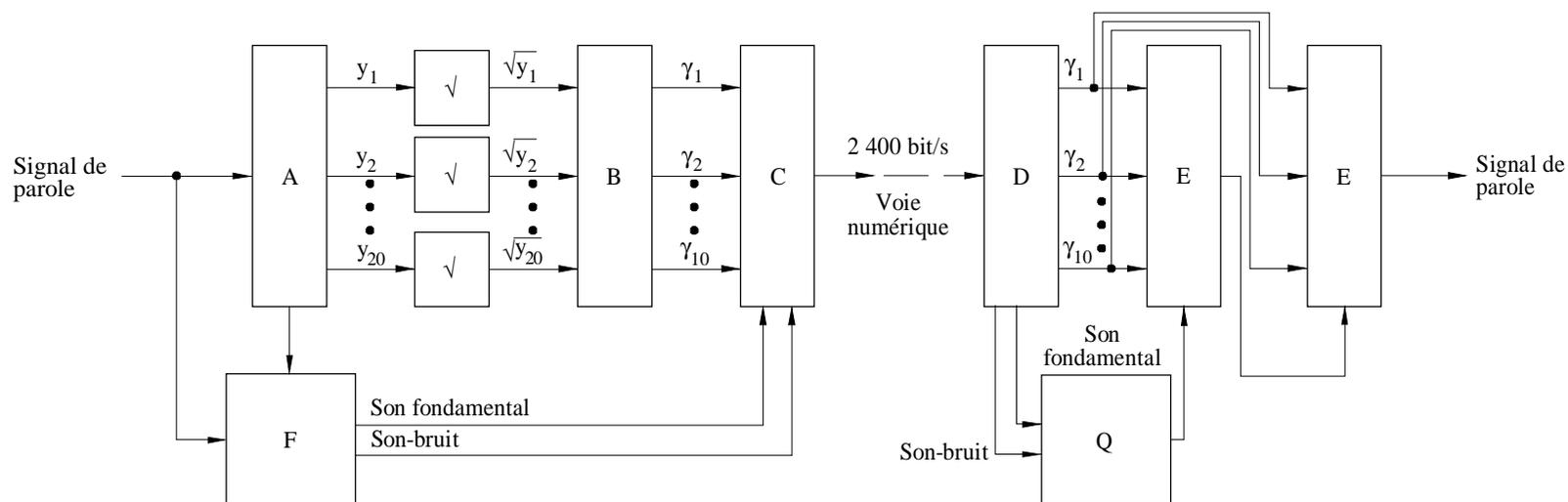
FIGURE 6

Schéma de principe d'un vocodeur orthogonal fonctionnant au débit de 4 800 bit/s



- | | |
|-----------------------------|----------------------------------|
| A: analyseur de spectre | E: synthétiseur |
| B: convertisseur orthogonal | F: séparateur du son fondamental |
| C: interface d'émission | Q: générateur d'excitation |
| D: interface de réception | |

FIGURE 7
Schéma de principe d'un vocodeur orthogonal
fonctionnant au débit de 2 400 bit/s



A: analyseur de spectre
B: convertisseur orthogonal
C: interface d'émission
D: interface de réception

E: synthétiseur
F: séparateur du son fondamental
Q: générateur d'excitation

D07

3 Modem

Le modem est un dispositif multivoie et à double modulation par déplacement de phase utilisant des signaux orthogonaux. Ses principales caractéristiques techniques sont les suivantes:

- rapidité de modulation dans les sous-voies: 100 ou 120 bit/s,
- nombre de voies: 20,
- intervalle entre les fréquences des voies: 142 Hz,
- intervalle d'orthogonalité: 1/142 s,
- durée de l'intervalle de protection: 1,29 ou 2,29 ms,
- mode de réception des signaux de voie: optimal, non cohérent.

Le modem permet le fonctionnement aux débits suivants:

- débit de voie: 4 800 bit/s – débit d'information: 4 800 bit/s,
- débit de voie: 4 800 bit/s – débit d'information: 2 400 bit/s (doublement de l'information par couples de sous-voies à espacement maximal en fréquence),
- débit de voie: 2 400 bit/s – débit d'information: 2 400 bit/s (modulation simple par déplacement de phase).

En outre, au débit de 4 800 bit/s, le code de Halley permet de passer au débit d'information de 2 400 bit/s.

Le modem permet également la réception avec deux antennes à diversité d'espace.

4 Essais

4.1 Méthodologie

Le fonctionnement du RTN a été étudié pendant plusieurs mois à différentes périodes de l'année sur des trajets d'une longueur de 1 500-3 000 km et jusqu'à 10 000 km avec relais, établis sur des parallèles et sur des méridiens de la Terre.

L'évaluation de la qualité de voie d'un RTN a été effectuée de deux façons:

- en mesurant l'intelligibilité de la parole transmise, à l'aide de tables de netteté, et la fiabilité de transmission du flux binaire en mesurant le taux d'erreur pendant les intervalles dans la transmission des tables;
- en recueillant une évaluation statistique subjective de la qualité de parole auprès des abonnés après un entretien de vive voix.

En tout, ce sont plus de 50 mesures de l'intelligibilité et de la fiabilité de transmission qui ont été effectuées.

4.2 Analyse des résultats

Des essais ont été effectués à différentes périodes de l'année et du jour, cela pendant plusieurs mois, sur des liaisons en ondes décimétriques d'une longueur de 1 500 à 10 000 km. Au cours des essais, on a mesuré, à l'aide de tables de netteté, l'intelligibilité syllabique de la parole transmise, et évalué en même temps la fiabilité de transmission du flux numérique. Une évaluation statistique subjective de la qualité générale de la parole a également été effectuée auprès des abonnés.

L'analyse des résultats permet de conclure ce qui suit:

- Un paquet d'erreurs sur un circuit à ondes décimétriques entraîne une grande dispersion de l'intelligibilité syllabique pour un même taux d'erreur moyen.
- Pour un faible taux d'erreur (inférieur à 1×10^{-3}), un paquet d'erreurs a une plus grande incidence sur l'intelligibilité syllabique qu'une répartition uniforme des erreurs. Cela s'explique par le fait que, dans le premier cas, les éléments des tables de netteté se détériorent en règle générale non pas à cause d'une erreur mais de plusieurs; de ce fait, ils se dégradent complètement du point de vue de leur reconnaissance sonore. Cependant, pour un même taux d'erreur moyen, l'évaluation de la qualité de parole fournie par les abonnés est nettement supérieure dans le cas d'un paquet d'erreurs que dans le cas d'une répartition uniforme des erreurs; en effet, dans le cas d'un flot de paroles régulier, un paquet d'erreurs se perçoit à l'ouïe comme une simple erreur isolée et la parole, de ce fait, s'entend mieux.
- Lorsque le taux d'erreur est plus élevé (5×10^{-3} et au-delà), le paquet d'erreurs affecte isolément des portions de parole mais en préserve des segments suffisamment longs pour permettre l'établissement d'une liaison radiotéléphonique stable, même avec un taux d'erreur moyen de 1×10^{-1} environ.

- L'utilisation de la double réception, avec antennes à diversité d'espace, améliore l'intelligibilité syllabique reçue de 3% à 5% de syllabes en moyenne et rehausse sensiblement la qualité générale de parole.
- L'intelligibilité de la parole transmise par RTN est supérieure à celle de la parole transmise dans une voie analogique classique en ondes décamétriques. La qualité générale de la parole transmise par RTN est, selon des évaluations subjectives, grandement améliorée, du fait de l'absence des phénomènes caractéristiques d'un circuit en ondes décamétriques – évanouissements sélectifs et brouillages provenant d'autres stations.
- Les RTN exploités sur de longs circuits utilisant des relais, où les flux numériques sont régénérés au point de retransmission permettent d'établir une liaison radiotéléphonique stable là où les systèmes classiques, y compris le Lincompex, ne garantissent pas une transmission de parole de qualité satisfaisante.
- L'exploitation d'un RTN doté de vocodeurs conçus pour des débits de 2 400 et 4 800 bit/s a montré que, bien qu'un vocodeur ayant un débit de 4 800 bit/s transmette une parole de meilleure qualité, dans une voie sans bruit, que celui à 2 400 bit/s, on observe cependant, dans les conditions réelles d'exploitation d'un circuit de radiocommunication en ondes décamétriques, que le fait de combiner un vocodeur à 2 400 bit/s avec un modem ayant un débit de voie de 4 800 bit/s pouvant se réduire à 2 400 bit/s donne dans la majorité des cas une meilleure qualité de parole. Dans ces conditions de fonctionnement, le RTN a des caractéristiques de seuil bien définies: la qualité de la parole transmise dans le cas d'une détérioration des conditions de propagation demeure élevée presque jusqu'au moment de l'interruption complète de la communication.
- Le coefficient de transmission du signal vocal dans le RTN est stable aussi bien pour l'ensemble du signal que pour les diverses composantes de fréquence. Les temps de propagation élevés du signal vocal transmis par RTN accroissent fortement la perception des signaux d'écho, d'où la nécessité d'introduire dans le système un suppresseur d'écho.

ANNEXE 4

Système de radiotéléphonie numérique utilisant un vocodeur à prédiction linéaire LPC 10

1 Description du système

Le système utilise deux vocodeurs ayant des débits respectifs de 2 400 et 800 bit/s.

Le premier, à 2 400 bit/s, fournit une parole de qualité suffisante pour procurer une bonne intelligibilité à débit réduit.

Le taux d'erreur admissible en ligne peut atteindre 1% à 2%, ce qui suffit quand la liaison est de bonne qualité.

Lorsque la qualité de la liaison se dégrade, c'est-à-dire lorsque le taux d'erreur à 2 400 bit/s dépasse 1,5%, on utilise le deuxième vocodeur à 800 bit/s. On met alors en œuvre un procédé de détection et correction d'erreur qui ramène le taux d'erreur effectif à 2 400 bit/s à une valeur compatible avec l'utilisation du vocodeur à 800 bit/s, à savoir environ 1%.

Ce second vocodeur procure une qualité légèrement inférieure à celle du vocodeur à 2 400 bit/s, ce qui peut être considéré comme le compromis indispensable pour continuer à communiquer sur un canal qui interdirait une liaison avec le vocodeur de départ.

Les deux vocodeurs comportent les fonctions suivantes:

- l'analyse du signal de parole, destiné à extraire de celui-ci les jeux de paramètres permettant de le représenter de façon pertinente;
- la quantification qui transforme les valeurs numériques desdits paramètres en un flux numérique susceptible d'être transmis;
- l'adjonction éventuelle de données complémentaires (redondance) pour détecter et corriger les erreurs de transmission;
- la déquantification pour régénérer les paramètres originaux;
- la synthèse qui, à partir des paramètres reçus, reconstitue un signal de parole destiné à produire une impression acoustique aussi proche que possible de celle qu'aurait produite le signal original, sans chercher à reproduire le signal lui-même.

Les fonctions d'analyse et de synthèse sont identiques pour les deux vocodeurs.

Seules diffèrent les fonctions de quantification, adjonction de redondance et déquantification qui, si elles traitent toutes deux un débit en ligne à 2 400 bit/s, concernent un débit utile de 2 400 ou 800 bit/s.

2 Principe des vocodeurs LPC 10

2.1 Modélisation du signal de parole

Tout d'abord, le signal de parole peut être considéré comme (presque) stable sur des durées faibles: on peut alors le segmenter en *trames* de durée constante (ici, 22,5 ms) où l'on suppose que ses caractéristiques ne changent pas. On transmet donc à intervalles réguliers toutes les données nécessaires pour synthétiser une ou plusieurs trames de parole, indépendamment de celles qui les précèdent ou les suivent.

Ensuite, comme le débit disponible est très faible, il convient de *modéliser* le signal de parole, c'est-à-dire le représenter par un jeu de paramètres aussi proches que possible de la réalité physique, à savoir des phénomènes responsables de sa production.

Pour cela, on considère deux cas, suivant que le signal est périodique (voisé) ou non périodique (non voisé).

Pour les sons périodiques, qui correspondent aux voyelles, on admet que le signal de parole est obtenu par mise en forme spectrale (filtrage) d'une excitation périodique, ce qui en fait représente bien la réalité physique où les vibrations acoustiques des cordes vocales se propagent dans le conduit vocal (larynx, bouche, etc.) qui joue le rôle de filtre dont la fonction de transfert dépend de la voyelle prononcée.

Quant aux sons non périodiques, il en existe de deux sortes:

- les sons stables ou semi-stables (tels que les «sifflantes», comme le «S»);
- les sons transitoires (tels que les «plosives», comme le «P»).

Ces deux cas ne se distinguent que par leur durée, ainsi que par la vitesse d'évolution du niveau sonore.

On les représente tous deux comme la sortie d'un filtre dont l'entrée est une excitation aléatoire (bruit blanc) à évolution plus ou moins rapide: ceci est encore conforme à la réalité, où ce type de son non structuré est produit par des turbulences ou des occlusions brutales dans le conduit vocal.

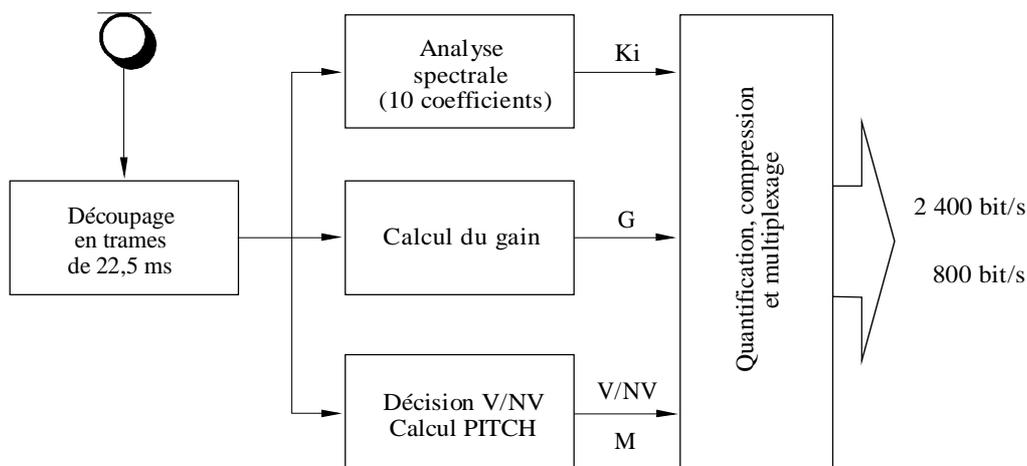
Finalement, le signal de parole dans son ensemble est toujours représenté comme une excitation de nature (périodique ou aperiodique) et de niveau variables, filtrée de façon à reproduire au mieux le spectre en fréquence du signal original.

Le calcul du filtre est basé sur les principes de l'analyse prédictive, d'où le nom de LPC (Linear Predictive Coding) donné au procédé.

2.2 Procédés d'analyse

Les procédés d'analyse extraient du signal de parole les paramètres nécessaires à sa modélisation. Ils sont schématisés sur la Fig. 8.

FIGURE 8
Analyse (côté émission)



Ki: coefficients de réflexion
 G: gain pour le signal synthétique
 V: son voisé (voyelles)
 NV: son non voisé (consonnes)
 M: valeur du pitch

Le premier paramètre à déterminer est la nature (voisé ou non voisé) du signal. Son évaluation se fait sur des critères de périodicité, qui varient suivant les cas. On calcule en général l'autocorrélation à long terme du signal de parole (on recherche si, à intervalles réguliers, il est identique à lui-même), ce qui donne une première estimation. Cette estimation est affinée sur d'autres critères, par exemple le niveau instantané, l'autocorrélation à court terme, le rapport entre l'énergie du signal dans les fréquences basses et son énergie dans les fréquences hautes.

Le résultat est un *indicateur de voisement* qui vaut 1 si le signal est jugé voisé, 0 autrement.

Quand le signal est estimé voisé, il convient d'estimer sa période, pour traduire sa hauteur. Cette période, appelée *pitch*, est elle aussi évaluée en général à partir de l'autocorrélation à long terme: l'intervalle de temps au bout duquel le signal de parole se reproduit identique à lui-même est la valeur du pitch.

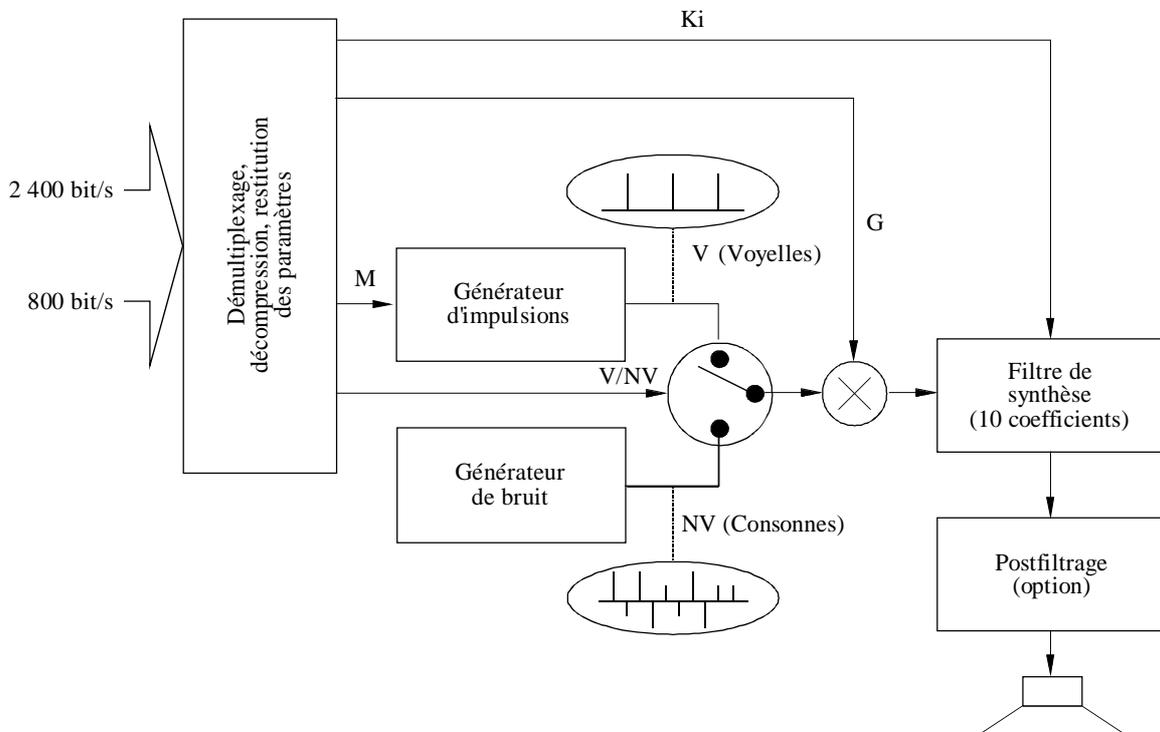
Quant au filtre dont la réponse en fréquence est le plus proche possible du spectre du signal de parole, on le calcule à partir des coefficients d'autocorrélation à court terme dont on déduit les coefficients du filtre par des procédés classiques en traitement du signal. Le résultat de cette analyse est un jeu de 10 coefficients de réflexion (d'où le nom de LPC 10) qui traduisent avec une fidélité suffisante les variations de section du conduit vocal qui, à l'origine, a filtré (coloré) l'excitation spectralement neutre de départ.

Enfin, pour chaque trame, on évalue le niveau du signal de parole, pour piloter le gain du synthétiseur placé côté synthèse.

2.3 Procédé de synthèse

Les algorithmes mis en œuvre pour synthétiser le signal de parole reflètent le modèle de production supposé de la parole. Ils sont représentés Fig. 9.

FIGURE 9
Synthèse (côté réception)



K_i : coefficients de réflexion
 M : valeur du pitch
 V : son voisé (voyelles)
 NV : son non voisé (consonnes)
 G : gain pour le signal synthétique

On trouve successivement:

- un générateur de bruit, utilisé pour les sons non voisés;
- un générateur de signal périodique, à qui l'on fournit le pitch, pour les sons voisés;
- un commutateur, permettant de sélectionner l'un ou l'autre générateur, suivant la nature du signal de parole à produire dans la trame courante;
- un filtre d'ordre 10, qui filtre l'excitation choisie pour lui donner sa coloration: c'est à ce niveau que se fait la distinction entre les différentes voyelles et les différentes consonnes;
- un système de régulation de gain, qui donne au signal synthétique le volume adéquat;
- optionnellement, un système de «post-filtrage» qui a pour but à la fois de masquer certaines imperfections du synthétiseur et d'améliorer l'agrément du signal synthétisé.

3 Vocodeur à 2 400 bit/s

Pour le vocodeur à 2 400 bit/s, les différents paramètres sont quantifiés indépendamment, trame par trame.

Le pitch et le voisement sont quantifiés conjointement sur 7 bits, le niveau sur 5 bits, et les 10 coefficients du filtre sur 41 bits pour les trames voisées, un peu moins pour les trames non voisées (les bits restants servent à tester la qualité de la liaison).

Après adjonction d'un bit de servitude (synchronisation), on obtient des trames de 54 bits.

Ce vocodeur a fait l'objet d'un accord de normalisation OTAN (STANAG 4198) et d'un standard fédéral US (Fed. Std. 1015).

4 Vocodeur à 800 bit/s

4.1 Principes utilisés pour la réduction du débit

La réduction du débit d'un facteur trois est obtenue en observant que, non seulement les valeurs des différents paramètres sont corrélées d'une trame à la suivante, mais aussi que la fidélité de quantification nécessaire varie en fonction du contexte.

Par exemple, il est rare que le niveau observé sur des trames consécutives varie beaucoup: on a donc intérêt à coder le niveau par blocs.

De plus, si le son est stable, le filtre de prédiction l'est aussi, et il est alors inutile de fournir un nouveau filtre de prédiction à chaque trame. Bien que le débit disponible soit réduit d'un facteur trois, il reste possible de quantifier finement un filtre commun pour (par exemple) trois trames consécutives. A contrario, si l'on est dans une portion transitoire, la notion même de spectre en fréquence tend à s'estomper, et on peut alors quantifier de façon grossière un filtre par trame, sachant que l'erreur de représentation commise sera masquée par la variation brutale de niveau sonore.

Par conséquent, le principe de base utilisé pour réduire le débit consiste à regrouper les trames de parole par paquets de trois (par exemple) et coder en bloc chaque paquet de paramètres.

4.2 Procédés de quantification

Chaque trame de données comporte 54 bits représentatifs de trois trames de parole. On décrit ci-après l'un des procédés de quantification envisagés, auquel, pour l'instant, des droits de propriété sont rattachés.

Un total de 10 bits sont utilisés pour quantifier à la fois le pitch et le voisement. On tient compte du fait qu'il ne peut y avoir plus d'une transition, voisé → non voisé (ou l'inverse) dans un paquet de trois trames et que, s'il y a plus d'une trame voisée, les pitches de trames adjacentes sont voisins: il suffit de transmettre une valeur de référence et un incrément.

Pour le niveau, 9 bits sont utilisés, à savoir:

- 4 bits pour définir un niveau de référence, commun aux trois trames,
- 5 bits pour décrire (par lecture d'une table appelée dictionnaire) l'évolution du niveau sur les trois trames.

Enfin, en ce qui concerne le filtre de prédiction, on utilise un total de 35 bits comportant deux champs.

Le premier champ, de 32 bits, est considéré comme relatif à:

- soit un filtre unique commun aux trois trames,
- soit un filtre commun à deux trames successives et un incrément pour obtenir le filtre restant,
- soit deux filtres.

Le deuxième champ, de 3 bits, est chargé de décrire le schéma de codage choisi, parmi huit possibilités.

Le meilleur schéma de codage est choisi côté analyse sur un critère de minimisation de distance spectrale pondérée par le niveau du signal dans chacune des trames: une trame de faible niveau est moins bien perçue qu'une trame voisine de niveau plus élevé.

4.3 Codage correcteur

Le code correcteur retenu pour passer de 800 à 2 400 bit/s est un code en blocs comportant 54 bits d'information utile pour chaque paquet de $3 \times 54 = 162$ bits transmis.

On peut choisir, par exemple, un code Reed-Solomon raccourci à symboles de 6 bits, comportant 27 symboles au total pour 9 symboles utiles. Ce code, noté RS(27,9), permet de corriger 9 erreurs au total, soit une proportion *maximale* de 33% de symboles erronés, ou une proportion *moyenne* de l'ordre de 20%.

5 Résistance aux erreurs de transmission

La résistance aux erreurs de transmission se fait en comparant les intelligibilités respectives des deux vocodeurs en fonction du taux d'erreur à 2 400 bit/s, débit transmis par le modem sur voie HF. Cette intelligibilité est tracée Fig. 10.

On constate qu'elle est légèrement plus faible à 800 bit/s sans erreur de transmission.

Mais la dégradation lorsque le taux d'erreur augmente est plus lente qu'à 2 400 bit/s, pour les deux raisons suivantes:

- le code correcteur diminue fortement le taux d'erreur lors du passage de 2 400 à 800 bit/s;
- lorsque le taux d'erreur est élevé à 2 400 bit/s, cela se traduit par des trames effacées à 800 bit/s, trames effacées ayant un effet moindre que des trames fausses.

6 Modem

La transmission s'effectue pour les deux vocodeurs à un débit de 2 400 bit/s, en utilisant un modem standard, par exemple le modem décrit dans la Recommandation UIT-R F.763 ou celui décrit par le STANAG 4197.

FIGURE 10
Intelligibilité des vocodeurs LPC 10

