

Zeitschrift:	Technische Mitteilungen / Schweizerische Post-, Telefon- und Telegrafenbetriebe = Bulletin technique / Entreprise des postes, téléphones et télégraphes suisses = Bollettino tecnico / Azienda delle poste, dei telefoni e dei telegraфи svizzeri
Herausgeber:	Schweizerische Post-, Telefon- und Telegrafenbetriebe
Band:	64 (1986)
Heft:	3
Artikel:	Procédés de codage des signaux de la parole et de la musique et leur évaluation à l'aide de l'ordinateur
Autor:	Mury, Roland
DOI:	https://doi.org/10.5169/seals-875021

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 07.01.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Procédés de codage des signaux de la parole et de la musique et leur évaluation à l'aide de l'ordinateur

Roland MURY, Berne

Codierverfahren von Sprach- und Musiksignalen und ihre Beurteilung mit Computer

Zusammenfassung. Der vorliegende Artikel soll in gedrängter Form einen allgemeinen Überblick über die hauptsächlichen Verfahren der digitalen Kompression von Sprach und Musiksignalen geben, mit Hinweisen auf die Sprachsynthese, wie sie bei der Automatisierung der Auskunftsdiene der PTT angewandt wird. Nach der Erläuterung grundsätzlicher Begriffe der in der Fernmelde-technik am meisten angewandten Codierungs-Methoden werden kurz die Hardware und die Programmstruktur untersucht, die für das Studium der Signal-kompressions-Algorithmen verwendet werden (ADDIO-2). Der Artikel schliesst mit einer nicht erschöpfenden Beschreibung der Spezifikationen der subjektiven Versuche und der Bedingungen für die Aufnahme, die Verarbeitung und das Abhören zur qualitativen Beurteilung von Sprachmustern.

Résumé. L'objectif de la présente publication est de donner un aperçu général succinct des principaux procédés de compression numérique des signaux de la parole et de la musique en passant par la synthèse de la parole appliquée à l'automatisation des services des renseignements de l'Entreprise des PTT. Après avoir rappelé quelques notions fondamentales des méthodes de codage les plus couramment appliquées dans les télécommunications, on examine brièvement l'architecture matérielle de l'équipement et la structure du logiciel utilisées pour l'étude des algorithmes de compression du signal (ADDIO-2). L'article s'achève par une description non exhaustive des spécifications des méthodes d'essais subjectifs et des conditions d'enregistrement, de traitement et d'écoute pour l'évaluation qualitative des échantillons de parole.

Procedimenti di codifica dei segnali vocali e musicali e loro valutazione mediante elaboratore

Riassunto. Nel presente articolo sono presentati sommariamente i principali procedimenti di compressione numerica dei segnali vocali e musicali, con alcune indicazioni sulla sintesi della voce, adottati per automatizzare i servizi delle informazioni delle PTT. Dopo aver spiegato i concetti basilari dei metodi di codifica più utilizzati nella tecnica delle telecomunicazioni, l'autore esamina brevemente l'hardware e la struttura del programma impiegati per lo studio degli algoritmi di compressione dei segnali (ADDIO-2). L'articolo termina con una descrizione non esaustiva delle specificazioni delle prove soggettive e delle condizioni di registrazione, di trattamento e di ascolto per la valutazione qualitativa dei campioni di segnali vocali.

1 Introduction

Du point de vue économique, il est souhaitable de réduire sensiblement le débit binaire du signal codé sous forme numérique. A cet effet, la compression-extension permet de diminuer le nombre de bits transmis pour chaque échantillon. Il y a encore quelques années, on ne pouvait étudier et comparer l'efficacité des procédés de compression-extension qu'en ayant recours à des montages appropriés. Dès lors, il était très souvent difficile de modifier un paramètre ou d'optimiser les caractéristiques de codage ou de compression sans une perte de temps appréciable. Aujourd'hui, l'ordinateur peut s'acquitter en temps réel ou différé, suivant la complexité de l'algorithme, de ces opérations de codage ou de toute autre procédure de traitement du signal. On peut ainsi modifier à volonté les différents paramètres du système à simuler. Les spécialistes des laboratoires de recherche de l'Entreprise des PTT suisses ont réalisé une installation programmable pour l'introduction et l'extraction des signaux audiofréquences dans un gros ordinateur. Ce système, connu sous l'abréviation d'ADDIO-2, qui correspond à la dénomination anglaise d'«Analog-Digital-Data-Input-Output», constitue aujourd'hui un outil de travail indispensable à l'ingénieur confronté aux difficultés de la reconnaissance du meilleur compromis parmi des solutions très voisines. Les études sur le codage à 64 kbit/s de la parole à 7 kHz de largeur de bande ainsi que sur les procédés de compression-extension des signaux de programmes radiophoniques de haute qualité à 384 kbit/s ont largement bénéficié des possibilités de l'ADDIO-2. Parmi les nombreux programmes disponibles dans le système, et qui ont servi à des expériences de simulation ou à des opérations de calcul, citons, par exemple:

- Simulation des lois de compression-extension instantanée ou quasi instantanée pour le codage et la transmission des programmes radiophoniques de haute qualité.
- Expériences sur différents algorithmes de codage de la parole.
- Simulation des filtres miroir en quadrature pour le codage en sous-bandes MICDA.
- Contrôle de l'efficacité des codes correcteurs d'erreurs appliqués aux systèmes de transmission numériques des programmes radiophoniques.
- Mesure et correction du niveau efficace (RMS) du signal de parole pour les essais subjectifs.
- Simulation de certaines opérations de transcodage.
- Production et montage des sources sonores pour les séances d'essais subjectifs, etc.

Dans ce contexte, quelques aspects du codage sont examinés ci-après et plus particulièrement ceux qui font l'objet d'un récent programme d'études et de recherches dans le cadre d'une application spécifique aux besoins des PTT suisses.

2 Méthodes de compression numérique du signal

L'opération de codage consiste à prendre en compte la valeur du signal analogique à des instants discrets et à la représenter par un nombre choisi parmi un ensemble fini de valeurs discrètes. Cette dernière opération, dite de quantification, peut prendre plusieurs formes. Les principales techniques de compression les plus couramment utilisées pour réduire le débit numérique du signal

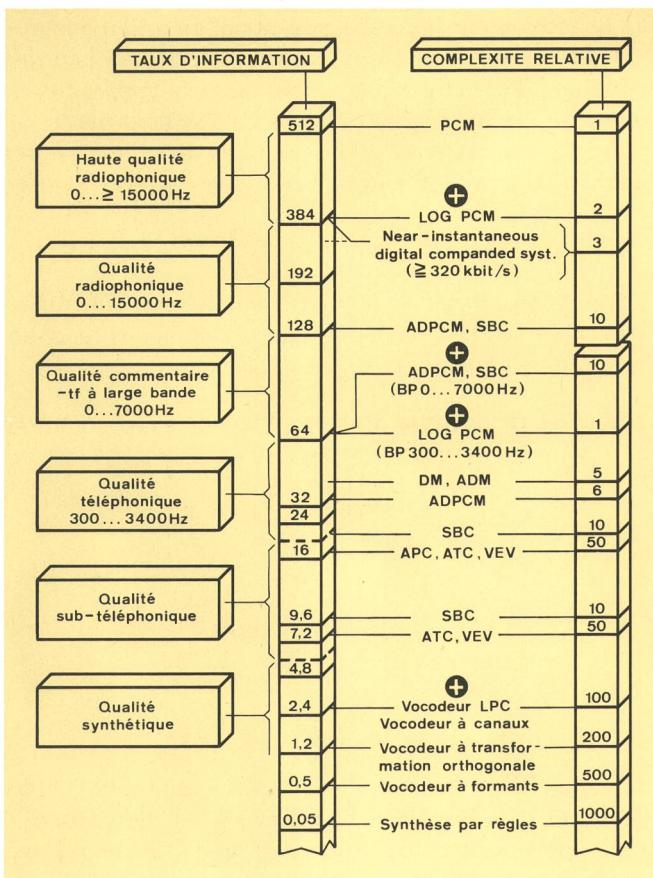


Fig. 1
Méthodes de compression-extension numérique classées suivant la qualité, le taux d'information et le degré de complexité

ADPCM Codage MIC différentiel adaptatif
SBC Codage en sous-bandes
DM Codage par modulation delta
ADM Codage adaptatif par modulation delta
APC Codage adaptatif par prédition
ATC Codage adaptatif par transformée
VEV Codage à excitation par bande de base
Projet de recherche des PTT

sont indiquées à la figure 1. Dans un article publié dans [1], J. Flanagan distingue quatre grandes catégories de méthodes de compression suivant la qualité qu'elles permettent d'obtenir, à savoir:

- Qualité radiophonique (30 Hz...8000 Hz)
- Qualité téléphonique (300 Hz...3400 Hz)
- Qualité subtéléphonique
- Qualité synthétique

Les progrès de ces dernières années permettent d'affiner cette classification en introduisant la notion de haute qualité et de commentaire pour la radiophonie et de parole à large bande pour la téléphonie.

La première catégorie englobe les méthodes qui assurent au signal décodé une haute qualité pour la transmission des programmes radiophoniques et une qualité dite «commentaire» correspondant à une largeur de bande d'environ 7 kHz pour les besoins de reportage, par exemple. Le débit numérique varie de l'ordre de 64 kbit/s à 512 kbit/s par voie son.

Dans la seconde catégorie, on classe les méthodes qui garantissent la largeur de bande téléphonique mais aussi celles, toutes récentes, qui permettent de doubler

cette même largeur de bande sans augmentation du débit numérique. Ces méthodes, fondées sur le codage adaptatif, trouvent une application nouvelle dans l'audio et la vidéoconférence, mais probablement aussi dans le futur RNIS (réseau numérique avec intégration des services). Les débits nécessaires sont de l'ordre de 16 kbit/s à 64 kbit/s et la qualité obtenue est dite transparente.

La troisième catégorie comprend les méthodes qui permettent un degré de compression élevé mais avec une distorsion audible du signal décodé qui reste toutefois très intelligible. Cette catégorie, souvent appelée subtéléphonique, requiert des débits de l'ordre de 4,8 kbit/s à 16 kbit/s environ.

Les techniques de codage de ces trois catégories sont appelées techniques de codage temporel parce qu'elles cherchent avant tout à préserver l'allure temporelle du signal.

Enfin, la dernière catégorie fait intervenir les techniques d'analyse et de synthèse où l'on extrait du signal de parole un petit nombre de paramètres que l'on transmet sous forme numérique. Le signal décodé a perdu une partie de son intelligibilité et de son naturel. Son allure temporelle est sensiblement différente de celle du signal original, ce qui se traduit subjectivement par une qualité dite synthétique.

Les débits propres aux techniques dites de codage de source ou vocodeur varient de quelques centaines de bits à 4,8 kbit/s environ.

Le choix d'une méthode de compression particulière sera non seulement fonction de la qualité et du débit nécessaires à la transmission mais aussi du degré de complexité du codec. Ces trois paramètres fondamentaux sont mis en relation dans la figure 1. Cette représentation n'est évidemment pas exhaustive mais suffisamment explicite pour apprécier le rapport coût-qualité des méthodes de codage les plus appropriées au domaine des télécommunications. Les croix reportées sur la figure 1 indiquent l'existence récente d'un projet de recherches des PTT suisses.

3 Codage de la parole

L'introduction des techniques numériques de transmission dans les réseaux de télécommunications s'est effectuée par le biais du codage MIC à 8 bits avec échantillonnage de fréquence 8 kHz. Le choix de cette technique s'est imposé pour les raisons suivantes:

- Ce type de codage simple conduit à une qualité indépendante du signal à coder.
- Il permet de conserver une qualité suffisante malgré un nombre raisonnable de codages-décodages à prendre en considération pour une communication internationale.

Toutefois, l'évolution toujours plus rapide de la technologie permet d'envisager d'incorporer aux équipements de multiplexage des procédés de traitement de l'information plus sophistiqués afin qu'il y soit possible de profiter au mieux des caractéristiques de redondance des signaux à coder. Par exemple, avec le codage MIC

différentiel adaptatif (MICDA) à 32 kbit/s, on peut utiliser plus efficacement la capacité de transmission disponible d'un conduit numérique à 2048 kbit/s, sans qu'il y ait dégradation sensible de la qualité de la parole.

C'est ainsi que les Administrations des télécommunications montrent un intérêt évident pour une hiérarchie numérique de codecs MIC différentiel fondée sur le débit de 64 kbit/s et de ses sous-multiples (32, 16, 8 kbit/s). Une alternative qui soulève également un intérêt croissant est le codage à bande étroite à 64 kbit/s. L'objectif de qualité visé dans ce cas est de doubler la largeur de bande d'une voie téléphonique et d'augmenter sensiblement la qualité de transmission par rapport au codage MIC à 64 kbit/s pour des applications telles que l'audio et la vidéoconférence. Dans cette hypothèse, le codage en sous-bandes présente certains avantages comme on le verra plus loin. Des procédés plus performants que le MICDA s'imposent lorsqu'il s'agit de coder le signal de parole à des débits compris entre 16 kbit/s et 5 kbit/s environ. Le codage adaptatif par prédition (APC), le codage adaptatif par transformée (ATC), le codage en sous-bandes (SBC) et le codage à excitation par bande de base (VEV) sont les quatre méthodes les plus connues et les plus efficaces opérant dans la gamme des débits considérés.

31 Principe du codage MIC différentiel

Les études de codage différentiel ont débuté il y a plusieurs années déjà par le codage delta, procédé dans lequel le signal de différence est quantifié par son signe. Cette quantification très simple est améliorée ou compensée par une fréquence d'échantillonnage élevée. L'adaptation du quantificateur consiste alors à agir sur le gain de la boucle de prédition à partir des signes précédents (Pissard, 1977). Les performances obtenues par ce type de codeur sont insuffisantes et encore trop éloignées de la limite de qualité admissible dans le réseau téléphonique.

La structure de codage différentiel adaptatif permet, par contre, d'améliorer sensiblement la qualité du signal codé. Le principe du codage différentiel consiste à quantifier la différence $e(n)$ entre le signal et une prédition $p(n)$ de sa valeur à partir des valeurs précédentes $s(n-1), s(n-2), \dots$. Le codage MIC différentiel exploite la redondance temporelle présente dans le signal à coder en construisant une prédition $p(n)$ de chaque échantillon à venir, à partir des informations précédemment codées et transmises. Le bruit de quantification est d'autant plus faible que la différence $e(n)$ à quantifier est petite. Comme le signal de différence $e(n)$ a une distribution spectrale presque uniforme, il exploite beaucoup mieux les possibilités du codeur que le signal d'entrée initial. La transmission du signal de différence quantifié $e_q(n)$ requiert donc moins de bits, à puissance de bruit de quantification égale.

Le principe de fonctionnement du codeur différentiel représenté à la figure 2 est le suivant:

Connaissant la prédition $p(n)$ du signal $\hat{s}(n)$ reconstitué à partir des échantillons précédents on calcule l'erreur de prédition $e(n)$:

$$e(n) = s(n) - p(n)$$

Le quantificateur Q compare le signal de différence $e(n)$ à des seuils de décision. A chaque valeur de $e(n)$ correspond un numéro de la plage dans laquelle apparaît cette valeur. Un quantificateur inverse Q^{-1} transforme le numéro de la plage en la valeur médiane de celle-ci. Après quantification on obtient donc:

$$e_q(n) = e(n) + q(n)$$

Le prédicteur P calcule la prédition $p(n)$ de l'échantillon $s(n)$ du signal d'entrée à partir des échantillons reconstitués précédents $\hat{s}(n-1), \hat{s}(n-2), \dots$

Le signal reconstitué $\hat{s}(n)$ est égal au signal d'entrée $s(n)$, au bruit de quantification près:

$$\hat{s}(n) = e_q(n) + p(n)$$

ou

$$\hat{s}(n) = s(n) + q(n)$$

311 Quantificateur adaptatif

Le signal de différence $e(n)$ à l'entrée du quantificateur présente des variations importantes de dynamique. Il est donc nécessaire d'adapter le quantificateur à ces variations et de rendre les seuils ou échelons de quantification dépendants du niveau du signal de différence.

Dans la méthode d'adaptation du quantificateur proposée par Goodman et Wilkinson, figure 3, le signal de différence $e(n)$ est divisé par une grandeur $\Delta(n)$ proportionnelle à l'énergie du signal. Ce signal $e_s(n)$ est ensuite comparé aux paliers d'un quantificateur à seuils fixes.

Si $\hat{e}_s(n)$ est la valeur de la sortie du quantificateur correspondant au palier $I(n)$, il vient:

$$\hat{e}(n) = \hat{e}_s(n)\Delta n$$

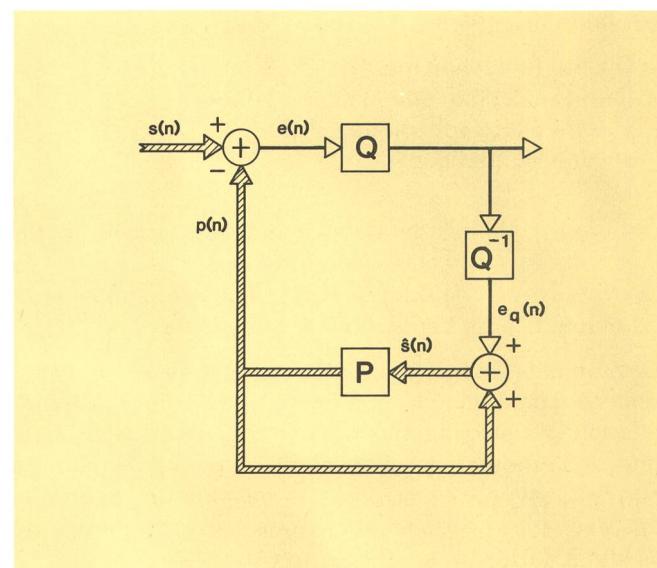


Fig. 2
Schéma de principe du codeur différentiel

Q Quantificateur
 Q^{-1} Quantificateur inverse

P Prédicteur

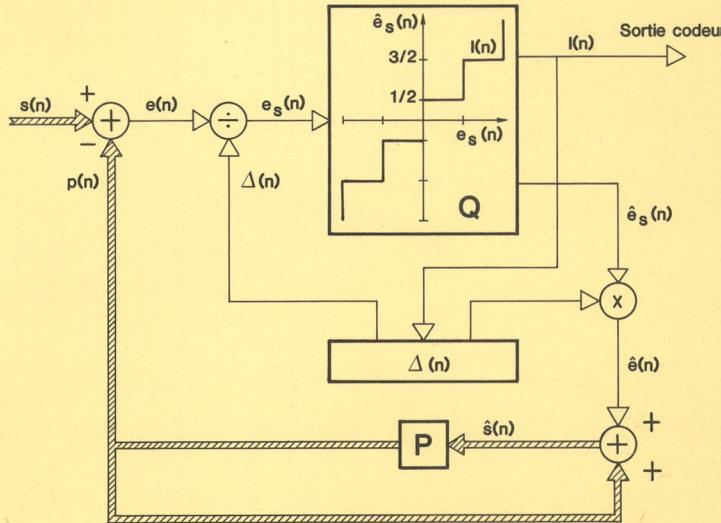


Fig. 3
Adaptation du quantificateur

L'adaptation du quantificateur consiste à modifier le seuil de saturation en fonction du niveau transmis $\hat{e}_s(n)$ à l'instant précédent, d'où l'algorithme d'adaptation:

$$\Delta(n+1) = [\Delta(n)]^y \cdot M[I(n)]$$

dans lequel le facteur M est calculé une fois pour toute pour chaque palier $I(n)$ et $y = 0,98$ (valeur typique).

Cet algorithme est plus facile à implanter sur un processeur de signal lorsque l'on fait appel à la forme logarithmique de $\Delta(n)$ et M , soit [2]:

$$d(n+1) = y \cdot d(n) + m[I(n)]$$

312 Prédicteur adaptatif

Le prédicteur adaptatif a pour fonction essentielle de calculer la valeur estimée $p(n)$ à partir du signal de différence quantifié $\hat{e}(n)$. L'adaptation du prédicteur consiste à rendre variable dans le temps les coefficients d'un filtre linéaire suivant l'évolution de la fonction de corrélation ou du spectre du signal $s(n)$. Le filtre de prédiction P a pour fonction de transfert la fonction $P(Z)$ telle que:

$$P(Z) = \sum_{i=1}^N a_i \cdot Z^{-i}$$

L'ordre du filtre N et celui de ses coefficients a_i doivent être choisis pour minimiser la puissance du signal de différence $e(n)$. Les coefficients sont calculés par l'expression [3]:

$$R(k) = \sum_{i=1}^N a_i \cdot R(k-i), \quad 0 \leq k \leq N$$

qui fait intervenir les éléments $R(k)$ de la fonction d'autocorrélation du signal $s(n)$. On admet en général que $N=4$ est une valeur convenable pour l'ordre du filtre prédicteur. Il existe plusieurs structures de prédicteur et un grand nombre d'algorithmes d'adaptation. Une structure couramment utilisée dans le traitement du signal de parole est la prédiction linéaire par filtre en treillis [4].

32 Codage en sous-bandes

Ce type de codage consiste à appliquer le signal d'entrée à un banc de filtres passe-bande adjacents couvrant toute la bande utile et de sous-échantillonner les signaux à la sortie de chaque filtre pour les ramener en bande de base. Finalement, les signaux ainsi filtrés sont codés et quantifiés séparément [5]. L'opération inverse est appliquée à la réception où les signaux sont décodés, suréchantillonnes, filtrés par des filtres passe-bande identiques à ceux de l'émission et enfin additionnés pour reconstituer le signal de parole. Cette procédure est illustrée par la figure 4 avec la représentation du schéma synoptique d'un système de codage à deux sous-bandes réalisé à l'aide de deux filtres miroirs en quadrature (QMF). Pour garantir une réponse globale uniforme dans toute la bande, les filtres doivent se couper à environ -3 dB. L'échantillonnage provoque des repliements de spectre d'une sous-bande à l'autre, ce qui se traduit par une distorsion audible plus ou moins prononcée. Même l'utilisation de filtres à pente très raide ne résout pas le problème, car d'autres phénomènes d'échos difficiles à maîtriser apparaissent dans ce cas [6]. La meilleure solution consiste à utiliser un banc de filtres miroirs en quadrature dont la propriété la plus intéressante est d'éliminer les repliements de spectre d'une sous-bande à l'autre. En l'absence du quantificateur, l'utilisation des filtres miroirs en quadrature per-

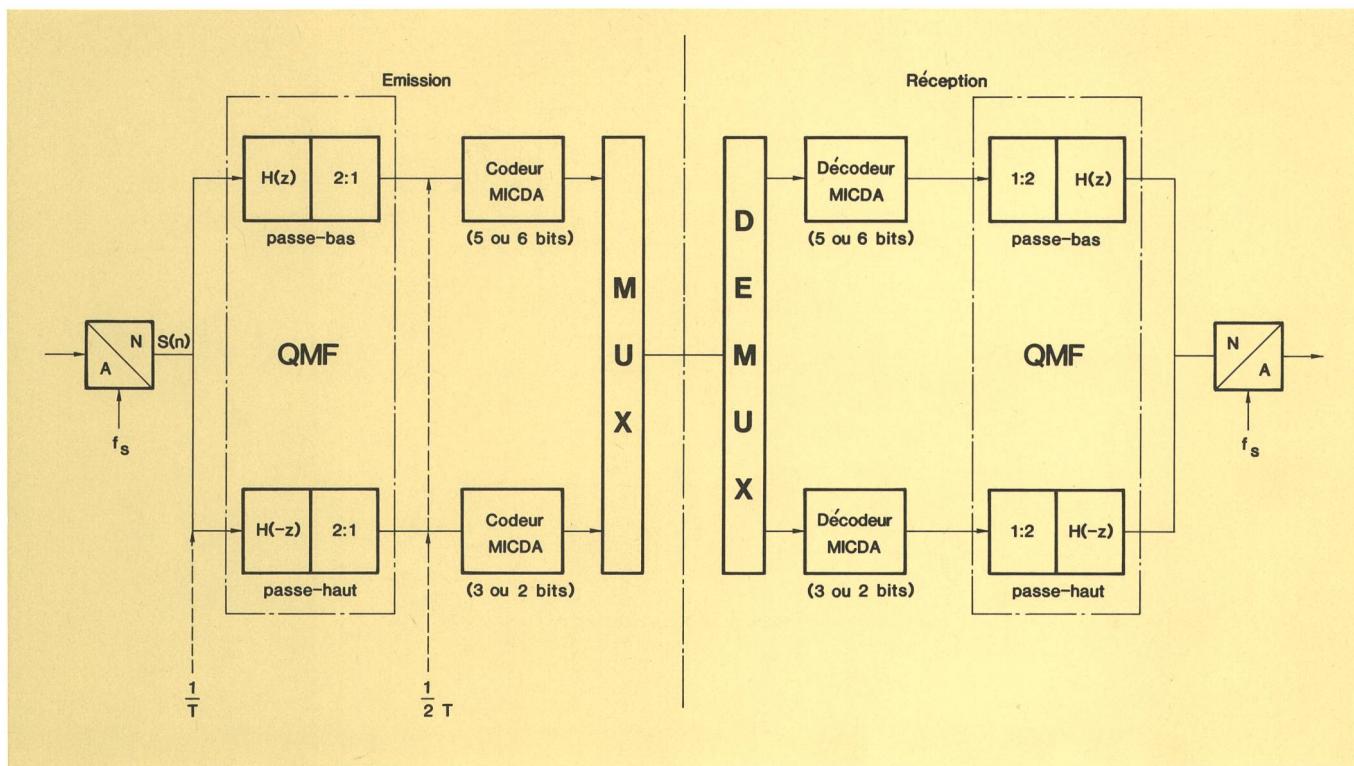


Fig. 4
Schéma de principe du codage à deux sous-bandes par une structure de filtres miroir en quadrature

met de reconstituer le signal d'entrée de façon quasi parfaite.

Le codage en sous-bande présente l'avantage de permettre l'ajustement du nombre de bits utilisés pour quantifier chaque sous-bande suivant l'énergie du signal. Pour tenir compte de la densité spectrale de puissance moyenne du signal de parole, le nombre de bits attribué à la sous-bande basse fréquence est plus important que celui généralement réservé à une sous-bande haute fréquence. La décomposition du signal d'entrée peut être augmentée à l'aide d'un banc de filtres ayant une structure en arbre dans laquelle chaque sous-bande peut être décomposée en deux nouvelles sous-bandes et ainsi de suite. L'utilisation du codage en sous-bandes à 64 kbit/s conduit à une excellente qualité permettant de prendre en compte une largeur de bande de fréquences plus étendue que la bande téléphonique. Pour un taux d'information de 16 kbit/s à 24 kbit/s, il est encore possible d'obtenir un signal de qualité transparente.

33 Autres principes de codage à bas débits

Lorsque le taux d'information disponible se situe entre 8 kbit/s et 16 kbit/s environ, il est indispensable de recourir à des méthodes de codage plus sophistiquées pour que la parole reste hautement intelligible, avec toutefois quelques distorsions perceptibles. Parmi ces méthodes, le codage adaptatif par prédition (APC), le codage adaptatif par transformée (ATC), et le codage à excitation par bande de base (VEV: Voice Excited Vocoder) sont, après le codage en sous-bandes, les plus connues. On n'entrera pas dans le détail de ces codeurs

qui ont fait l'objet de nombreux travaux cités dans [5] et dont les applications dans le domaine des télécommunications sont encore à l'étude, en particulier sur le codage des signaux de la parole à 16 kbit/s pour les systèmes de communication radio mobile. Rappelons simplement sur quelles hypothèses reposent ces trois procédés de codage.

- L'objectif recherché avec le *codage adaptatif par prédition* est de réduire sensiblement la redondance du signal en lui retranchant ce qui peut être prédit à l'aide d'un bloc d'échantillons précédents. Le signal de différence ainsi obtenu est quantifié, puis transmis. La prédition exploite, d'une part, la corrélation entre les échantillons successifs du signal et, d'autre part, tire profit de la nature quasi périodique des sons voisés pour éliminer une partie de leur densité spectrale. Le gain de prédition total est d'environ 17 dB. En d'autres termes, le codage adaptatif par prédition est une méthode où l'on recherche à minimiser l'énergie de prédition, ce qui a pour corollaire la réduction sensible de l'énergie du bruit de quantification. A 16 kbit/s, le codage APC permet d'obtenir une très bonne qualité dans la classe subtéléphonique. A 9,6 kbit/s, la qualité de la parole est encore très acceptable.
- Dans le *codage adaptatif par transformée*, le signal est échantillonné puis analysé par blocs de N échantillons ($N = 128 \dots 256$). Chaque bloc est traité suivant la transformée discrète en cosinus et les échantillons qui sont quantifiés de façon indépendante sont transmis en même temps que les paramètres des quantificateurs [5]. Cette technique assez complexe est en pleine évolution et on peut s'attendre à obtenir une bonne qualité subtéléphonique à 9,6 kbit/s.

- Le codage à excitation par bande de base est une solution hybride entre celle du codeur et du vocodeur qui permet un débit d'information inférieur à 9,6 kbit/s tout en conservant une bonne qualité subtéléphonique. Comme pour le codage adaptatif par transformée son optimalisation n'est pas terminée [5].

Dans un proche avenir, il est possible que l'une ou l'autre de ces méthodes trouve une application pratique dans le codage à bas débit qui figure au nombre des prochaines études du CCITT.

34 Application au codage à 32 kbit/s

De nombreuses Administrations ont engagé des études sur le codage MIC différentiel adaptatif, en vue d'améliorer le rendement des supports de transmission numériques. On s'est rapidement rendu compte que les performances des systèmes à paramètres adaptatifs étaient supérieures à celles des systèmes à paramètres fixes pour la transmission de la parole. C'est ainsi que le CCITT vient de publier une nouvelle recommandation sur le codage MIC différentiel adaptatif (MICDA) à 32 kbit/s qui indique les caractéristiques à employer pour la conversion d'une voie MIC utilisant la loi A ou la loi μ à 64 kbit/s en une voie à 32 kbit/s.

Le schéma synoptique d'un transcodeur MIC-MICDA est indiqué à la figure 5. Dans ce système, on trouve, après les interfaces MIC, un organe démultiplexeur qui sépare les signaux des voies des signaux contenus dans les intervalles de temps ITO et IT16 (signal de verrouillage de trame, signalisations, alarmes, etc.). Les 2×30 signaux de voie sont alors appliqués à une calculatrice numérique qui effectue l'extension du code MIC, loi A à 8 bits, en un code linéaire à 16 bits. Le train de bits résultant est ensuite converti à l'aide d'une technique de transcoding conforme à l'algorithme défini par le CCITT. Les signaux des intervalles de temps ITO et IT16 sont acheminés à un multiplexeur qui forme avec la suite des

mots codes MICDA à 4 bits une trame à 2048 kbit/s compatible avec les équipements de lignes à 2 Mbit/s. Les opérations inverses sont effectuées pour la conversion MICDA-MIC [7].

Pour les Administrations qui ont choisi un débit de base de 64 kbit/s pour la commutation, l'application du codage à 32 kbit/s se limitera aux transmissions par satellites ou pour doubler temporairement la capacité d'une ligne de transmission surchargée.

35 Applications au codage en bande élargie à 64 kbit/s

L'audioconférence et la visioconférence représentent les applications les plus importantes à moyen terme du codage des signaux vocaux à large bande, mais certaines Administrations envisagent aussi d'autres applications, en particulier pour le téléphone à haut-parleur dans un RNIS de haute qualité ou pour les canaux de commentaires, voire de musique en radiodiffusion. Les travaux de normalisation au CCITT sont assez avancés et il semble qu'il y a lieu de donner la préférence à un algorithme de codage MIC différentiel adaptatif en sous-bandes (MICDA-SB), avec 5 ou 6 bits pour le quantificateur de la sous-bande inférieure et 3 ou 2 bits pour celui de la sous-bande supérieure. La fréquence d'échantillonnage est de 16 kHz. Les essais subjectifs ont également montré la supériorité d'une quantification et prédition adaptative. Le système devrait être conçu en tenant compte du besoin d'une voie auxiliaire de données (8 kbit/s...16 kbit/s), par exemple, pour la transmission d'images à défilement lent ou la télécopie.

4 Synthèse de la parole

La synthèse vocale commence à faire une apparition de plus en plus remarquée dans le domaine des télécommunications, le microprocesseur parlant n'est plus une

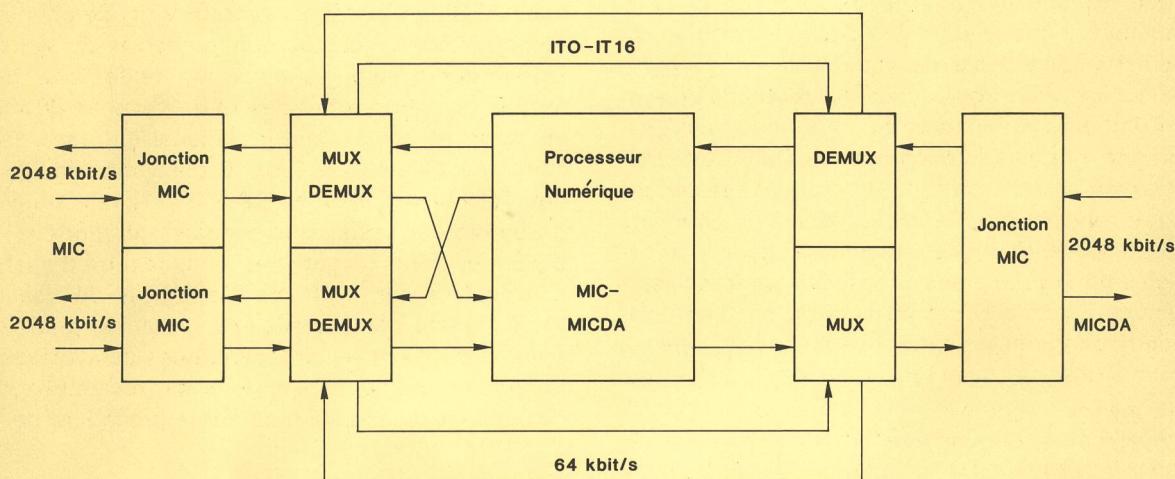


Fig. 5
Schéma du transcodeur MIC-MICDA à 60 voies

vue de l'esprit, mais bel et bien la réalité de demain. La croissance rapide du réseau téléphonique numérique (RNIS) entraîne réciproquement une mutation sensible des besoins en renseignements de tout genre. La demande de renseignements augmente déjà beaucoup plus rapidement que celle des abonnements, ce qui se traduira tôt ou tard par de multiples problèmes au niveau des services de renseignements des PTT. Or, les moyens mis en œuvre aujourd'hui ne pourront plus être indéfiniment extrapolés. L'objectif sera, à l'avenir, de limiter cette croissance en utilisant des systèmes opérationnels de synthèse automatique de la parole, fonctionnant en temps réel et produisant une parole non seulement intelligible mais autant que possible «naturelle». On envisage pour l'avenir l'automation partielle, voire intégrale, des services suivants des télécommunications:

- annuaire téléphonique parlant
- service du réveil
- taxation: l'abonné compose un numéro, la machine le renseigne sur le prix de la communication correspondante
- changement du numéro de téléphone
- banque de données
- centre de documentation automatique
- service des renseignements, etc.

Parmi les nombreuses possibilités d'application qui sont offertes aux administrations des PTT, la synthèse vocale automatique, à partir du texte, soulève un intérêt croissant et les motivations sont suffisantes pour que des recherches de plus en plus nombreuses soient engagées avec pour objectif final la réalisation d'équipements de réponse vocale automatique pour les services d'informations et de renseignements avec accès au réseau téléphonique.

41 Synthèse par éléments phonétiques

La production artificielle d'un signal de parole peut être réalisée selon différentes méthodes dites paramétriques. Ces méthodes visent à coder non plus le signal $f(t)$ lui-même, mais les paramètres d'un modèle de production reposant sur le principe du produit d'une fonction d'excitation par une fonction de transfert. En général, les applications dans le domaine des télécommunications requièrent une taille de vocabulaire très importante. Il est alors préférable d'utiliser la méthode de synthèse par diphonèmes qui permet, en théorie, une qualité supérieure à celle d'autres procédés. Dans cette méthode, les éléments constitutifs du dictionnaire rangés en mémoire sont des fragments de mots. Si le phonème est la plus petite unité phonétique d'un langage – on dénombre plus de 37 phonèmes dans la langue française – le diphonème désigne l'ensemble formé par l'association de deux phonèmes. Les diphonèmes sont séparés les uns des autres par des états quasi stationnaires qui ne portent pas d'information. Pour reconstituer un mot, il suffit d'assembler les diphonèmes par leur extrémité commune. La synthèse par diphonèmes est séduisante par sa simplicité mais exige, en contrepartie, une bibliothèque assez importante d'éléments de parole. Théoriquement, il faudrait pouvoir disposer pour la langue française d'au moins (37)² combinaisons soit 1369 diphonèmes. L'assemblage des diphonèmes permet d'obtenir

une bonne intelligibilité de la parole, mais malheureusement totalement dépourvue de «naturel». Par conséquent, il est indispensable d'y ajouter une intonation selon une procédure de traitement dite prosodique. En d'autres termes, il s'agit d'élaborer des schémas mélodiques sur chaque syllabe d'un mot et de calculer le rythme sur le mot considéré. Lorsqu'il existe des discontinuités à la fonction des syllabes, il est parfois nécessaire d'utiliser des indices prosodiques pour y remédier. Cette variation fine de la mélodie est couramment désignée par traitement de la micromélodie. Conjointement à ces méthodes, on utilise, de préférence, la compression par codage à prédiction linéaire comme type de synthétiseur car c'est aujourd'hui la technique la mieux appropriée à la synthèse à bas débit. Evidemment, la bibliothèque d'éléments phonétiques doit être adaptée au type de synthétiseur. Cette combinaison d'une technique de codage-compression LPC avec la synthèse par diphonèmes et traitement prosodique a donné lieu à des réalisations dont les performances sont encourageantes, ce qui a incité plusieurs laboratoires à retenir ce choix technique dans la réalisation de divers projets spécifiques sur la synthèse vocale automatique à partir du texte [10].

42 Synthèse automatique de la parole à partir du texte

La synthèse automatique de la parole à partir du texte consiste à passer de l'écriture orthographique au signal de parole. Elle est un sujet privilégié d'applications dans le secteur des services de renseignements des télécommunications. Le schéma synoptique de la conversion texte-parole est montré à la *figure 6*. Le système opère suivant trois niveaux successifs:

- la transcription orthographique – phonétique
- le calcul de la prosodie et l'assemblage des trames qui constituent la commande du synthétiseur
- le synthétisation de parole proprement dite.

421 Transcription orthographique-phonétique

Le texte entre dans l'unité correspondant au niveau 1 en code ASCII, puis réapparaît traité en code phonétique, y compris les marques de ponctuation et de syntaxe que l'opérateur a insérées dans son texte. Une des méthodes de transcription les plus efficaces développées en commun par le Centre National d'Etudes des Télécommunications (CNET) et l'Ecole Nationale Supérieure des Télécommunications (ENST) [11] fait intervenir un ensemble de tests arborescents qui portent sur les contextes droits ou gauches d'une fenêtre d'analyse que l'on déplace sur la phrase. Ce logiciel de transcription est complété par des sous-programmes pour la transcription des chiffres et des marques de syntaxes en caractères phonétiques ainsi que par divers programmes de services destinés à modifier la procédure de test.

422 Calcul de la prosodie

Le traitement prosodique s'effectue au niveau 2 de la figure 6, il consiste à analyser la chaîne phonétique, c'est-à-dire à détecter le nombre de syllabes du mot et la na-

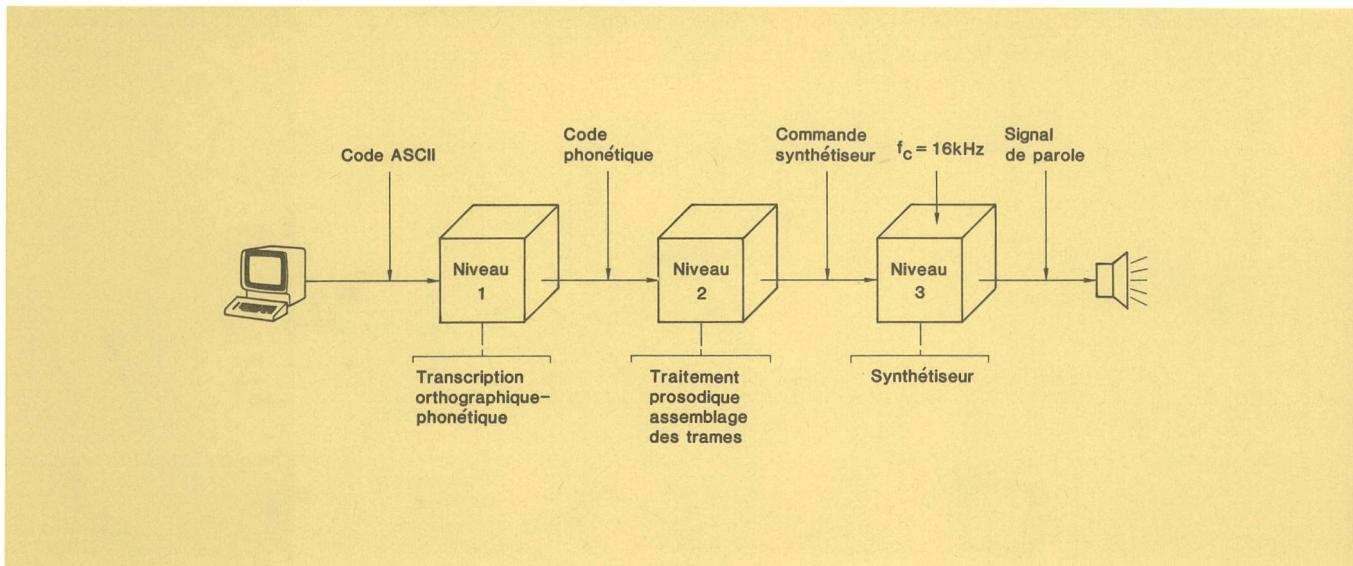


Fig. 6
Conversion texte-parole

ture des débuts et fins de mots, à calculer le rythme et la mélodie pour finalement élaborer des trames à partir des schémas précédemment dégagés.

423 Assemblage des trames

L'assemblage des trames qui constitue la commande du synthétiseur est effectué à l'issue de tous les traitements précédents. Finalement, le synthétiseur utilise les paramètres de commande pour générer la parole. En général, on calcule les paramètres qui caractérisent la valeur du signal de parole toutes les 10 ms à 25 ms.

43 Procédures pour optimiser la parole

Dans l'hypothèse d'une application qui requiert une certaine qualité de la parole et dans laquelle le vocabulaire est pratiquement illimité, il est parfois nécessaire que l'architecture du système de synthèse permette d'effectuer des essais d'optimisation de la parole. A cet effet, les trois procédures d'optimisation suivantes sont très souvent appliquées:

- correction de la prononciation
- correction du rythme
- modification des caractéristiques de la voix.

431 Correction de la prononciation

En général, les systèmes les plus performants prononcent correctement les mots d'un test orthographique, mais il peut arriver qu'un mot ou un secteur de phrase soit prononcé d'une manière insatisfaisante. Dans ce cas, il est possible d'optimiser la qualité mélodique ou la compréhensibilité du mot ou du secteur de phrase en faisant appel, par exemple, à une suite de symboles phonétiques en lieu et place de l'épellation orthographique.

432 Correction du rythme

L'introduction de pauses dans la phrase s'effectue grâce au choix de divers signes de ponctuation. Il est donc également possible, avec des marques de ponctuation et des règles de syntaxe, d'améliorer le rythme et l'accent de la séquence parlée.

433 Modification des caractéristiques de la voix

En général, un système de synthèse automatique de la parole lit un texte au rythme moyen de 180 mots par minute. Ceci est un taux de lecture normal qui est bien adapté à la plupart des situations et au genre d'informations envisagé. Dans certaines conditions, il est toutefois souhaitable d'augmenter ou éventuellement de diminuer ce taux à des valeurs comprises entre 120 et 300 mots par minute. A cet effet, le logiciel de commande permet de modifier la durée des pauses définies par la ponctuation et d'affecter ainsi le débit de la parole et le timbre de la voix.

44 Le projet SAPOR

SAPOR (*synthèse automatique de parole par ordinateur*) est un projet de développement sur la synthèse vocale à partir du texte. Il devrait concrétiser les études menées par une équipe de chercheurs de l'Ecole Polytechnique Fédérale de Zurich, avec le soutien et la collaboration des laboratoires de recherches et du développement des PTT. On envisage une première application expérimentale de la synthèse dans le cadre de l'exploitation du service des machines parlantes d'un réseau de télécommunication interne des PTT. L'objectif déclaré comporte les deux volets principaux suivants:

Axe de recherche A

- Evaluation qualitative globale de la parole résultante pour la langue allemande
- Appréciation des niveaux intonatifs et du degré des contrastes (mélodie)

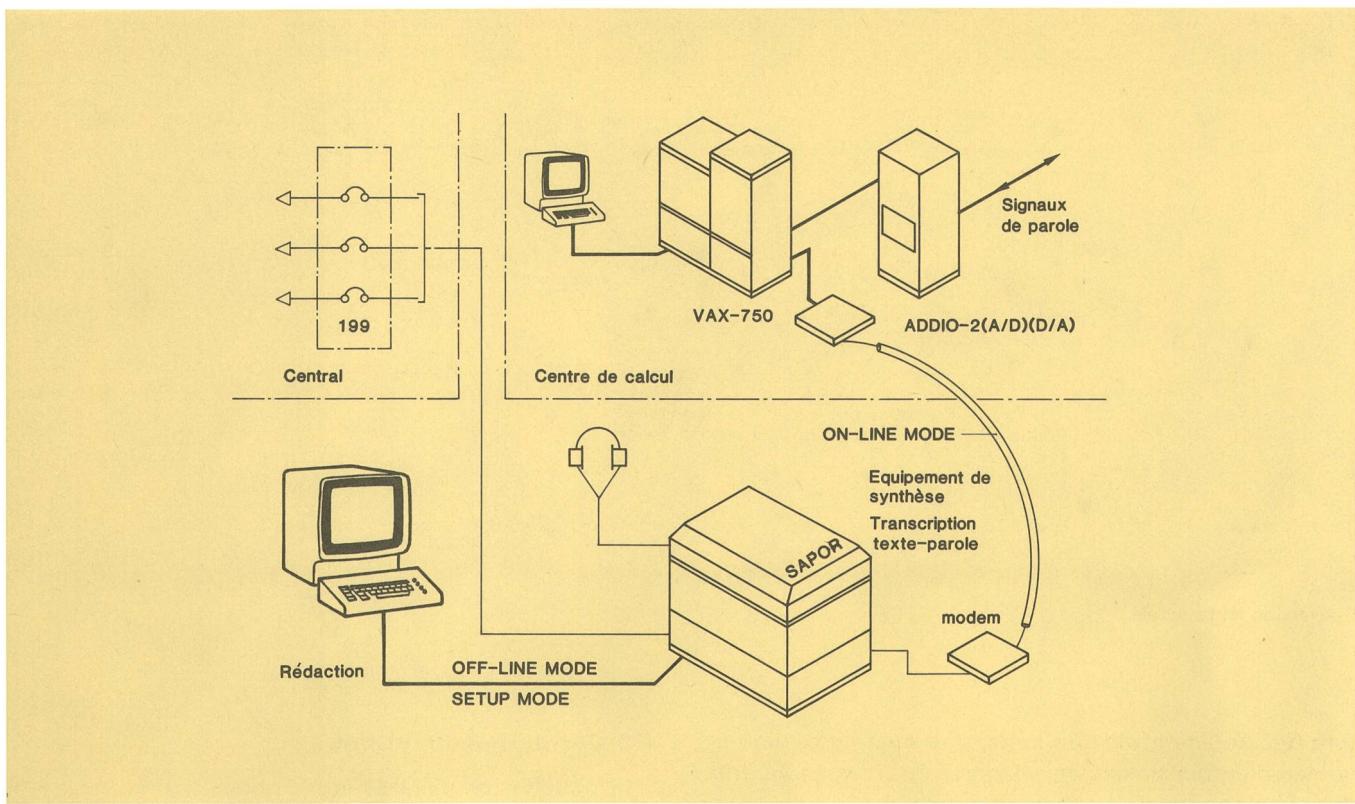


Fig. 7
Mode d'exploitation de l'unité expérimentale de synthèse vocale «SAPOR»

- contrôle des performances de la transcription du code orthographique au code phonétique.

Axe de recherche B

- Dégager les principales restrictions pratiques dans l'exploitation des systèmes de synthèse de la parole
- analyser les réactions et comportements individuels des usagers à l'égard d'une voix «quasi naturelle».

La figure 7 montre la configuration provisoire et les modes d'exploitation du système expérimental de synthèse vocale à partir du texte. Le cœur du système est constitué par un convertisseur texte-parole dont le texte provient, soit d'un processeur hôte (VAX 750) installé au centre technique de la Division des recherches des PTT, soit d'un terminal installé au local de rédaction du service des machines parlantes. Dans sa phase finale de développement, l'installation SAPOR permettra une exploitation selon les trois modes de fonctionnement suivants:

- En mode on line, SAPOR est relié avec un ordinateur hôte. Dans cet état, il est possible d'agir directement sur les différents algorithmes de transcription et de synthèse pour les améliorer progressivement, ce qui est indispensable dans un tel projet.
- A l'issue des développements, SAPOR devra fonctionner en mode off line, c'est-à-dire comme un système autonome de conversion texte-parole. Il devra être en mesure de lire les commandes en provenance d'un terminal et de les traduire en un langage parlé.
- En mode SET UP, l'opérateur a la possibilité de modifier certains paramètres internes de SAPOR pour optimiser la transcription texte-parole.

Le projet SAPOR devrait apporter une contribution importante à la connaissance et à l'appréciation des possibilités et des limites de la synthèse vocale. A priori, les avantages et les inconvénients peuvent se résumer comme il suit:

- renouvellement rapide des informations et des renseignements
- l'abonné pourra avoir accès au début du texte de l'information
- diminution sensible des coûts d'exploitation du service des renseignements des PTT
- grande flexibilité lors de l'introduction de nouveaux services
- incertitude au sujet de l'accueil réservé par le public à une voix «quasi naturelle». Ce facteur dépend dans une large mesure des progrès réalisés sur les algorithmes de synthèse et de la richesse des bibliothèques d'éléments phonétiques.

5 Codage des signaux musicaux

Les systèmes de multiplexage MIC des signaux sonores de haute qualité font en général appel à trois procédés de codage différents:

- codage linéaire de 13 bits à 16 bits
- codage à compression-extension instantanée à 11 segments
- codage à compression-extension quasi instantanée à 5 échelles.

De nombreuses expériences réalisées sous forme d'essais subjectifs ont montré qu'une résolution maximale de 14 bits par mot code était nécessaire lorsque l'on ap-

pliquait pour des signaux audio de haute qualité un codage linéaire. Une résolution de 16 bits est aujourd'hui très souvent utilisée dans l'industrie du disque numérique et les nouveaux studios de radiodiffusion. L'importante largeur de bande que requiert ce procédé est un lourd handicap lorsqu'il s'agit de l'appliquer aux systèmes de transmission. Dans ce cas, il est préférable de recourir à une forme de compression-extension qui n'introduit pas de dégradation audible.

51 Principe du codage à compression-extension instantanée à 11 segments

Le procédé de codage à compression instantanée est appliqué en utilisant la loi dite «A» approchée par 11 segments de droite et 11 bits par mot code, ce qui correspond à une résolution minimale de 9 bits d'un codage linéaire. La caractéristique de compression logarithmique est définie en détail dans le tableau I. Cette loi présente une caractéristique de transfert discontinue avec des niveaux de quantification peu espacés ou très espacés, suivant que les signaux ont une amplitude faible ou grande. On obtient ainsi une diminution du bruit de quantification qui affecte les signaux faibles au prix d'un accroissement du bruit qui accompagne les signaux forts. Grâce à l'effet de masque de ces derniers, la qualité subjective du signal n'est pratiquement pas altérée. De cette manière, on réduit le nombre de bits à transmettre pour chaque échantillon. Le 11^e bit (libre) du segment central peut servir à la transmission d'un canal auxiliaire de données dont le débit binaire dépend de la distribution d'amplitude du signal radiophonique. Ce n'est que pendant les pauses du programme que le débit binaire atteint sa valeur maximale de 32 kbit/s. De plus, un bit de parité protège contre les erreurs les cinq bits les plus importants de chaque échantillon. Le débit binaire total d'une voie audio est donc de 384 kbit/s et

Tableau I. Loi de codage A à 11 segments

Signe	Quantification		Segment N°	Nombre de bits	Codage à 11 bits										
	Uniforme	Non uniforme			1	2	3	4	5	6	7	8	9	10	11
Positif	8192 4097	896 769	1	9	1 1	1 1	1 1	1 1	1 0						
	4096 2049	768 641			1 1	1 1	1 1	0 0	1 0						
	2048 1025	640 513			1 1	1 0	0 1	1 0							
	1024 513	512 385			1 1	1 0	0 0	0 0	1 0						
	512 257	384 257			1 1	0 1	1 1	1 0							
	256 129	256 129			1 1	0 0	1 1	0 0	1 1	1 0	1 0	1 0	1 0	1 0	1 0
	128 65	128 65			1 1	0 0	0 1	1 0	x x						
	64 1	64 1			1 1	0 0	0 0	0 0	1 0	1 1	1 1	1 1	1 1	1 1	x x
Négatif	1 64	1 64	6	14	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	x x
	65 128	65 128			0 0	0 0	0 0	1 1	0 1	1 1	1 1	1 1	1 1	1 1	x x
	129 256	129 256			0 0	0 0	1 1	0 1	0 0						
	257 512	257 384			0 0	0 0	1 1	1 1	0 1	0 1	0 1	0 1	0 1	0 1	0 0
	513 1024	385 512			0 0	1 0	0 0	0 0	0 1	0 1	0 1	0 1	0 1	0 1	0 0
	1025 2048	513 640			0 0	1 0	0 0	1 0	0 1	0 1	0 1	0 1	0 1	0 1	0 0
	2049 4096	641 768			0 0	1 1	1 0	0 1	0 0	0 1	0 1	0 1	0 1	0 1	0 0
	4097 8192	769 896			0 0	1 1	1 1	1 1	0 1	0 1	0 1	0 1	0 1	0 1	0 0

Les 11^{es} bits, indiqués par un x, peuvent servir à la transmission d'un signal de données

correspond exactement à celui de six voies téléphoniques au débit binaire de 64 kbit/s, définie par le CCITT. Cette loi est traitée en détail dans [8].

52 Principe du codage à compression-extension quasi instantanée à cinq gammes

L'hypothèse sur laquelle se fonde le principe de la loi de compression-extension numérique quasi instantanée est la suivante [9]: lorsque l'on observe les éléments binaires d'une succession d'échantillons numériques d'un signal audio, on s'aperçoit que les bits de poids élevé varient d'une manière plus ou moins lente suivant la nature du signal. Dans ces conditions, il est apparu avantageux de grouper périodiquement ces échantillons en un bloc dans lequel les bits n'ayant peu ou pas évolué ne seront transmis qu'une seule fois. Le procédé à appliquer est donc d'observer le signal à intervalle régulier et d'adapter chaque fois la loi de quantification à la caractéristique d'amplitude du groupe d'échantillons.

La figure 8 définit plus précisément la loi de compression quasi instantanée à cinq gammes. Le codage s'effectue de la manière suivante: on code tout d'abord le signal analogique à l'aide d'un convertisseur à quantification uniforme. Après avoir mis en mémoire un certain nombre d'échantillons successifs, on prend en considération l'échantillon qui possède l'amplitude maximale en valeur absolue et on détermine, sur la base de cette observation, la gamme appropriée sur laquelle seront codés tous les autres échantillons du bloc considéré. Cette gamme est identifiée par un facteur d'échelle qu'il faudra transmettre avec chaque groupe d'échantillons.

L'analyse empirique de cette loi de compression met en évidence une performance égale à $(n-1)$ bit par rapport à la loi de compression semi-logarithmique.

53 Codage des signaux audio à 128 kbit/s

Au cours des séances d'écoute subjective et lors des travaux d'évaluation des algorithmes de codage des signaux vocaux en bande étroite à 64 kbit/s, il est apparu que le codage MIC différentiel adaptatif en sous-bandes fournit une meilleure qualité que le codage MIC à 128 kbit/s. Il semble donc intéressant de chercher à connaître les possibilités d'un codage MICDA en sous-bandes à 128 kbit/s, applicable aux signaux audiofréquences mais aussi dans l'éventualité d'une solution de remplacement numérique de la télédiffusion à haute fréquence. Un codage à 128 kbit/s permettrait une application compatible avec le RNIS (réseau numérique avec intégration des services). Les PTT suisses s'intéressent à cette question et étudient les hypothèses suivantes:

- Optimisation de l'algorithme de codage MICDA-SB des signaux vocaux en bande étroite à 64 kbit/s pour les signaux de radiodiffusion à 7 kHz de largeur de bande.
- Codage MICDA-SB à 128 kbit/s et 7 kHz de largeur de bande. La quantification est suffisamment fine pour que la qualité de codage se rapproche de celle obtenue avec des systèmes à 192 kbit/s fonctionnant selon une loi de compression-extension instantanée à 11 segments ou quasi instantanée à cinq gammes.
- Codage MICDA-SB à 128 kbit/s et 15 kHz de largeur de bande. Cette hypothèse soulève indiscutablement le plus de difficultés en rapport avec l'augmentation du bruit de quantification et de granulation. Cette dégradation peut être partiellement masquée en utilisant un signal composite additionnel de vibration (dither) formé de la superposition d'un bruit blanc et d'un signal carré à fréquence égale à la moitié de la fréquence d'échantillonnage. D'autres méthodes sont à l'étude comme, par exemple, celle qui consiste à utiliser le spectre résiduel de repliement d'une sous-bande pour améliorer l'impression subjective d'écoute des signaux de fréquences élevées.
- Codage MICDA-SB à 128 kbit/s avec une largeur de bande qui résulte du meilleur compromis subjectif entre le bruit de quantification et la bande passante. On peut envisager, par exemple, une résolution globale à 10 bits et une fréquence d'échantillonnage de 25,6 kHz, ce qui correspondrait à la largeur de bande actuelle de la télédiffusion à haute fréquence (12 kHz). Toutefois, pour des raisons de synchronisation et de compatibilité dans la hiérarchie des systèmes numériques, cette solution semble offrir certains inconvénients.

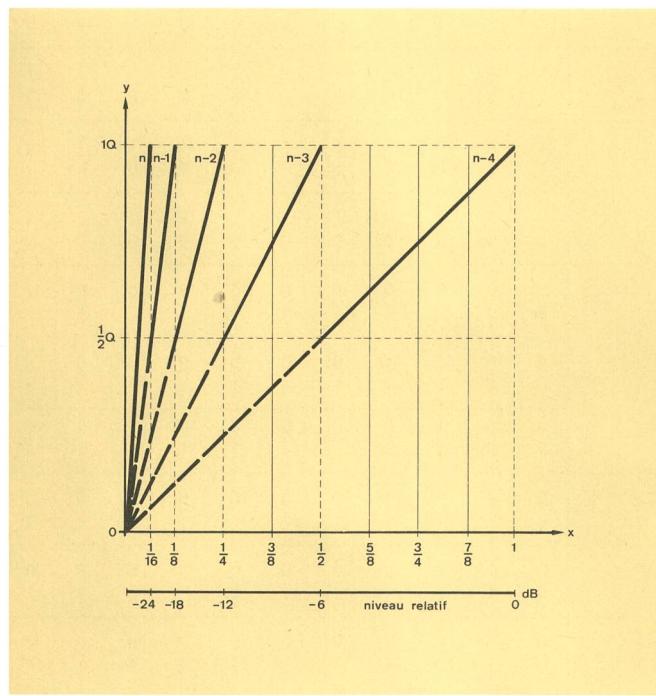


Fig. 8
Caractéristique de compression quasi instantanée à cinq gammes

Q Nombre total d'échelons de quantification contenus dans le quadrant positif
x Amplitude normée du signal d'entrée
y Niveau de quantification
(n-1...4) Résolution par gamme

6 Evaluation des lois de codage grâce à des logiciels

L'évaluation d'une loi de codage consiste à rechercher constamment la relation existant entre les valeurs mesurées et l'impression subjective produite par les imper-

fections de l'algorithme de codage ou par les perturbations dues à la transmission. Cette recherche pour optimiser les caractéristiques de codage est liée à de nombreuses modifications de paramètres qui ne peuvent se concevoir sans l'aide de l'ordinateur. A cet effet, on utilise toujours le même montage dont un logiciel diversifie les modes d'emploi, ce qui évite de perdre du temps à concevoir et à mettre au point chaque fois un appareillage différent. Aujourd'hui, il existe toute une série de processeurs rapides et spécifiques au traitement numérique du signal qui possèdent une capacité de calcul telle que même les algorithmes les plus complexes peuvent être exécutés en temps réel. Toutefois, au cours des premières phases d'étude et de recherche, la simulation en temps réel n'est pas essentielle car les tests d'évaluation subjective ne portent que sur des échantillons de parole ou séquences musicales de durée très limitée (20...30 s). Une mémoire de masse suffisante au profit d'une procédure de simulation simple et très souple sont les exigences fondamentales pour garantir une certaine efficacité dans le développement. Il est évident que pour optimiser l'architecture matérielle du produit définitif, il est essentiel d'utiliser ensuite des processeurs rapides permettant un traitement en temps réel de l'algorithme développé. Pour répondre aux besoins de la recherche en matière de codage, les PTT suisses ont réalisé une installation qui permet de simuler et d'étudier les systèmes de codage et de transmission des signaux de la parole et de la musique à l'aide d'un ordinateur de grande capacité et d'un convertisseur A/D-D/A à haute résolution.

61 Architecture matérielle de l'ADDIO-2

L'infrastructure de l'installation comporte trois sections représentées de manière synoptique à la figure 9. Elles assurent chacune les fonctions suivantes:

- *Section analogique*: elle comprend l'ensemble des appareils de production et de réception des signaux (microphones, amplificateurs, enregistreurs numériques, synthétiseurs, haut-parleurs, etc.) installé dans des locaux d'acoustique adéquats tel, par exemple, celui décrit au paragraphe 75.
- *Section des convertisseurs A/D-D/A*: elle rassemble tous les circuits de conversion analogique-numérique-analogique, de filtrage et de mise en forme du signal, ainsi que les équipements des jonctions numériques et analogiques. Le convertisseur proprement dit est un DSC-200 produit par *Digital Sound Corporation*. Sa résolution est de 16 bits, ce qui correspond à une dynamique théorique de 96 dB. La fréquence d'échantillonnage maximale est de 50 kHz, permettant ainsi le traitement de signaux quelconques dans une bande de fréquence de plus de 20 kHz programmable à l'aide d'un banc de filtres passe-bande. Le nombre de voies pouvant être traitées simultanément est de 8 aux maximum.
- *Section processeur*: elle comprend un ordinateur VAX-11/750 et une mémoire de grande capacité RA81 (456 MB) permettant d'enregistrer ou de reproduire en temps réel une séquence musicale de haute qualité en stéréophonie, échantillonnée à 50 kHz, d'une durée de plus de 30 minutes. On trouve naturellement dans

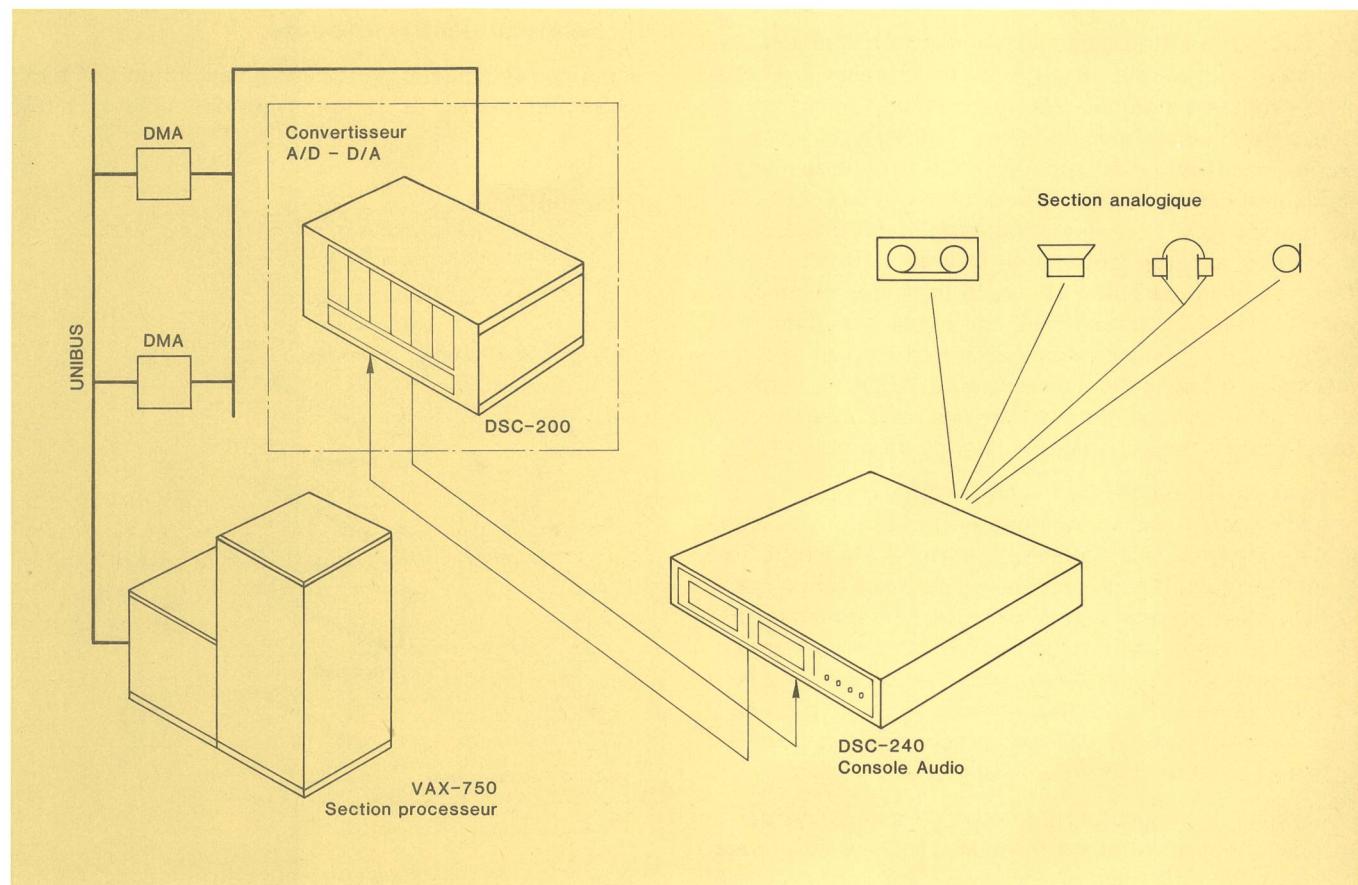


Fig. 9
Schéma synoptique de l'ADDIO-2

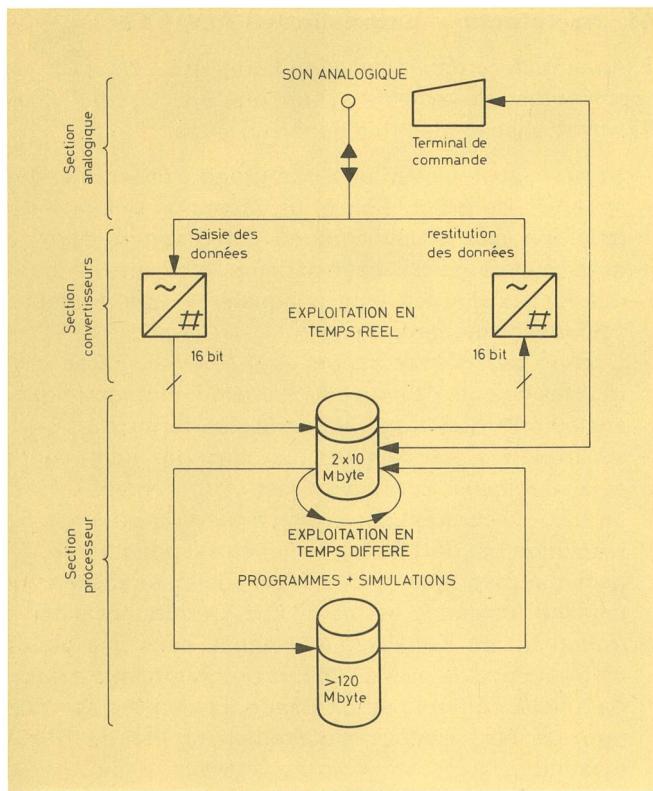


Fig. 10
Principe de transfert des signaux numériques

cette section tous les périphériques nécessaires aux travaux et au développement du logiciel.

62 Procédures de simulation

La configuration du système permet le traitement des signaux de la parole ou de la musique selon la procédure simplifiée indiquée à la figure 10. Le signal analogique, préalablement mis en forme (réglage du niveau, filtrage, préaccentuation, etc.) est codé à partir de niveaux de quantification uniformes pour obtenir une succession de mots à 16 bits. Le signal numérique est ensuite mémorisé sur disque. L'opération inverse s'effectue d'une façon similaire. La saisie et la restitution des données se font toujours en temps réel, mais pas simultanément (half-duplex). La description suivante donne un exemple classique d'application du système ADDIO-2 dans le domaine du codage et de la transmission numérique des programmes radiophoniques. Enoncé du problème:

- Dans l'éventualité d'un multiplexage de deux voies audiofréquences numériques en mode d'exploitation plésiochrome, la différence de rythme à absorber est réalisée à l'aide d'un procédé dit de justification négative, nulle ou positive. A cet effet, il s'agit d'insérer dans un signal numérique des indications de justification ainsi qu'un mot de synchronisation, par une substitution de certains bits, et de juger, à la reproduction, l'effet de dégradation subjectif des changements forcés des bits d'information du signal original.

L'organisation des programmes de simulation par le logiciel est présentée à la figure 11. Dans un premier temps, le signal audiofréquence subit certaines opérations de réglage. Il est ensuite codé en un signal numérique à l'aide du convertisseur analogique-nu-

mérique à 16 bits de résolution fonctionnant à une fréquence récurrente d'échantillonnage de 32 kHz. Le signal ainsi formé est alors directement mémorisé en temps réel (programme Pr. 1 ADDIO) sur un disque du VAX. Le programme Pr. 2 est destiné à effectuer une opération de transcodage à partir d'un signal numérique à 16 bits, de manière à obtenir un signal comprimé à 11 bits seulement, grâce à une loi de compression instantanée à 11 segments. Cette opération simule une voie audio numérique à 384 kbit/s. Le programme Pr. 3 est prévu pour réaliser les simulations sur le signal numérique issu du programme Pr. 2. Dans le cas de l'exemple choisi, il effectue le remplacement à des intervalles de temps définis d'un bit de moindre signification (LSB) par un autre faisant partie d'un mot code indiquant s'il y a lieu d'effectuer une opération de justification ou non. Le programme Pr. 4 effectue l'opération inverse de Pr. 2, c'est-à-dire qu'il reconvertis le signal comprimé en un signal linéaire à 16 bits qui est ensuite acheminé sur Pr. 5 et le convertisseur numérique-analogique.

Chaque pas de simulation modifie donc le contenu du signal entrant qui est ensuite mémorisé dans sa nouvelle forme. En principe, la procédure peut être appliquée pour d'autres programmes, tels que la simulation des erreurs de transmission en ligne, et reconstituer à l'aide du logiciel une situation réelle d'exploitation. L'utilisation d'un ordinateur programmé pour traduire avec précision des algorithmes de codage permet de les comparer en éliminant toutes sources d'erreurs dues à des structures matérielles différentes. Le jugement ne porte effectivement que sur le choix du logiciel.

7 Essais subjectifs d'écoute

Face à certains phénomènes liés à l'évolution des techniques numériques de compression des signaux, on ne

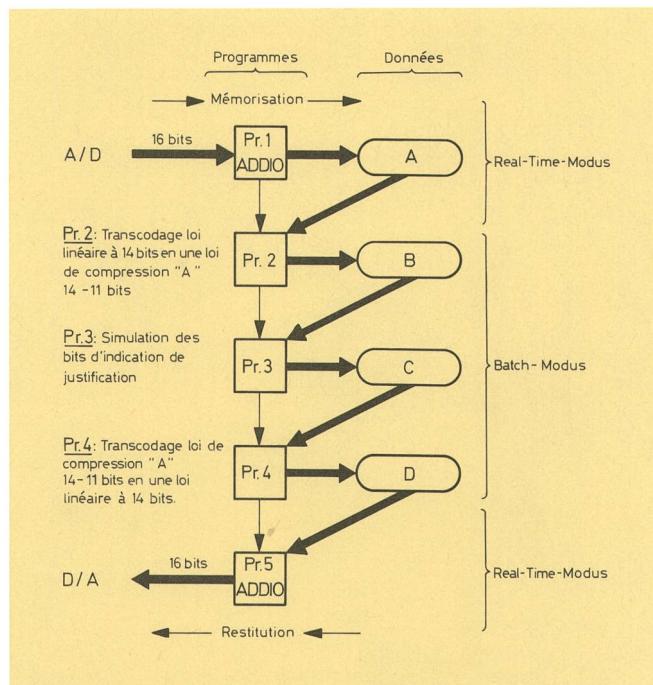


Fig. 11
Simulation par le logiciel

dispose pas toujours de références connues pour apprécier la qualité réelle d'un système de transmission. Il est donc indispensable, dans la plupart des cas, d'avoir recours à des essais auditifs à caractère nécessairement statistique pour compléter les mesures objectives faites avec des appareils habituels et permettre de tirer de l'ensemble des conclusions sûres [12].

A cet effet, les essais subjectifs d'écoute ont pour but d'évaluer à l'aide d'échelles de qualité et de dégradation le degré de gêne subjective engendrée par des défauts ou perturbations affectant le signal utile au cours des différentes opérations de codage-décodage. Selon la nature et l'objet des essais, on choisira quelle échelle d'évaluation subjective convient le mieux. Le *tableau II*

Tableau II. Echelle de qualité et de dégradation à cinq notes.

Echelle de qualité	Echelle de dégradation
5 Excellente	5 Imperceptible
4 Bonne	4 Perceptible mais non gênante
3 Assez bonne	3 Légèrement gênante
2 Médiocre	2 Gênante
1 Mauvaise	1 Très gênante

indique les deux échelles à cinq notes les plus couramment utilisées pour l'évaluation de la qualité intrinsèque d'un système, d'une part, et pour juger des effets subjectifs d'une dégradation, d'autre part.

Pour certaines catégories d'essais subjectifs, il peut être préférable d'utiliser une échelle dite de comparaison. Elle comprend les sept notes et appréciations suivantes:

- 3 bien meilleur
- 2 meilleur
- 1 légèrement meilleur
- 0 identique
- 1 légèrement plus mauvais
- 2 plus mauvais
- 3 beaucoup plus mauvais

Les essais doivent être organisés et conduits de manière à obtenir le maximum de renseignements sur certaines formes d'imperfections propres aux systèmes de codage de la parole et de la musique.

Parmi les méthodes d'essais d'écoute subjective les plus couramment utilisées pour la parole, citons la méthode des essais comparatifs par paires (paired-comparison procedure) de deux échantillons de parole A et B et la méthode classique qui fait appel à un jugement ou une note moyenne (mean opinion score) exprimé à l'aide des cinq notes de l'échelle de qualité. Ces deux méthodes ont été utilisées avec succès par les laboratoires qui ont participé, dans le cadre des travaux du CCITT, à la définition du meilleur algorithme MICDA pour le codage de la parole à large bande à 64 kbit/s. A cet effet, les procédures suivantes ont été appliquées.

71 Méthode des essais comparatifs par paire

Cette méthode consiste à comparer deux échantillons de parole A et B d'un même locuteur, mais qui ont été traités numériquement par deux systèmes différents de

codage. L'auditeur appelé à donner son appréciation indique s'il préfère l'échantillon A ou l'échantillon B. En général, on considère que cette procédure de test est plus sensible que les méthodes d'évaluation globale, mais elle a aussi l'inconvénient de conduire rapidement à un nombre de réponses assez élevé. En pratique, il peut être préférable de faire appel aux deux méthodes mentionnées ci-dessus, mais en diminuant le nombre des conditions à tester pour la méthode des essais comparatifs par paire. Les principales étapes qui conduisent au test subjectif proprement dit sont les suivantes:

- Préparation des sources sonores avec différents locuteurs dans des conditions d'enregistrement bien définies.
- Enregistrement numérique des signaux originaux traités (codage-décodage) par les systèmes à évaluer. Les signaux de sortie résultants sont encore mis dans un format adéquat et constituent le montage final de la bande magnétique utilisée pour les tests subjectifs.
- Conduite des séances d'écoute subjective.

A titre d'exemple, 32 auditeurs représentant plus de 10 000 réponses auront été nécessaires à l'évaluation statistique de quatre codecs MICDA à 64 kbit/s dans cinq conditions de débit binaire, de taux d'erreurs et de transcodage différentes.

72 Méthode des notes moyennes d'opinion

Dans ce cas, l'auditeur est convié à juger successivement un groupe d'échantillons constitués par deux courtes phrases, sans relation entre elles, lues par deux locuteurs. Lorsque le test porte sur quatre codecs par exemple, il est raisonnable d'utiliser 480 phrases différentes correspondant à 240 échantillons traités numériquement pour obtenir un degré de vérité suffisant. Dans cette hypothèse, les échantillons sont regroupés suivant les conditions de test en huit segments de 30 échantillons chacun qui seront ensuite présentés aux auditeurs dans un ordre correspondant à un carré latin (8×8). Avec 16 auditeurs au total, il est donc possible de recueillir 3840 réponses. L'utilisation de quelques échantillons d'une séquence musicale est parfois recommandé pour départager deux codecs aux qualités comparables avec les signaux de la parole.

73 Production des enregistrements numériques pour les tests subjectifs d'écoute

Le choix judicieux des sources sonores, la qualité de l'enregistrement original, les conditions d'écoute, les réglages du niveau de modulation, les auditeurs qui ont participé aux essais sont d'une grande importance lorsqu'il s'agit de comparer les résultats avec d'autres laboratoires. Il convient donc de définir avec exactitude tous les paramètres à respecter et d'indiquer d'une manière exhaustive dans quelles conditions les essais ont été réalisés.

731 Signaux musicaux

Les séquences de programme radiophonique doivent être choisies spécialement pour leur comportement très

critique à l'égard des défauts particuliers apportés par le système soumis aux essais. Le choix représentatif d'éléments de programme types dépend très souvent du genre de dégradation recherché. L'expérience a montré que les sources sonores suivantes sont particulièrement sensibles à la compression numérique du signal:

- Tous les signaux audiofréquences qui se caractérisent par une grande transparence avec un minimum d'harmoniques sont en général très sensibles au test subjectif du bruit de quantification et de granulation. Par exemple, le piano et la basse électronique sont parmi les instruments les plus critiques.
- Lorsqu'il s'agit d'évaluer la résistivité aux erreurs d'un système de transmission radiophonique numérique, il est recommandé d'utiliser comme source sonore de test les sons de la trompette ou du «Glockenspiel» qui comportent beaucoup d'harmoniques.
- On a observé que l'introduction de la préaccentuation/désaccentuation peut donner lieu à des dégradations inacceptables lorsque l'essai porte, par exemple, sur des cymbales.

732 Signaux de parole

Selon l'objectif précis qui a été fixé pour l'évaluation d'un procédé de codage de la parole, différentes formes et structures du langage parlé dans plusieurs langues sont utilisées. Lorsqu'il s'agit de déterminer les limites d'intelligibilité, il est parfois recommandé d'utiliser la PROSE qui est une manière de s'exprimer qui n'est pas assujettie, comme la poésie, aux lois d'une mesure et d'un rythme. De plus, le test doit avoir une certaine longueur. Différentes distorsions et leurs conséquences peuvent être beaucoup mieux cernées par le test de RIMES. Fairbank propose, à cet effet, un ensemble de syllabes qui ne se différencient que par un élément de la composition invariable consonne-voyelle-consonne (CVC). Par exemple, pour la langue allemande la variante C_{var} VC peut se présenter comme il suit:

Bach, Fach, wach, Dach, Sack, Mach, etc.

Pour les tests préférentiels, en particulier, on recommande d'utiliser des phrases dites *phonétiquement équilibrées*. P. Combescure propose un recueil de 20 listes de 10 phrases chacune dont les fréquences relatives pour chaque phonème reflètent celles qui existent dans la langue française [13], par exemple:

- Il se garantira du froid avec ce bon capuchon
- Annie s'ennuie loin de mes parents
- Les deux camions se sont heurtés de face, etc.

Il existe également des études sur la structure phonologique de l'allemand réalisées par J. Sotscheck.

Le test dit de «Berlin» est le plus connu, il comporte 12 groupes de phrases à huit phrases chacun donnant une bonne approximation de la distribution de fréquence des phonèmes dans la langue allemande, par exemple:

- Heute ist schönes Frühlingswetter
- Die Sonne lacht
- Am blauen Himmel ziehen die Wolken, etc.

Pour l'anglais, signalons la liste «1965 Revised List of Phonetically Balanced Sentences (Harvard Sentences)».

Ces listes sont en général employées pour les tests de qualité et pour les jugements par isopréférence ou par catégories de dégradation.

En pratique, l'enregistrement des voix doit être réalisé avec des locuteurs masculins et féminins ayant un timbre de voix sensiblement différent.

733 Enregistrement numérique

Dans le domaine du codage de la parole à large bande comme dans celui du son de haute qualité, il est indispensable de recourir à l'enregistrement numérique pour effectuer les tests d'évaluation subjective. La qualité obtenue dépasse largement celle des systèmes analogiques. Les avantages apportés par cette nouvelle technique sont liés à l'amélioration des paramètres de qualité suivants:

- Le rapport signal/bruit est supérieur à 85 dB.
- Les fluctuations de vitesse et l'effet d'écho sont totalement éliminés.
- La distorsion harmonique et l'intermodulation ne sont introduites que par les circuits de codage et de décodage.

Toutefois, c'est au niveau de la copie que le magnétophone numérique offre la caractéristique la plus intéressante. En effet, lors d'un montage de séquences sonores ou lors de la simulation de plusieurs codecs connectés en tandem, il est parfois indispensable de faire des copies, sans pour autant passer par l'intermédiaire analogique et sans effectuer de nouveaux traitements. Cela impose en particulier d'utiliser la même fréquence d'échantillonnage pour les deux machines et de n'effectuer aucune correction de gain, de tonalité, etc. Dans ces conditions, le signal ne subit rigoureusement aucune dégradation tant que les procédés de correction contre les erreurs mis en œuvre dans tous les magnétoscopes numériques fonctionnent correctement.

Le tableau III indique les principales caractéristiques d'un magnétoscope numérique.

Tableau III. Caractéristiques principales d'un magnétoscope numérique

MIC	2 voies
Fréquence d'échantillonnage	48,50,4 kHz
Résolution	16 bits linéaire
Bande de fréquence	20 Hz...20 kHz
Dynamique	90 dB
Facteur de distorsion	0,05 %
Diaphonie	80 dB
Niveau d'entrée	+6 dBm
Niveau de sortie	+6 dBm
Vitesse de bande	38 cm/s
Largeur de la bande	6,3 mm

En ce qui concerne la console de prise de son, il semble qu'à court terme elle ne soit pas directement concernée par les nouvelles techniques numériques, car la fonction de réglage analogique du gain sera toujours nécessaire avant le processus de codage.

74 Traitement par ordinateur des enregistrements

Pour l'enregistrement de la parole, il est très difficile, voire impossible, de produire, au cours du premier enregistrement, une bande originale qui puisse être utilisée sans retouche, correction ou arrangement pour les séances de tests subjectifs. En général, les trois opérations suivantes se révèlent indispensables lors de la préparation des bandes de test:

- Réglage du niveau efficace de la parole.
- Suppression de certains défauts lors des séances d'enregistrement.
- Arrangement des échantillons de parole dans un ordre bien défini.

Toutes ces opérations peuvent être effectuées avec l'ordinateur grâce aux possibilités pratiquement illimitées du logiciel de traitement numérique du signal.

741 Mesure et réglage du niveau de la puissance vocale

La notion de niveau électrique d'un signal non permanent est toujours partiellement ambiguë, car il existe autant de résultats différents que de constantes de temps du réseau intégrateur. En pratique, on utilise deux catégories d'instruments de mesure du niveau sonore. La première comprend les appareils dont la priorité est donnée au niveau subjectif, comme le Vu-mètre à constantes de temps assez longues, et la seconde regroupe tous les appareils donnant la priorité au niveau de crête du signal.

Les recherches se sont toutefois orientées vers un appareil plus spécifique à la parole qui mesure le niveau de la puissance vocale dans les parties actives d'une conversation. Ce niveau se calcule au moyen d'un algorithme qui a été établi à partir de la distribution des niveaux de puissance du signal vocal. Exception faite des exemplaires de laboratoire, peu d'appareils de ce type sont disponibles sur le marché. Toutefois, le problème peut être résolu d'une façon élégante, en mettant en œuvre un logiciel qui se charge de calculer la valeur efficace des différents échantillons de parole enregistrés numériquement puis, si nécessaire, de les corriger de manière que le niveau de la puissance vocale de chaque phrase demeure constant. Par définition, le niveau efficace de la parole est de 15 dB inférieur à celui d'un signal sinusoïdal. Compte tenu d'une marge de protection de 1 dB, il y a donc lieu de régler le signal de parole à 19 dB au-dessous du point de surcharge du codeur.

Le choix de la limite entre les parties actives et les pauses d'une conversation doit être soigneusement défini. Les pauses d'une certaine longueur ainsi que celles entre les phrases doivent être exclues de la mesure du niveau efficace de la parole. Par contre, on doit inclure dans le calcul les brefs silences qui font partie de la structure d'un mot. Par exemple, dans la prononciation usuelle du mot «acte», il apparaît un silence d'environ 200 ms entre le «c» et le «t», qui doit être pris en compte, car il est perçu par un auditeur comme un son continu.

742 Montage des bandes de test et correction des défauts d'enregistrement

En général, les méthodes proposées pour l'évaluation subjective de la qualité de la parole exigent souvent un arrangement bien défini des phrases dites par des locuteurs différents. Pour des raisons pratiques évidentes, cet arrangement ne correspond pas toujours à celui réalisé à l'occasion de la séance d'enregistrement au studio. Dès lors, il est indispensable de procéder à un montage des bandes de test et de recourir à la copie numérique à l'aide de deux enregistreurs compatibles. Bien qu'il n'en résulte aucune dégradation supplémentaire du signal, cette opération comporte deux inconvénients majeurs: le premier résulte du prix d'acquisition encore fort élevé des enregistreurs numériques professionnels et le second tient au temps nécessaire à la recherche des phrases sur la bande originale et à la réalisation du montage des bandes de test. Toutes ces opérations de recherche et d'assemblage peuvent être réalisées automatiquement avec l'aide du logiciel. Celui-ci se charge de disposer les phrases dans un certain ordre et de garantir, à l'échantillon près, la valeur des espaces entre deux phrases consécutives et pauses quelconques. Il est évident que pour préserver la qualité du signal de parole, l'enregistrement original doit être réalisé directement sans l'intermédiaire d'un enregistreur numérique. D'autre part, au cours des séances d'enregistrement parfois longues, voire pénibles, comportant la lecture d'un grand nombre de phrases, on ne peut pratiquement pas éviter que des bruits parasites provenant de l'environnement de la salle ou même des erreurs discrètes de langage passent inaperçus, mais apparaissent à l'occasion du contrôle final de la bande originale. Dans des conditions de test sévères, ces défauts mineurs peuvent, malgré tout, se révéler gênants pour l'auditeur participant à l'évaluation subjective et, par ailleurs, pour des raisons pratiques, il n'est pas toujours possible de recommencer une nouvelle séance d'enregistrement. C'est aussi dans ces conditions particulières que le système ADDIO est une fois de plus un outil très performant. L'exemple qui suit permet d'illustrer la remarquable souplesse et l'efficacité du système. Il est tiré des nombreux travaux de préparation des bandes de test pour l'évaluation subjective des codecs MICDA 64 kbit/s

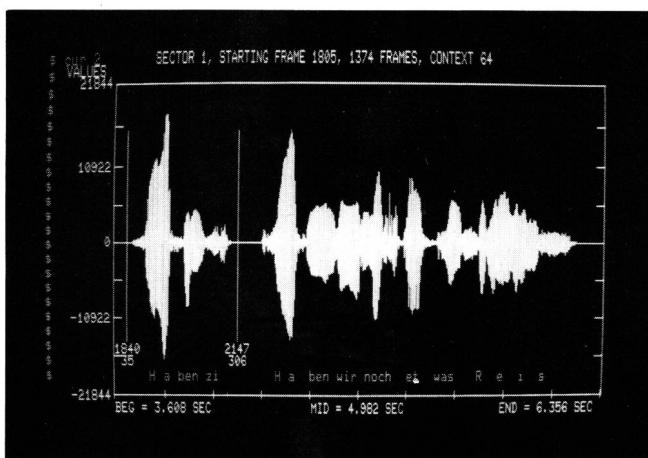


Fig. 12
Signal de parole à corriger
En encadré, la partie à effacer

à large bande (7 kHz), au cours desquels il a été nécessaire d'apporter quelques corrections à la bande d'enregistrement originale. Le signal de parole représenté dans la partie encadrée de la figure 12 doit être éliminé, car il correspond à une erreur de langage ou plus précisément à la prononciation inachevée d'un mot. La localisation de cette partie du signal passe naturellement par une écoute attentive de la phrase à corriger, suivie d'un repérage visuel de l'erreur sur l'écran d'un terminal. Lorsque les limites du signal à éliminer sont connues et que le choix de la zone de raccordement est précisée, il ne reste plus qu'à effacer le signal en erreur à l'aide d'un curseur que l'on déplace sur l'écran pour déterminer exactement le début et la fin des domaines du signal à traiter. La figure 13 montre le résultat final obtenu à l'issue de ces opérations qui, pour l'auditeur, se traduira par une écoute exempte de défaut.

75 Propriétés acoustiques des salles d'écoute [14, 15]

En général, lorsqu'il s'agit de connaître les limites des possibilités techniques d'un système, les séquences de programme utilisées pour les essais subjectifs doivent être exemptes de tout défaut audible. On s'efforcera donc de réduire au maximum l'effet de masque produit dans la salle d'écoute par le bruit ambiant ou par d'autres fréquences parasites. Les propriétés acoustiques de la salle d'écoute sont d'une très grande importance et la forme de la salle, surtout lorsqu'il s'agit de petites pièces, peut engendrer quelques problèmes sérieux de réverbération si les proportions ne sont pas judicieusement choisies. L'expérience acquise en matière d'acoustique des salles d'écoute en radiodiffusion permet de recommander le respect des principales caractéristiques physiques et acoustiques suivantes:

- Volume de la salle:
40 m³...120 m³, mais de préférence 120 m³ ± 30 m³
- Forme de la salle:
le rapport optimal entre les côtés de la salle est 1,9:1,4:1,0. Le rapport 1/2 entre deux dimensions quelconques de la salle est à éviter.
- Temps de réverbération:
250 ms...400 ms (au-dessus de 100 Hz)

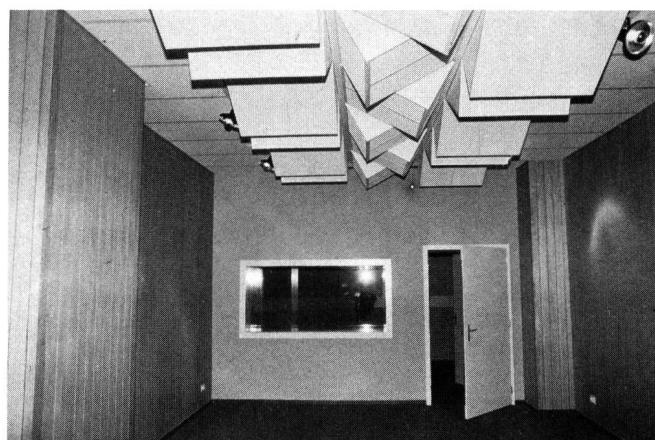


Fig. 14
Vue de la nouvelle salle d'acoustique des PTT

- Niveau de bruit:

il doit être suffisamment bas pour éviter de masquer les signaux les plus faibles. L'objectif le plus souvent retenu est de ramener le bruit provenant de l'extérieur à un niveau inférieur à 25 dB pondéré selon la courbe dite «A».

La figure 14 montre la nouvelle salle d'acoustique de la Division des recherches et du développement des PTT en cours de construction. Elle repose entièrement sur des coussinets de 3,5 cm d'épaisseur d'une matière synthétique qui absorbe efficacement les vibrations et fréquences basses au-dessus de 30 Hz. Les spécifications de ce studio d'écoute sont très sévères, avec pour objectif un niveau de bruit de fond inférieur à 20 dBA. L'isolement acoustique placé sur le pourtour de la salle a été calculé de manière à ramener les perturbations provenant d'autres zones à un niveau inférieur à celui qui existe dans la salle en l'absence de perturbations, et cela dans toutes les bandes de fréquences. L'ensemble des équipements d'enregistrement et de reproduction, les instruments de mesure et les unités de calcul, processeurs, etc., qui représentent une source de bruit non négligeable, sont groupés dans une salle de régie soigneusement isolée phoniquement de la salle d'écoute.

8 Conclusions

Les techniques de codage et de compression décrites se rapportent plus ou moins directement à un projet de recherches ou à une étude particulière des PTT dans le domaine des télécommunications. Ces travaux ne peuvent se concevoir sans le concours de l'informatique, c'est-à-dire de l'ordinateur spécialement organisé pour permettre l'acquisition et la restitution des signaux de la parole et de la musique en temps réel. A cet effet, l'AD-DIO s'est rapidement révélé comme un outil d'une flexibilité remarquable lorsqu'il s'agit de mettre au point une méthode de compression ou de modifier un algorithme de codage.

Le soin apporté à l'enregistrement des sources sonores, au choix des méthodes de tests subjectifs et aux conditions d'écoute sont de la plus grande importance, pour se garantir contre les erreurs d'appreciation et de jugement imputables à des distorsions étrangères au système de codage proprement dit.

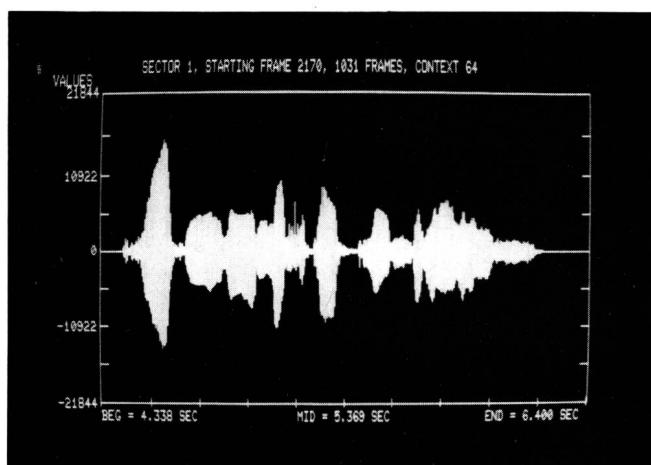


Fig. 13
Signal de parole après la correction

Bibliographie

- [1] Flanagan J. L., Schroeder M. R., Atal B. S., Crochière R. E., Jayant N. S., Tribolet J. M. Speech Coding IEEE Trans. on Communications. New York, COM-27 (1979), p. 710.
- [2] Boddie J. R., Johnston J. D., Mc Gonagle C. A., Upton J. W., Berkley D. A., Crochière R. E., Flanagan J. L. Adaptive Differential Pulse-Code-Modulation Coding. Bell system technical journal, Murray Hill 60 (1981) 7, p. 1547.
- [3] Bellanger M. Traitement numérique du signal. Paris 1981.
- [4] Le Guyader Gilloire A. Codage différentiel adaptatif de la parole pour le réseau téléphonique. 7^e colloque du GRETSI, Nice 1979.
- [5] Galand C., Esteban D., Menez J. Techniques de codage de la parole à débit moyen (5 à 16 kbit/s). L'onde électrique, Paris 61 (1981), 8/9.
- [6] Esteban D., Galand C. Application of Quadrature Mirror Filters to Splitband Voice Coding Schemes. Int. Conf. on Acoustics, Speech and Signal Processing. Hartford, 1977.
- [7] Raulin J. M., Bonnerot G., Jeandot J. L., Lacroix R. A 60 Channel PCM-ADPCM Converter. IEEE Trans. on Communications, New York, Com. 30 (1982) 4.
- [8] Mury R. Progrès réalisés dans le domaine de la transmission numérique des programmes radiophoniques. Bull. techn. PTT, Berne 56 (1978) 10, p. 376.
- [9] De Lamare E. Le codage numérique des signaux sonores de haute qualité: propositions pour un système entièrement compatible avec les normes téléphoniques. Revue de radiodiffusion-télévision, Paris (1976) 44, p. 12.
- [10] Emerard F. Les diphones et le traitement de la prosodie dans la synthèse de la parole. Bulletin de l'Institut de Phonétique de Grenoble, vol. VI, 1977, p. 103.
- [11] Courbon J. L. Les systèmes de production de la parole. Ecole Nationale Supérieure des Télécommunications, formation continue, Paris 1984.
- [12] CCIR Méthode pour l'évaluation subjective de la dégradation due aux caractéristiques des systèmes pour transmission radiophonique. Programme d'études 19A/CMTT, vol. XII, Genève 1982.
- [13] Combescure P. 20 listes de dix phrases phonétiquement équilibrées. Revue d'Acoustique (1981) 56.
- [14] CCIR Proposition de révision du rapport 797, doc. 10/5023. Commissions d'études du CCIR, période 1978-1982.
- [15] Propriétés fondamentales d'une salle d'écoute et de vision destinée à permettre l'évaluation subjective des essais du traitement des signaux sonores et de télévision. DG PTT, Berne. Rapports de la Division des recherches et du développement VL 24. 172U, VD15. 1141U, 1983.