



②
ETUDE D'UN NOUVEAU PRINCIPE DE
COMPRESSION DYNAMIQUE INSTANTANEE DU
SIGNAL DE PAROLE:
APPLICATION A LA RADIO-MOBILE

Rapport final

In
J
B I
In

CENTRE DE RECHERCHES SUR LES COMMUNICATIONS

FACULTÉ DES SCIENCES APPLIQUÉES

UNIVERSITÉ DE SHERBROOKE

TÉL.: 819-565-4450

TÉLEX 05-836149

P
91
C655
S639
1983

SHERBROOKE, QUÉBEC, CANADA, J1K 2R1

checked 11

P
91
C655
S639
1983

CENTRE DE RECHERCHES SUR LES COMMUNICATIONS

Faculté des sciences appliquées

Université de Sherbrooke

②
ETUDE D'UN NOUVEAU PRINCIPE DE
COMPRESSION DYNAMIQUE INSTANTANEE DU

SIGNAL DE PAROLE:

APPLICATION A LA RADIO-MOBILE

Industry Canada
LIBRARY
JUL 20 1998
BIBLIOTHEQUE
Industrie Canada

Rapport final

pour le

Ministère des Communications
Ottawa, Ontario
contrat no. OST82-00083

Dans le cadre du programme des
Centres d'Excellence

30 mars 1983

COMMUNICATIONS CANADA
OCT 1 1984
LIBRARY - BIBLIOTHEQUE

Responsable du projet:

Joël Soumagne, ing.

Sarto Morissette, ing.
Directeur CRCS

P
91
C655
S639
1983

DD4807242
DL4819825

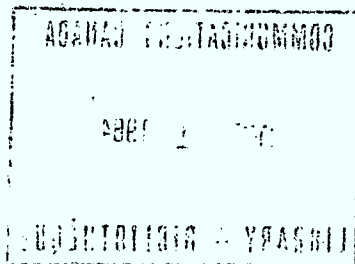


TABLE DES MATIERES

PREAMBULE -	1
CHAPITRE 1- PRINCIPE DE CODAGE NUMERIQUE.....	3
1. Transposition de signal - Principe de base.....	3
1.1 Modèle de codage.....	4
1.2 Algorithme de codage.....	5
1.2.1 Cercle de codage normalisé.....	5
CHAPITRE 2 - ORGANISATION DE L'ALGORITHME DE CODAGE.....	8
1. Principe de l'algorithme de codage.....	9
1.1 Normalisation des paramètres.....	10
1.1.1 Calcul des paramètres "temps et gain" normalisés...	15
1.1.2 Conditions de codage - Tests de décision.....	18
1.2 Détermination de l'angle $\phi_n(t)$	23
1.2.1 Cas particulier des silences.....	23
1.2.2 Cas des faibles variations d'angle.....	24
1.3 Quantification de l'angle $\phi_n(t)$	26
1.3.1 Algorithme de la K-moyenne.....	26
1.3.2 Débit numérique.....	28
1.4 Quantification du temps.....	28
1.5 Organigramme de codage.....	30
CHAPITRE 3 - ORGANISATION DU DECODEUR.....	32
1. Mémoire du décodeur	33
2. Organe de calcul du décodeur.....	34
2.1 Initialisation: calculs des paramètres adaptatifs du codage	36
2.2 Reconstitution du signal à partir de l'angle ϕ_n	37
2.2.1 Rappel des calculs à la réception.....	38
2.2.2 Adaptation et normalisation des paramètres.....	40
2.3 Cas particulier des silences ou bruit de fond.....	42
2.4 Réinitialisation des paramètres de codage.....	43
2.5 Organigramme de décodage.....	44

CHAPITRE 4 - RESULTATS DE CODAGE NUMERIQUE.....	46
1. Performance de basé du codage - qualité-débit.....	46
1.1 Quantification linéaire à 12 bits.....	48
1.2 Quantification linéaire à 8 bits.....	48
1.3 Quantification à 6 bits linéaire et optimalisée par une K-moyenne.....	49
1.3.1 Quantification linéaire.....	49
1.3.2 Quantification optimale.....	50
1.3.3 Conclusion.....	51
1.4 Quantification à 4 bits linéaire et optimalisée par une K-moyenne.....	52
1.4.1 Quantification linéaire et moyenne.....	52
1.4.2 Quantification optimale (K-moyenne) à 4 bits, qualité globale de codage.....	52
2. Effet de la taille de la mémoire sur qualité.....	53
3. Caractéristiques du signal transmis $\phi_n(t)$	57
3.1 Analyse du spectre et de l'autocorrélation du signal.....	57
3.2 Signal $\phi_n(t)$ "alterné".....	65
CHAPITRE 5 - COMPRESSION DE DYNAMIQUE - TRANSMISSION ANALOGIQUE DE $\phi_n(t)$ -.....	73
1. Signal analogique - simulation.....	73
2. Procédé de synchronisation du récepteur avec l'émetteur.....	76
2.1 Codage particulier en période de silence: $\phi_n =$ constante	77
2.2 Synchronisation initiale en période de silence.....	78
2.2.1 Application au signal continu analogique $\phi(t)$	79
2.2.2 Utilisation des silences pour la synchronisation...	81
2.3 Spectres alternés et synchronisation.....	90
2.3.1 Spectre de la séquence artificielle $\{\pm\phi_n\}$ alternée	90
2.3.2 Spectre des signaux $\phi(t)$ alternés en bande réduite 6,4 KHz.....	90
3. Effet du bruit.....	100
CHAPITRE 6 - CODAGE SPATIO-TEMPOREL	101
1. Principe général du codage spatio-temporel.....	101
1.1 Principe de base.....	103
1.2 Choix du représentant dans la fenêtre.....	105
1.3 Rôle du filtrage.....	105

2. Application du principe - Essais de codage.....	106
2.1 Rôle du filtrage.....	107
2.2 Résultats du codage préliminaire.....	108
CONCLUSION GENERALE.....	113
ANNEXE - CASSETTE ENREGISTREE.....	115

PREAMBULE

La présente étude de traitement numérique du signal de parole a initialement été envisagée dans le but de développer un nouveau procédé de codage (par quantification numérique) de la parole. La transmission du signal est alors envisagée elle aussi sous forme numérique.

Cependant, grâce à l'algorithme inventé dans ce procédé, il est apparu une ouverture nouvelle de traitement de signal. En effet la "transposition" du signal effectué, conduit, sous réserve d'un algorithme approprié, à la définition d'un nouveau signal, toujours numérisé (c'est-à-dire échantillonné et quantifié), et qui présente une particularité très intéressante: l'absence quasi totale de fluctuations dynamique de niveau. Or la parole habituelle nécessite au moins 60 dB de dynamique, ce qui représente un signal peu efficace lorsque l'on envisage les modulations analogiques (modulation de fréquence "MF", modulation à bande latérale unique "SSB-AM"). On se propose donc dans le rapport d'étude de présenter:

En première partie: Les concepts généraux de traitement de signal, avec une description détaillée de l'algorithme choisi; de plus l'orientation d'étude retient ici surtout les aspects de codage - quantification - transmission totalement numériques. Des résultats intéressants au niveau du compromis classique qualité/débit numérique dégagent une orientation d'étude à compléter vers un développement de ce nouveau mode de codage numérique. Il sera très utile de vérifier jusqu'à quel débit minimal on peut envisager une transmission avec une qualité encore satisfaisante. Cette partie est composée des chapitres 1 à 4.

En deuxième partie: Un développement spécifique à l'application de compression de dynamique. Bien sûr, le traitement de signal lui-même sera totalement numérique: seulement, après conversion en analogique, on générera un signal continu, qui sera modulé (FM ou SSB-AM), puis transmis. Le récepteur distant, devra, préalablement ré-échantillonner le signal démodulé. Il apparaît alors une difficulté certaine de "synchronisation" entre les trames d'échantillonnage au niveau de l'émetteur (codeur) et du récepteur (décodeur). Des idées de solutions sont proposées et ont été simulées dans les conditions de laboratoire. L'influence du bruit sur ce signal est aussi abordée. Une méthode permettant une modification du spectre du signal analogique transmis est suggérée: cela permet d'utiliser au mieux la bande téléphonique 300-3300 Hz allouée à la transmission du signal analogique. C'est l'étude présentée au chapitre 5.

Une troisième partie: Sensiblement liée aux idées des travaux précédents a aussi été abordée en parallèle sur cette étude. Le concept d'échantillonnage à rythme variable est retenu pour une application dite de "codage numérique spatio-temporel". Il s'agit là de sélectionner dans une "fenêtre d'échantillonnage" adaptative (en temps et en amplitude) le meilleur représentant pour le signal codé tout en imposant, une bonne qualité de codage et le débit minimal. L'avancement de l'étude ne permet pas une conclusion définitive sur cette technique, mais laisse voir des possibilités intéressantes pour obtenir des bas débits, particulièrement dans un contexte de transmission d'un ensemble de signaux de parole sur des voies téléphoniques. Le chapitre 6 présente cette technique de base.

CHAPITRE 1

PRINCIPE DE CODAGE NUMERIQUE

Dans ce chapitre, nous envisagerons une présentation de l'algorithme de codage numérique et de ses contraintes. Les résultats, pour divers débits numériques inférieurs à 64Kbit/s (référence: débit PCM 8 bits => 64 Kbit/s sur ligne téléphonique actuelle), sont présentés sous forme d'évaluation en rapport Signal sur Bruit calculé, et enregistrés sur cassette en comparaison avec le signal de référence et le codage PCM (ou MIC) conventionnel.

1. Transposition de signal - Principe de base

Dans le rapport intermédiaire (octobre 1982), les idées de bases ont été présentées. Depuis le concept est resté essentiellement le même mais l'algorithme a évolué vers une bien meilleure qualité subjective de la parole obtenue par les codages effectués. On rappellera donc les éléments de départ:

- Générer un signal intermédiaire $\phi_n(t)$ numérique,
- Prélever ces échantillons à un rythme variable et les mettre en mémoire,
- Transmettre ceux-ci à taux fixe de 8 KHz après une quantification avec un code restreint à m bits par échantillon,
- (Ultérieurement: deuxième partie) Effectuer un retour à un signal analogique $\theta(t)$ après conversion D/A du message original $\phi_n(t)$ constitué au rythme imposé de 8 KHz.

1.1 Modèle de codage

On rappellera sur la figure 1.1 le principe de la définition, développée dans le rapport intermédiaire:

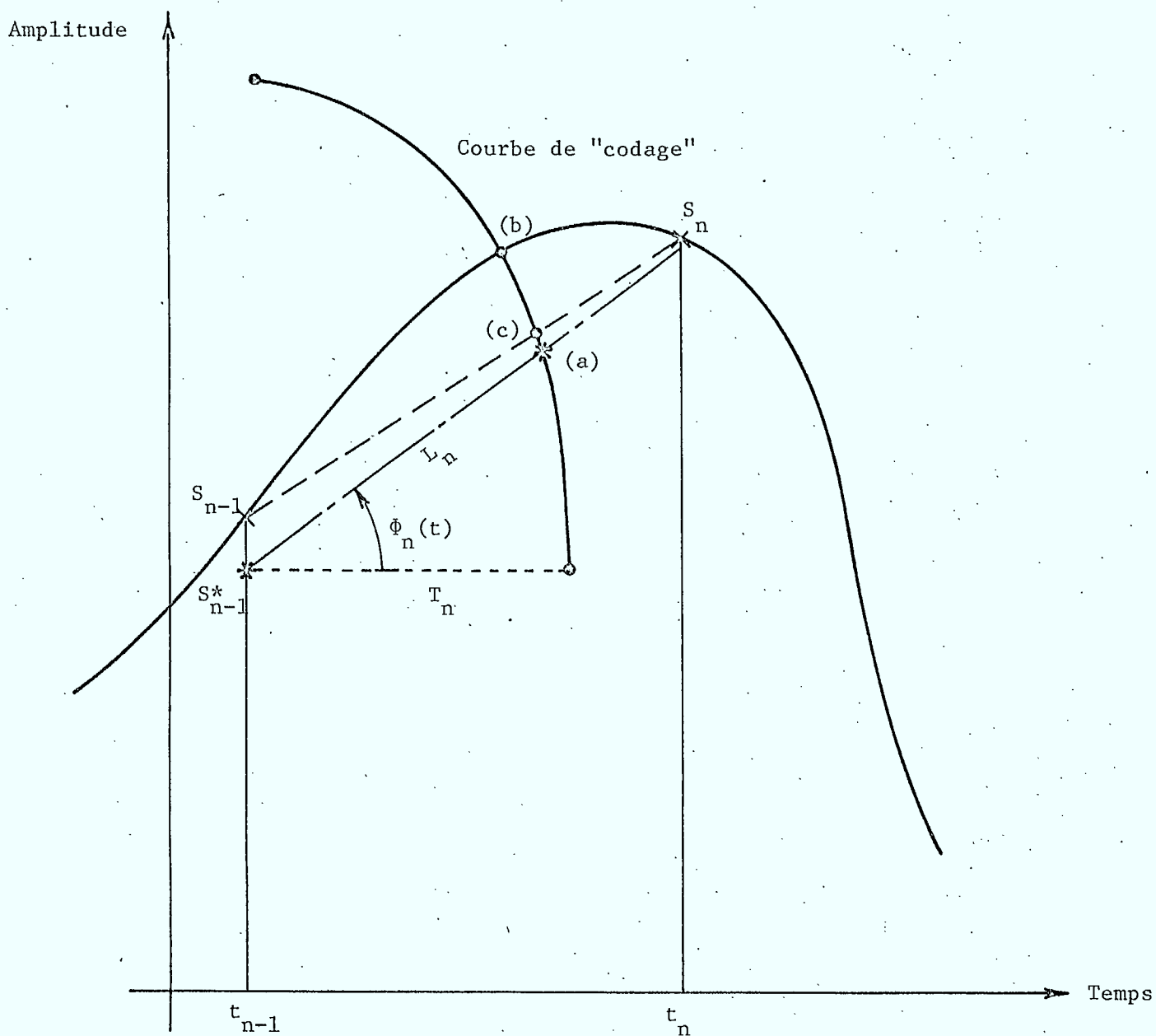


Fig. 1.1 Modèle de codage proposé : définition d'un nouveau signal $\phi_n(t)$.

- d'un angle de codage $\phi_n(t)$,
- d'une courbe associée de codage.

Sur cette figure les * représentent des échantillons codés/décodés à un rythme variable suivant une progression dans le temps:

$$T_n = L_n \cos \phi_n$$

On conçoit donc que selon ce format de codage:

- on prélève une information $\phi_n(t)$
 - à rythme variable T_n ,
 - d'amplitude bornée $\left] \frac{-\pi}{2} \frac{+\pi}{2} \right[$

1.2 Algorithme de codage

L'algorithme est conçu en fonction d'une distribution de valeurs discrètes de l'angle ϕ_n réparties sur une "courbe de codage" associée à chaque échantillon prélevé (angle ϕ_n) et adaptée à chaque nouvel instant de prélèvement T_n . (Rappelons que T_n n'est pas égal au temps de base T_0 : 125 μ s ou période d'échantillonnage de 8 KHZ du signal de parole simultanément échantillonné sous forme numérique).

1.2.1 Cercle de codage normalisé

Le logiciel de traitement de signal (version I.L.S.: Interactive Laboratory System, développé en une version plus élaborée ILS-US: Université de Sherbrooke), permet une visualisation continue du processus de codage.

La mise au point de l'algorithme est essentiellement tributaire de ce logiciel exploité en parallèle avec les possibilités d'écoute de parole continue: phrases dites tant par des voix d'hommes que de femmes.

Les fichiers de parole testés sont issus d'un laboratoire spécialisé dans l'évaluation de qualité de parole. Ces phrases, en français, sont phonétiquement équilibrées, ce qui assure au traitement proposé un seuil de validité pour l'ensemble de la parole continue.

Dans le rapport intermédiaire, il avait été suggéré une normalisation de l'écart d'amplitude entre échantillons consécutifs et aussi du temps. Cette fonction de normalisation, à partir du signal transmis $\phi_n(t)$ s'est avéré donner des résultats satisfaisants dans le comportement de l'algorithme développé.

Ainsi dans l'espace à deux dimensions normalisées, la courbe de codage demeure un arc de cercle unitaire, et l'angle $\phi_n(t)$ est calculé sur la figure 1.2 par

$$\phi_n = \text{arctg} \frac{E_n}{\tau_n}$$

où E_n est l'écart normalisé d'amplitude,
 τ_n est l'écart de temps normalisé,
 r_{n-1}^* est le dernier échantillon (passé) codé et décodé,
 S_n est l'échantillon du signal de parole suivant à coder.

Dans l'opération de dénormalisation, l'évolution du signal reconstitué dépendra d'une courbe de codage déformée en ellipse dans une représentation géométrique associée au logiciel d'exploitation. Cette présentation est celle qui apparaîtra sur les figures décrivant les résultats du codage par "temps élastique", qui seront incluses dans les chapitres ultérieurs.

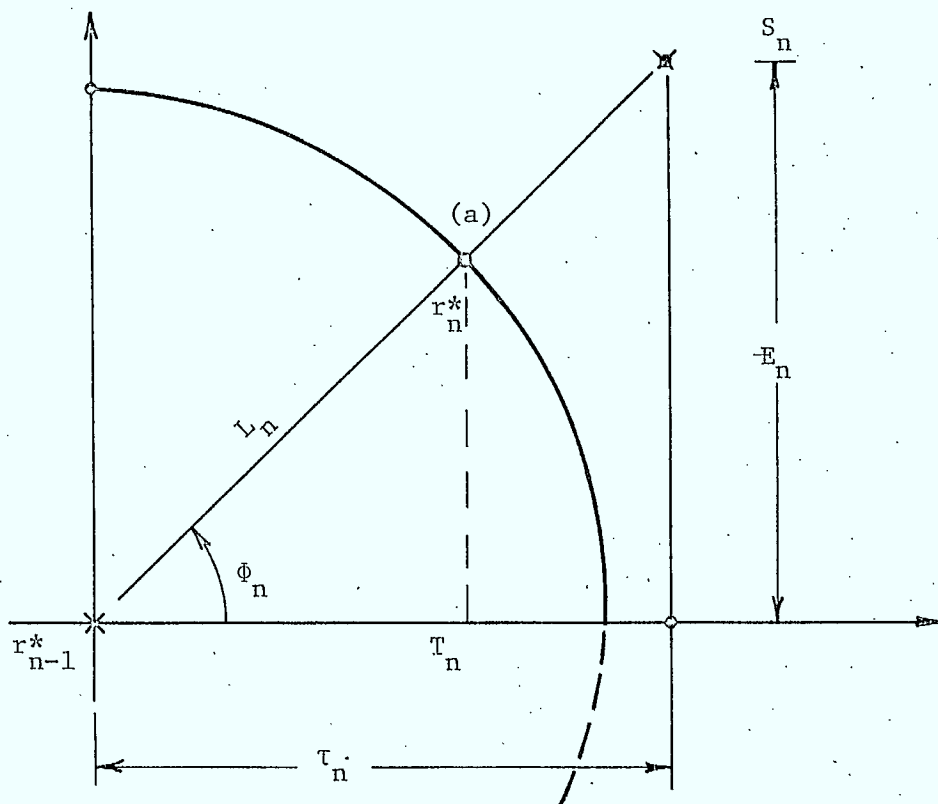


Fig. 1.2 Cercle de codage normalisé.

CHAPITRE 2

ORGANISATION DE L'ALGORITHME DE CODAGE

Le signal de parole, initialement analogique, c'est-à-dire continu, est préalablement filtré dans une bande limitée puis numérisé par un convertisseur limité à 12 bits de quantification linéaire.

Deux options de filtrage d'entrée et de fréquence d'échantillonnage seront suggérées dans cette étude pour des raisons développées par ailleurs (chapitres 3 et 5).

- La première option retient le filtrage téléphonique 300-3300 Hz et la fréquence d'échantillonnage conventionnelle de 8 KHz.
- La seconde option consiste en une réduction, et de la bande utile à 300-3000 Hz, et de la fréquence d'échantillonnage à 6400 Hz seulement.

Une banque de phrases (10) phonétiquement équilibrées, prononcées par 6 personnes différentes choisies pour leur équilibre de timbre (3 hommes, 3 femmes) constitue un fichier source stocké sur disque sous deux formats différents d'échantillonnage (8 KHz et 16 KHz). Le filtrage en bande réduite est réalisé par un filtre numérique simulé par nos soins.

Le logiciel ILS (Interactive Laboratory System) autorise la lecture et l'analyse par blocs: ici le format retenu pour les illustrations est celui d'un bloc ou fenêtre de 32 échantillons (c'est-à-dire 4 millisecondes de signal à 8 KHz d'échantillonnage...).

1. Principe de l'algorithme de codage

La lecture d'un bloc de 32 échantillons est effectuée puis suivie d'une analyse continue point à point.

Le fichier source correspond à une progression "en temps naturel" de 125 μ s par échantillon (à 8 KHz qui est l'exemple type retenu). Le fichier intermédiaire contenant l'information codée et transmise au récepteur sera constitué de blocs de N mots codés mais avec N égal, supérieur ou inférieur à 32 étant donné la progression dite "en temps élastique", ce qui correspond pour chaque mot à une détermination par calcul (au niveau du décodeur) et à chaque mot reçu d'un nouveau temps appelé "temps réel" du codage.

La différence de progression en temps entre le signal original de parole (temps naturel) et le signal intermédiaire codé/décodé (temps réel) impose la présence d'une "mémoire tampon". Cette mémoire sera remplie au rythme imposé par l'accumulation d'informations (temps réel plus court que le temps naturel), et vidée progressivement dans la condition opposée et maintenue à un niveau de remplissage stable quand les deux progressions se font au même rythme.

Il faut comprendre dès lors que:

- Le signal intermédiaire, stocké partiellement dans cette mémoire sera lui transmis au rythme numérique fixe de 8 KHz avec un retard de transmission correspondant à la taille de la mémoire.
- En cas de transmission analogique, c'est ce fichier, reconstruit à 8 KHz, qui est préalablement passé par un convertisseur numérique/analogique.

- Dans les deux cas ci-dessus, l'information reconstituée à un rythme de 8 KHz d'horloge (temps naturel), représente en réalité pour le décodeur distant un message à reconstruire à rythme numérique variable (temps réel), ce qui impose une mémoire de stockage au récepteur identique à celle du codeur. Cependant le délai total imposé correspond à la capacité de l'une de ces mémoires (codeur ou décodeur comme l'illustre la figure 2.1).

1.1 Normalisation des paramètres

Comme il l'a déjà été décrit dans le rapport d'étape (octobre 1982), le principe de codage est basé sur une transposition de signal.

Le nouveau signal représente alors un angle de codage entre le dernier échantillon connu par le décodeur r_{n-1}^* (signal de parole codé puis décodé) et le premier échantillon suivant pour la parole S_n .

La figure 2.2 rappelle le principe général proposé.

L'algorithme simple (rapport 1, page 13) alors proposé, s'est avéré donner un fonctionnement efficace et on retient les formules d'adaptation d'un facteur de gain et d'un temps de codage T_n associé à un rayon de codage L_n de la courbe associée:

$$S_n \text{ normalisé} = \frac{s_n}{G_n}$$

avec

$$G_n = G_{n-1} \cdot \text{tg}(\phi_n)$$

et

$$L_n = L_{n-1} / \text{tg}(\phi_n)$$

Soit alors $T_n = L_n \cos(\phi_n)$.

Ainsi dans les cas extrêmes où par exemple l'angle ϕ_n devient grand ($\rightarrow \pi/2$) et qui correspond au cas de la figure 2.3.

Signal à fréquence f_e
d'échantillonnage

a : remplissage si f_e supérieur à 8 KHz

b : vidage si f_e inférieur à 8 KHz

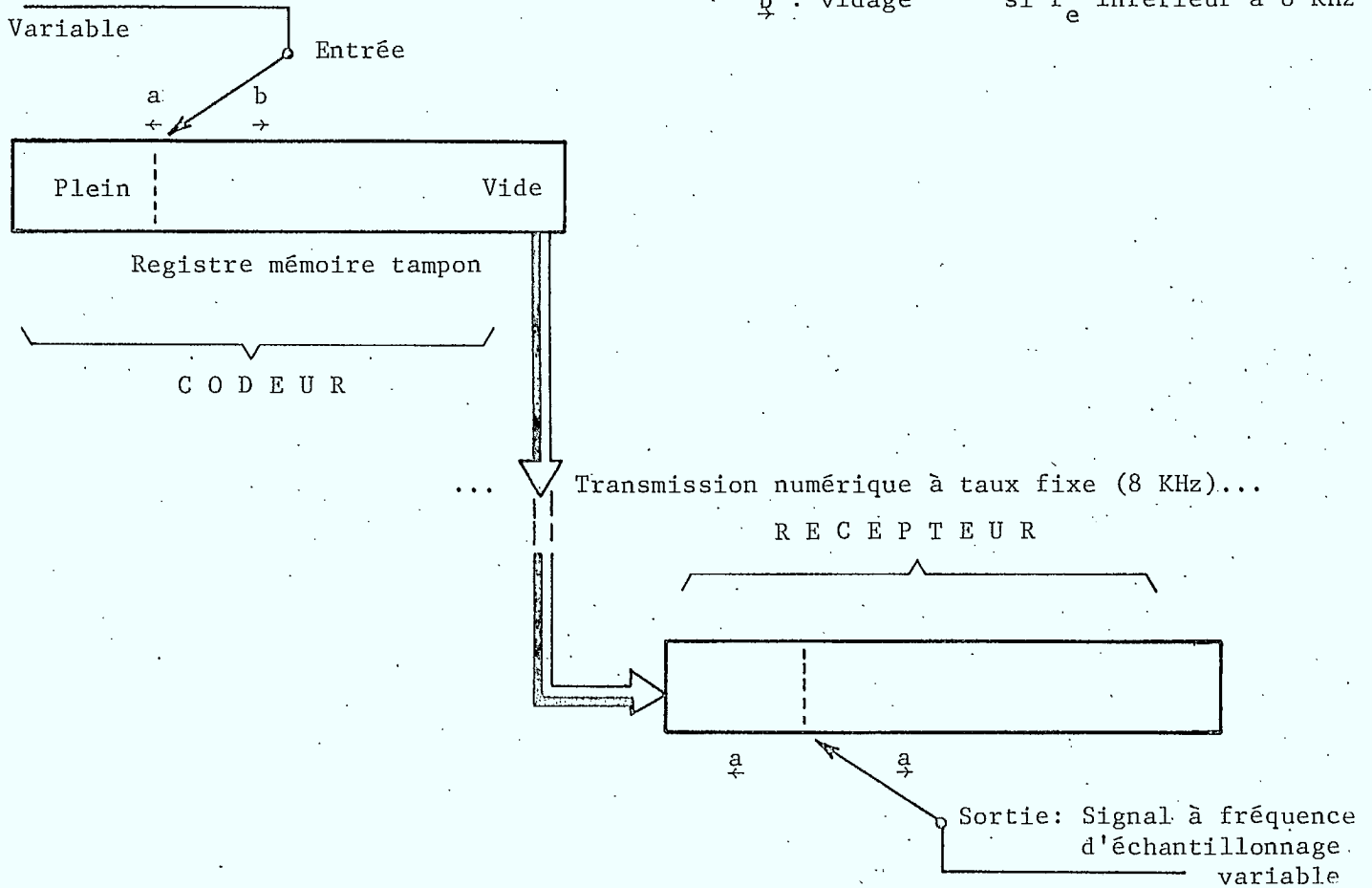


Fig. 2.1 Rôle des mémoires au codeur et au décodeur

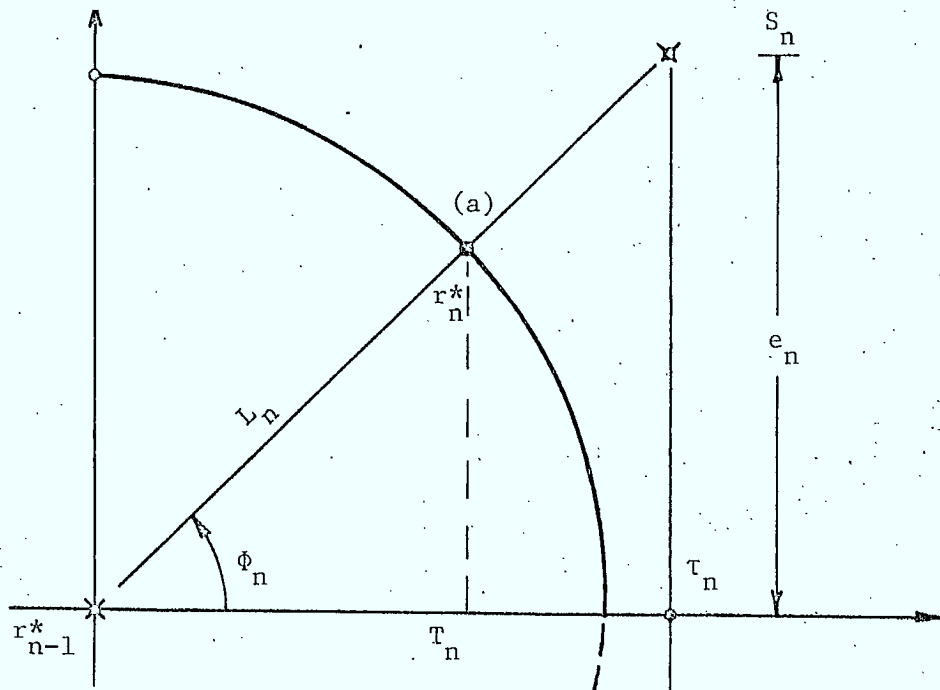


Fig. 2.2 Exemple de codage de signaux normalisés sur un cercle de rayon unité $L_n = 1$.

L'algorithme choisi impose:

- Une augmentation du gain de normalisation de la différence E_n ,
- Une réduction sur l'axe des temps du rayon de codage (T_n et L_n)

ce qui se traduit à l'instant de codage suivant par une modification de la courbe de codage, à partir de r_{n-1}^* cette fois sur la figure 2.4.

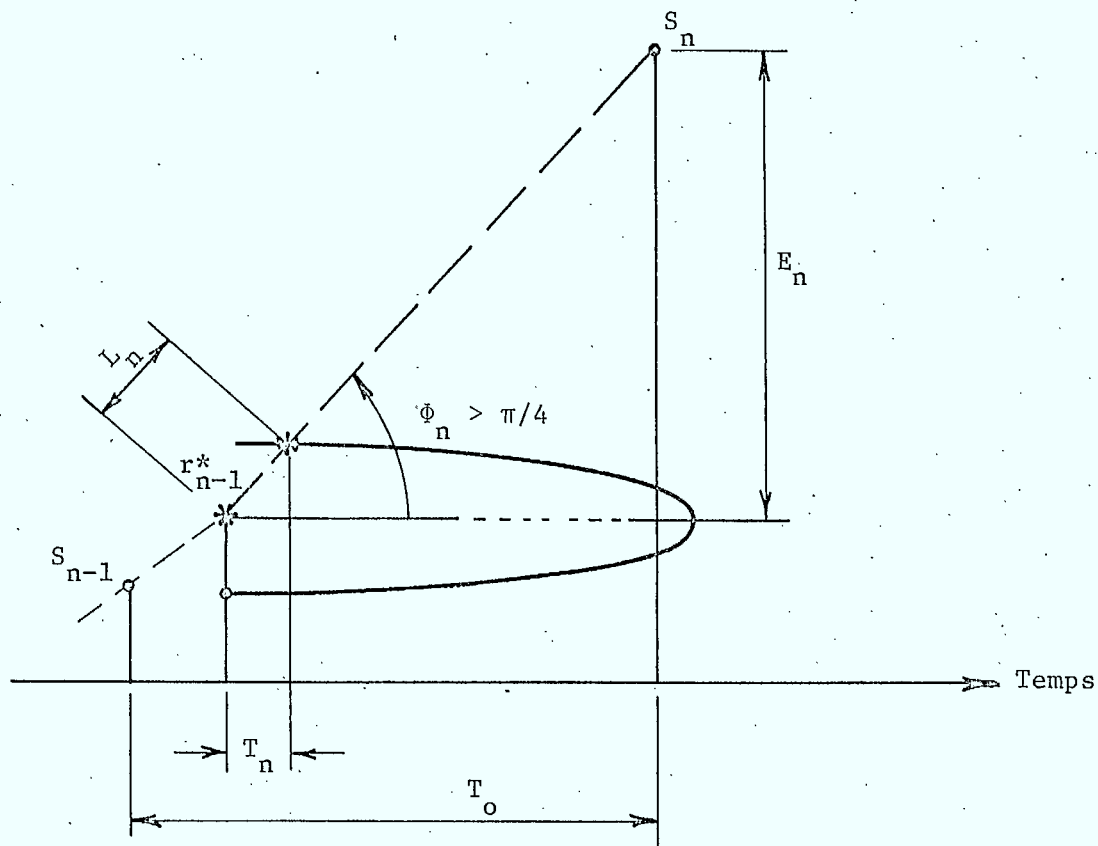


Fig. 2.3 Cas des angles supérieurs à $\pi/4$.

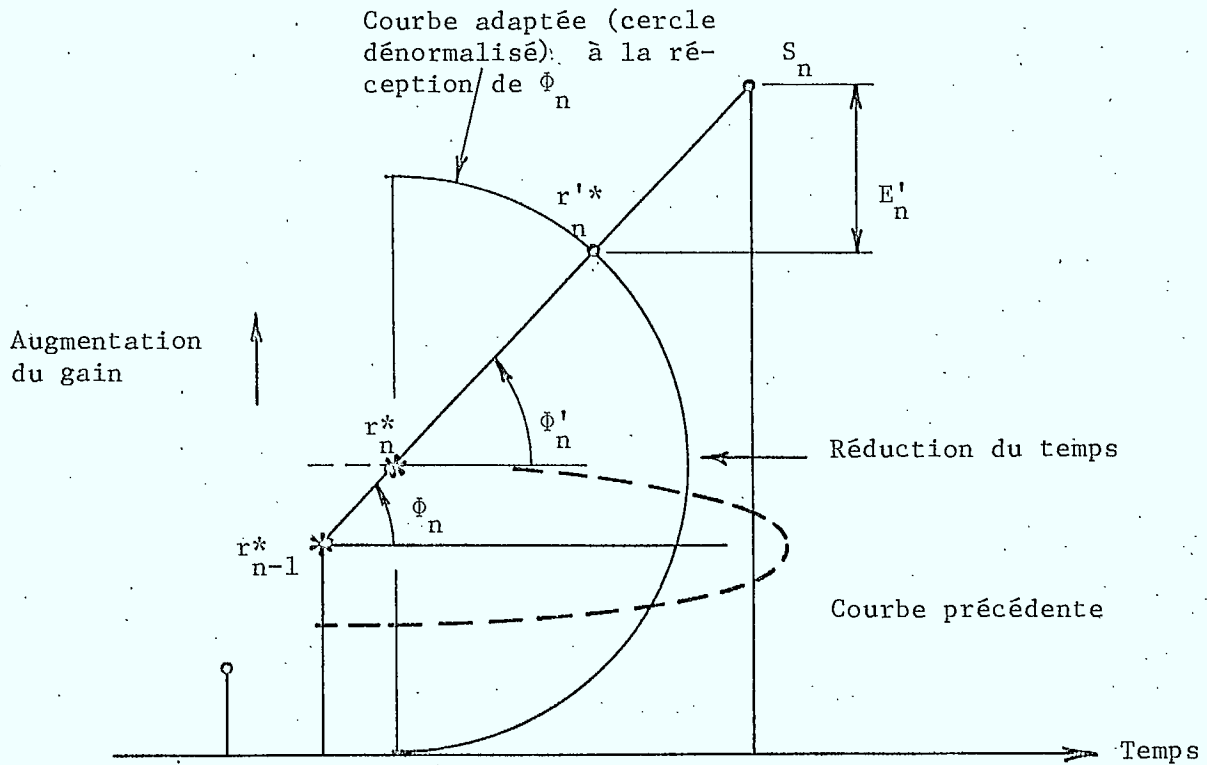


Fig. 2.4 Adaptation de la courbe de codage selon l'angle ϕ_n .

A ce niveau, il est utile de rappeler que, si les calculs intermédiaires sont assurés à partir de variables normalisées, la représentation à l'échelle normale du signal de parole traité par ce procédé fait correspondre au "cercle normalisé de codage" une représentation géométrique en "ellipse" dont l'un des axes dépend du temps et l'autre directement du gain de normalisation.

Il en résulte que, si sur la figure 2.4 précédente, il apparaît "géométriquement" que les angles ϕ_n et ϕ_n^* sont identiques, de fait, dans l'espace normalisé de calcul où cette ellipse est redéformée en cercle, il s'agit de valeurs (très) différentes.

Le cas des faibles valeurs de ϕ_n contribue avec l'algorithme choisi à réduire le gain donc l'amplitude verticale de la courbe de codage et à allonger le temps de progression comme l'illustre la figure 2.5.

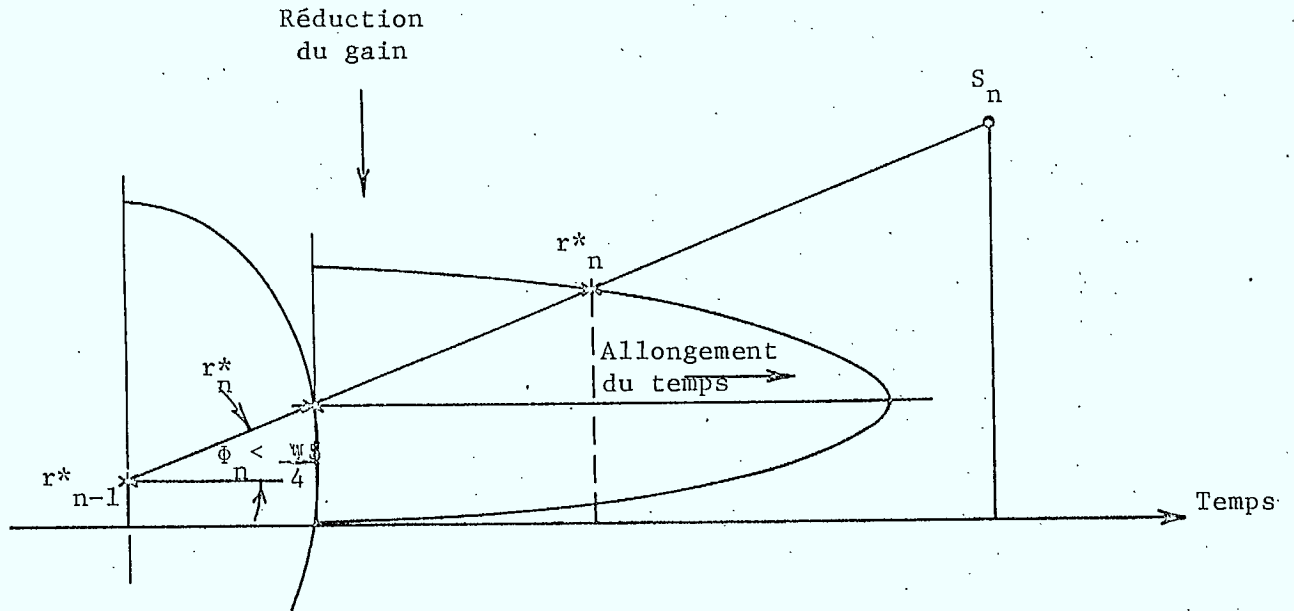


Fig. 2.5 Adaptation due aux faibles angles.

En espace normalisé, des angles voisins de $+\pi/4$ ou $-\pi/4$ ne modifient pas sensiblement les paramètres de codage. L'algorithme est donc essentiellement conçu pour adapter la courbe de codage en un cercle de rayon unité, prélevant un angle normalisé aussi proche que possible de $\pi/4$. En pratique cette "convergence" idéale de l'algorithme n'est jamais atteinte, ce qui impose de transmettre une valeur de l'angle ϕ_n à chaque instant (élastique: variable selon les conditions).

Ce nouveau signal devra à son tour être quantifié et transmis. Du point de vue pratique la quantification de ϕ_n évitera la définition d'angles trop proches de $\pi/2$ où la tangente devient infinie. De fait la simulation limite les valeurs de $\text{tg}(\phi_n)$ à un intervalle numérique de $[1/10 \ 10]$.

1.1.1 Calcul des paramètres "temps et gain" normalisé

Dans la simulation effectuée chaque fichier primaire ou phrase a pour origine du temps le premier échantillon numérique.

Un fichier secondaire contiendra les valeurs de l'angle $\phi_n(t)$ prélevées à un rythme variable selon les conditions de codage: sa longueur pourra donc être différente.

- Ecart de temps: Normalisation et adaptation du temps

Chaque mot numérique du fichier primaire correspond à une progression en "temps naturel" de 125 μ s pour la fréquence d'échantillonnage de 8 KHz. Selon le mode de codage proposé le "temps réel" progresse à chaque mot du fichier secondaire ($\phi_n(t)$) d'un intervalle de temps variable T_n comme illustré sur la figure 2.2.

Du point de vue pratique utilisé dans le programme de simulation, la valeur du temps exprimée en abscisse sur la figure 2.2 peut être considérée comme le rapport de l'écart de temps réel (t_n) divisé par la période de base de l'échantillonnage (t_0).

$$T_n = \frac{t_n}{t_0} \quad \text{avec} \quad t_0 = 125\mu\text{s à } 8 \text{ KHz}$$

T_n est alors un nombre sans dimension.

Ainsi, en échelle normalisée, le rayon du cercle L_n à l'instant n sera sur la figure 2.6; calculé de façon adaptative par la relation

$$L_n = L_{n-1} / \text{tg}(\phi_n)$$

d'où

$$T_n = L_n \cos(\phi_n),$$

est alors un temps normalisé variable.

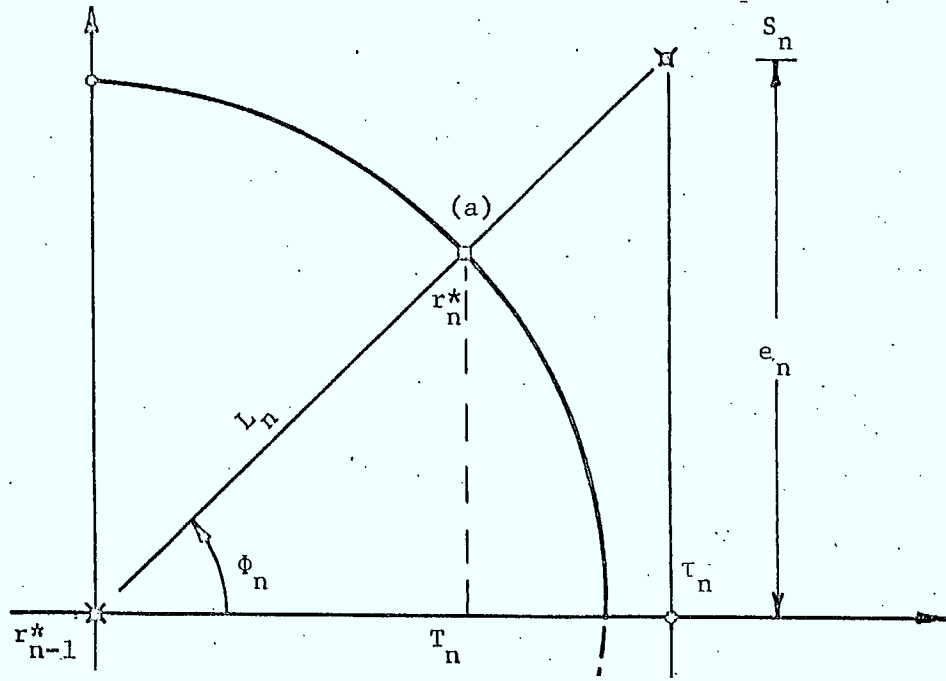


Fig. 2.6 Rappel du principe de la courbe de codage.

- A chaque progression du "temps réel" d'un incrément égal à T_n le processus de codage de l'échantillon suivant du signal original de parole (en temps "naturel") se répète tant que, depuis l'instant d'origine

$$\sum_m T_m < n t_0$$

où nt_0 représente, sur n échantillons de parole, la durée depuis le début du codage de la phrase échantillonnée avec la période (naturelle) t_0 . Le nombre m de mots intermédiaires codés peut être supérieur à n et en ce cas, les mémoires du codeur (et du récepteur) sont partiellement remplies.

- Gain d'adaptation du codeur: normalisation d'amplitude

L'écart ou différence d'amplitude entre deux échantillons du signal de parole ou plus précisément:

entre le dernier échantillon décodé (à rythme variable r_{n-1}^*) et l'échantillon naturel (S_n , à 8 KHz par exemple) à coder immédiatement après est normalisé par un facteur de gain adaptatif, c'est-à-dire, selon l'exemple de la figure 2.6.

$$e_{n \text{ normalisé}} = (S_n - r_{n-1}^*)_{\text{normalisé}}$$

avec, par rapport au signal de parole non normalisé

$$e_n = \frac{E_n}{G_n}$$

E_n écart réel entre amplitudes

où G_n se calcule par une formule adaptative, dépendant seulement de l'angle transmis et de la valeur précédente soit:

$$G_n = G_{n-1} \cdot \text{tg}(\phi_n)$$

De cette façon les deux paramètres d'adaptation et de normalisation sont déterminés par un algorithme simple dépendant strictement de la seule information transmise, à savoir l'angle $\phi_n(t)$.

On notera qu'aucune information, sur la dynamique (variation d'amplitude de la parole) n'est directement transmise. Ce gain en est a priori totalement indépendant.

- Distances de codage: rayon de la courbe de codage

Comme le montre la figure 2.6 du principe de codage, on définit la distance D_n entre échantillons (normalisés bien sûr) par la relation euclidienne

$$D_n^2 = \left(S_n - r_{n-1}^* \right)^2 + \left(nT_o - \sum_m T_m \right)^2$$

et le rayon de codage, connu du récepteur à l'instant n étant alors

L_{n-1} :

où $S_n - r_{n-1}^* = e_n$ représente l'écart d'amplitude normalisé
 $nT_0 - \sum T_m$ représente l'écart de temps normalisé

respectivement entre l'échantillon à coder (normalisation) soit S_n et le dernier élément codé/décodé connu au récepteur, soit r_{n-1}^* (normalisé bien entendu).

Il est alors aisé d'identifier la position relative de S_n par rapport à la courbe de codage:

si $D_n^2 > L_{n-1}^2$

l'échantillon à coder est en dehors de la courbe et le résultat de ce codage donnerait r_n^* (voir figure 2.6).

Dans certains cas, même si $D < L_{n-1}$, où S_n est à l'intérieur de ce cercle, l'échantillon sera codé malgré tout, et fournira toujours un point de codage r_n^* sur le cercle lui-même par une extrapolation linéaire: ce cas sera précisé ultérieurement au chapitre 3

1.1.2 Conditions de codage - Tests de décision

La décision de coder un échantillon de parole par un ou plusieurs échantillons intermédiaires représentant un ou des angles $\phi_n(t)$ dépend de certaines caractéristiques instantanées (Rapport 1, figure 7, page 18).

- Etat de remplissage de la mémoire

Lorsque celle-ci est remplie au-dessus d'un seuil proche du remplissage maximal, il ne peut être admis de sur-échantillonner le signal intermédiaire $\phi_n(t)$. L'écart de "temps réel" imposé doit être choisi au moins égal à l'écart de temps naturel (125 μ s) (Rapport 1, figure 8, page 19). Une dégradation de codage inévitable en résulte alors. Toutefois, l'ensemble de paramètres choisis en simulation permettent

de ne jamais rencontrer effectivement cette situation avec les mémoires utilisées entre 256 et 1024 échantillons (à 8 KHz le retard de transmission s'établirait entre 32 et 128 millisecondes).

- Etat des échelles de temps - "Temps réel" et "temps naturel"

Le "temps naturel" relié à la période d'échantillonnage (125 μ s à 8 KHz) est comparé à tout instant de codage au "temps réel" de durée variable. Seuls sont codés les échantillons du signal de parole correspondant par leur position dans le temps à la relation

$$\begin{array}{rcl} T_{\text{naturel}}(n) & > & T_{\text{réel}}(n-1) \\ \text{total de l'échan-} & & \text{total du dernier échan-} \\ \text{tillon à coder} & & \text{tillon codé et décodé} \end{array}$$

- Cas des échantillons très "rapprochés"

L'algorithme de codage peut conduire à une reconstitution d'un échantillon r_n^* sur la courbe de codage très proche (en distance euclidienne) du premier échantillon suivant à coder S_n . La figure 2.7 illustre ce cas.

L'application à partir de r_n^* de la courbe de codage peut conduire, par extrapolation linéaire, à la reconstitution d'un échantillon $r_n'^*$ qui peut imposer une forte erreur de codage sur l'échantillon ultérieur à coder, soit S_{n+1} .

Dans ce cas, compte tenu des limites choisies pour la distance L_n (ou L'_n), il a été possible d'abandonner le codage local de l'échantillon S_n si celui-ci est à l'intérieur d'un arc de courbe auxiliaire défini par son rayon λ_n soit en simulation

$$\lambda_n^2 = \frac{1}{8} L_n^2$$

Un compromis sur cette distance doit être assuré pour garder une qualité subjective de codage très satisfaisante.

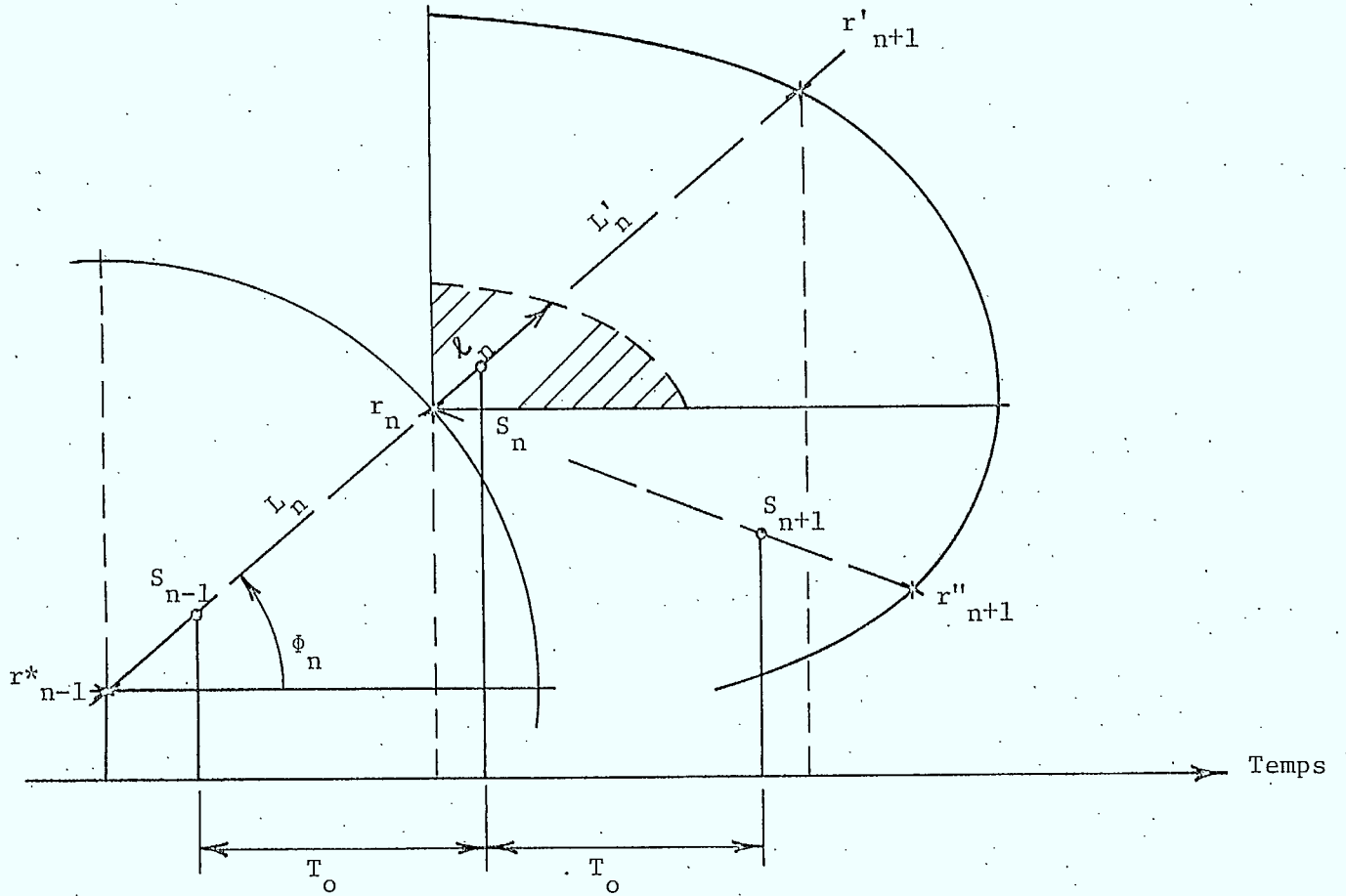


Fig. 2.7 Cas de proximité des échantillons en hachuré: zone d'exclusion de codage.

Cas des silences ou bruit de fond

En l'absence de signal de parole, la forme d'onde à coder est généralement un bruit de fond dont les caractéristiques spectrales (nombre de passages par zéro du signal et de la dérivée de celui-ci) et de dynamique (niveau faible peu modulé) permettent une identification rapide par des techniques numériques simples (réf. Rapport Jean-Pierre Adoul, CRC-Ottawa*.) Ces techniques déjà développées et utilisées permettent, au codeur, de remplacer ce bruit de fond par du silence (signal nul ou très faible: inaudible) et de développer un "algorithme de codage en temps élastique" utilisant ces intervalles de silence de manière à augmenter l'efficacité globale. De la sorte on attend d'une part:

- Une réduction du niveau de remplissage de la mémoire par un allongement très net du "temps réel" de base: c'est-à-dire dans les exemples simulés, la "longueur" du rayon L_n de la courbe en cercle de codage est alors portée à 8 fois la durée de base ou période d'échantillonnage à 8 KHz. Cela signifie que, tant que la mémoire n'est pas totalement vidée, un seul échantillon (toutes les millisecondes: $8 \times 125 \mu s = 1 \text{ ms}$) est effectivement codé.

Ceci affecte peu la qualité du signal reçu car il s'agit d'un codage très inefficace d'un bruit de fond sans intérêt.

d'autre part:

- Dans une version avec transmission analogique de l'angle $\phi_n(t)$,

ADOUL, J.P., & al, - "Conception et réalisation d'un détecteur de parole identifiant, pour un canal vocal, silence/données/paroles voisées/paroles non-voisées", Rapport de l'Université de Sherbrooke dans le cadre du contrat CRC no. 02SU-36100-79518, 51 pages, (avril 1978).

une possibilité d'utiliser ces intervalles de silences pour suggérer une méthode permettant de récupérer le synchronisme rigoureux entre:

- l'échantillonnage numérique initial du signal intermédiaire $\phi_n(t)$ au codeur et l'échantillonnage à reconstituer au décodeur à 8 KHz toujours,
- l'échantillonnage numérique à 8 KHz obtenu sur le signal de parole finalement reconstitué au décodeur et qui doit correspondre exactement à la relation existant au codeur entre le fichier source de parole (à 8 KHz) et l'échantillonnage à rythme variable (temps réel) construit à partir d'échantillons intermédiaires (r_{n-1}^* , r_n^* , voir figure 2.2 par exemple). Pour cette étude voir chapitre 5.

Pour cette analyse des silences le codeur bénéficie d'une détermination adaptative du niveau de signal reconnu comme seuil de décision de la nature Parole/silence. En effet l'énergie moyenne des échantillons de parole codés puis décodés est pondérée par un facteur numérique proportionnel au coefficient de remplissage de la mémoire.

Ainsi avec ce coefficient noté B_f , variant linéairement de -1 à +1 selon que la mémoire est vide ou pleine, le seuil de décision est calculé par:

$$S_o = E \cdot (2 + B_f)$$

où E est l'énergie moyenne filtrée déterminée sur les 16 derniers échantillons (2 millisecondes à 8 KHz) du signal décodé (donc disponible au récepteur). Un seuil plus élevé, en cas de mémoire pleine, contribue à vider plus rapidement celle-ci avant que les signaux ne soient devenus très faibles. Le

gain en place mémoire est obtenu par une détérioration localisée du signal codé. Le compromis obtenu semble subjectivement très satisfaisant.

1.2 Détermination de l'angle $\phi_n(t)$

Dans le cas général de fonctionnement de l'algorithme, les conditions de la figure 2.2 sont le plus souvent remplies. L'angle ϕ_n est simplement calculé, en échelles normalisées de temps et d'amplitude par:

$$\phi_n = \text{artg} \frac{e_n}{\tau_n}$$

Ce type de codage conduit généralement à un sur-échantillonnage local du signal et contribue à un remplissage de la mémoire. Les phases de silence (ou bruit de fond) permettent, à l'opposé de vider la mémoire.

1.2.1 Cas particulier des silences

Dans ce cas spécifiquement, seul le premier échantillon (à venir) en dehors du cercle de codage normalisé, peut être codé, ce qui conduit à des configurations illustrées sur la figure 2.8.

Les valeurs intermédiaires (mal) codées sont restituées par interpolation linéaire directe.

Quand la mémoire est totalement vidée le rythme intermédiaire d'échantillonnage de $\phi_n(t)$ au codeur est obligatoirement figé à la valeur de base du temps naturel soit t_0 . Ceci se maintient jusqu'à l'arrivée de signaux suffisamment important pour remplir au moins un peu la mémoire par codage imposant un sur-échantillonnage.

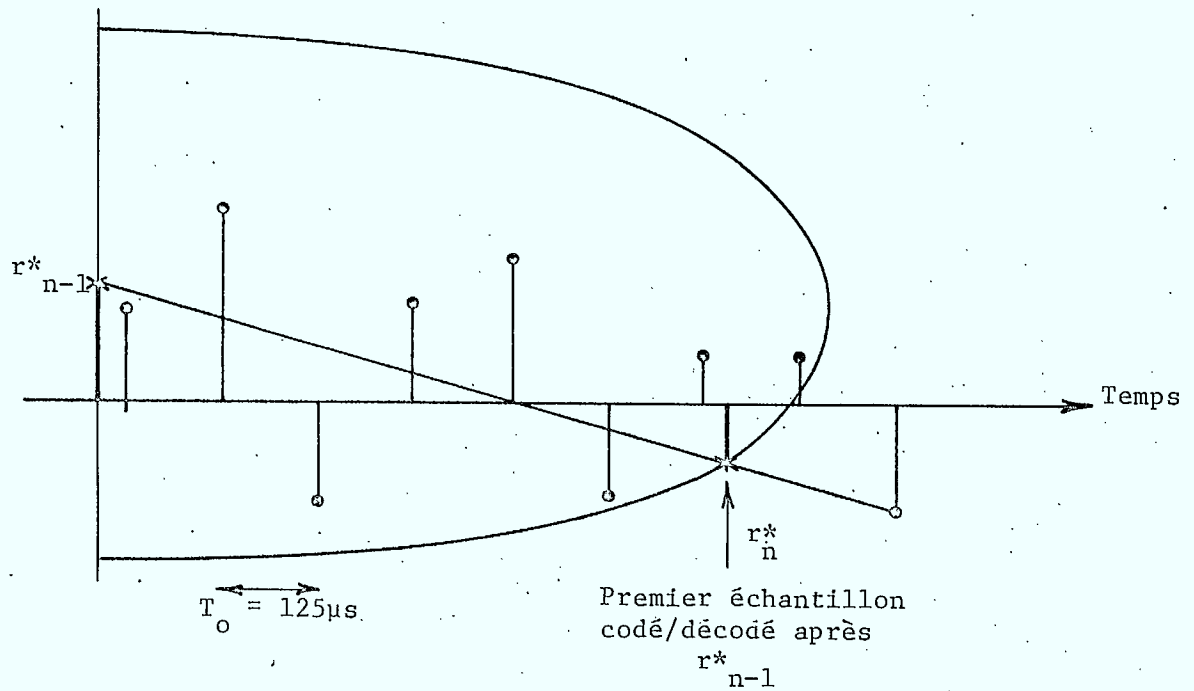


Fig. 2.8 Codage en phase de silence (petits signaux).

1.2.2 Cas des faibles variations d'angle

Dans les phases de codage où le rayon de codage est important, la courbe de codage peut englober un ou plusieurs échantillons consécutifs comme sur la figure 2.9.

Dans de tels cas le choix, au codeur entre ϕ_α ou ϕ_β permettant de reconstituer au décodeur r^*_α ou r^*_β à la place de S_n , peut permettre un certain gain au niveau de remplissage de la mémoire et de la précision du signal reconstitué. Notons que le nouveau point de départ du codage devient soit r^*_α , soit r^*_β .

Un test de décision est donc effectué sur l'écart angulaire:

$$|\phi_\alpha - \phi_\beta| \leq \text{Seuil}$$

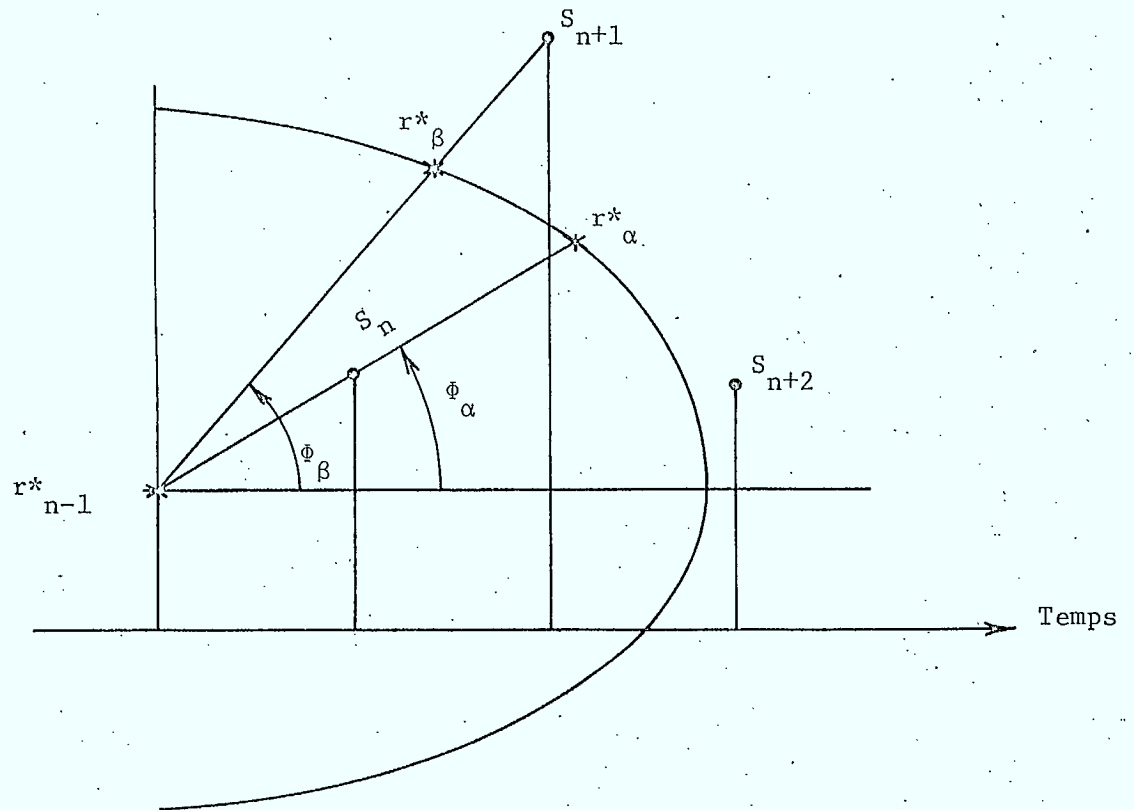


Fig. 2.9 Cas particulier des faibles variations d'angle.

Dans ce cas l'angle retenu est celui conduisant au meilleur codage soit ϕ_{β} sur cette figure. Sinon, la première valeur donc ϕ_{α} serait la seule retenue et transmise.

Ce seuil est de fait adapté en amplitude par le coefficient normalisé de remplissage de la mémoire autorisant ainsi un codage plus précis quand la mémoire est vide. La relation d'adaptation est:

$$\text{Seuil adapté} = (\text{Seuil fixe}) \cdot (1,25 + B_f)$$

Un seuil fixe de 5° à 10° est très satisfaisant au niveau compromis qualité de codage/degré de remplissage de la mémoire.

1.3 Quantification de l'angle $\phi_n(t)$

Le processus de traitement numérique envisagé jusqu'alors permet de générer un nouveau signal prélevé à un rythme variable (temps réel), temporairement mémorisé et transmis à rythme fixe (8 kHz; temps naturel). Les calculs suggérés des paramètres de codage utilisent une précision de 16 bits (quantification linéaire) du calculateur PDP-11/60.

Une réduction à 12 bits peut être envisagée sans difficulté ni perte de qualité. Il est de plus vraisemblable qu'une longueur de 8 bits serait même suffisante compte tenu des essais de quantification ultérieure.

La quantification de l'angle $\phi_n(t)$ à transmettre peut être effectuée de manière optimale par une détermination des paliers de quantification en fonction de la distribution statistique de niveau d'amplitude du signal $\phi_n(t)$.

1.3.1 Algorithme de la K-moyenne

Il s'agit d'un algorithme permettant sur l'ensemble du fichier de parole traité de déterminer par convergence un vecteur de représentants optimaux à partir d'un vecteur initial quelconque.

Cet algorithme déjà utilisé dans des travaux précédents [*] et décrit dans son principe est utilisé à partir d'un vecteur original de quantification linéaire.

Les figures 2.10 et 2.11 représentent pour deux phrases différentes la distribution statistique de répartition de niveaux de l'angle $\phi_n(t)$ (limité à 64 niveaux possibles, donc quantifié et codé avec 6 bits).

[*] ADOUL, J.P., -"Transmission et traitement numérique de l'information parlée" Phase III, Rapport final, chapitre 3, contrat no. OSU80-00A8 pour CRC.

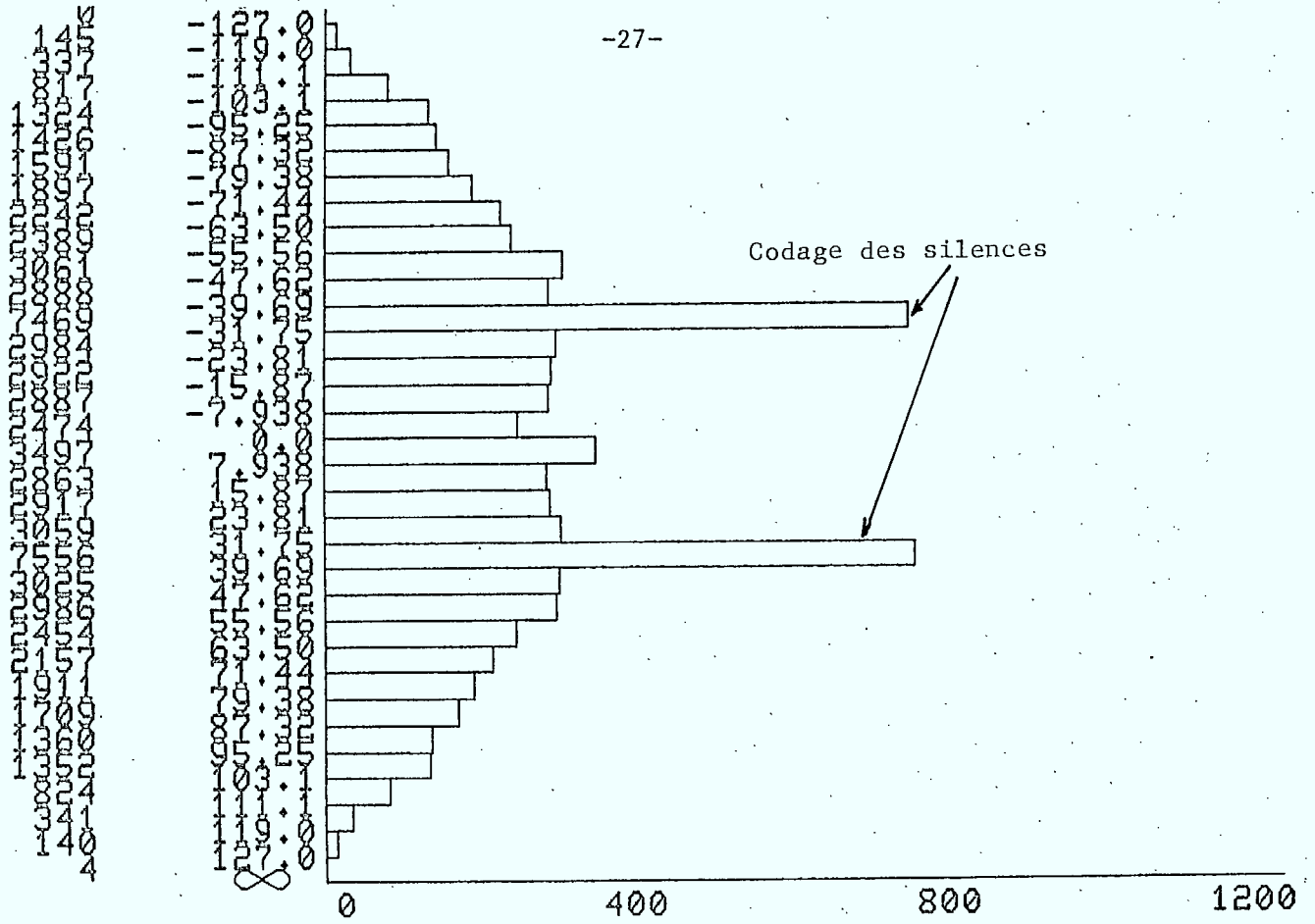


Fig. 2.10 Distribution statistique des niveaux de l'angle $\phi_n(t)$ sur l'ensemble des 10 phrases.

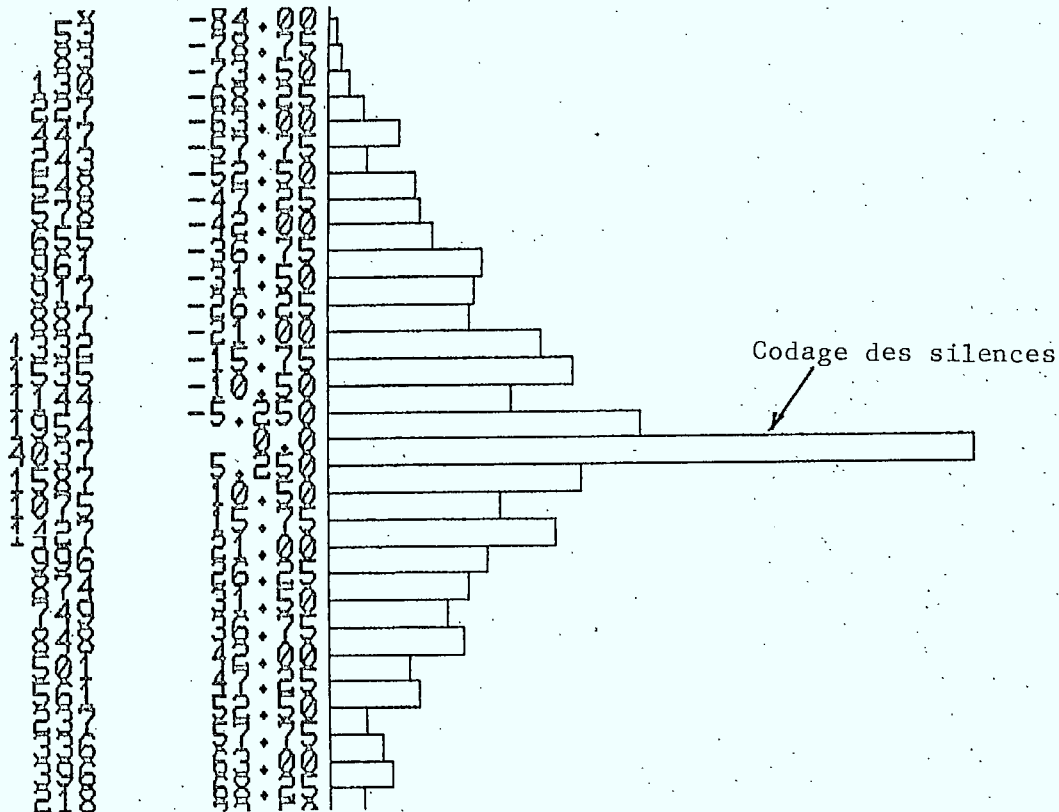


Fig. 2.11 Distribution statistique des niveaux de l'angle $\phi_n(t)$ sur une phrase.

1.3.2 Débit numérique

Le signal de référence MIC (ou PCM) est constitué de 8 bits à 8000 Hz, soit un débit de 64,000 bits/s. Le but premier d'un nouveau codage numérique est, à qualité comparable, de réduire ce débit.

Avec pour référence de qualité la valeur du rapport Signal sur Bruit de 37 à 38 dB, le MIC représente l'étalon de comparaison. Dans le chapitre 4 montrant des résultats et sur la cassette enregistrée et jointe à ce rapport, on peut vérifier le niveau de qualité élevé obtenu par ce type de codage.

A débit comparable le rapport S/N atteint l'ordre de grandeur de 38 dB en ce qui concerne le bruit "granulaire", c'est-à-dire comparable au bruit de quantification du PCM.

Toutefois la réduction du nombre de bit de codage diminue ce chiffre mais la nature du bruit généré, assez différent du bruit de quantification classique, permet une évaluation encore très favorable, même avec seulement 4 bits (32 Kbits/s) de quantification.

La cassette présentera une comparaison entre les deux méthodes de codage à 4 bits: le MIC (PCM) 4 bits comparé au codage en temps élastique à 4 bits.

1.4 Quantification du temps

La progression en temps, à partir du rayon de codage L_n (figure 2.2) calculée par:

$$T_n = L_n \cos \phi_n$$

impose d'une part des approximations de calcul mais en plus une variation constante qui doit être rigoureusement identique au décodeur. En particulier ce principe ne peut tolérer la moindre erreur de transmission puisque tout le codage dépend de la seule information transmise ϕ_n à partir de laquelle le décodeur "reconstruit" le rayon L_n donc le temps T_n .

D'autre part une quantification de ce temps intermédiaire, associé à un procédé de resynchronisation et de réinitialisation pendant les silences (voir chapitre 5), limitera l'effet d'éventuelles erreurs de transmission.

De plus la solution proposée tend à limiter le remplissage de la mémoire en imposant un temps minimal (donc interdisant un sur-échantillonnage trop rapide).

Ainsi chaque intervalle de temps naturel d'origine sera quantifié en multiples entiers de $\frac{T_0}{4}$ où T_0 est la période (normalisée) de l'échantillonnage d'origine soit 125 μ s à 8 KHz. De plus, au codeur, la trame d'échantillonnage intermédiaire (à $T_0/4$) sera construite de manière synchrone avec l'échantillonnage de base.

La figure 2.12 représente l'effet combiné:

- d'une quantification de l'angle ϕ_n ,
- d'une quantification du temps T_n .

Ce principe introduit sûrement de légères erreurs de codage, affaiblissant le rapport S/N mais l'effet subjectif à l'écoute s'avère très peu sensible. Cependant le gain très important dans le domaine de la réduction du remplissage de la mémoire et de la minimisation des effets des erreurs de transmission justifie ce compromis.

En toute rigueur la courbe de codage devient un espace curviligne de codage où chaque représentant n'est plus un point fixe, mais dépend des

Signal "quantifié" en
temps et en angle

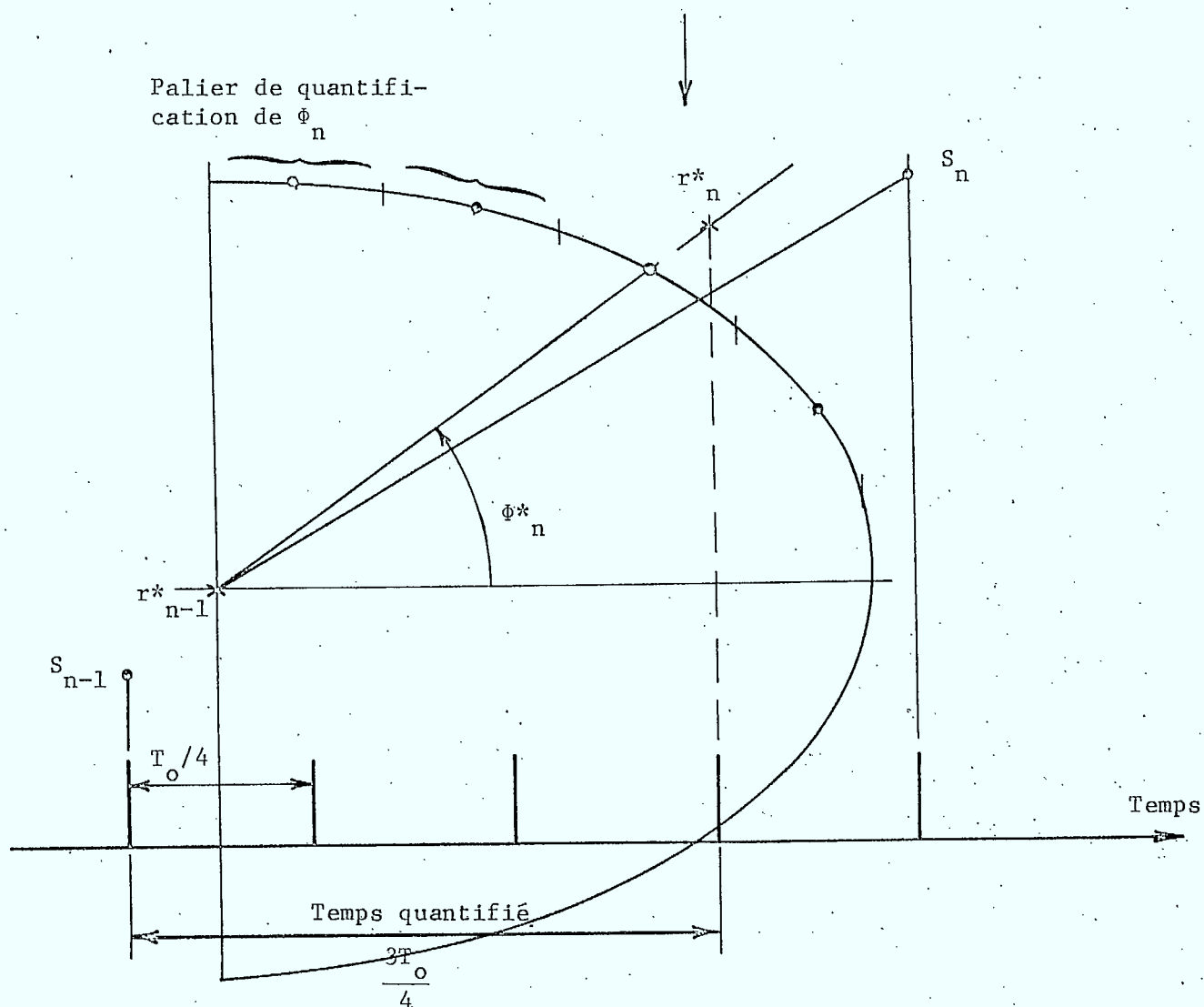


Fig. 2.12 Quantification simultanée de l'angle et du temps.

conditions mêmes de codage (angle ϕ_n , position relative des r^*_{n-1} et S_n qui sont les échantillons de parole codé/décodé et à coder).

L'application des algorithmes de K-moyenne permet de choisir les valeurs quantifiées de l'angle $\phi_n(t)$ optimale pour le codage de la parole moyenne.

1.5 Organigramme de codage

Chaque paragraphe présenté dans le texte du chapitre 2 peut être retrouvé sur l'organigramme de simulation présenté ici.

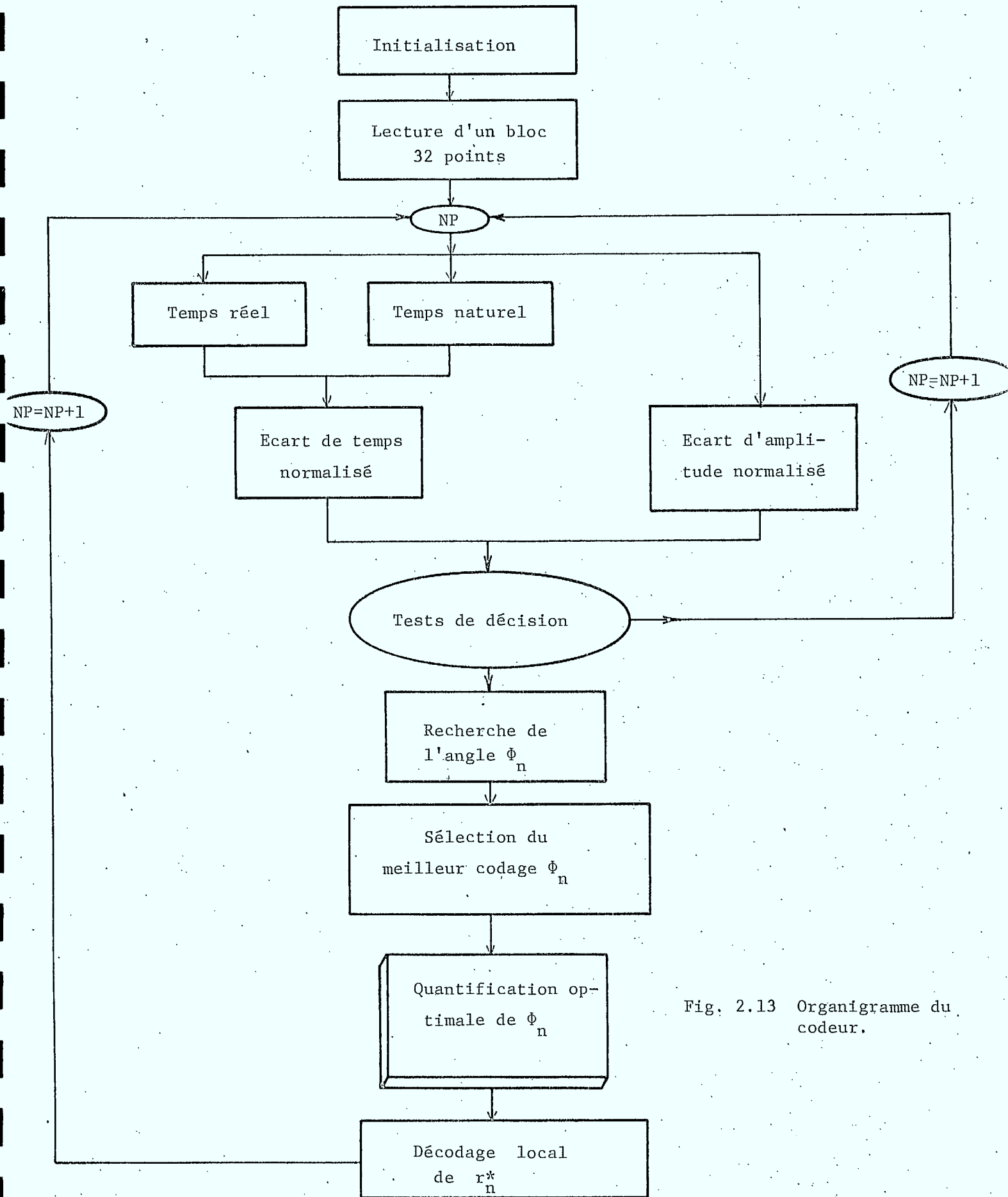


Fig. 2.13 Organigramme du codeur.

CHAPITRE 3

ORGANISATION DU DECODEUR

Le signal numérisé $\phi_n(t)$ représentant un angle de codage est quantifié sous un nombre restreint (m) de bits afin de diminuer le débit de transmission.

Le canal de transmission est supposé sans erreur et le décodeur reçoit donc pour seule information l'angle $\phi_n(t)$ à un débit fixe correspondant à un échantillonnage de 8 KHz (débit = 8.m Kbit/s).

Il faut rappeler que ce signal correspond à une reconstitution d'une forme d'onde (signal de parole) prélevée à rythme variable (temps réel) mais par la suite restructurée sous la forme conventionnelle en temps naturel (8 KHz de fréquence d'échantillonnage).

Le processus général de décodage comprendra donc:

- Une mémoire d'entrée (identique à celle du codeur) où le rythme d'entrée (horloge) est de 8 KHz. Toutefois les échantillons extraits de la mémoire le seront à rythme variable, calculé directement à partir de l'information déjà reçue. On rappelle que le temps (réel) est quantifié en multiple entier de $T_0/4$; T_0 période naturelle du 8 KHz soit 125 μ s.
- Un organe de calcul comprenant la détermination des paramètres du codage (gain et temps normalisés)
- Une analyse de la nature du signal reçu: segmentation en parole ou silence (bruit de fond) pour améliorer, comme au codeur, l'efficacité globale du traitement.

1. Mémoire du décodeur

On rappellera la figure présentée au chapitre 2.

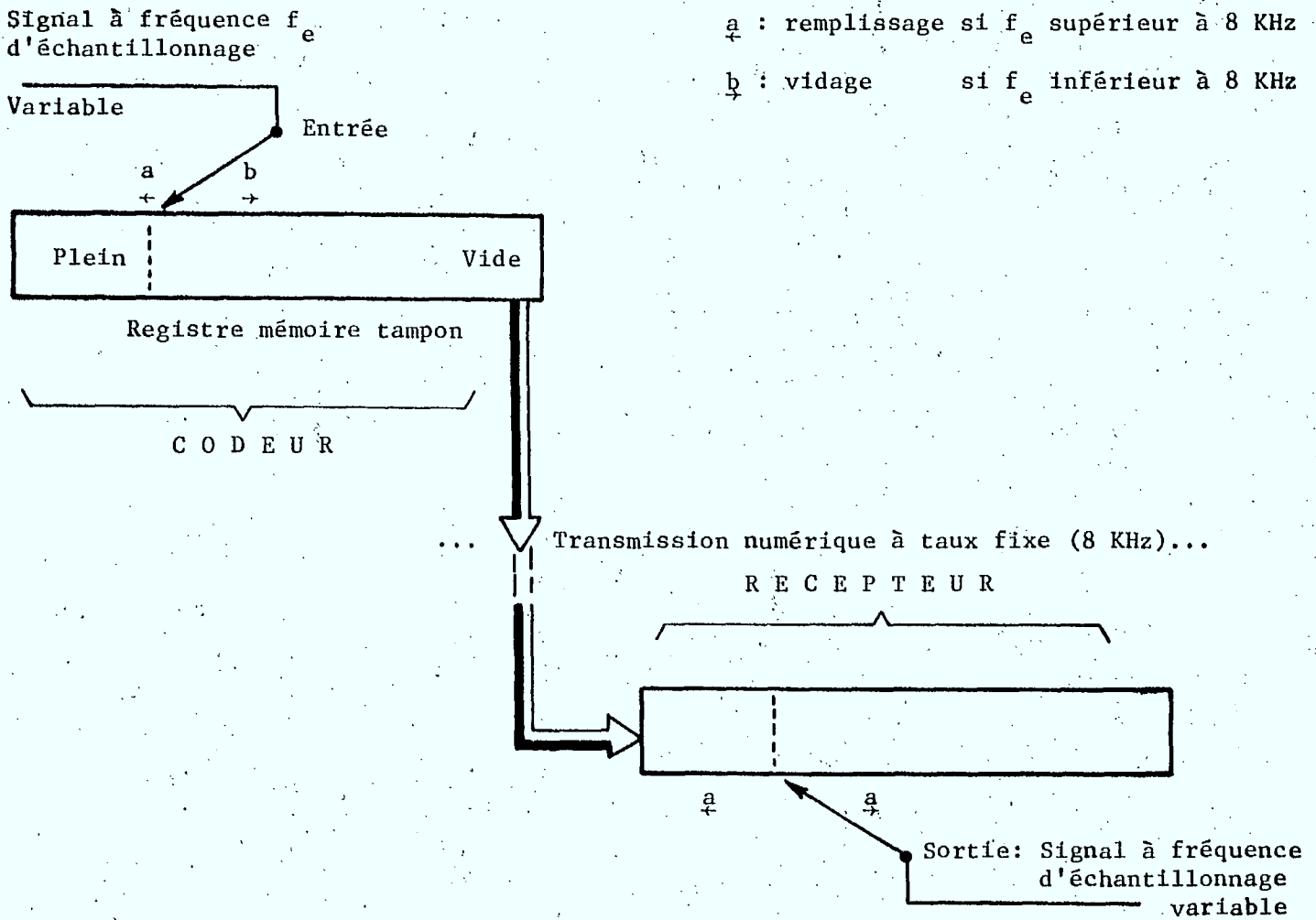


Fig. 3.1 Rôle des mémoires au codeur et au décodeur.

On constate que la durée totale du retard de transmission correspond à la longueur de l'une ou l'autre des mémoires soit par exemple:

$$1024 \times 125 \mu s = 128 \text{ millisecondes}$$

pour une mémoire de 1024 échantillons d'angle (1,024 K). Cette mémoire peut être réduite assez largement à 512, 256 ou même 128 mots, réduisant ainsi le retard de transmission. Toutefois l'adaptation des paramètres de calcul du décodeur dépendra du degré de remplissage de cette mémoire. La qualité de parole transmise diminue avec le remplissage de la mémoire, ce que montrera le paragraphe d'analyse suivant.

2. Organe de calcul du décodeur

Le décodeur existe tant à l'émetteur (c'est-à-dire associé au codeur d'origine) qu'au récepteur suivant le schéma général de la figure 3.2.

Le codeur, connaissant le dernier échantillon codé/décodé, soit r_{n-1} , reçoit, à l'instant n , le nouvel échantillon à coder soit S_n :

A partir, de paramètres internes de décodage, calculés tant au codeur qu'au décodeur, soit:

- Le rayon de la courbe de codage L_n ,
- Le gain normalisé G_n ,
- L'état parole/silence,
- et autres points analysés au chapitre 2.

Le codeur et le décodeur inclus dans l'émetteur ont en commun certaines fonctions de calculs qui ont été incluses en simulation dans un sous-programme utilitaire.

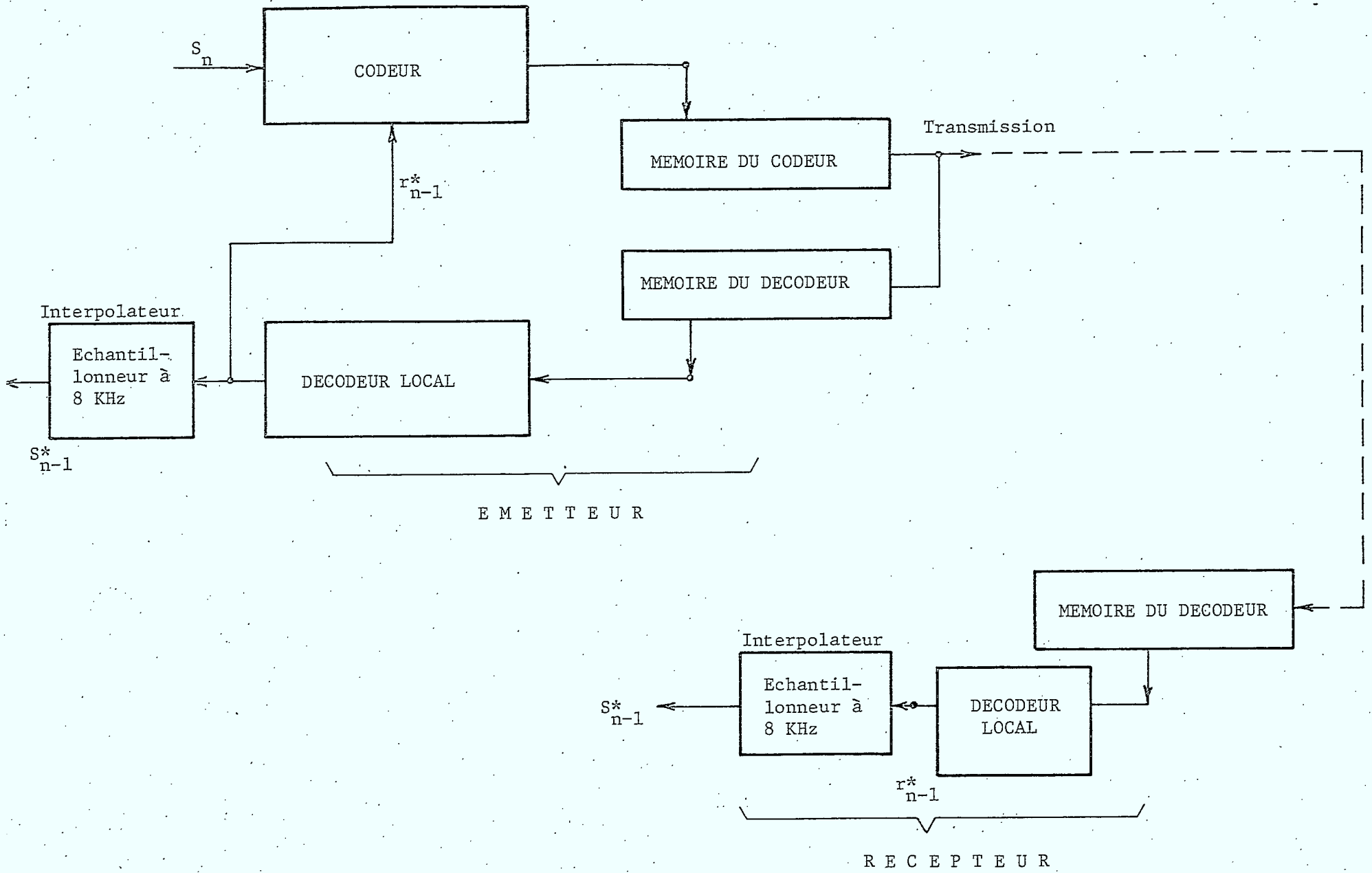


Fig. 3.2 Principe général de codage/décodage.

2.1 Initialisation: calculs des paramètres adaptatifs du codage

- L'état de remplissage de la mémoire est normalisé de façon linéaire par un facteur B_f compris entre -1 et +1.
- Les maxima acceptés pour le codeur et le décodeur sont liés à ce facteur par les relations simples:

$$\begin{aligned} \text{Gain maximal } G_{\max} &= E_n \cdot (2+B_f) \\ \text{Gain minimal } G_{\min} &= \alpha E_n \cdot (2+B_f) \\ \alpha &= 1/4 \end{aligned}$$

où E_n est une amplitude moyenne des derniers échantillons de parole S^*_{n-1} (8 KHz, temps naturel) reconstitués au décodeur, et est calculée sur une fenêtre de 16 échantillons (2 millisecondes à 8 KHz).

$$E_n = E_{n-1} + \frac{1}{16} [|S^*_{n-1}| - E_{n-1}]$$

Cette formule très simple de filtrage est utilisée plusieurs fois pour divers paramètres, ce qui, dans une réalisation expérimentale, autoriserait un multiplexage temporel des opérations avec une seule unité

- Le rayon de codage L_n est aussi borné dans ses variations pour les relations:

$$L_{\min} = L_0$$

où L_0 correspond au rayon normalisé unité qui caractérise le temps naturel (125 μ s), et,

$$L_{\max} = L_0 \cdot \left(\frac{3+B_f}{2} \right)$$

$$\text{Mémoire vide } B_f = -1 \quad L_{\max} = L_0$$

$$\text{Mémoire pleine } B_f = +1 \quad L_{\max} = 2L_0$$

- Enfin pour l'identification du bruit de fond le nombre de passages par zéro N_z sur les 16 derniers échantillons (à 8 KHz) de parole est retenu.

La nature du signal considéré comme un bruit est donc défini par deux conditions:

- $N_z \geq 10$ pour 16 échantillons (à 8KHz)
dans le cas de bruit de fond
- $E_n \leq \text{Seuil} \cdot (2 + B_f)$

Le seuil peut être imposé ou déterminé par un filtrage à bande étroite dans le cas de bruit de fond très important.

Le bruit est identifié, au codeur, pour éviter d'imposer par un sur-échantillonnage important (qui remplit rapidement la mémoire de stockage) un codage très efficace qui n'a aucun intérêt du point de vue subjectif à l'écoute. Autrement dit on sacrifie délibérément la qualité de codage (rapport S/N) lorsque l'on sait que celle-ci n'influe pas ou peu sur l'impression sonore.

Il en résulte que, dans le chapitre 4 présentant des résultats, certaines phrases riches en type de "sons bruités" présentent un rapport de qualité (S/N) en dB relativement plus faible que d'autres; mais en contre partie les enregistrements sur cassette ne permettent pas de distinguer ces défauts localisés de codage où l'oreille n'est pas ou peu sensible.

Ainsi les signaux bruités ne sont pas éliminés mais seulement codés peu efficacement. Cette méthode de traitement s'avère essentielle pour modérer le remplissage des mémoires.

2.2 Reconstitution du signal à partir de l'angle ϕ_n

A l'émetteur et au récepteur, chaque décodeur détermine le signal de parole à reconstituer à partir:

- de l'angle $\phi_n(t)$ (reçu à 8 KHz) et mémorisé,
- des paramètres adaptatifs de codage, calculés sur les échantillons passés, c'est-à-dire principalement:
 - le rayon de la courbe de codage L_n ,
 - l'état silence/parole détecté.

2.2.1 Rappel des calculs à la réception

Toute valeur de ϕ_n stockée au récepteur dans la mémoire y est prélevée au rythme nécessité par le calcul du temps réel instantané.

$$T_n = L_n \cos \phi_n$$

compte tenu de la quantification faite sur la progression en temps en multiples entiers de $T_0/4$ ($T_0/4$ temps naturel: 125 μ s à 8 KHz).

Ainsi selon la figure 3.3, à partir du dernier échantillon de parole r_{n-1}^* obtenu à rythme variable (temps réel), le nouvel échantillon de parole intermédiaire r_n^* est obtenu par: (voir figure 3.3).

- Un déplacement dans la direction de ϕ_n noté en fait ϕ_n^* en raison de la quantification effectuée au codeur.
- Une quantification auxiliaire de la longueur de ce déplacement calculé par la distance L_n de codage (rayon du cercle normalisé)

$$L_n = L_{n-1} / \text{tg } \phi_n^*$$

mais en réalité étendue à la première valeur qui soit un multiple entier de $T_0/4$ (temps naturel) et qui conserve un synchronisme (modulo $T_0/4$) avec le rythme naturel d'origine d'échantillonnage.

Ceci se traduit sur la figure par la longueur L_n^* au lieu de L_n pour reconstruire le point r_n^* .

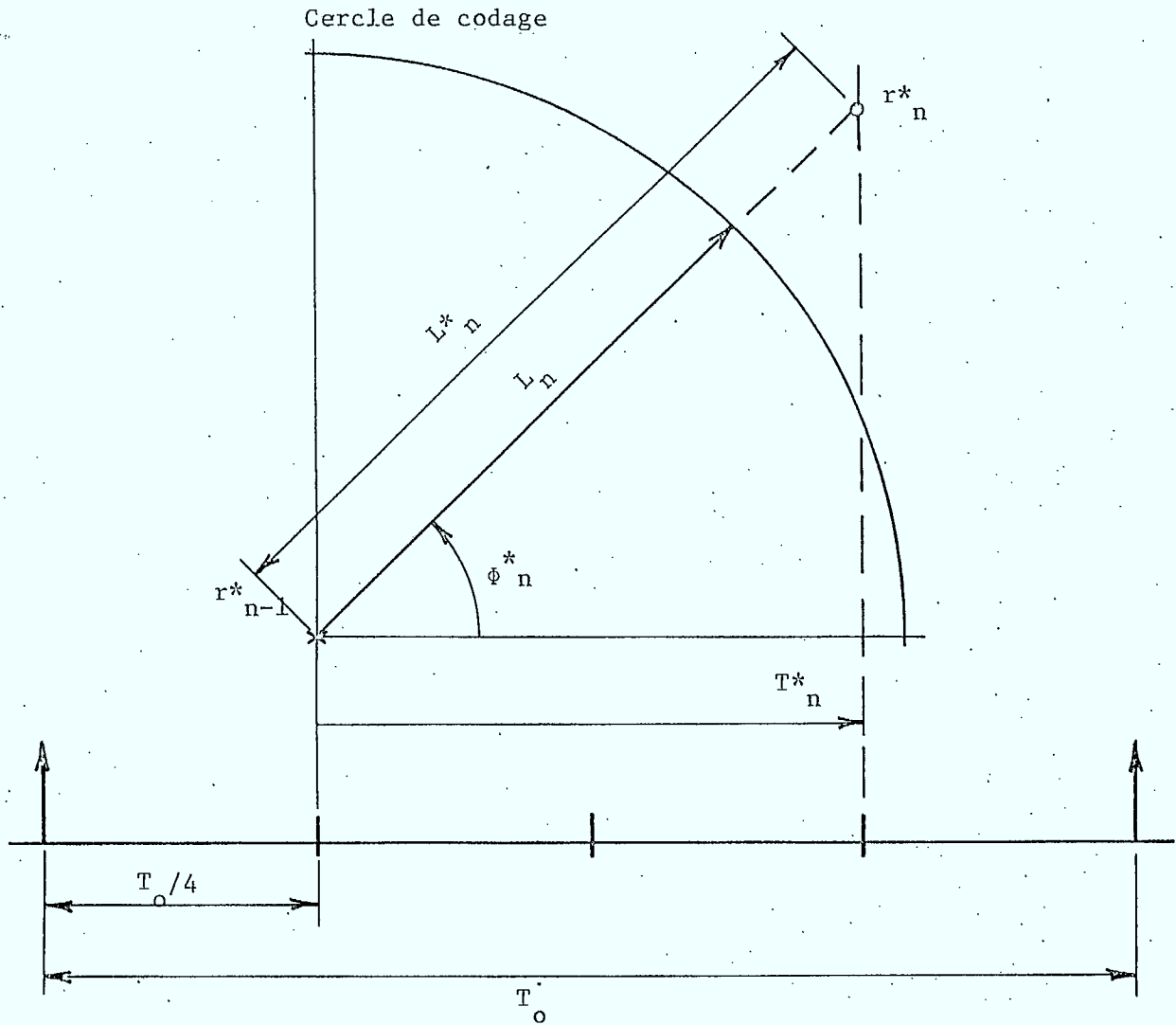


Fig. 3.3. Reconstruction du signal de parole au récepteur, à partir de variables quantifiées.

Le temps variable ou temps réel T_n^* sépare alors r_n^* de r_{n-1}^* . De fait, de cette façon, tous les échantillons intermédiaires, à rythme variable, r_i^* ($\forall i$) sont placés en synchronisme avec une trame intermédiaire d'échantillonnage constituée à 32 KHz (4 x 8 KHz).

- Enfin il faut retrouver un signal définitif de parole construit au rythme d'origine de 8 KHz. Ceci peut être fait en sous-échantillonnant le signal intermédiaire construit à 32 KHz (par interpolation linéaire entre points espacés d'un temps variable). Des précautions, analysées au chapitre 5 sur l'approche dite analogique, permettront de choisir le sous-échantillonnage (32 KHz à 8 KHz) qui correspond exactement aux instants d'échantillonnage du signal de parole avant codage. Les périodes de silence ou de bruit de fond sont utilisées à cet effet.

Si toutefois on désire simplifier le codage en évitant cette resynchronisation, le décalage de $k.T_0/4$ que l'on peut obtenir entre les trames d'origine et codée/décodée se traduit par un bruit de codage assez perceptible (voir aussi rapport 1, figure 6, page 17).

2.2.2 Adaptation et normalisation des paramètres

A la réception d'une valeur ϕ_n^* (on supprime le symbole * de quantification pour simplifier la suite).

- Le gain de normalisation d'amplitude est modifié,

$$G_n = G_{n-1} \cdot \text{tg}(\phi_n)$$

- Le rayon de codage est modifié,

$$L_n = L_{n-1} / \text{tg } \phi_n$$

- La progression en temps (temps réel) est déterminée par:

$$T_n = L_n \cdot \cos \phi_n$$

en tenant compte ultérieurement des opérations de quantification déjà décrites.

Cependant, ces nouvelles valeurs (G_n et L_n) calculées en fonction du passé ont nécessité un lissage avant de reconstruire le signal de parole pour éviter l'influence de variations d'angle trop brutales.

- Filtre de lissage

Le filtre déjà signalé est utilisé pour calculer, au codeur et au décodeur, les paramètres G_n et L_n selon la relation entrée (x_n), sortie (y_n) générale:

$$y_n = y_{n-1} + \frac{1}{P} (|x_n| - y_{n-1})$$

Cette formule revient sensiblement à définir un calcul de moyenne d'amplitude sur une fenêtre de P échantillons consécutifs. Des valeurs de P égales à 4 ou 8 sont retenues et s'avèrent satisfaisantes.

2.3 Cas particulier des silences ou bruit de fond

Dès la détection de ce cas par les deux seuils de décision

$$N_z > 10$$

$$E_n < \text{seuil} \cdot (2+B_f)$$

le rayon de codage est rendu fixe

$$T_n = 8 T_0$$

et correspond à un cercle de codage de rayon (sur l'axe du temps) 1 milliseconde. Tous les échantillons dont l'amplitude normalisée est à l'intérieur

du cercle normalisé ne seront pas codés. Seul le premier échantillon hors du cercle peut être codé. La figure 3.4 montre ce résultat, sur une figure "dénormalisée" où le cercle est une ellipse de codage.

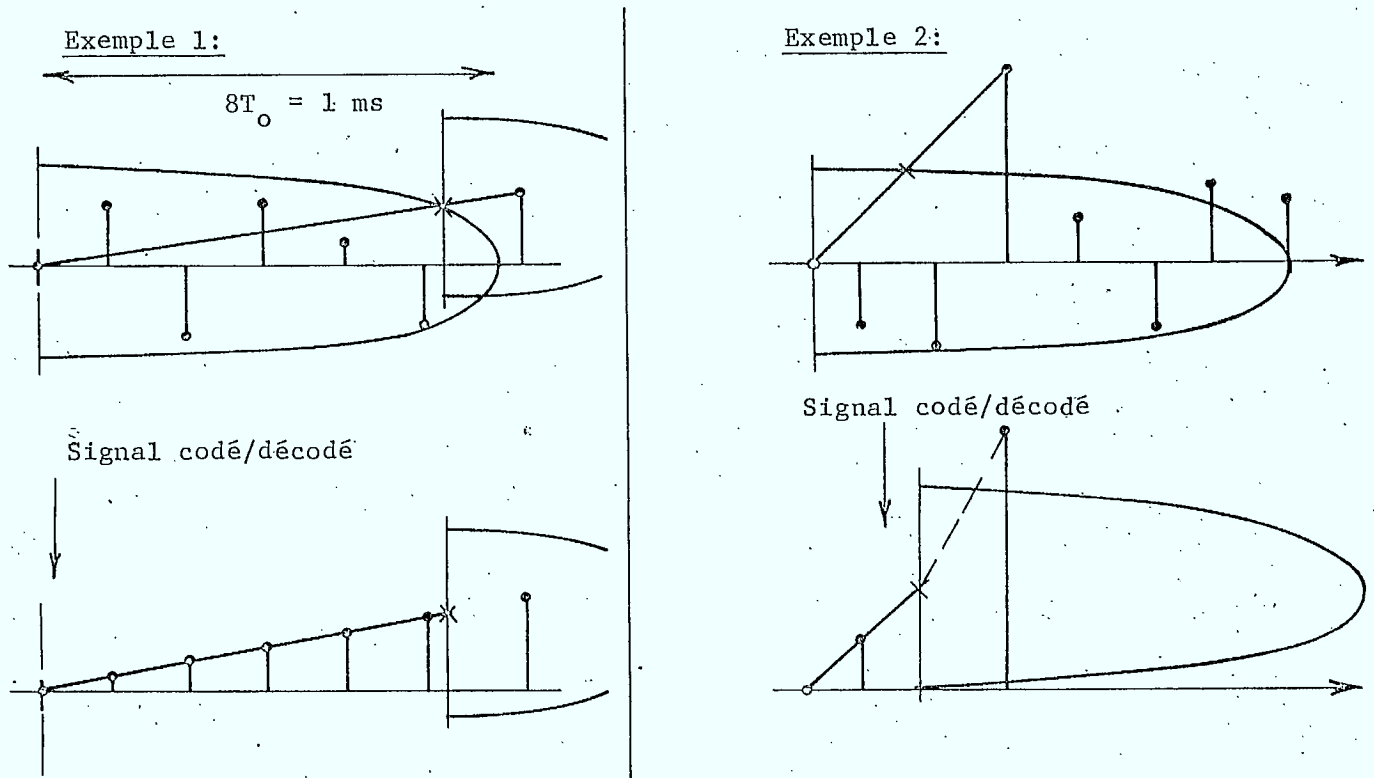


Fig. 3.4 Traitement des silences.

Cette technique permet de vider rapidement la mémoire, tout en autorisant le codage immédiat des pointes de signal qui, par exemple, pourraient représenter une consonne plosive, très brève et d'énergie supérieure au seuil de silence.

2.4 Réinitialisation des paramètres de codage

Lors du codage du signal de parole, incluant les silences, certains tests sont effectués pour assurer un codage le plus efficace possible

- Dans le cas d'une mémoire remplie à capacité maximale tout sur-échantillonnage est évidemment interdit, quelque soit la dégradation obtenue sur le signal codé/décodé.
- A chaque changement de signe de l'angle $\phi_n(t)$, à condition que l'écart dépasse un certain seuil (adaptatif: dépendant du remplissage de la mémoire), le codeur et donc le décodeur réinitialise le rayon de codage L_n à sa valeur minimale L_0 correspondant, sur l'axe du temps à la période d'échantillonnage naturelle d'origine.
- Si en plus un état particulier de signal est reconnu au nombre des passages par zéro élevé $N_z > 10$ le gain est simultanément figé à une valeur adaptée fixée à

$$G_n = \frac{1}{2} G_{\max} \cdot (3 + B_f)$$

avec G_{\max} déjà déterminé par l'amplitude moyenne du signal reconstitué à 8 KHz. Ce cas est représentatif des phonèmes tels que /CH/:

- fortement énergétique,
- de nature bruitée (N_z élevée),

et qui pose le pire problème de codage en remplissant très rapidement la mémoire par sur-échantillonnage permanent.

Cette solution permet de se soustraire à cette difficulté que le rapport 1 illustre, page 27-28, figure 12.

2.5 Organigramme de décodage

Le décodeur, simulé par un sous-programme en langage FORTRAN, est représenté par un organigramme, (figure 3.5) où l'on reconnaîtra les divers points analysés dans chacun des paragraphes de ce chapitre.

Ce sous-programme est utilisé autant pour le décodeur local, associé à l'émetteur, que pour le récepteur distant. Il ne reçoit comme information nouvelle que la valeur quantifiée de manière optimale (K-moyenne) de l'angle $\phi_n(t)$ au rythme fixe de 8 KHz.

Tous les autres paramètres du codage sont calculés de manière interne au décodeur. Il s'agit donc d'un processus adaptatif dont la robustesse aux erreurs de transmission (non simulées dans cette étude) reste à étudier.

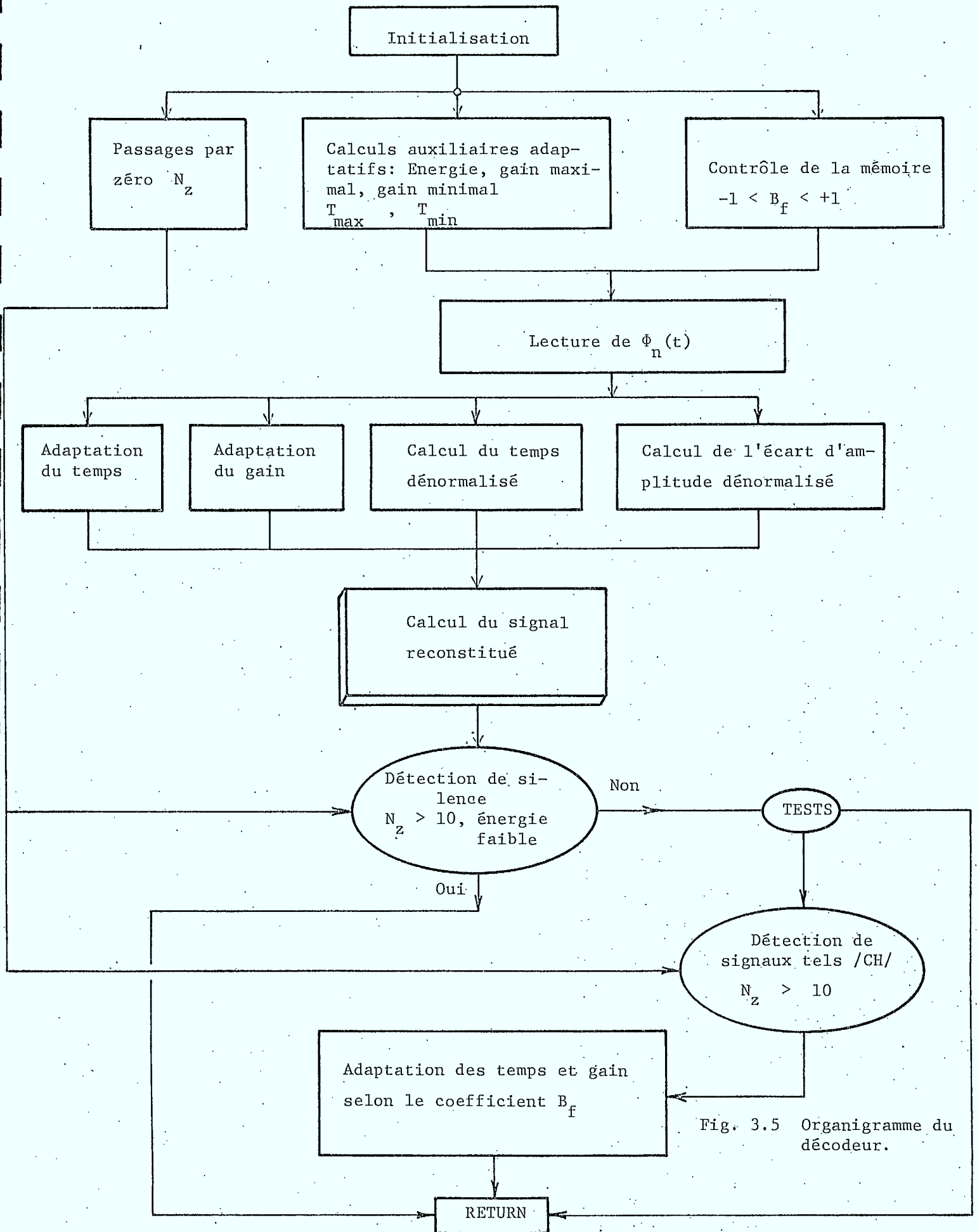


Fig. 3.5 Organigramme du décodeur.

CHAPITRE 4

RESULTATS DE CODAGE NUMERIQUE

On conviendra, au niveau objectif, de retenir la mesure du rapport "signal sur bruit" (S/N), comme élément de comparaison. Cependant en raison de certains codages particuliers des silences entre autres, et des phonèmes tels que /CH/ (forte énergie, nombreux passages par zéro) - voir chapitre 3 - on peut s'intéresser au:

"RAPPORT SIGNAL SUR BRUIT GRANULAIRE"

qui, de fait, dans les calculs de détermination de bruit limite les opérations aux séquences de signal où "le bruit se présente de manière comparable aux cas de codage conventionnel tels que le PCM ou MIC". Ce cas spécifique donnera un élément de comparaison plus valable dans les comparaisons effectuées avec le MIC ou PCM. On le notera S/N_g .

Un autre type de calcul (S/N) segmental a été proposé [*] et est présenté aussi en référence sur les figures selon la présentation déjà illustrée dans le rapport 1, page 30 et 31.

1. Performance de base du codage - Qualité-débit

Sur l'ensemble de phrases, on choisit d'abord une quantification linéaire à 12 bits de l'angle $\phi_n(t)$ pour définir les performances maximales en (S/N) du principe de codage.

[*] McDermott, B., Scagliola, C., Goodman, D., - "Perceptual and objective evaluation of speech processed by adaptative differential PCM", BSTJ, Volume 57, #5, pp. 1597-1618, (May-June 1978).

Avec les phrases suivantes

- Phrase 1 Il se garantira du froid avec ce bon capuchon,
- Phrase 2 Annie s'ennuie loin de ses parents,
- Phrase 3 Les deux camions se sont heurtés de face,
- Phrase 4 Un loup s'est jeté immédiatement sur la petite chèvre
- Phrase 5 Dès que le tambour bat, les gens accourent,
- Phrase 6 Mon père m'a donné l'autorisation,
- Phrase 7 Vous poussez des cris de colère,
- Phrase 8 Le petit canard apprend à nager,
- Phrase 9 La voiture s'est arrêtée au feu rouge,
- Phrase 10 La vaisselle propre est mise sur l'évier

et un ensemble de 3 locuteurs (hommes H1, H2, H4) et de 3 locutrices (femmes F3, F5, F6).

1.1 Quantification linéaire à 12 bits

On peut établir un tableau des rapports (S/N et (S/N)_g caractéristiques.

	1	2	3	4	5	6	7	8	9	10
H1	24.3									
H2		25.0								23.6
F3									27.3	
H4				21.9				25.1		
F5			27.7		27.5		26.5			
F6						25.7				

Tableau 1. Résultats objectifs, (S/N) avec 12 bits de quantification linéaire $\phi_n(T)$.

	1	2	3	4	5	6	7	8	9	10
H1	29.1									
H2		29.3								27.7
F3									30.8	
H4				28.5				28.0		
F5			31.7		29.0		30.0			
F6						28.2				

Tableau 2. Résultats objectifs $(S/N)_g$ dans les mêmes conditions.

1.2 Quantification linéaire à 8 bits

Pour comparer plus efficacement avec le MIC ou PCM à 8 bits, les tableaux 3 et 4 représentent une distribution des résultats à débit équivalent sans optimisation de la quantification par K-moyenne

	1	2	3	4	5	6	7	8	9	10
H1	24.5									
H2		25.6								23.7
F3									27.3	
H4				22.6				25.3		
F5			27.3		27.8		26.5			
F6						25.7				

Tableau 3. Résultats objectifs (S/N) avec 8 bits de quantification linéaire.

	1	2	3	4	5	6	7	8	9	10
H1	29.3									
H2		29.5								27.8
F3									30.4	
H4				28.39				27.8		
F5			30.8		30.1		29.2			
F6						28.0				

Tableau 4. Résultats objectifs $(S/N)_g$ avec 8 bits de quantification linéaire.

On notera que le codage MIC (PCM) attribue une qualité (S/N) et $(S/N)_g$ voisine de 37 dB à 38 dB pour une dynamique normale de la parole.

On constate la qualité raisonnable du principe.

1.3 Quantification à 6 bits linéaire et optimisée par une K-moyenne

1.3.1 Quantification linéaire

On limitera le processus de calcul à un échantillonnage de deux phrases et de 2 locuteurs (homme et femme).

Par exemple: Phrase 2
 Phrase 4
 Locuteur H4
 Locutrice F6

Les tableaux 5 et 6 représentent les résultats obtenus avec 6 bits (64 niveaux de quantification) représentant une répartition linéaire des valeurs codées de $\phi_n(t)$.

Locuteur/locutrice	Phrases	2	4
	H4	25.01	22.26
	F6	26.21	23.31

Tableau 5. Résultats objectifs: (S/N) avec 6 bits quantification linéaire.

Locuteur/locutrice	Phrases	2	4
	H4	27.15	26.35
	F6	28.26	26.90

Tableau 6. Résultats objectifs (S/N)_g avec 6 bits de quantification linéaire.

1.3.2 Quantification optimale

L'algorithme de K-moyenne, prévu pour un nombre maximal de 64 représentants est utilisé pour une détermination optimale des niveaux de quantification.

L'ensemble des fichiers de parole peut être utilisé pour les calculs de "convergence des échantillons" vers la position optimale finale.

On relève alors dans les tableaux 7 et 8 les résultats de mesures objectives.

Locuteur/locutrice	Phrases	2	4
K-moyenne →	H4	25.09	21.63
	F6	25.92	22.94

Tableau 7. Résultats objectifs, (S/N) avec 6 bits quantification optimale par K-moyenne

Locuteur/locutrice	Phrases	2	4
K-moyenne →	H4	27.84	26.46
	F6	28.49	26.46

Tableau 8. Résultats objectifs, $(S/N)_g$ avec 6 bits quantification optimale par K-moyenne.

1.3.3 Conclusion

On constate que la quantification de l'angle $\phi_n(t)$ à 8 bits ne dégrade pas sensiblement le niveau de qualité de codage par rapport à une définition de 12 bits et plus. Une réduction à 6 bits conserve même le niveau de qualité initial.

L'écoute sur cassette montre que ces codages se comparent avantageusement au MIC (PCM) de référence à 8 bits bien que les valeurs relevées du rapport S/N soient différentes. Ce fait provient de codages spécialisés (silence, /CH/...) qui contribuent à diminuer le rapport S/N mais n'ont pas d'influence subjective.

Les codages à 4 bits (32 Kbit/s) sont encore qualitativement raisonnables. L'optimisation pour K-moyenne ne semble pas apporter d'amélioration ce qui s'explique par les quantifications auxiliaires (des temps) et les techniques d'adaptation de l'algorithme.

1.4 Quantification à 4 bits linéaire et optimisée par une K-moyenne

Dans le but final d'une transmission à 32 Kbit/s, la quantification est réduite à 4 bits par échantillon $\phi_n(t)$.

1.4.1 Quantification linéaire et moyenne

On relèvera, dans un même tableau, les chiffres de rapport S/N et $(S/N)_g$ granulaire cité entre parenthèse:

Locuteurs	Phrases	2	4
	H4	19.18 (19.62)	17.00 (18.80)
	F6	19.70 (20.30)	18.28 (19.56)

Tableau 9. Résultats objectifs S/N et $(S/N)_g$ (entre parenthèses) avec 4 bits (16 états) de quantification linéaire.

1.4.2 Quantification optimale (K-moyenne) à 4 bits, qualité globale de codage

Le tableau 10 présente les résultats optimaux relevés. Si l'on réduit le codage PCM (MIC) de 8 à 4 bits, les rapports (S/N) et $(S/N)_g$ diminuent de 6 dB par bit abandonné.

Locuteurs	Phrases	2	4
K-moyenne →	H4	17.20 (21.85)	14.70 (19.79)
	F6	18.28 (21.73)	15.45 (20.31)

Tableau 10. Résultats objectifs, (S/N) et $(S/N)_g$ avec K-moyenne et 4 bits.

2. Effet de la taille de la mémoire sur qualité

La dimension de la mémoire influe directement sur la qualité de codage. En effet une mémoire de longue durée autorise, par sur-échantillonnage un codage de haute qualité et de plus, étant en niveau relatif (coefficient B_f de -1 à +1) moins souvent remplie, les coefficients intermédiaires du codage (gain maximal et minimal, seuil adapté de silence...) sont plus petits. Il en résulte une définition plus précise des échantillons codés.

La figure 4.1, 4.2 et 4.3 représentent, pour la même phrase (P4), avec la même locutrice (F3), l'évolution du coefficient de remplissage de la mémoire et de l'angle $\phi_n(t)$.

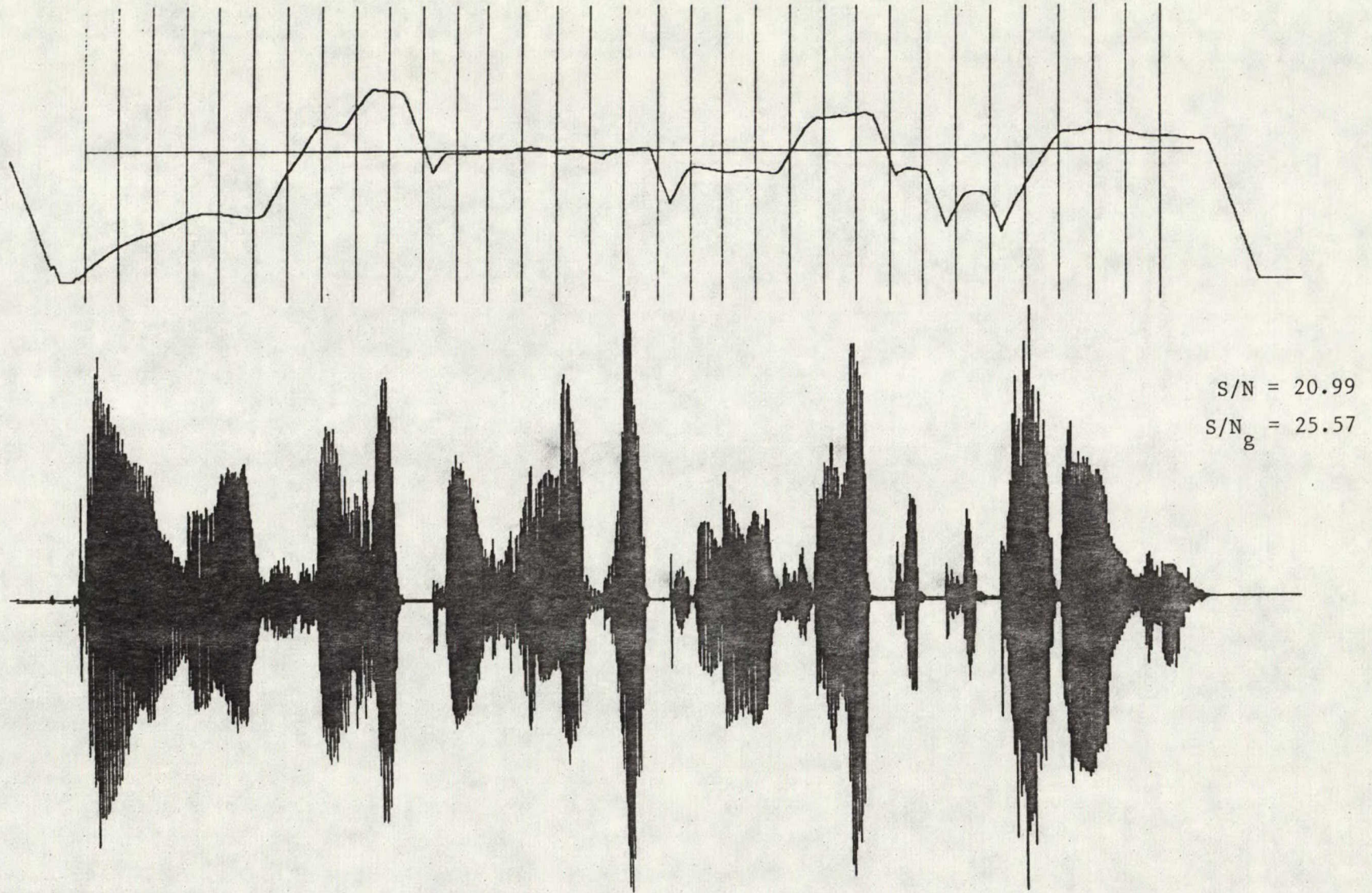


Fig. 4.1 Codage avec 6 bits de $\phi_n(t)$. Représentation du remplissage de la mémoire de 1024 mots et signal de parole.

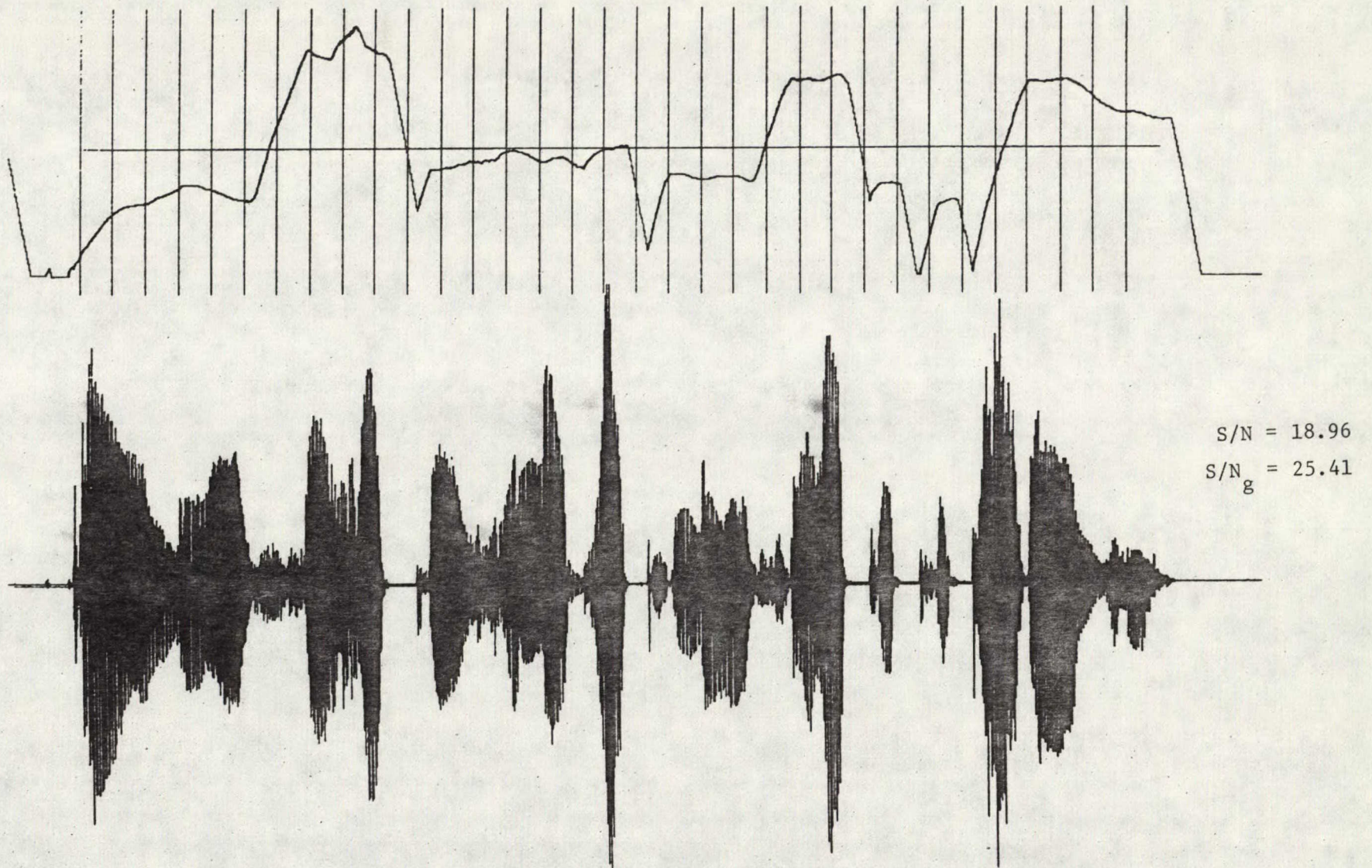


Fig. 4.2 Codage avec 6 bits de $\phi_n(t)$. Représentation du remplissage de la mémoire de 512 mots et angle $\phi_n(t)$.

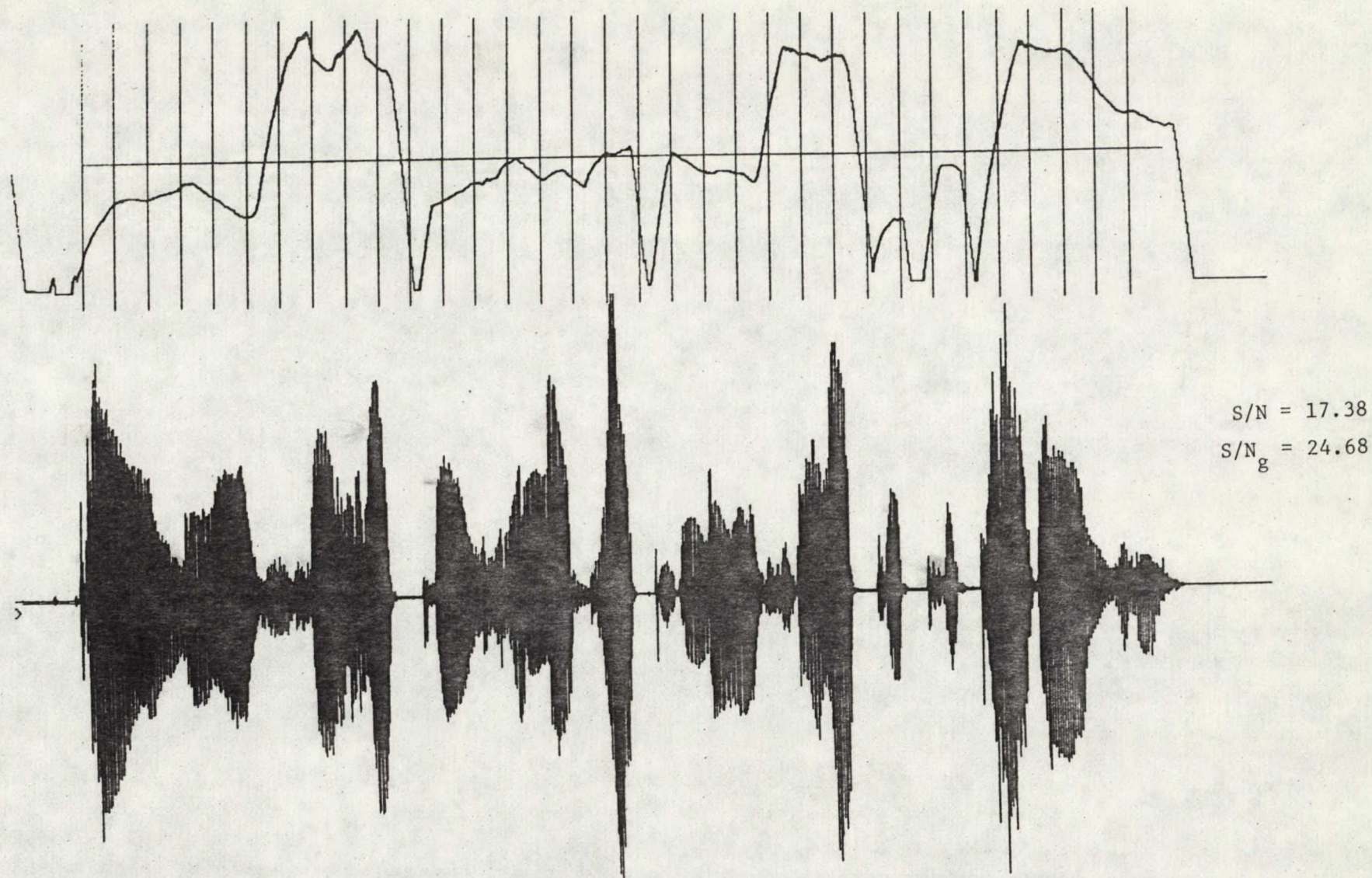


Fig. 4.3 Codage avec 6 bits de $\phi_n(t)$. Représentation du remplissage de la mémoire de 256 mots et angle $\phi_n(t)$.

3. Caractéristiques du signal transmis $\phi_n(t)$

Le signal d'angle est transmis sous forme numérique directe, après éventuellement, une détermination optimale (K-moyenne) des niveaux quantifiés. On peut analyser ce signal afin de déterminer si, à la place d'un codage direct, il ne serait pas préférable d'envisager un codage différentiel adaptatif pour réduire encore le débit transmis.

3.1 Analyse du spectre et de l'autocorrélation du signal

Sur quelques échantillons de parole caractéristique, soient les phonèmes /A/, /I/, /S/, /CH/, on a donc relevé, sur une séquence de longue durée, l'évolution du spectre calculée par "Transformée de Fourier Rapide" (F.F.T.) et représenté sur la série des figures 4.4 à 4.6 pour les /A/, /S/ et /CH/.

On y constate la présence de hautes fréquences, particulièrement dans la bande 3300-4000 Hz, soit hors de la bande téléphonique normalisée.

Simultanément, les figures 4.7, 4.8, 4.9 et 4.10 représentent les fonctions d'autocorrélation, montrent la "non-prédictibilité" du signal qui, de toute évidence, est très faiblement corrélé entre échantillons consécutifs.

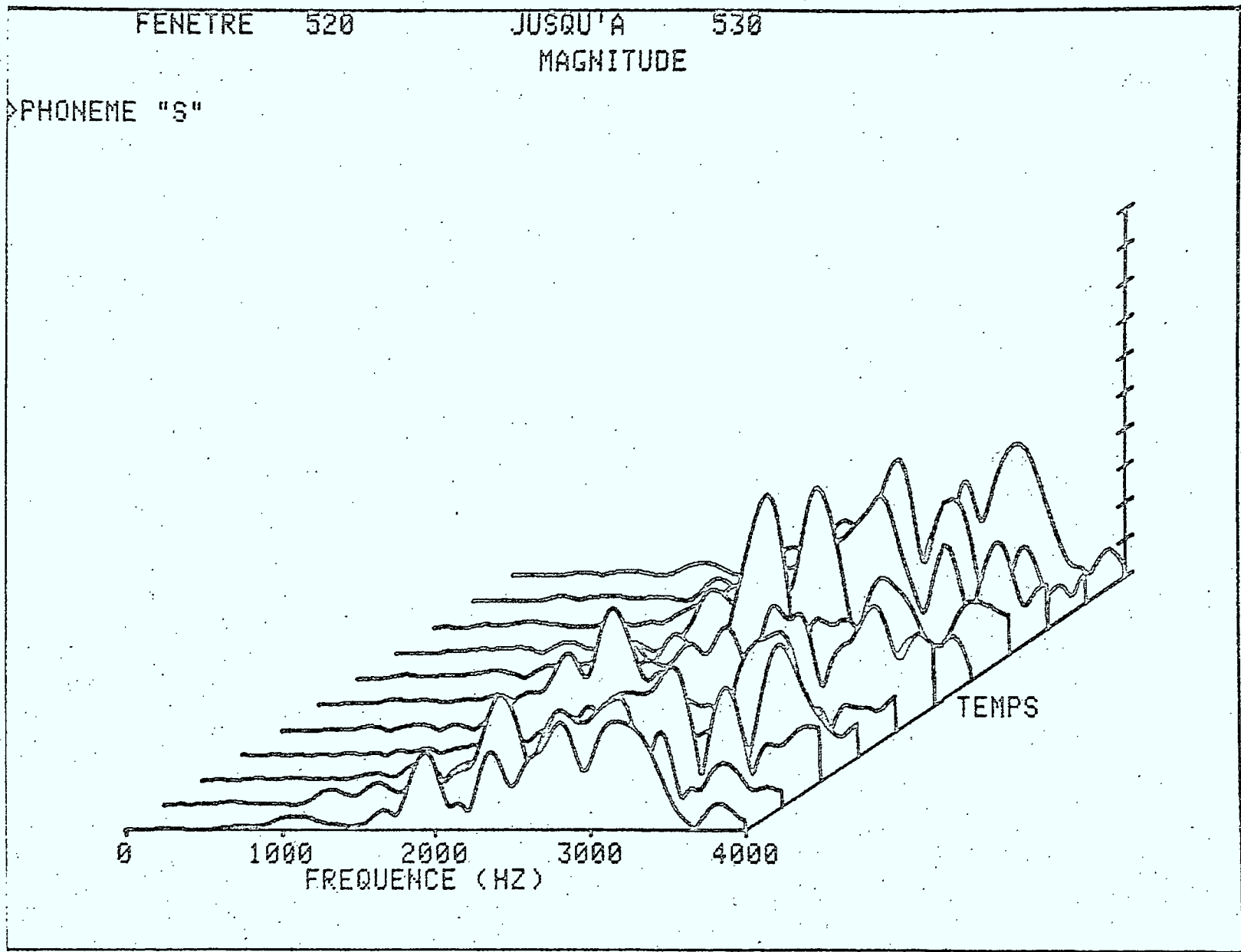


Fig. 4.4 Spectre de $\phi_n(t)$ pour le phonème /S/.

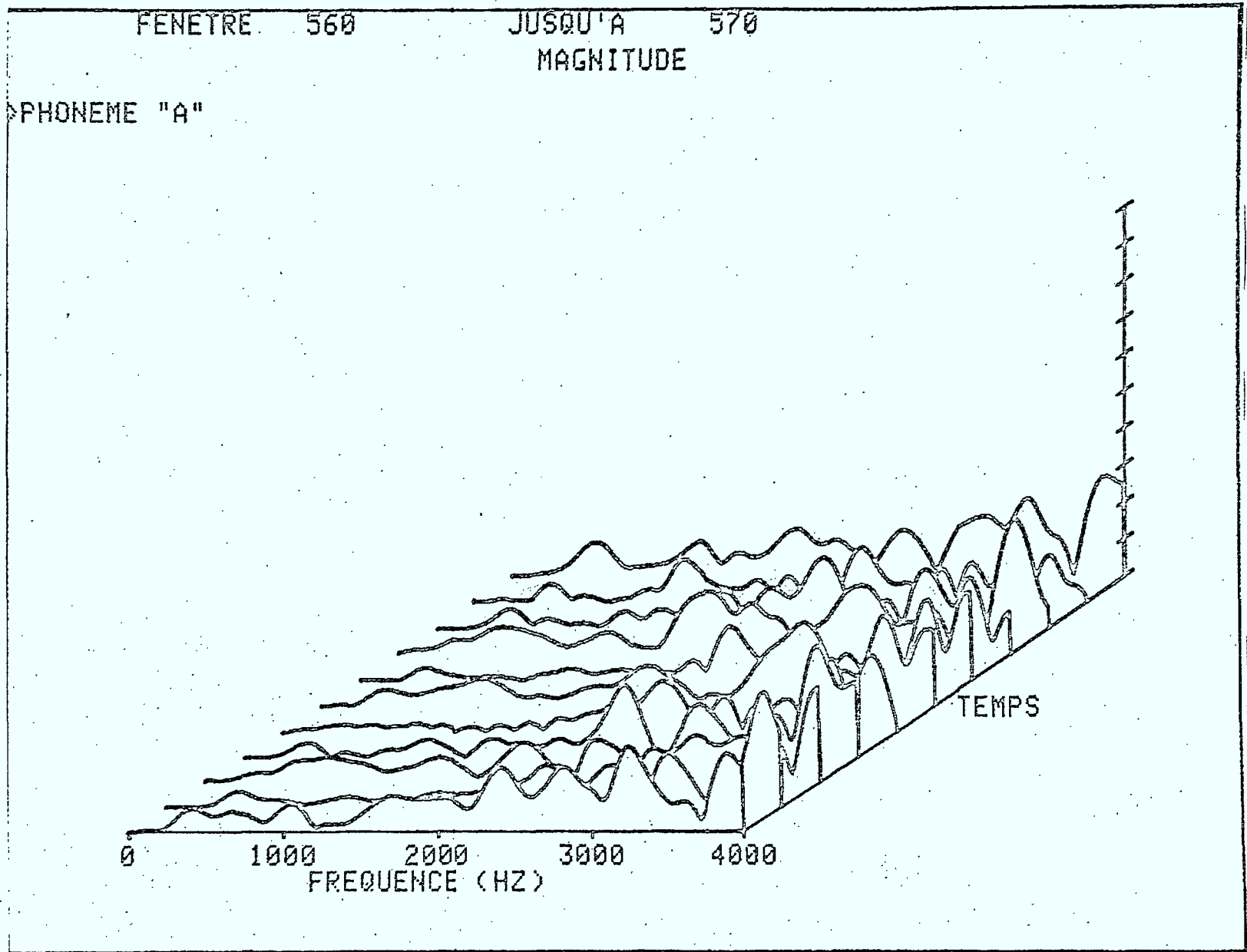


Fig. 4.5 Spectre de $\phi_n(t)$, phonème /A/.

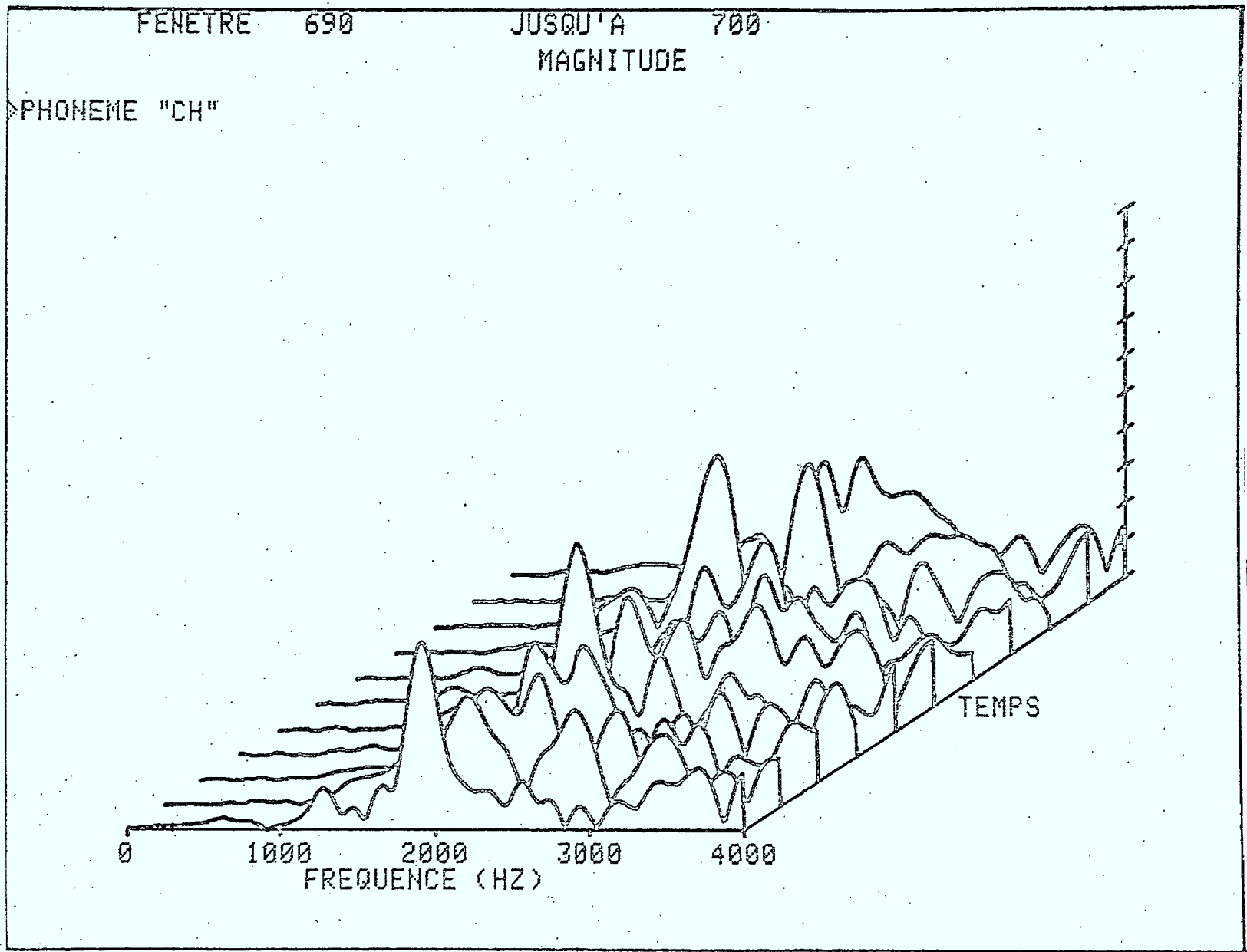


Fig. 4.6 Spectre de $\phi_n(t)$, phonème /CH/.

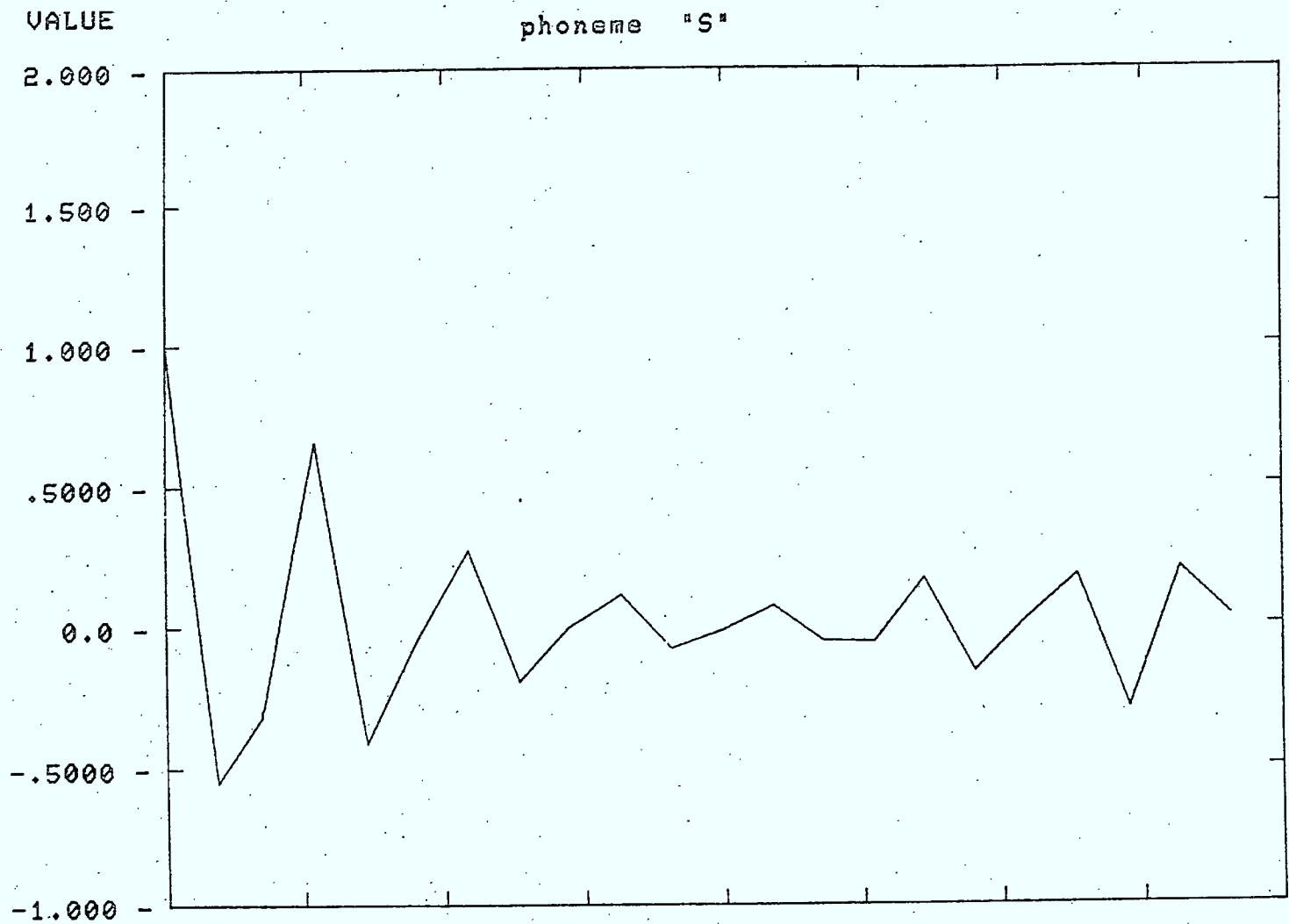


Fig. 4.7 Fonction d'autocorrélation de $\phi_n(t)$, phonème /S/.

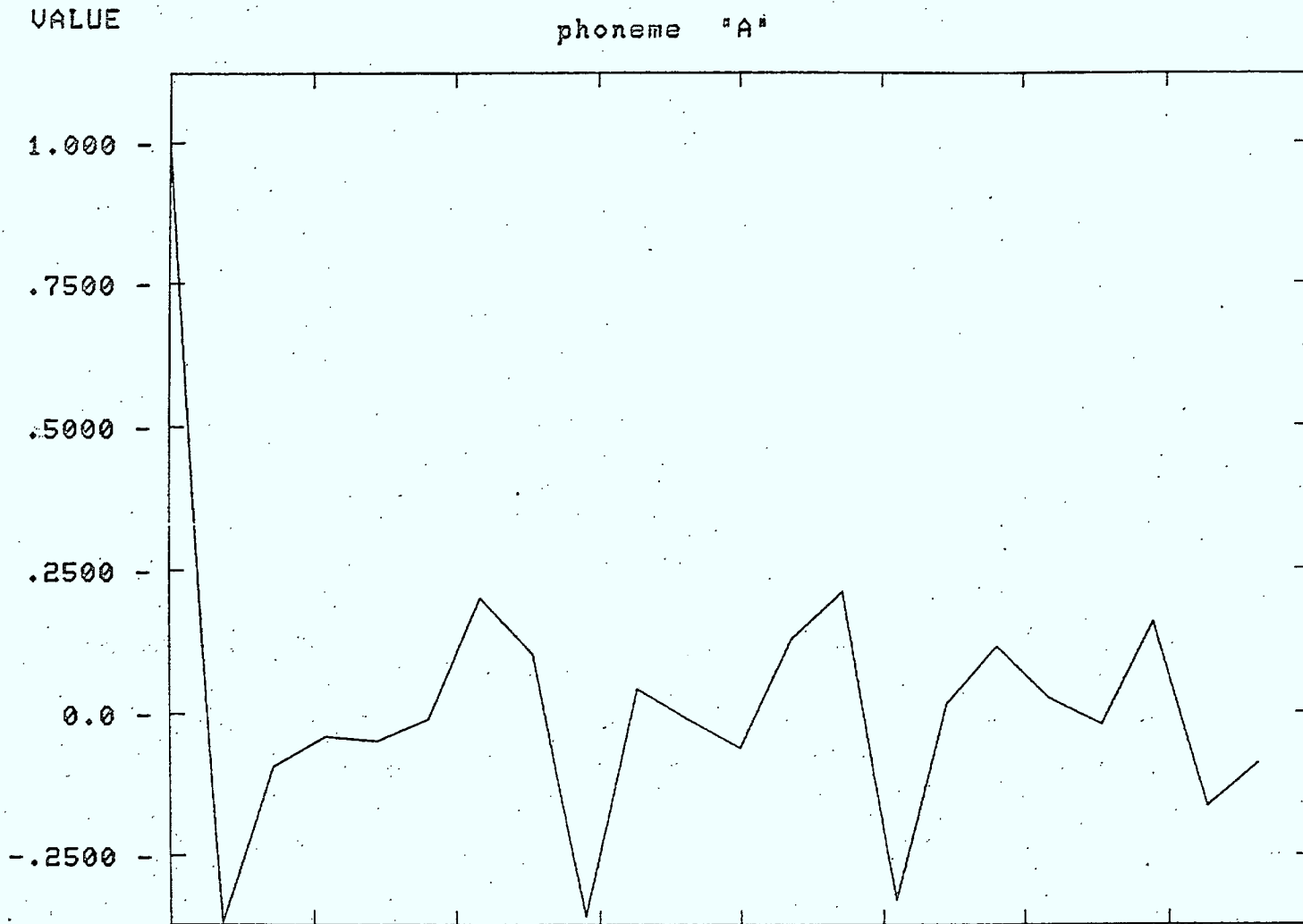


Fig. 4.8 Fonction d'autocorrélation de $\phi_n(t)$, phonème /A/.

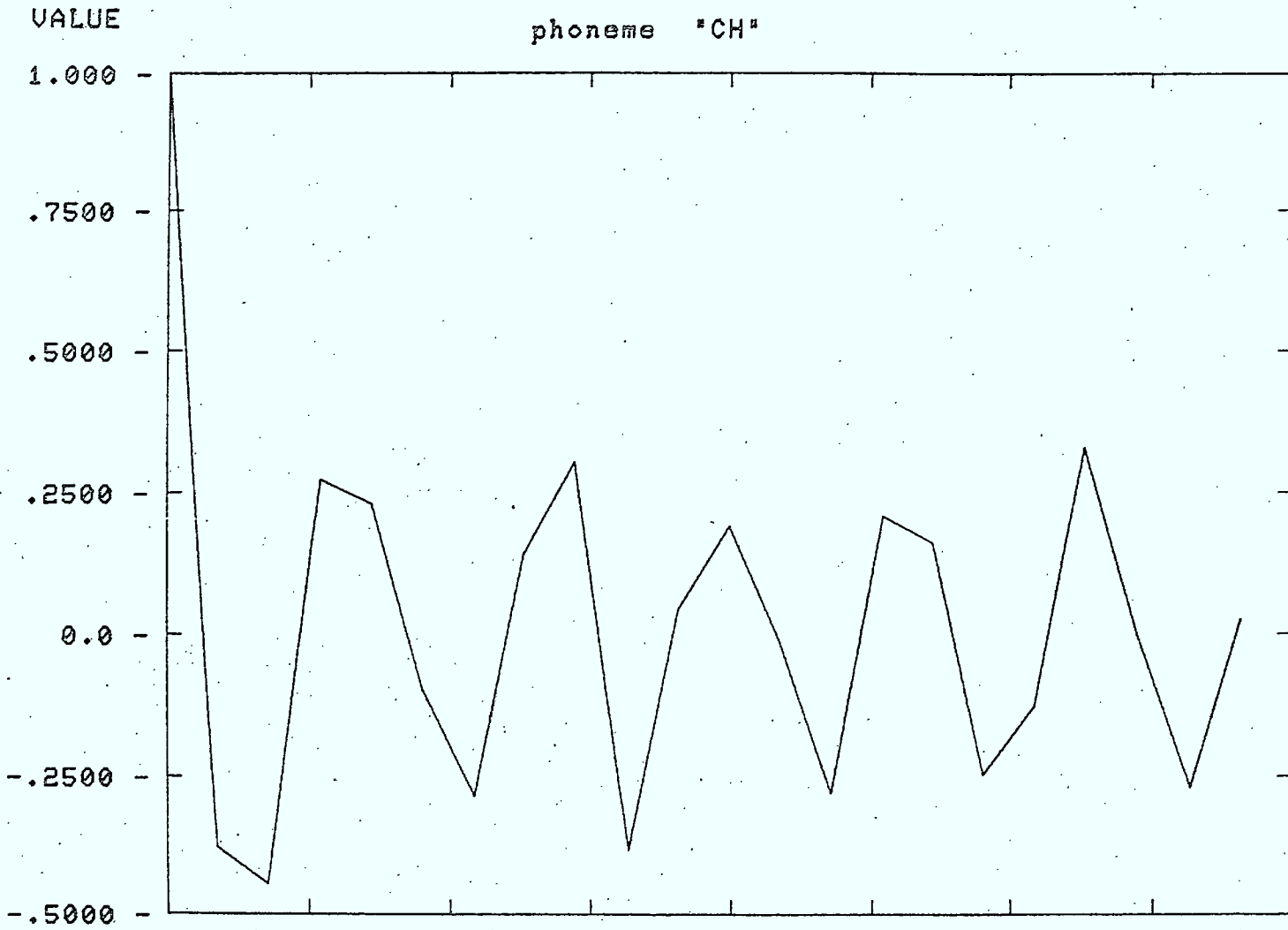


Fig. 4.9 Fonction d'autocorrélation de $\phi_n(t)$, phonème /CH/.

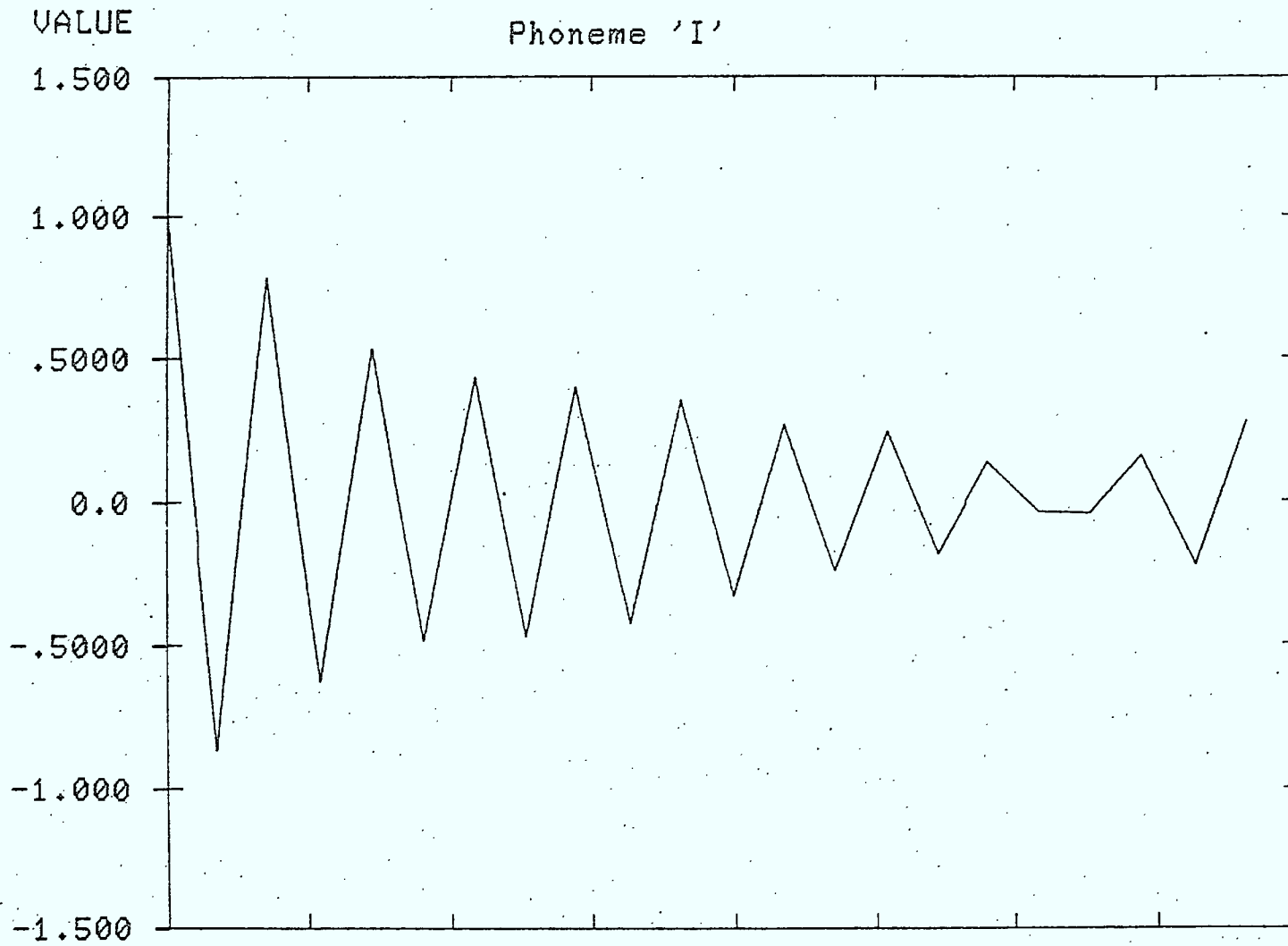


Fig. 4.10 Fonction d'autocorrélation $\phi_n(t)$, phonème /I/.

3.2 Signal $\phi_n(t)$, "alterné"

L'analyse du spectre à forte énergie du côté des hautes fréquences suggère une modification du signal dans le but de "retourner le spectre symétriquement par rapport au milieu de la bande passante". Cette opération est réalisée très simplement par la modification en alternance des signes du signal.

Plus précisément on génère, avant transmission, un nouveau signal:

$$(-1)^n \phi_n(t)$$

et le récepteur qui n'a pas besoin d'être synchronisé reconstitue $(-1)^m \phi_n(t)$. (Si $m = n+1$, le signal finalement décodé sera l'opposé en signe du signal de parole d'origine).

L'effet de l'alternance de signes contribue à donner une meilleure configuration du spectre, comme le montrent les figures 4.11 à 4.13.

L'énergie principale est concentrée sensiblement au milieu de la bande téléphonique utile. Cette configuration ne prend un intérêt capital que si l'on envisage une transmission sous forme analogique de l'angle $\phi_n(t)$. Cette étude est réservée au chapitre 5.

Malheureusement, le "retournement" du spectre n'améliore pas la "prédictibilité" du signal comme le montrent les figures 4.14 à 4.17, où sont représentées les fonctions d'autocorrélation du signal " $\phi_n(t)$ alterné" pour les phonèmes /S/, /A/, /CH/, /I/.

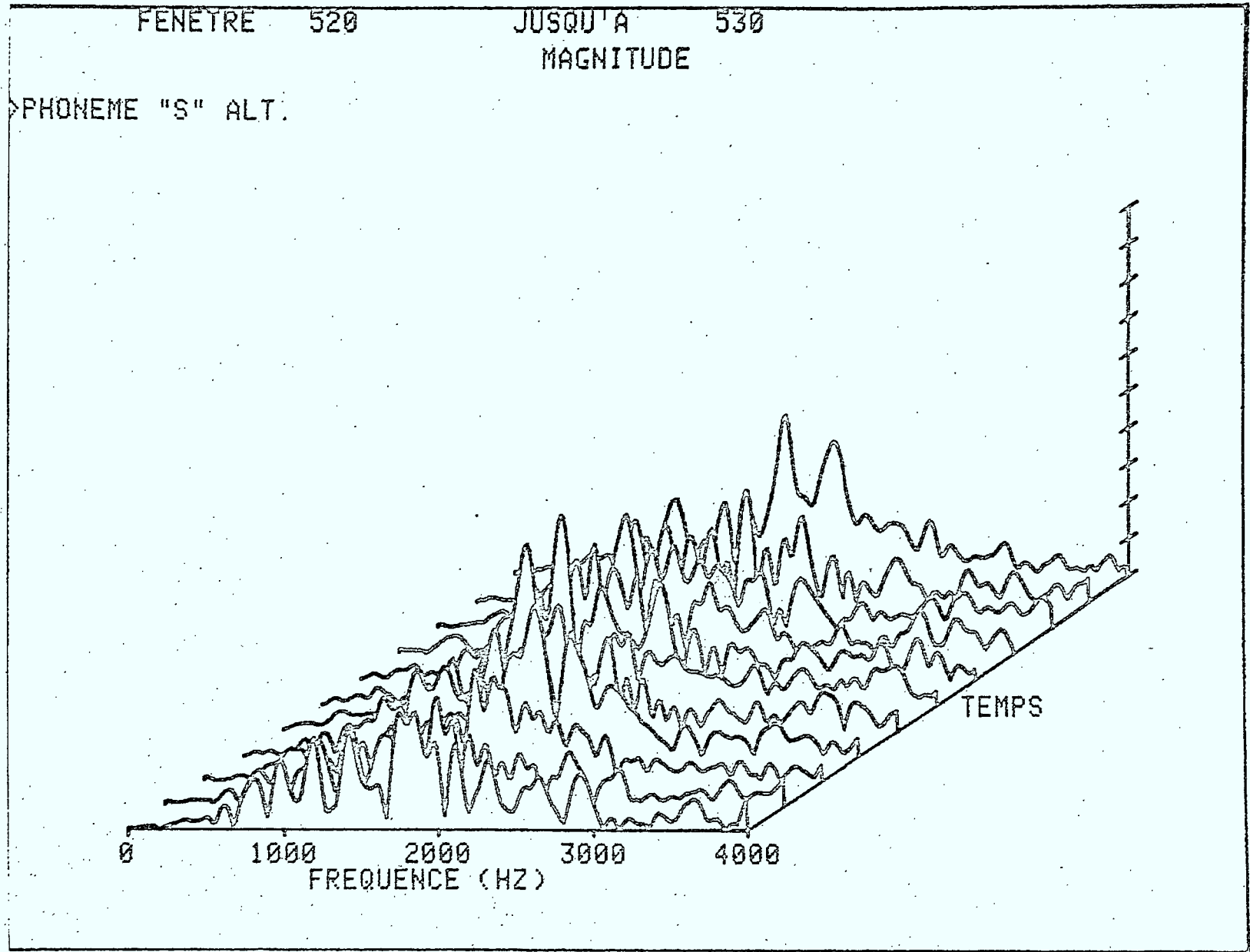


Fig. 4.11 Spectre de " $\phi_n(t)$ alterné", phonème /s/.

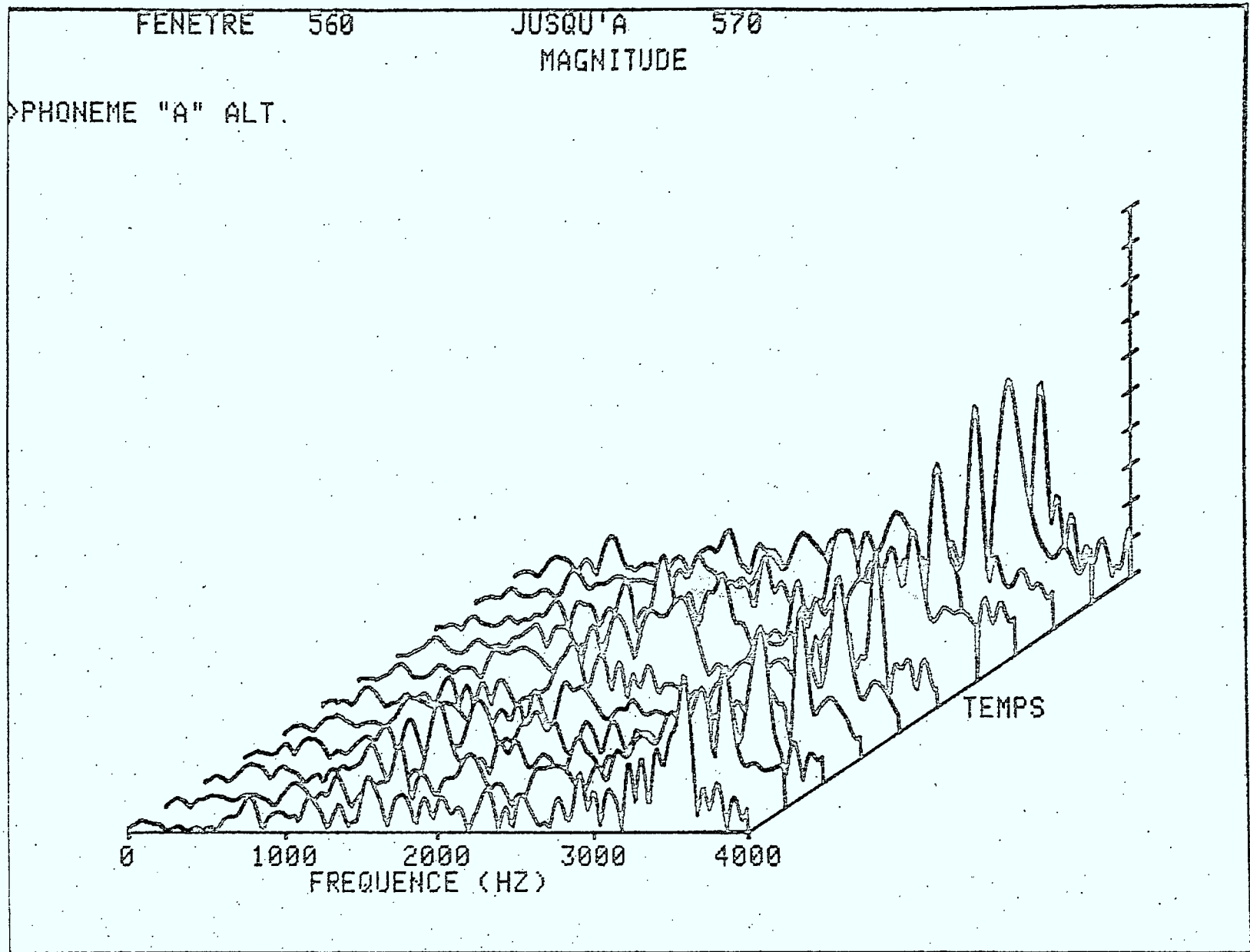


Fig. 4.12 Spectre de " $\phi_n(t)$ alterné", phonème /A/.

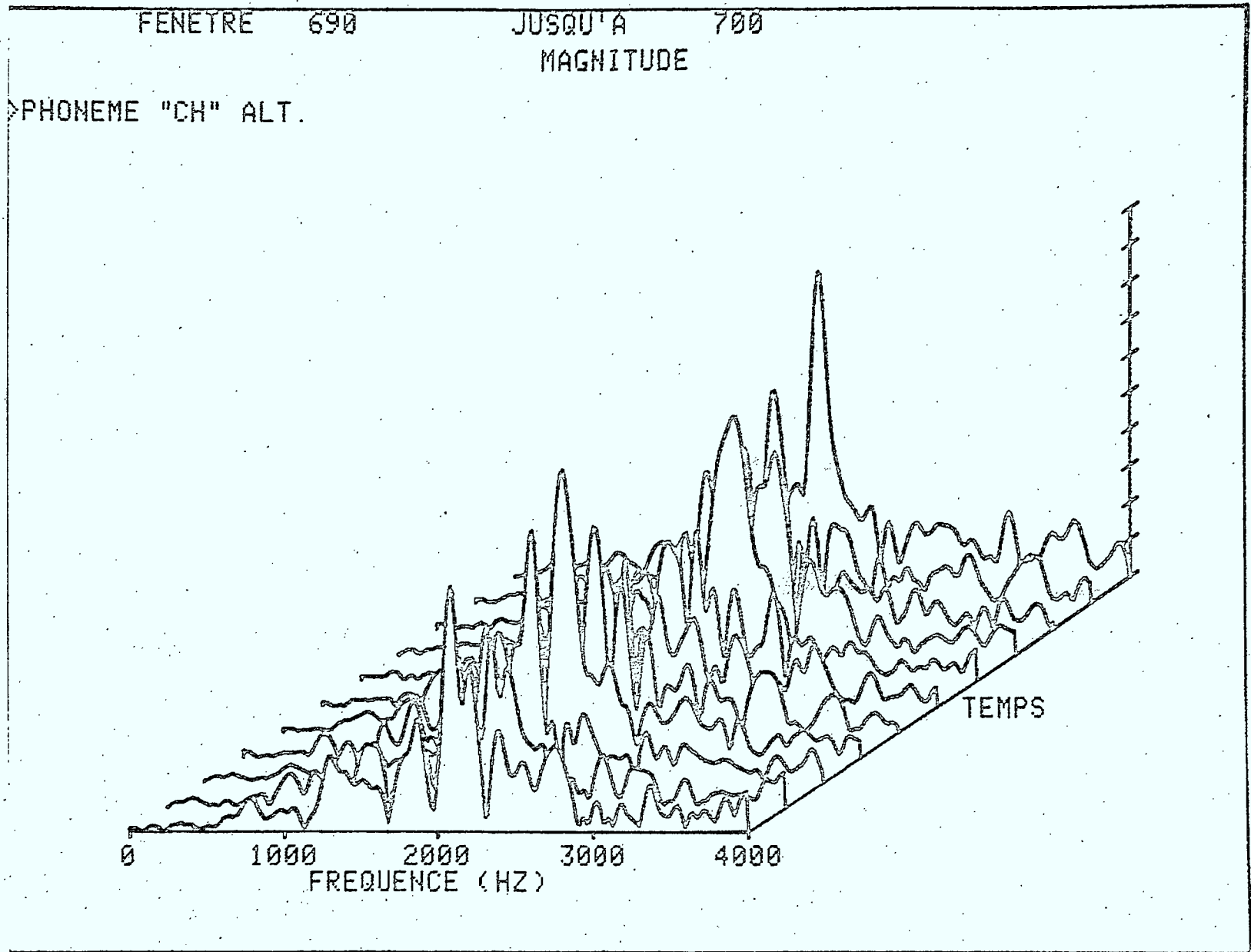


Fig. 4.13 Spectre de " $\phi_n(t)$ alterné", phonème /CH/.

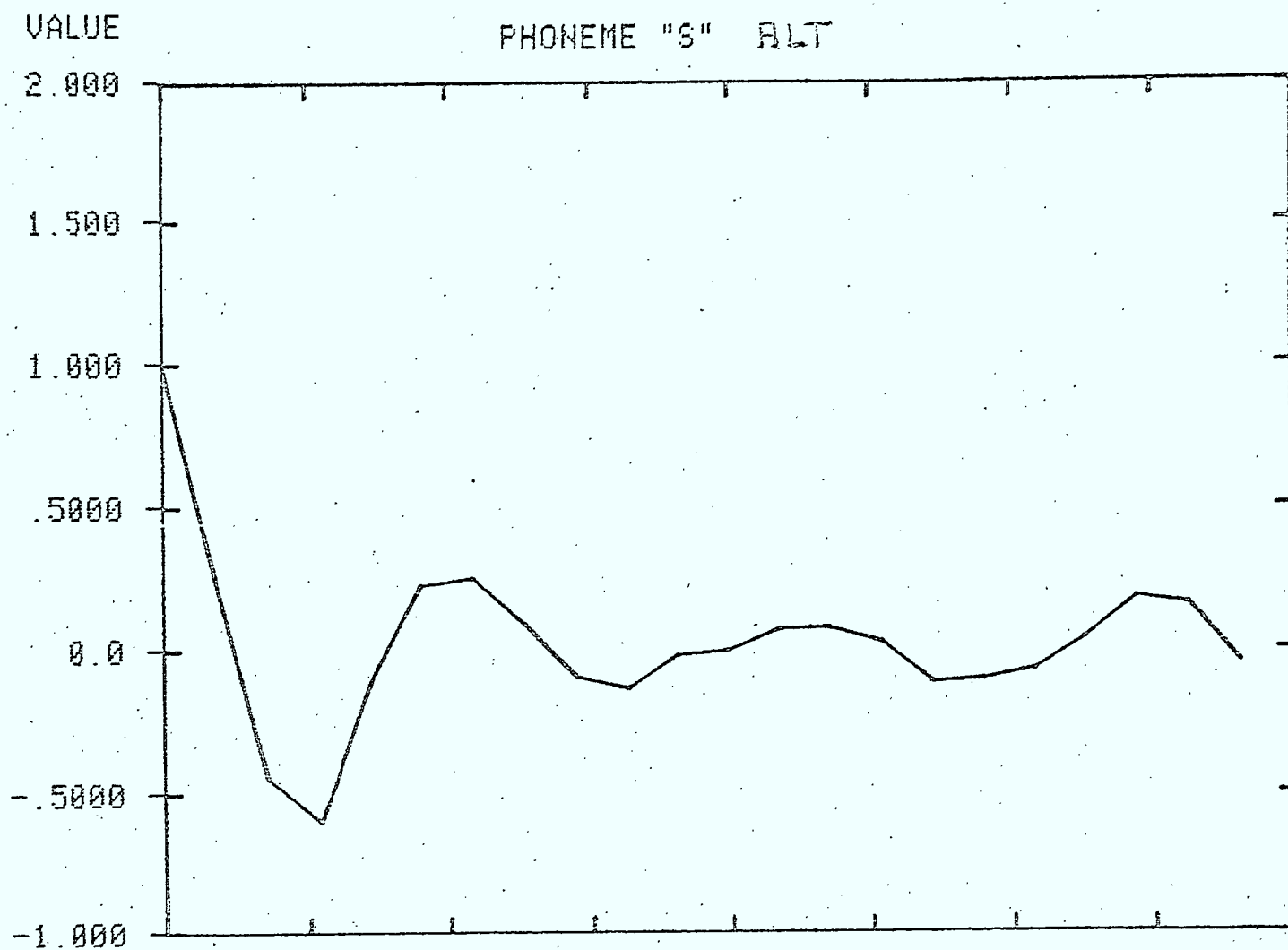


Fig. 4.14 Autocorrélation de " $\phi_n(t)$ alterné", phonème /S/.

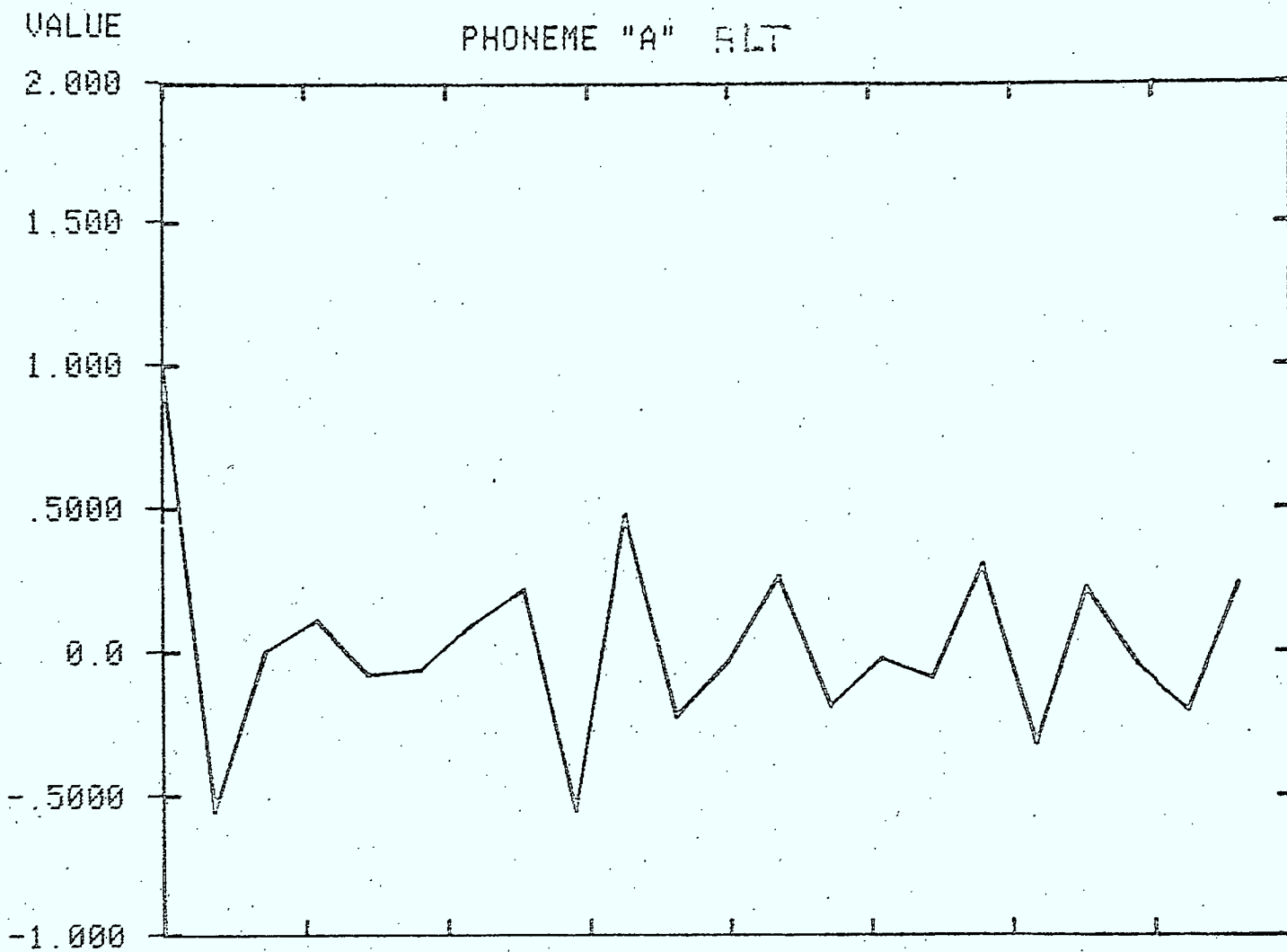


Fig. 4.15 Autocorrélation de " $\phi_n(t)$ alterné", phonème /A/.

VALUE

PHONEME "CH" ALT

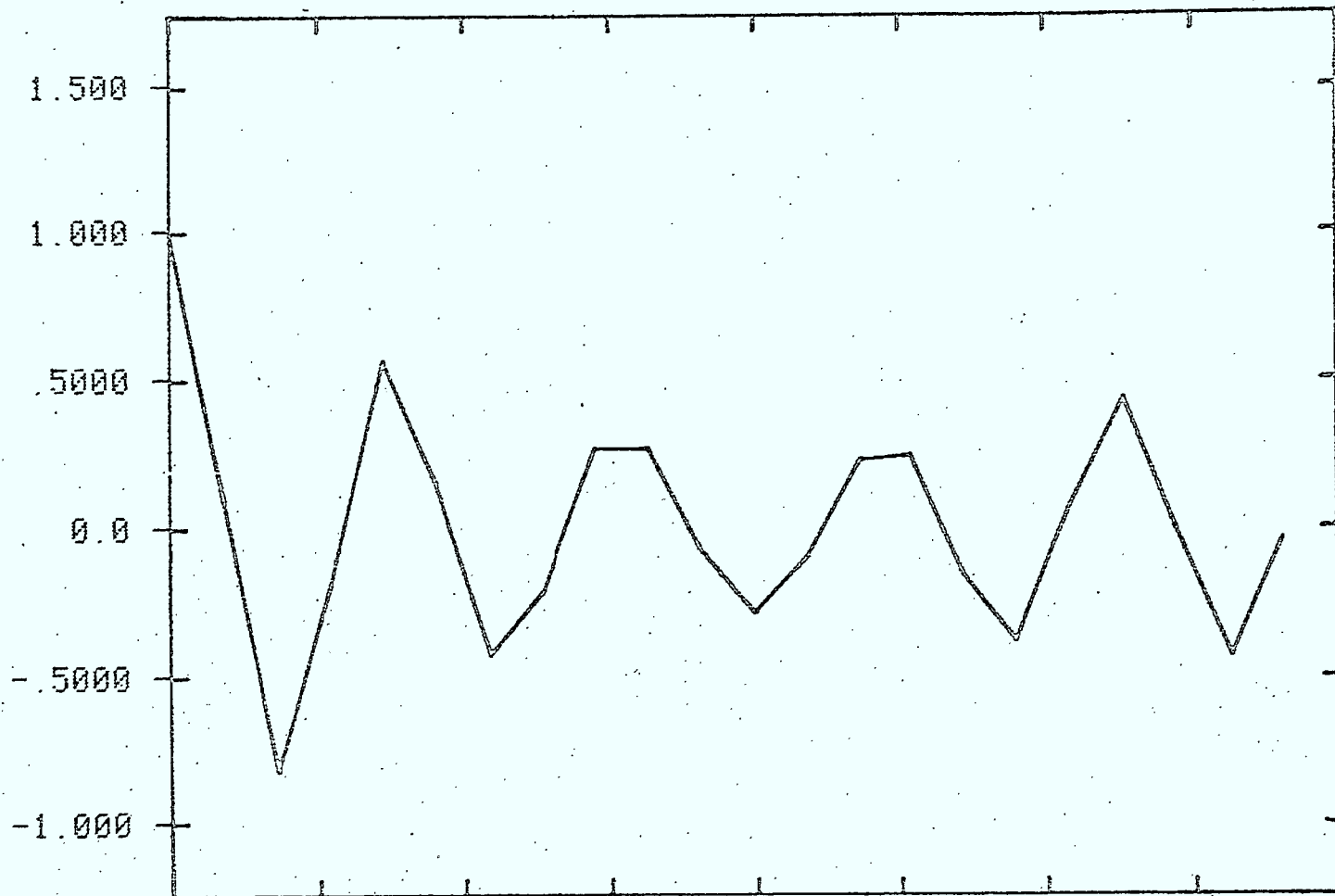


Fig. 4.16 Autocorrélation de " $\phi_n(t)$ alterné", phonème /CH/.

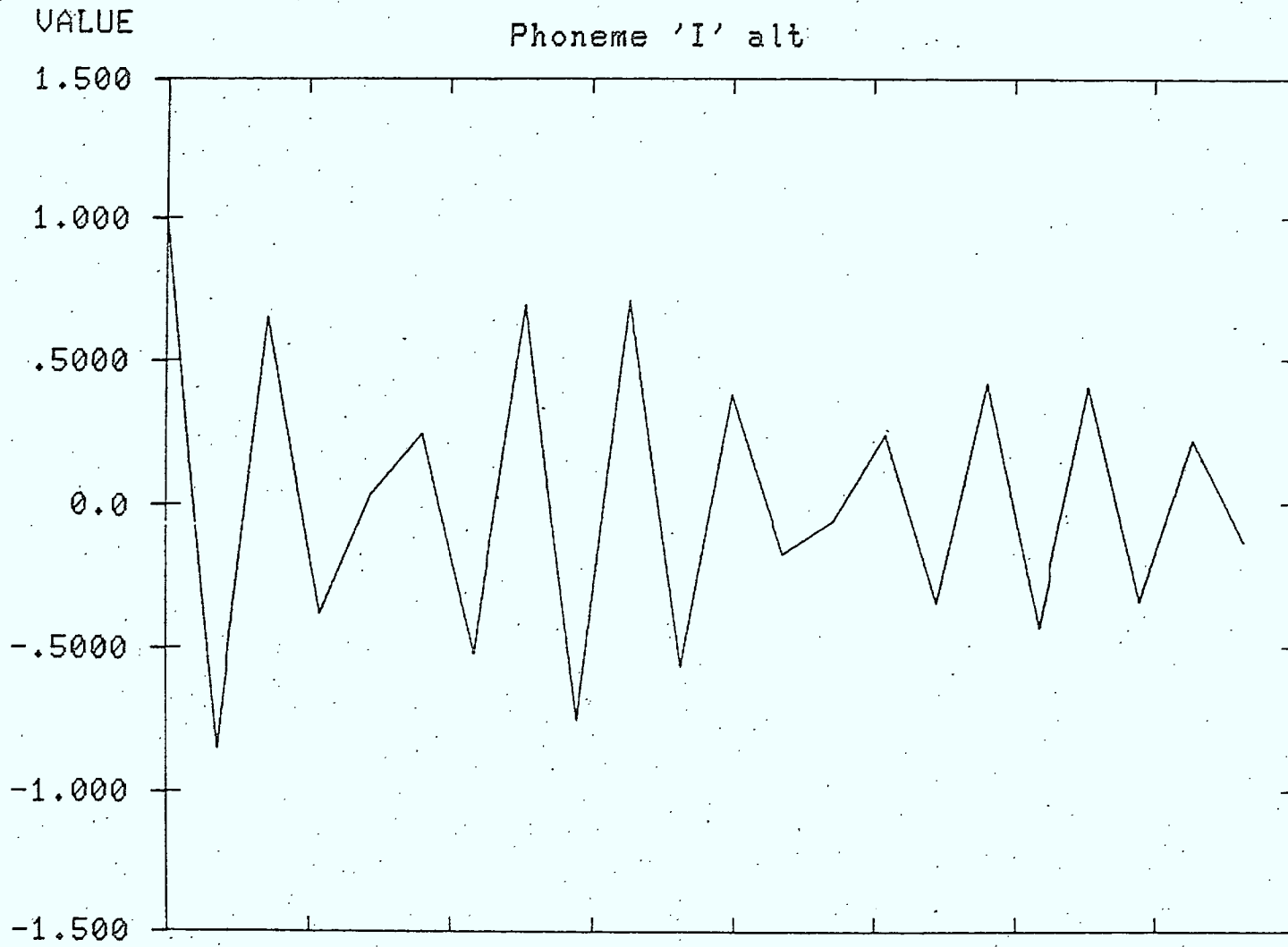


Fig. 4.17 Autocorrélation de " $\phi_n(t)$ alterné", phonème /I/.

CHAPITRE 5

COMPRESSION DE DYNAMIQUE

--TRANSMISSION ANALOGIQUE DE $\phi_n(t)$ --

L'angle $\phi_n(t)$ prélevé à partir du signal de parole, incluant les temps de silence et/ou de bruit de fond, représente un signal borné en amplitude occupant, de manière stable, toute la plage dynamique qui lui est allouée. Les figures 11a-11b, pp. 24-25 du rapport 1, donnaient une première approche de représentation. Dans une version finalisée du système de codage où l'angle $\phi_n(t)$ est "alterné" en signes, on conserve ce résultat général.

Les figures 5.1 et 5.2 montrent deux exemples associés à deux phrases distinctes .

Ces figures incluent le codage des temps de silence (bruit de fond) par l'algorithme complet (décrit aux chapitres 1 à 3).

1. Signal analogique - simulation

L'angle " $\phi_n(t)$ alterné" que l'on désigne à l'avenir par $\phi_n(t)$ est disposé dans une mémoire, avant transmission. Le débit numérique à la sortie est initialement choisi de 8 KHz. Pour simuler la conversion D/A, ce fichier à 8 KHz est "sur-échantillonné à 32 KHz" par exemple, par une adjonction de zéros entre échantillons puis filtré par un filtre numérique. Ce filtre est calculé pour une fréquence d'échantillonnage de 32 KHz et a pour bande passante la bande téléphonique 300-3300 Hz.

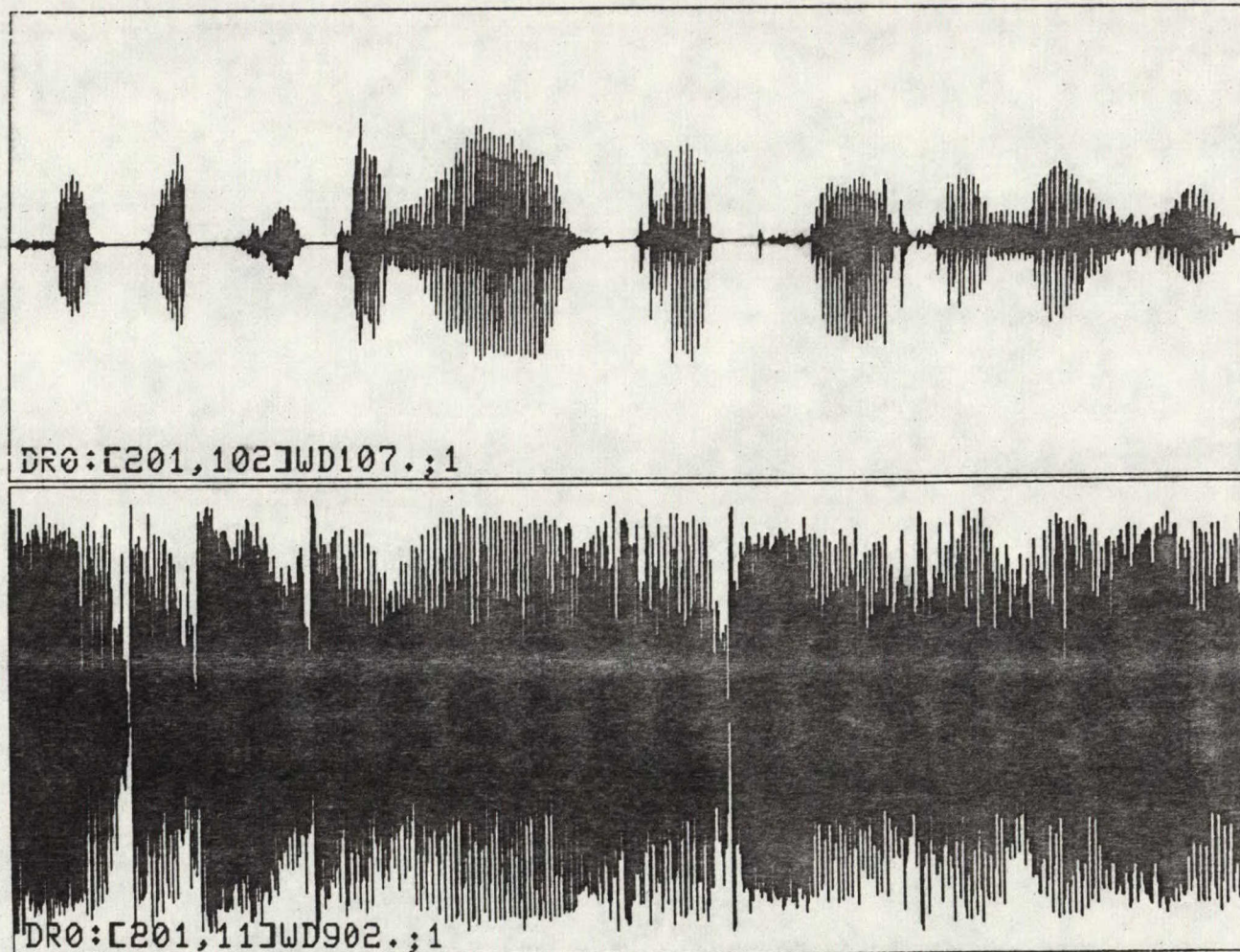


Fig. 5.1 Représentation de l'angle " $\phi_n(t)$ alterné" et du signal original de parole.
Phrase: "Le petit canard apprend à nager".

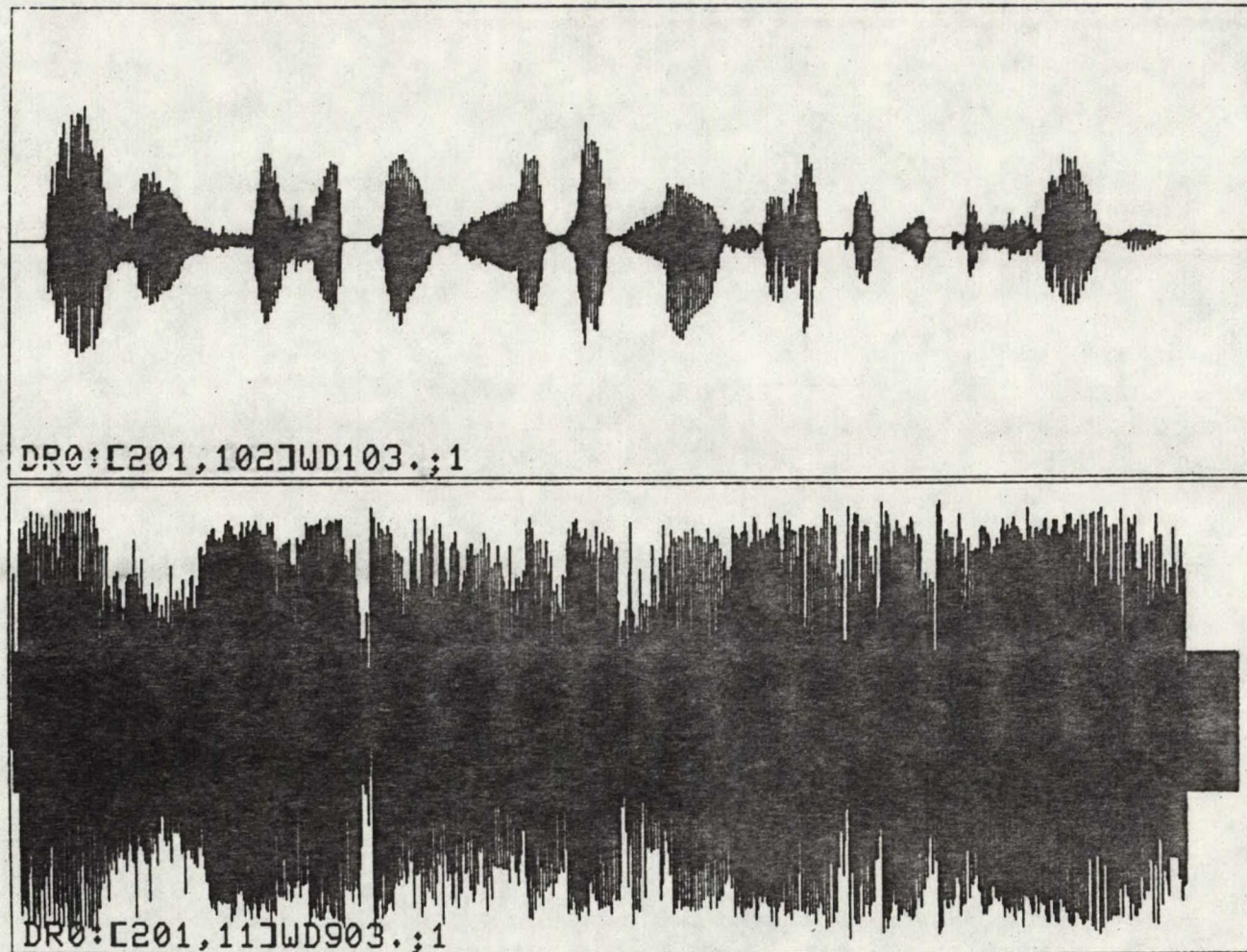


Fig. 5.2 Représentation de l'angle " $\phi_n(t)$ alterné" et du signal original de parole.
Phrase: "Un loup s'est jeté :....chèvre".

Le nouveau fichier à 32 KHz simule un signal continu (analogique) soumis à la contrainte de transmission (filtrage de canal).

Le problème, au récepteur, consiste

1. A ré-échantillonner ce signal (pseudo) continu avec une trame synchrone des échantillons sortis de la mémoire à 8 KHz sous forme numérique.
2. Une fois reconstitué (en synchronisme) le signal $\phi_n(t)$ numérique, à calculer par l'algorithme de décodage le signal de parole $S_n(t)$, mais aussi dans ce signal reconstruit en temps réel (temps d'échantillonnage variable) à retrouver la position de la trame 8 KHz d'origine. La figure 6, page 17, du rapport 1, illustre l'erreur de codage faite si ce synchronisme n'était pas assuré. Il en résulte un bruit de codage supplémentaire.

2. Procédé de synchronisation du récepteur avec l'émetteur

Au niveau de l'émetteur (codeur) on a déjà établi un mode de codage particulier des périodes de silence ou de bruit de fond. On rappelle que l'intérêt premier était de favoriser un vidage rapide de la mémoire. De la sorte en établissant le rayon de la courbe de codage à $8T_0$ (1 milliseconde), soit un échantillon sur 8 (à 8 KHz) jusqu'à ce que la mémoire soit totalement vide (au codeur et au récepteur), on assurait ensuite un codage sommaire de la phase de silence par un cercle normalisé de codage de rayon réduit à T_0 (125 μ s) et de gain minimal. L'angle normalisé prélevé alors, sur les petits niveaux du signal $S(t)$, continue à occuper la dynamique totale de l'intervalle d'angle $\left] \frac{-\pi}{2} \frac{+\pi}{2} \right[$ [voir figure 5.3.

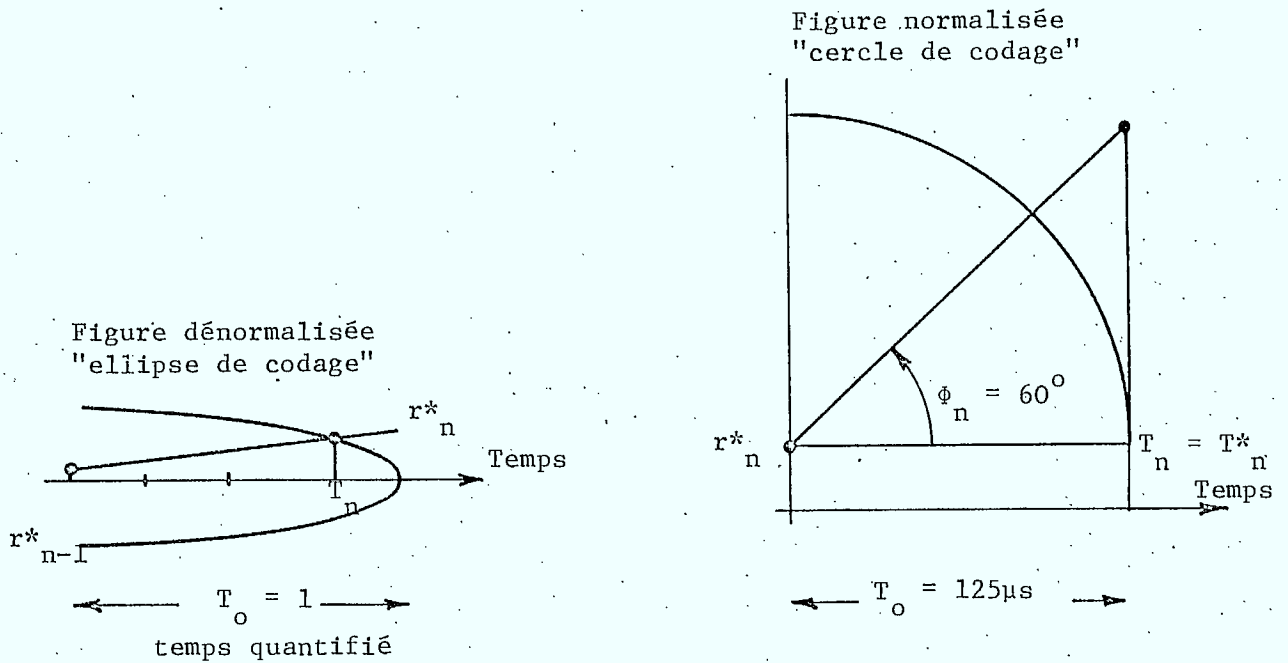


Fig. 5.3 Rappel de codage des silences.

La quantification du temps réel en multiples de $T_0/4$ assure que, si l'angle $\phi_n(t)$ est assez faible, l'échantillon reconstitué r_n^* se confond avec le véritable échantillon de parole S_n tant que la mémoire demeure vide. Ce cas se maintient aussi longtemps que ϕ_n est assez faible, donc que le signal n'est pas sur-échantillonné (ϕ_n faible $\rightarrow T_n \geq 3T_0/4$ est quantifié par la valeur égale à T_0).

2.1 Codage particulier en période de silence: $\phi_n = \text{constante}$

On envisagera donc, au codeur exclusivement de provoquer un codage spécifique (pendant le temps de silence) indépendant du signal S_n original. On limitera la valeur de $\phi_n(t)$ codé et quantifié à une amplitude qui, dans l'algorithme choisi (chapitres 2 et 3) ne provoque aucune variation des paramètres de codage que ce soit le gain G_n ou le rayon T_n .

A priori $\phi_n = \frac{\pi}{4}$ tel que $\text{tg } \phi_n = 1$ satisfait à cette condition. Mais

la quantification en $K.T_0/4$ de la progression en temps réel ne permet pas toujours de synchroniser

$$r_n^* \quad \text{avec} \quad S_n$$

Cela provoque un glissement (multiple de $T_0/4$) de la trame d'échantillonnage du récepteur par rapport à celle du codeur.

Par contre en prenant $\phi_n < 30^\circ$

$$\text{tg } \phi_n < 1$$

- Le gain $G_n = G_{n-1} \cdot \text{tg}(\phi_n)$ devrait diminuer mais il est limité par une valeur G_{\min} .
- Le rayon $L_n = L_{n-1}/\text{tg}(\phi_n)$ devrait augmenter mais, la mémoire étant vide, il est figé à la valeur T_0 (125 μ s).

Ainsi en période de silence on convient au codeur de ne prélever que des valeurs de ϕ_n d'amplitude constante soit ϕ_0 :

- assurant un temps réel quantifié $T_n = T_0$,
- un signal reconstitué S_n^* différent de l'original S_n mais de très petit niveau à cause du gain G_{\min} , et synchronisés entre eux,
- seul le signe de ϕ_n , en période de silence peut changer.

La figure 5.4 illustre un exemple de signal reconstitué au décodeur pendant un silence au codeur.

L'exemple précédent suppose que le premier échantillon est synchronisé. La suite, grâce à l'algorithme, maintient le synchronisme.

2.2 Synchronisation initiale en période de silence

Au niveau du codeur, il est toujours possible, lorsqu'une phase de silence est détectée, de choisir une valeur de $\phi_n(t)$ qui assure que l'échantillon suivant $\phi_{n+1}(t)$ sera prélevé à un instant où le temps réel quantifié est identique au temps naturel.

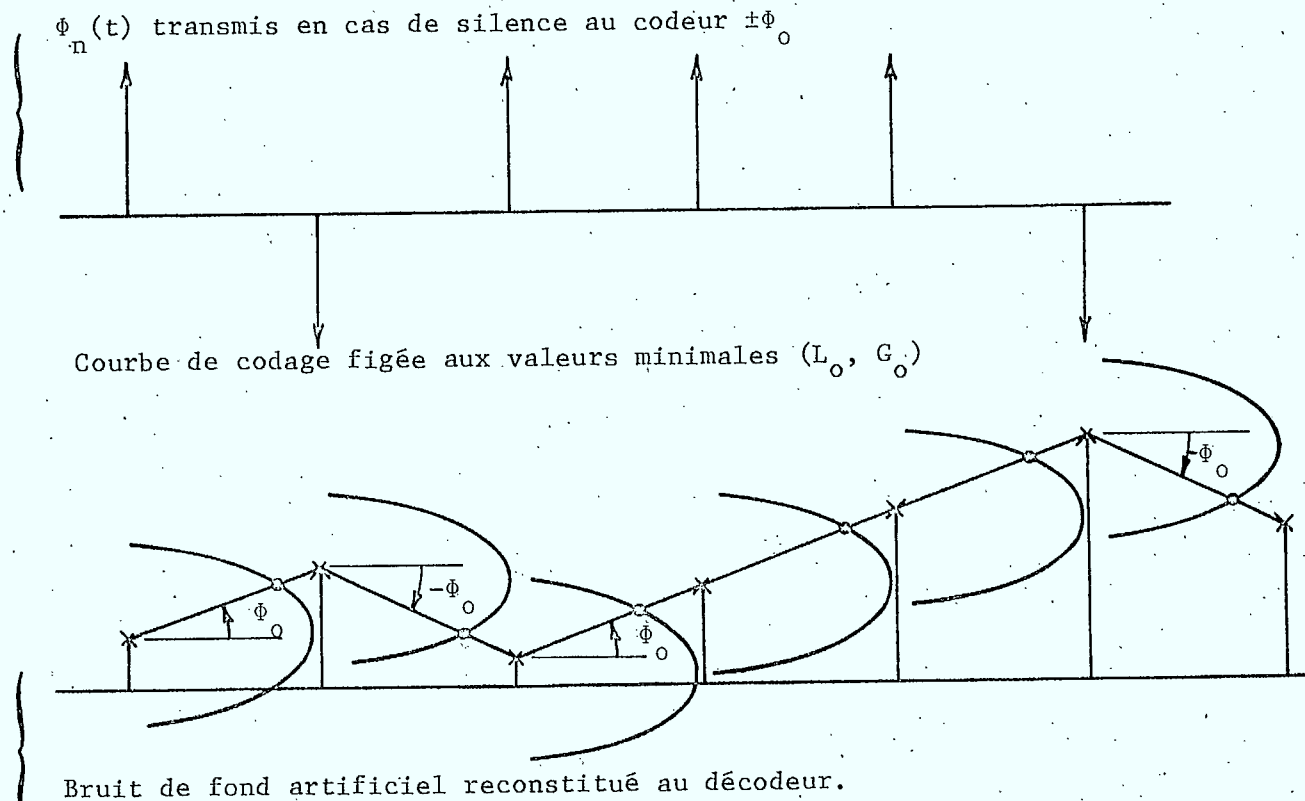


Fig. 5.4 Exemple de bruit de fond artificiellement reconstitué au décodeur, application à la synchronisation de trame.

La synchronisation sera possible si plusieurs conditions sont assurées au codeur (et au décodeur).

- La mémoire est totalement vide ($B_f = -1$).
- Le temps réel (variable) devient identique au temps naturel ($125 \mu s$).
- Un premier échantillon a été synchronisé: " r_n^* synchrone de S_n ".
- L'angle ϕ_n est figé à une amplitude constante adéquate.

2.2.1 Application au signal continu analogique $\phi(t)$

Le récepteur reçoit en fait une onde continue, sur laquelle il doit prélever à un rythme fixe (8 KHz) l'information numérique initiale $\phi_n(t)$.

La figure 5.5 montre que s'il n'y a pas de synchronisation avec le codeur la séquence d'origine $\{\phi_n\}$ devient $\{\phi'_n\}$. Les valeurs différentes feront converger l'algorithme du récepteur vers des valeurs des paramètres

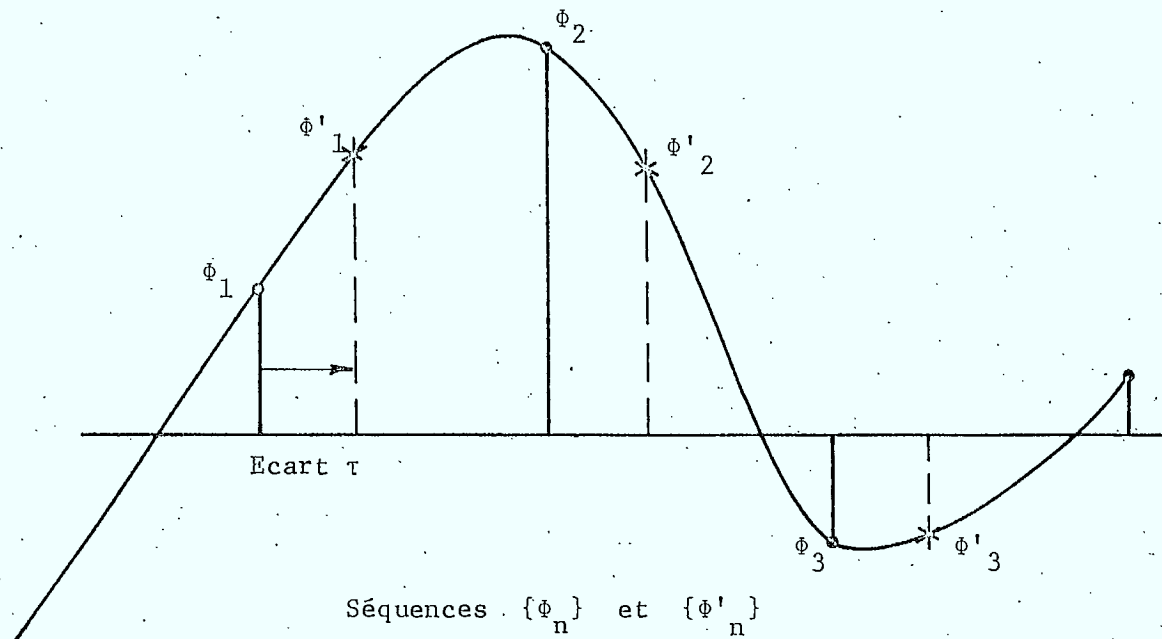


Fig. 5.5 Exemple de désynchronisation.

distinctes de celles du codeur. Le récepteur ne reconstituera plus un signal identique à celui de l'émetteur. En pratique le résultat est désastreux. (Les angles et les temps réels sont changés).

L'écart de temps τ doit être absolument déterminé si la séquence $\{\phi'_n\}$ est prélevée par le récepteur. Une fois cet écart connu, la trame d'échantillonnage au récepteur sera décalée de τ pour récupérer la séquence numérique $\{\phi_n\}$.

On notera aussi que ϕ_n a été quantifié en amplitude: donc au récepteur une légère erreur sur le calcul de τ n'influe pas sur la valeur de ϕ_n prise en compte par l'algorithme numérique de décodage.

La figure 5.6 illustre ce résultat particulier.

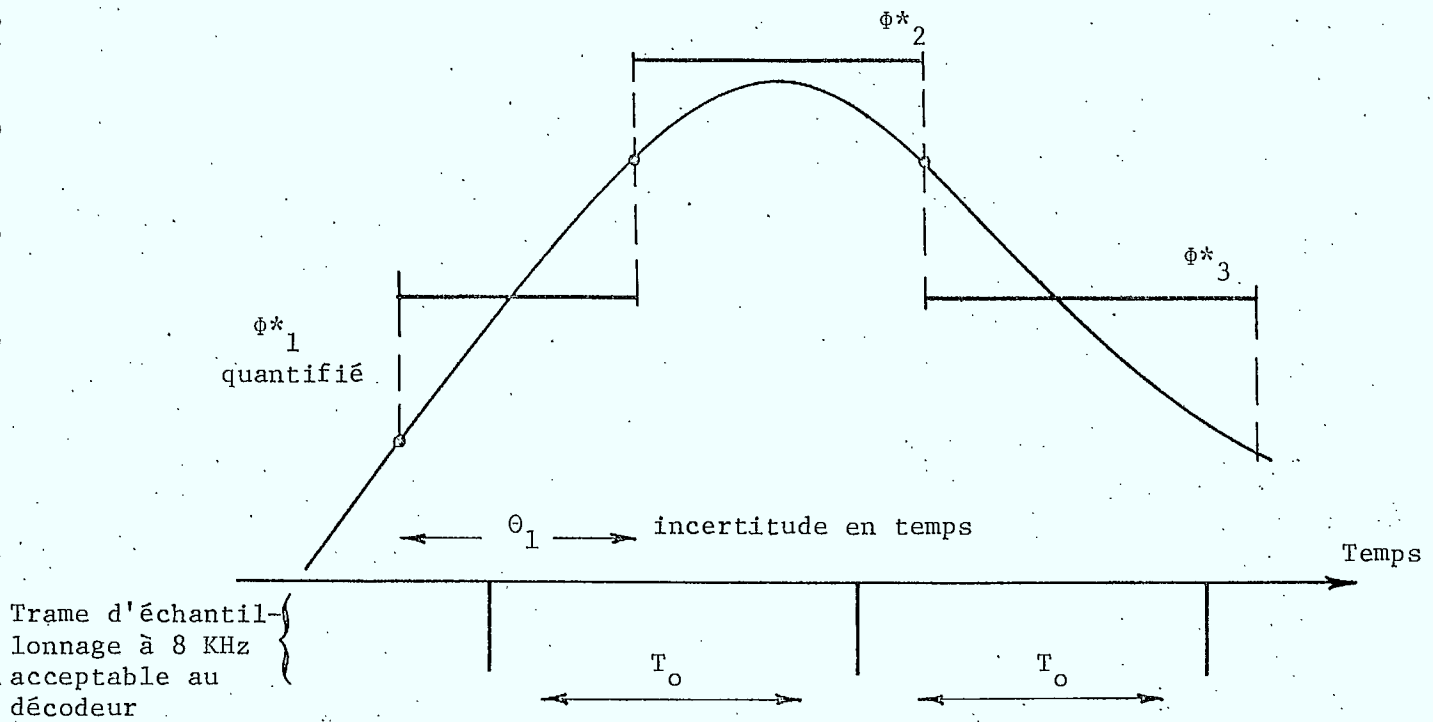


Fig. 5.6 Effet de la quantification d'origine (au codeur) de l'angle $\phi_n(t)$.

θ_1 représente l'incertitude en temps tolérée pour le calcul au récepteur du décalage τ de la figure 5.2. En pratique $\theta_1 \ll \tau$, et toute valeur numérique prélevée dans cet intervalle (θ_1) sera avant utilisation au décodeur transformée en une valeur quantifiée ϕ_1 , identique bien sûr à celle sortie du codeur.

2.2.2 Utilisation des silences pour la synchronisation

Une fois imposé au codeur les trois conditions essentielles caractérisant un état de silence utilisé pour la synchronisation numérique:

1. Mémoire totalement vidée $B_f = -1$,
2. Temps réel (quantifié) devenu identique au temps naturel

$$T_n = T_o = 125 \mu s$$

- 3 Un premier échantillon décodé r_n^* synchrone de l'échantillon réel S_n .

Tant que cet état persiste:

- Tous les r_n^* sont reconstitués aux instants synchrones des S_n quelque soit n.
- L'angle ϕ_n transmis est figé en amplitude, soit ϕ_o .

Cependant le récepteur doit pouvoir identifier que cette structure de codage du silence a été imposée au codeur.

- Angle "figé" de codage en période de silence

On propose donc de modifier la séquence transmise en imposant que l'angle $\phi_n(t)$ soit figé à une valeur ϕ_o normalisée de codage telle que montré sur la figure 5.7

$$T_n = L_o \cos \phi_o$$

avec L_o , valeur minimale, maintenue fixe par l'algorithme (du codeur et du décodeur). Il faut que T_n ainsi calculé et quantifié en multiple de $T_o/4$ reste constant et égal à T_o [note: $T_n > 3/4 T_o$ est quantifié en T_o]. Ceci impose

$$\phi_n \leq \arccos \frac{T_n}{L_o}$$

où la valeur calculée T_n est

$$\frac{3T_o}{4} \leq T_n \leq T_o$$

et en représentation normalisée $L_o = 1$, $T_o = 1$ dans l'algorithme du codeur et du décodeur, donc:

$$\phi_o \leq \arccos 3/4$$

Un angle de 25 à 28° (quantifié autour de ces valeurs) sera donc satisfaisant pour assurer le fonctionnement

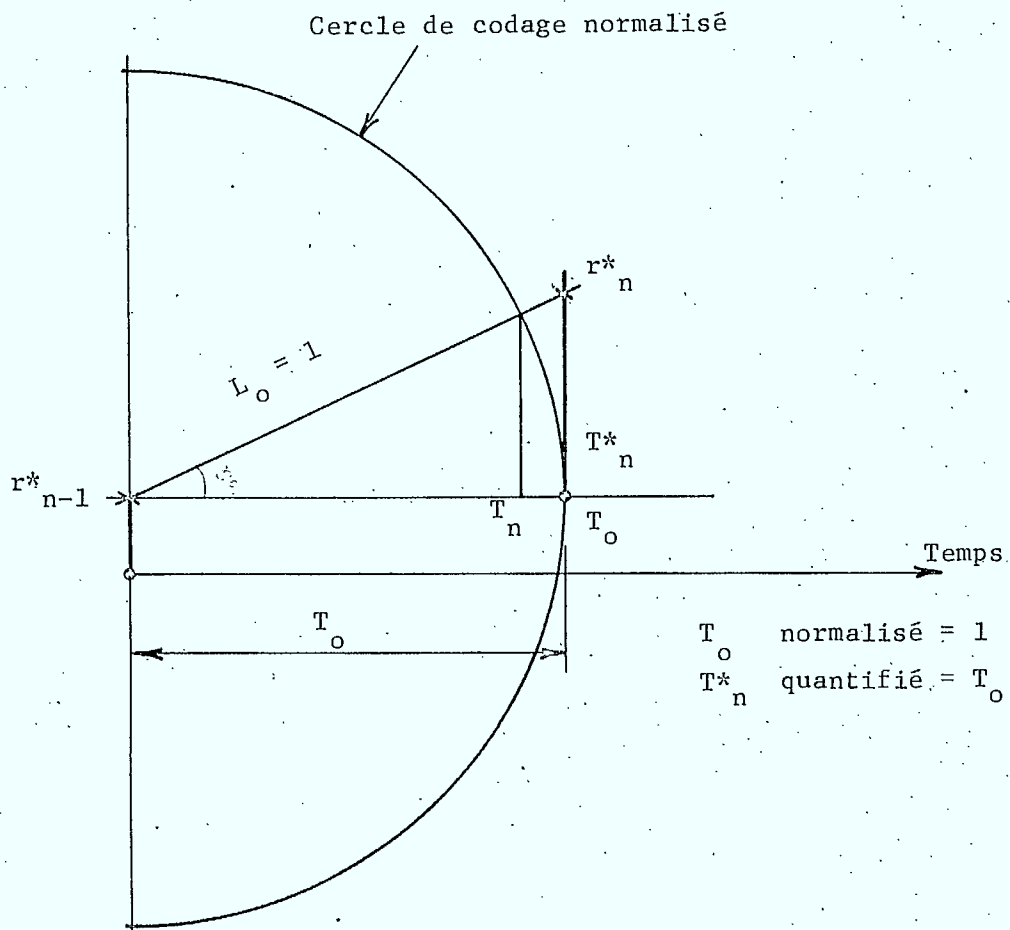


Fig. 5.7 Choix de l'angle figé ϕ_o en période de silence.

• Séquence "figée" de codage en période de silence

Au lieu, comme dans le traitement précédent, de coder le silence par:

- La valeur d'amplitude ϕ_o ,
- Le signe exact de l'angle de codage,

On décidera, une fois identifiées les trois conditions de décision sur l'état de silence (voir 2.2.2), de générer arbitrairement au codeur une séquence artificielle de codage du silence en cours par la transmission (mise en mémoire préalable, puis conversion D/A au rythme de sortie de la mémoire) de la séquence suivante:

$$\{+\phi_0, +\phi_0, -\phi_0, -\phi_0, +\phi_0, +\phi_0, -\phi_0, -\phi_0, \dots\}$$

Cette séquence avec une alternance par paire permet, en sortie analogique du convertisseur de générer une onde sinusoïdale à 2 KHz ($T_0 = 125 \mu s$, $f_e = 8 \text{ KHz}$) représentés à la figure 5.8 et d'équation $\phi(t) = \phi_0 \sqrt{2} \sin(2\pi f_e t)$ avec $f_e = 8 \text{ KHz}$.

Si comme sur la figure 5.8, le récepteur échantillonne cette même onde avec un "écart d'échantillonnage" noté τ sur la figure:

- On sait que le signal reconstitué r_n^* sera très différent de l'original de parole S_n .

Cependant, du fait que l'onde continue à 2 KHz constitue pour le récepteur une tonalité facilement détectée:

- A cause bien sûr de sa fréquence (nombre de passages par zéros rigoureusement constant dans une fenêtre donnée),
- De l'état de la mémoire au récepteur (vide) en état de silence pour le codage/décodage,
- Des autres conditions dues à cet état (voir 2.2.2).

On pourra sans erreur importante calculer au récepteur l'écart τ , et corriger, en la déplaçant, la trame d'échantillonnage de l'onde reçue. En effet cette séquence particulière $\{\phi_0\}$ n'a été générée dans ces conditions qu'avec des valeurs alternées par paires $[+\phi_0, +\phi_0]$ et $[-\phi_0, -\phi_0]$. Or cet échantillonnage à 2 KHz est unique et sa position relative, par rapport à l'onde sinusoïdale continue, absolument fixe permet de retrouver sur la sinusoïde pure les temps exacts d'échantillonnage d'origine.

Ce sont les valeurs où la fonction sinusoïdale $\phi(t) = \phi_0 \sqrt{2} \sin \omega t$ atteint le niveau quantifié ϕ_0 initialement imposé au codeur pour générer cette fonction.

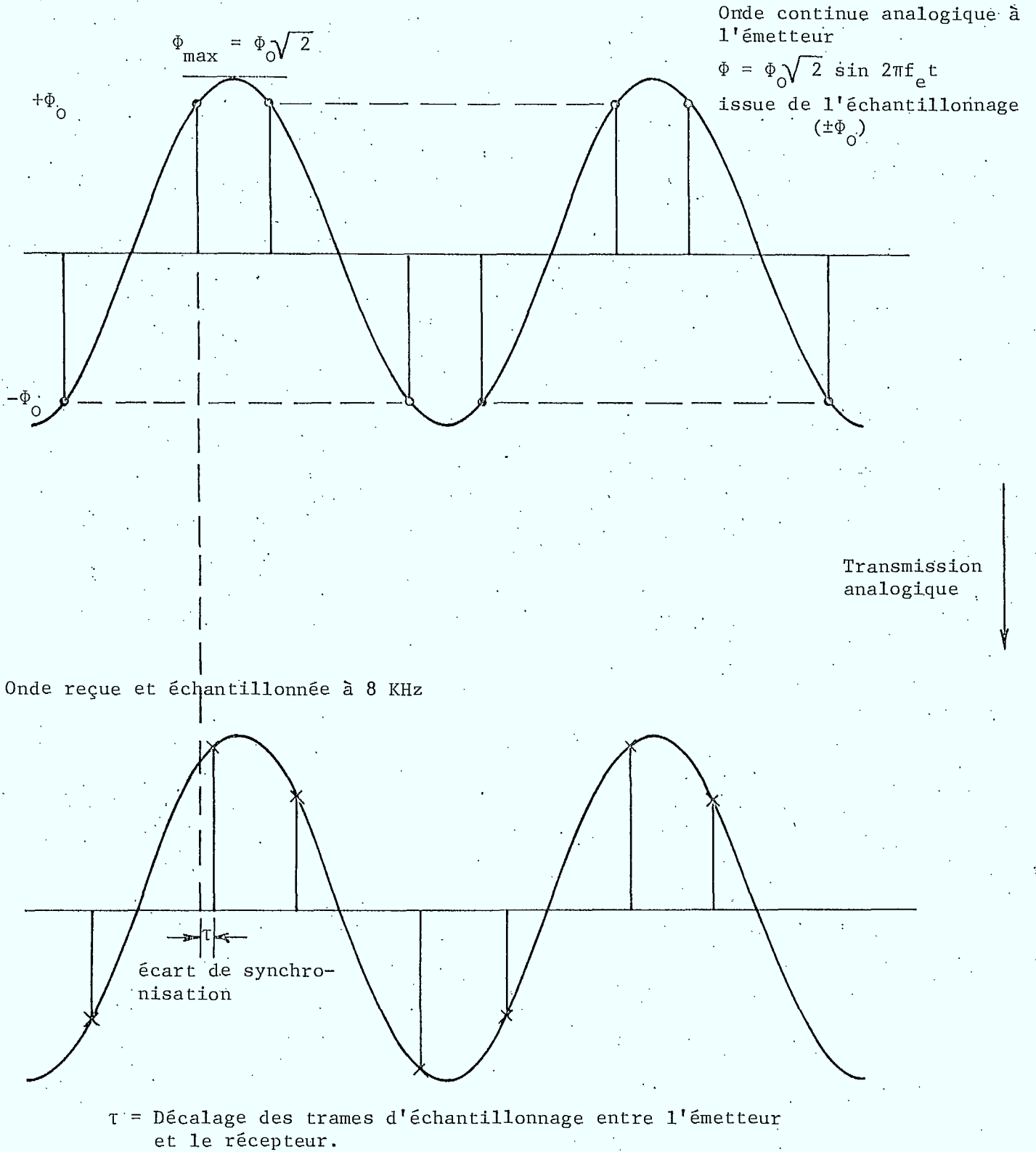


Fig. 5.8 Exemple d'ondes reçues, et, transmission en phase "silences".

Le sur-échantillonnage à 32 KHz suivi du filtrage en bande téléphonique assure un décalage quelconque simulant la transmission analogique.

Le récepteur, basé sur le principe précédemment décrit et représenté schématiquement sur la figure 5.9 assure le calcul de l'écart τ (ici quantifié en multiples entiers de la période à 32 KHz du signal dit "analogique").

A ce niveau, le processus semble fonctionner de manière satisfaisante.

Les figures 5.10 et 5.11 suivantes montrent en outre:

- Figure 5.10: un codage sous la forme $\phi_n(t)$ sans traitement particulier des silences pour une phrase donnée.
- Figure 5.11: un codage de la même phrase mais où l'on a imposé au codeur de générer la séquence en paires alternées $\{\pm\phi_0\}$.

On constate, que même au niveau de la représentation analogique de l'onde reçue, la détection des séquences, destinées à resynchroniser le signal pendant les silences, est aisément assurée par la présence de la "tonalité" à 2 KHz dont l'amplitude $(\phi_0\sqrt{2})$ est nettement distinguée du signal $\phi_n(t)$ codant la parole continue.

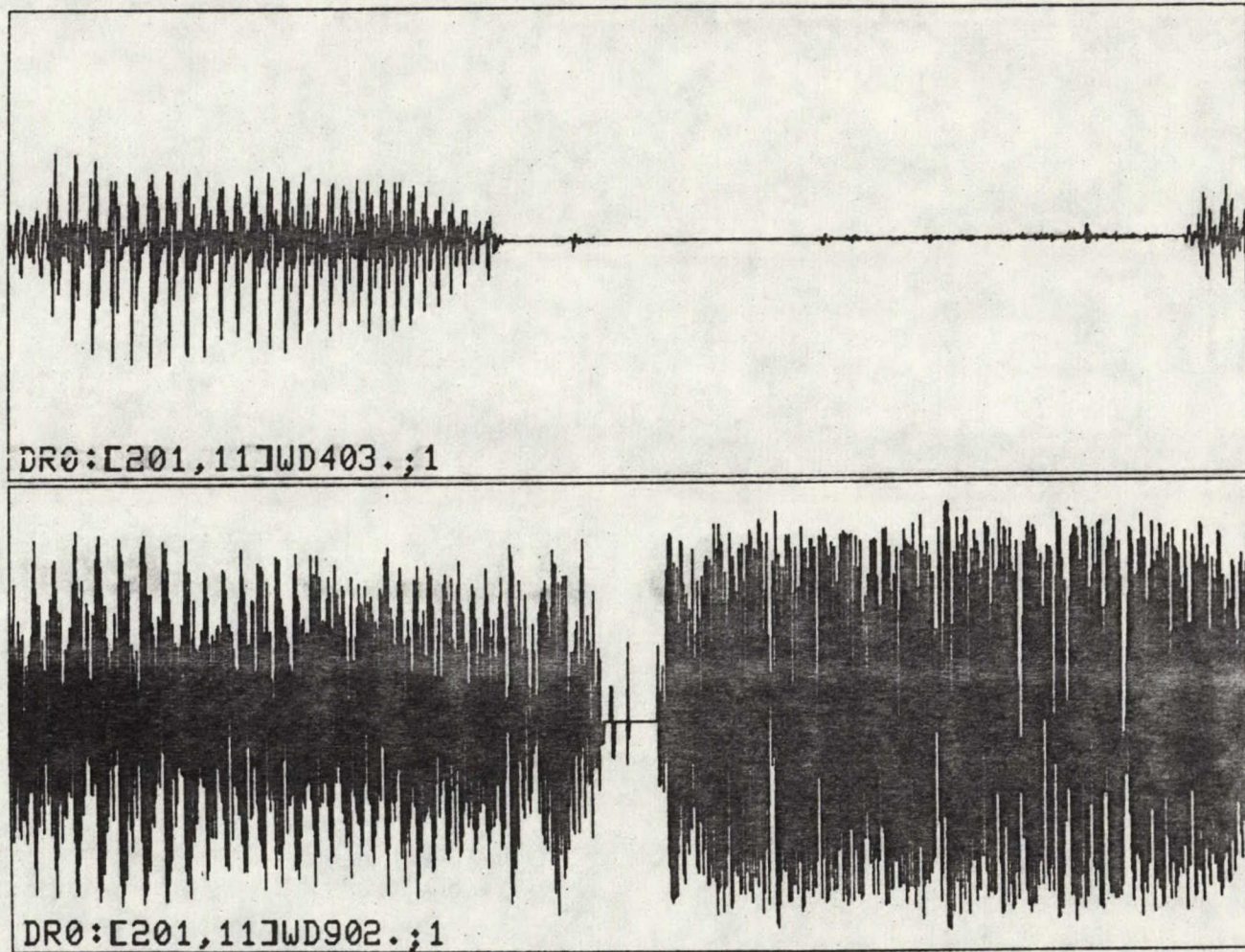


Fig. 5.10 Codage de parole: angle $\phi_n(t)$ sans action particulière pendant les silences.

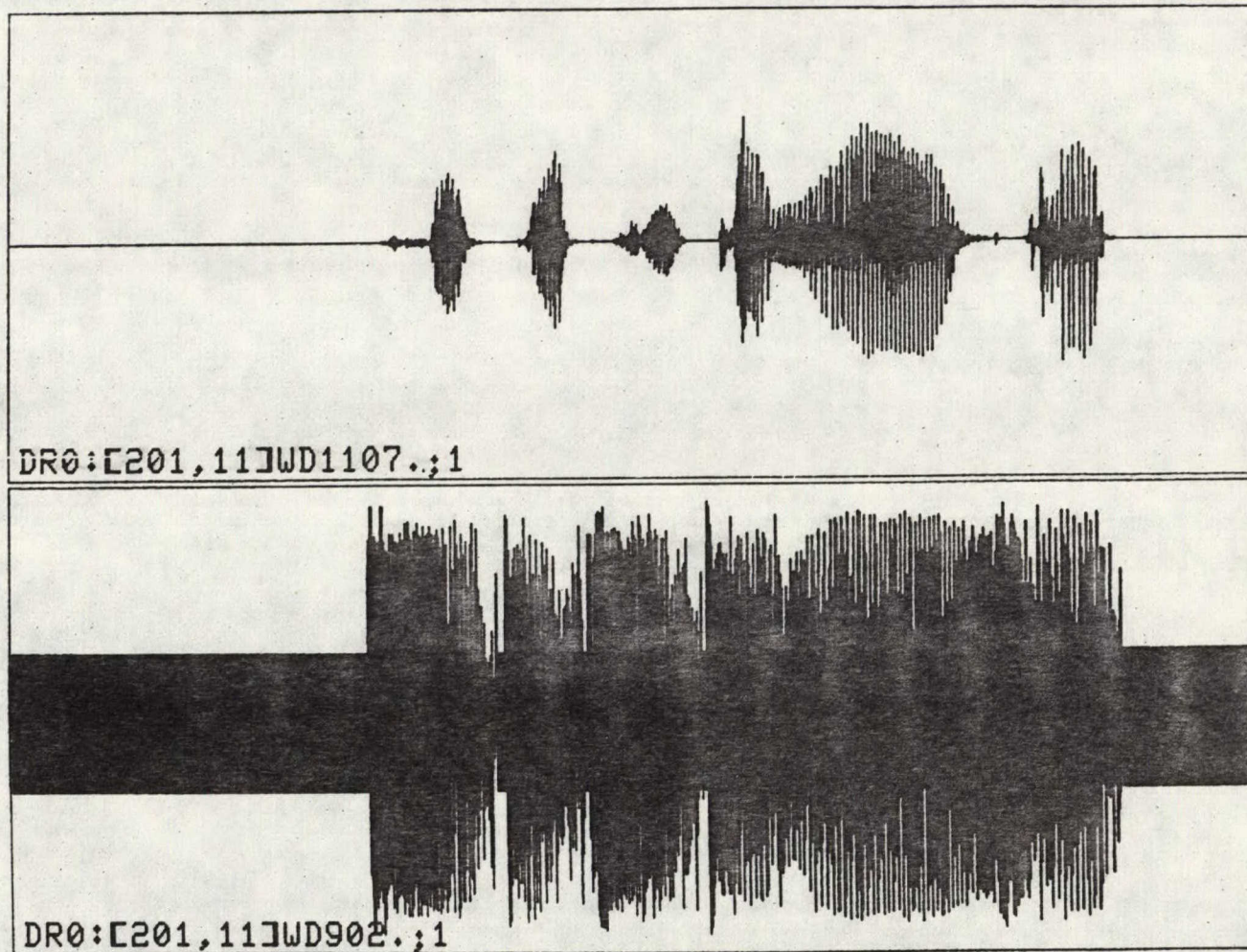


Fig. 5.11 Codage de parole: angle $\phi_n(t)$, génération de séquences $[\pm\phi_0]$ pendant les silences.

2.3 Spectres alternés et synchronisation

Au chapitre précédent il a été proposé un retournement du spectre par une alternance volontaire des signes du signal $\phi_n(t)$ numérique.

On conservera ce procédé qui élimine la présence de trop nombreuses composantes de hautes fréquences dans le spectre.

2.3.1 Spectre de la séquence artificielle $\{\pm\phi_0\}$ alternée

L'onde échantillonnée générée à 2 KHz n'est absolument pas modifiée par un changement de signe $(-1)^n$ tel que proposé. En effet cette fréquence est exactement le centre de la bande que l'on a retourné. Le récepteur reçoit donc toujours, en phase de silence une tonalité à 2 KHz représentant la séquence $\{\pm\phi_0\}$ d'amplitude maximale analogique $\phi_0/\sqrt{2}$.

Notons toutefois que cette tonalité à 2 KHz sur l'angle $\phi_n(t)$ ne représente pas pour le signal reconstitué une perturbation. Les paramètres de l'algorithme de base en "état de silence" sont à leurs valeurs minimales, et dans ce cas, le signal définitif est inaudible.

2.3.2 Spectre des signaux $\phi(t)$ alternés en bande réduite 6,4 KHz

Le chapitre 4 présentait des spectres des signaux $\phi(t)$ alternés. L'énergie entre 3000 et 4000 Hz, bien que très fortement réduite y était parfois encore présente.

Ce qui signifie, que en transmission analogique, les effets de filtrage éliminent ces composantes hors bande téléphonique. Le récepteur ne reçoit pas exactement la forme d'onde du codeur et des erreurs se traduisent par un bruit de codage sur le signal reçu.

Pour limiter ce problème, il sera suggéré de générer $\phi_n(t)$, non pas à 8 KHz (bande passante 0-4000 Hz en numérique), mais de limiter le prélèvement numérique $\phi_n(t)$ (donc aussi $\phi_n(t)$ alterné) à 6,4 KHz, et d'imposer un filtrage analogique à 3000 Hz (au lieu de 3300 Hz habituellement).

Le choix de cette fréquence est déterminé par des conditions de simulation:

- Un fichier d'origine à 16 KHz (non filtré) est sur-échantillonné à 32 KHz, filtré en bande réduite (filtre numérique à 3000 Hz, échantillonné à 32 KHz) puis sous-échantillonné en prélevant 1 échantillon sur 5

$$\left(\frac{32}{5}\right) = 6,4 \text{ KHz}$$

Les figures 5.12 à 5.19 montrent les calculs de l'évolution des spectres par F.F.T.:

- Figures paires: du signal $\phi_n(t)$ à 6400 Hz,
 - Figures impaires : du signal alterné en signe à 6400 Hz.
- pour les phonèmes /I/, /CH/, /A/ et /S/ pendant la durée de ces sons.

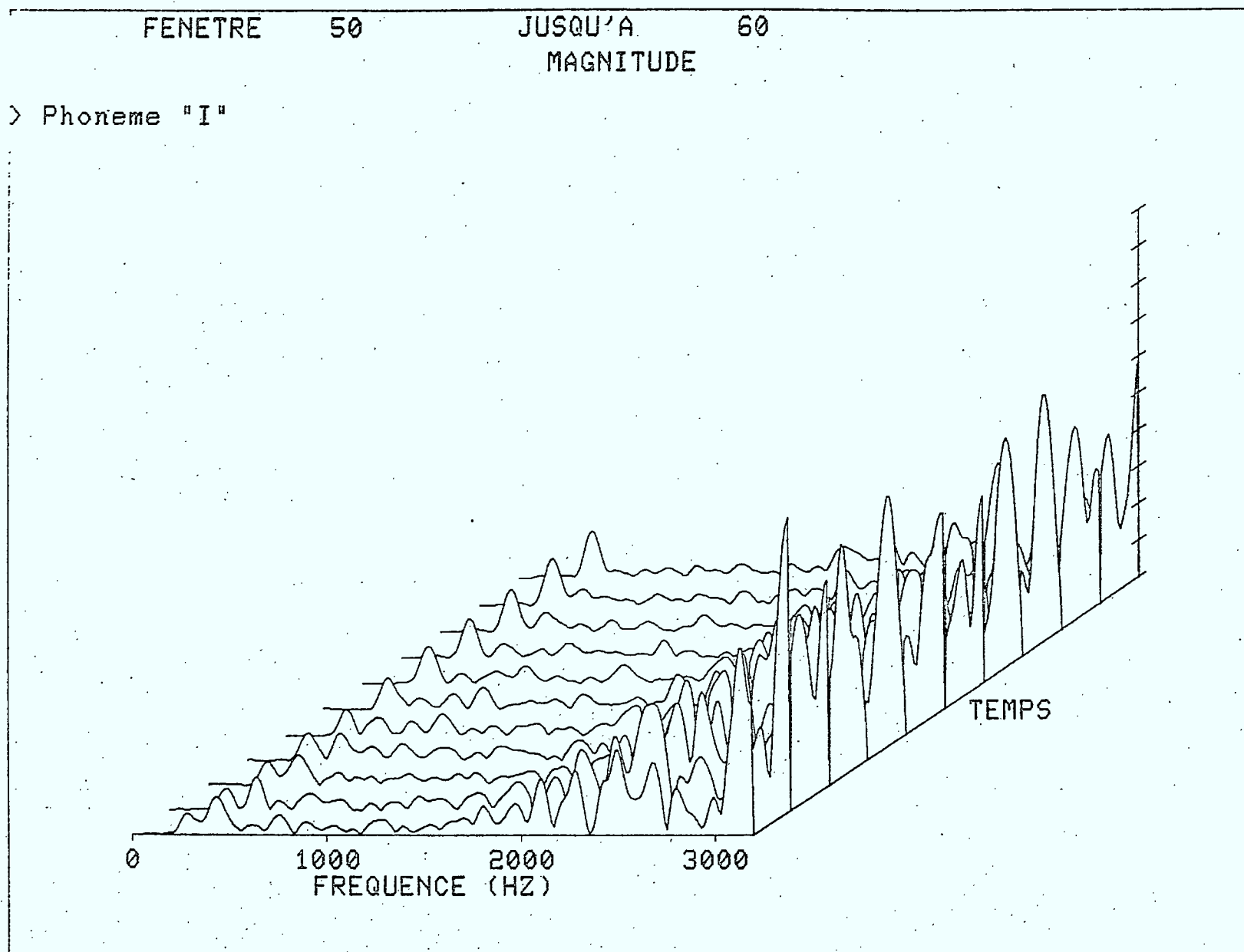


Fig. 5.12 Evolution du spectre de l'angle $\phi_n(t)$ à 6400 Hz, phonème /I/.

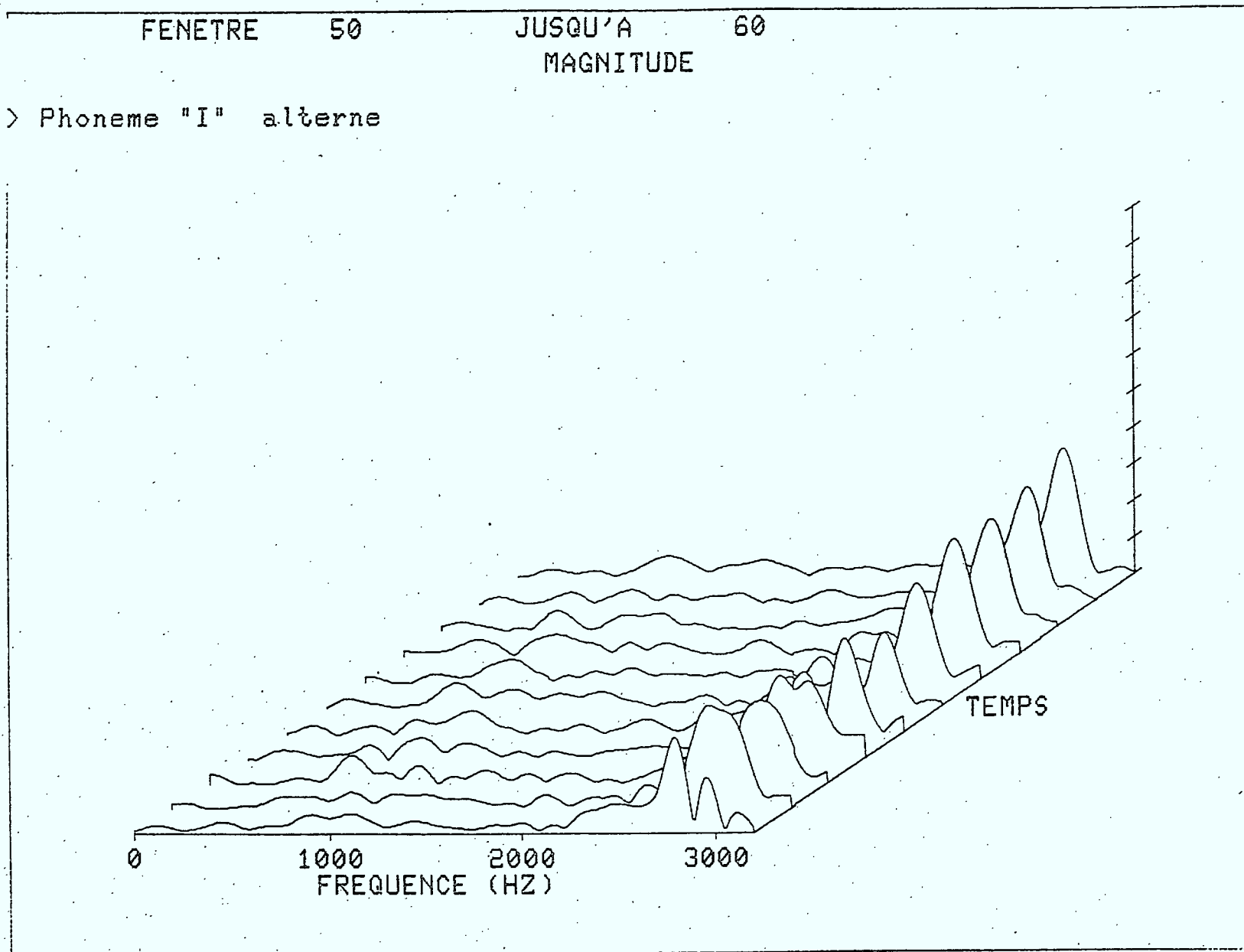


Fig. 5.13 Evolution du spectre de l'angle $\phi_n(t)$ alterné, phonème /I/.

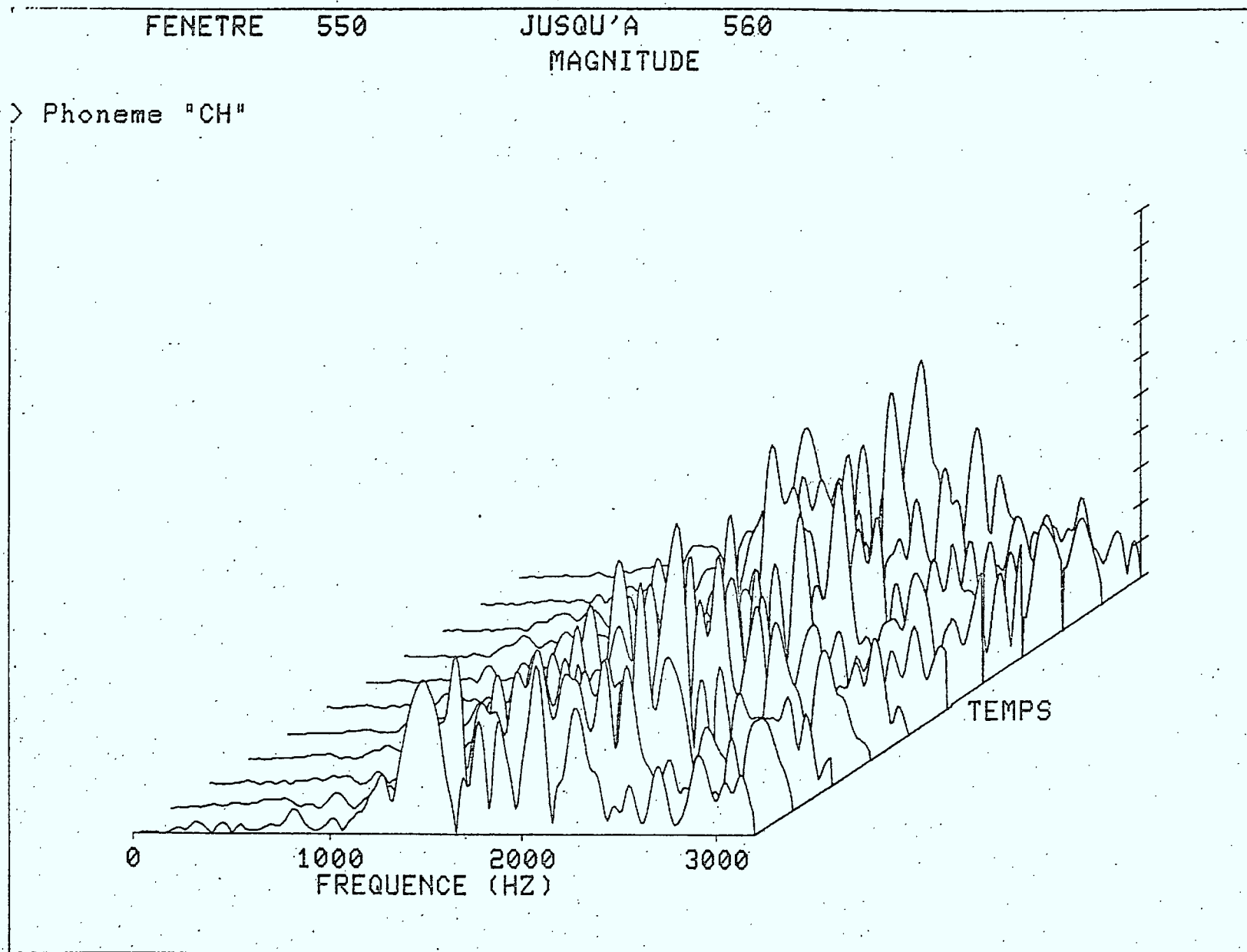


Fig. 5.14 Evolution du spectre de l'angle $\phi_n(t)$, phonème /CH/.

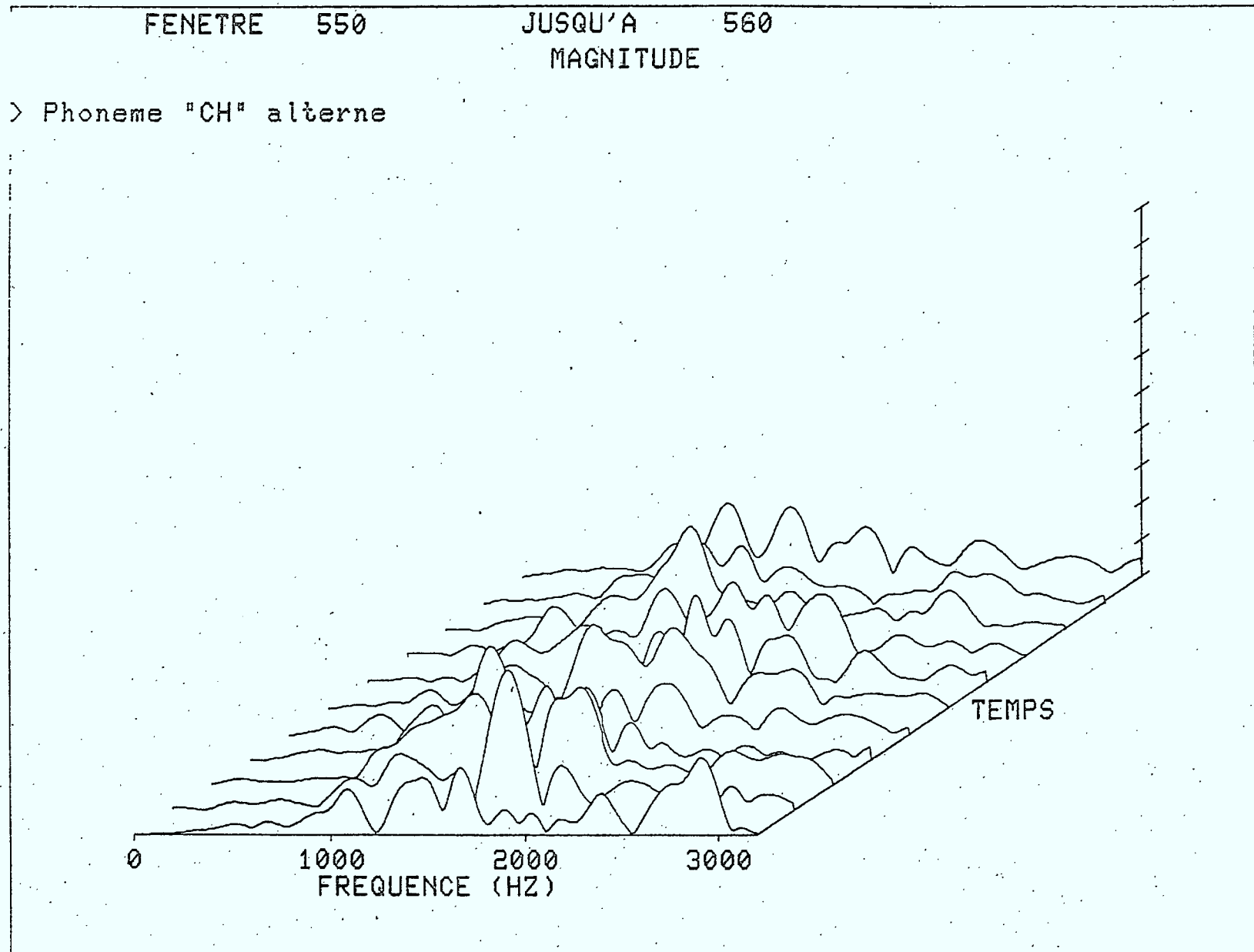


Fig. 5.15 Evolution du spectre de l'angle $\phi_n(t)$ alterné, phonème /CH/.

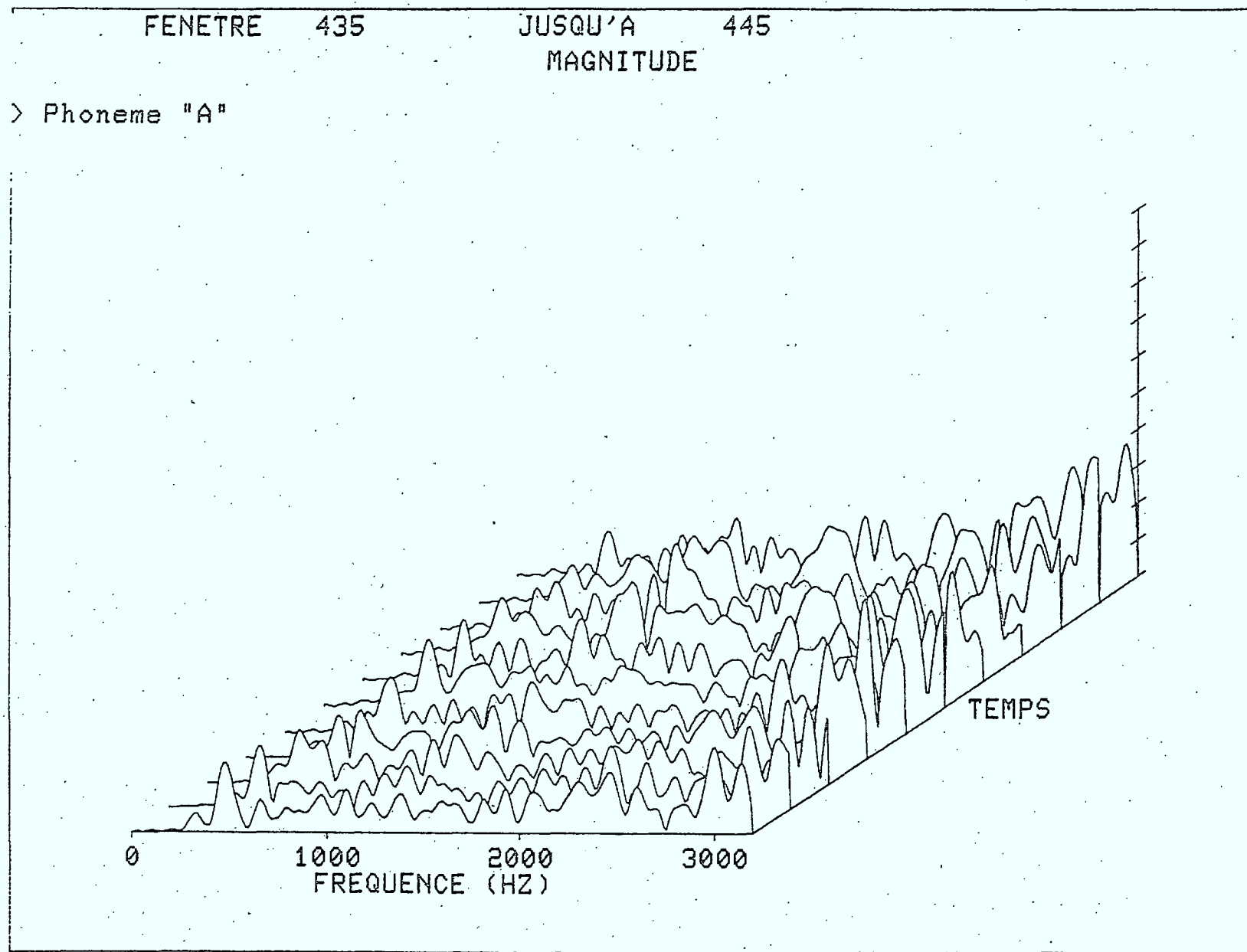


Fig. 5.16 Evolution du spectre de l'angle $\phi_n(t)$, phonème /A/.

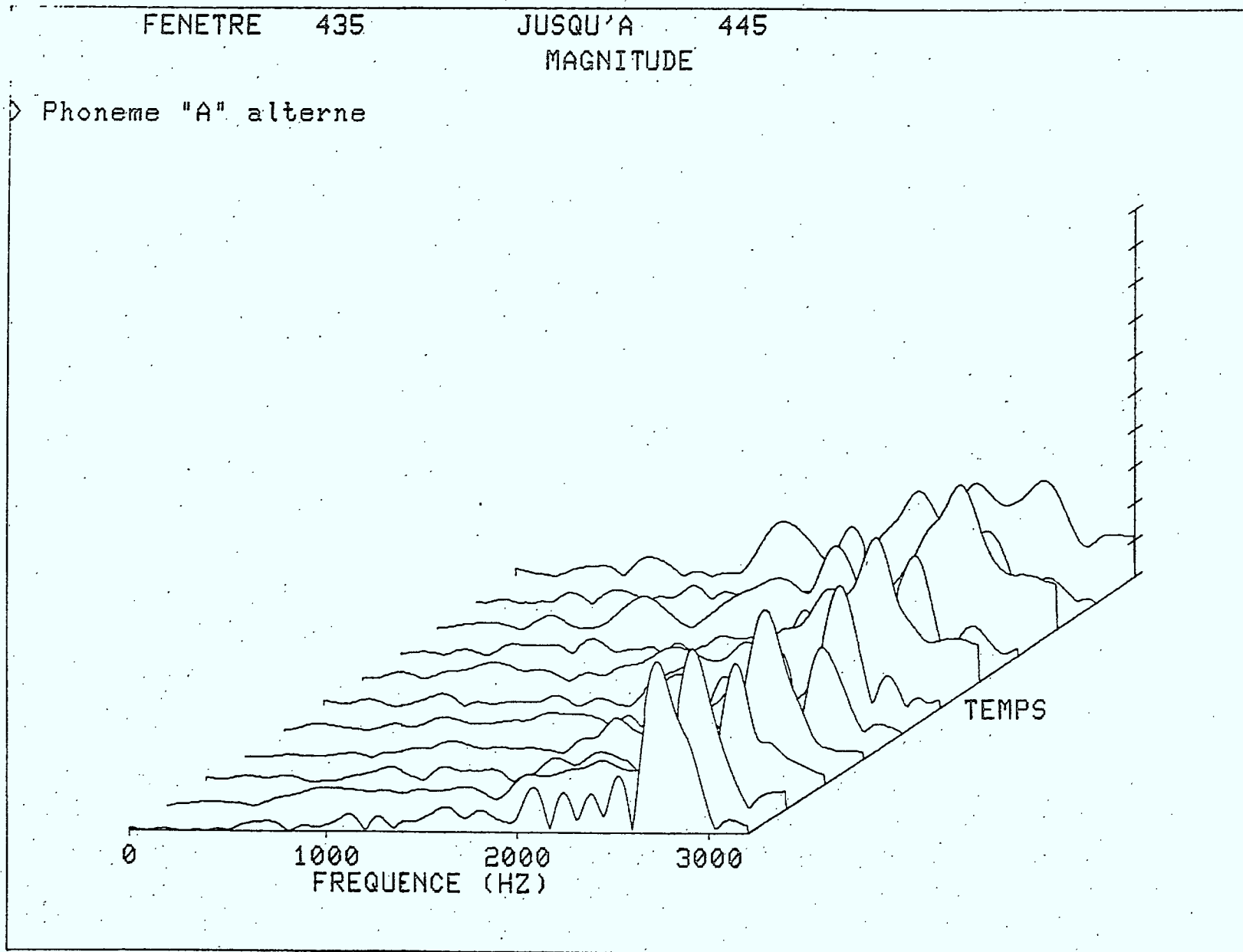


Fig. 5.17 Evolution du spectre de l'angle $\phi_n(t)$ alterné, phonème /A/.

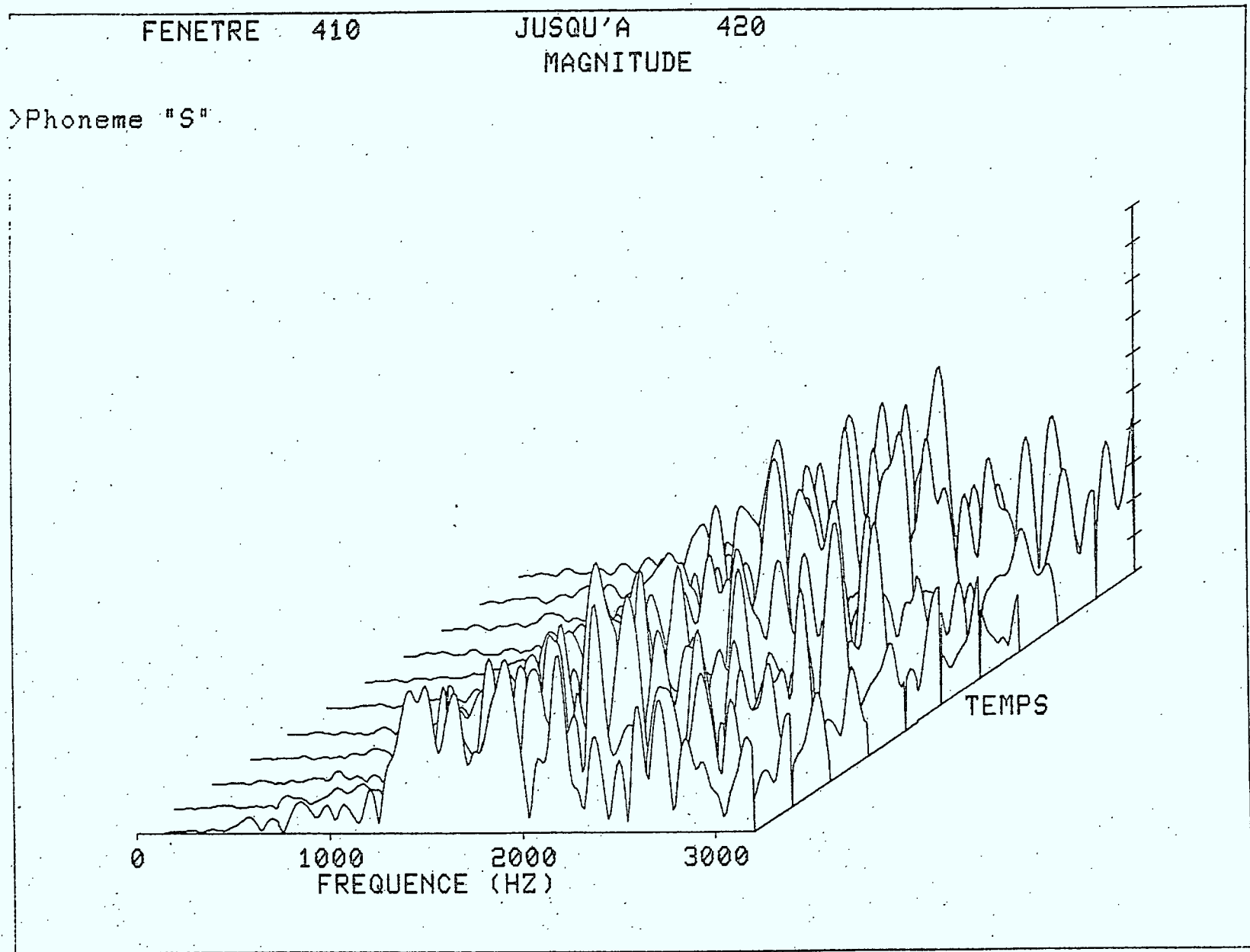


Fig. 5.18 Evolution du spectre de l'angle $\phi_n(t)$, phonème /s/.

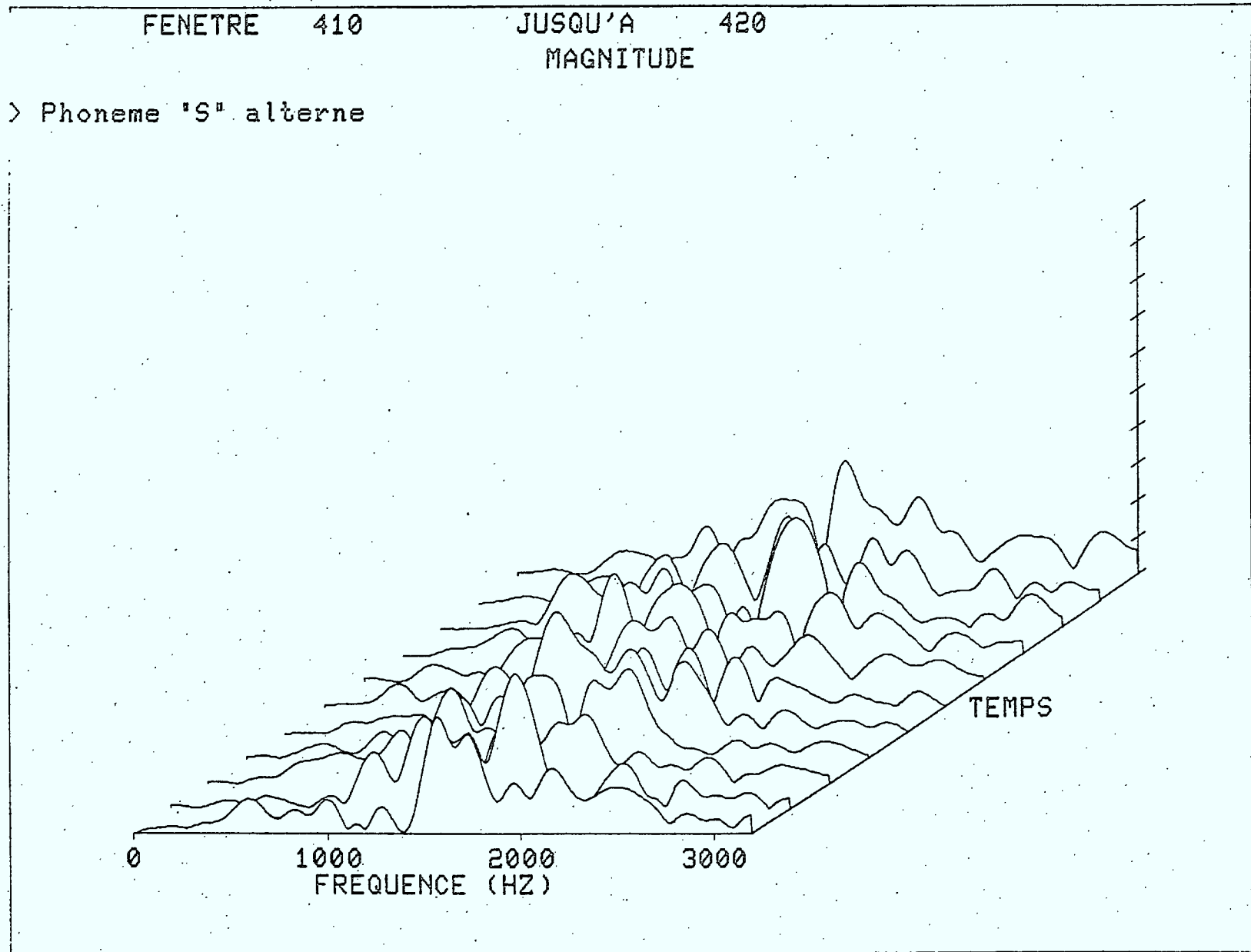


Fig. 5.19 Evolution du spectre de l'angle $\phi_n(t)$ alterné, phonème /S/.

La technique d'alternance des signes de tout le signal $\phi_n(t)$, limitée à une bande réduite utile de 3 KHz en analogique permettra de fournir une forme d'onde relativement peu dégradée par les filtres de modulation et de transmission et de restituer un signal de parole de qualité acceptable.

Un enregistrement sur cassette de quelques codages à bande réduite à 3000 Hz par le procédé décrit complète l'analyse de principe de ce chapitre.

3. Effet du bruit

Un bruit blanc additif est ajouté au fichier de parole. L'algorithme complet de codage/décodage y est appliqué sans développer de techniques particulières de traitement par élimination de bruit.

Un enregistrement sur cassette montrera l'influence de divers niveaux de bruit.

CHAPITRE 6

CODAGE SPATIO TEMPOREL

A partir d'idées générées dans l'étude du codage en temps "élastique", quelques essais d'une nouvelle méthode de codage ont été simulés.

Il s'agit de choisir un échantillon qui sera transmis et qui représente en fait un couple de données:

- une amplitude,
- un temps.

1. Principe général du codage spatio-temporel

Le schéma proposé est une généralisation du principe de codage différentiel à temps fixe d'échantillonnage.

En termes simples, l'idée d'un codage différentiel consiste à ne transmettre, en code à m bits, que la différence entre le dernier échantillon connu du signal (passé) et la nouvelle valeur de l'échantillon courant (c'est le cas d'un système sans prédicteur, appelé aussi prédicteur d'ordre "0" ou "bloqueur").

En pratique, si l'on réfère à la figure 6.1:

- Le temps d'échantillonnage T est fixe, habituellement $125 \mu s$ (8 KHz) en bande téléphonique,
- Le choix de la répartition des niveaux de quantification peut être

optimisé par un algorithme (de MAX) ou une K-moyenne adaptant la distribution des niveaux aux statistiques de distribution des écarts entre échantillons consécutifs, comme l'illustre la figure 6.2.

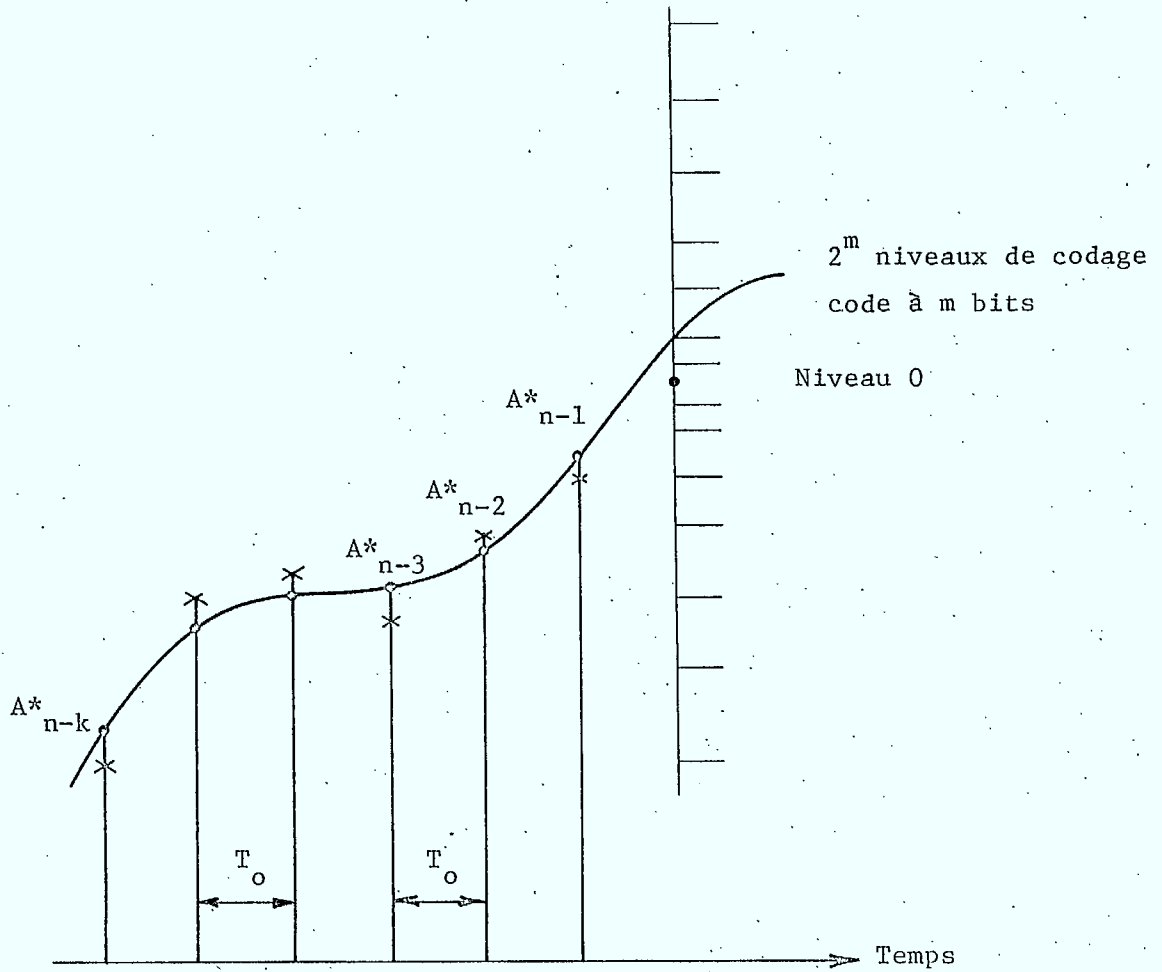


Fig. 6.1 Codage différentiel.

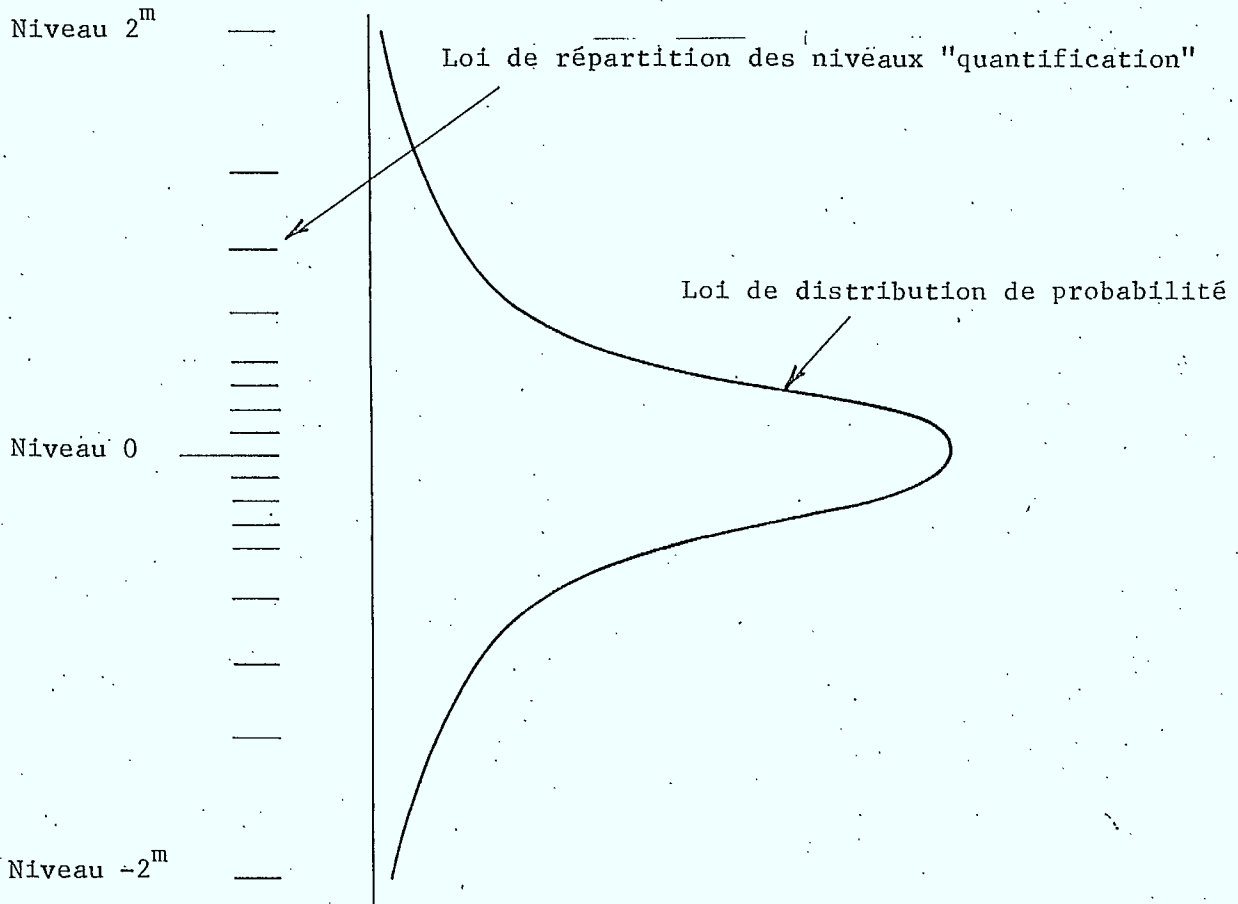


Fig. 6.2 Loi de quantification de codage différentiel.

A partir de ce modèle une nouvelle idée vise à introduire une extension du temps d'échantillonnage, et donc de répartir les niveaux de codage dans un domaine ou couple "Ecart-Temps" (E_n, T_n) auquel sera affecté un indice unique i_n .

La figure 6.3 illustre ce principe général.

1.1 Principe de base

- Le signal est reconstitué et connu jusqu'à l'échantillon A_{n-1} .
- Une fenêtre (rectangulaire en première approche du problème) est associée à l'échantillon A_{n-1} .
- Parmi tous les représentants possibles dans cette fenêtre, l'un deux, s'il est choisi (par exemple i)

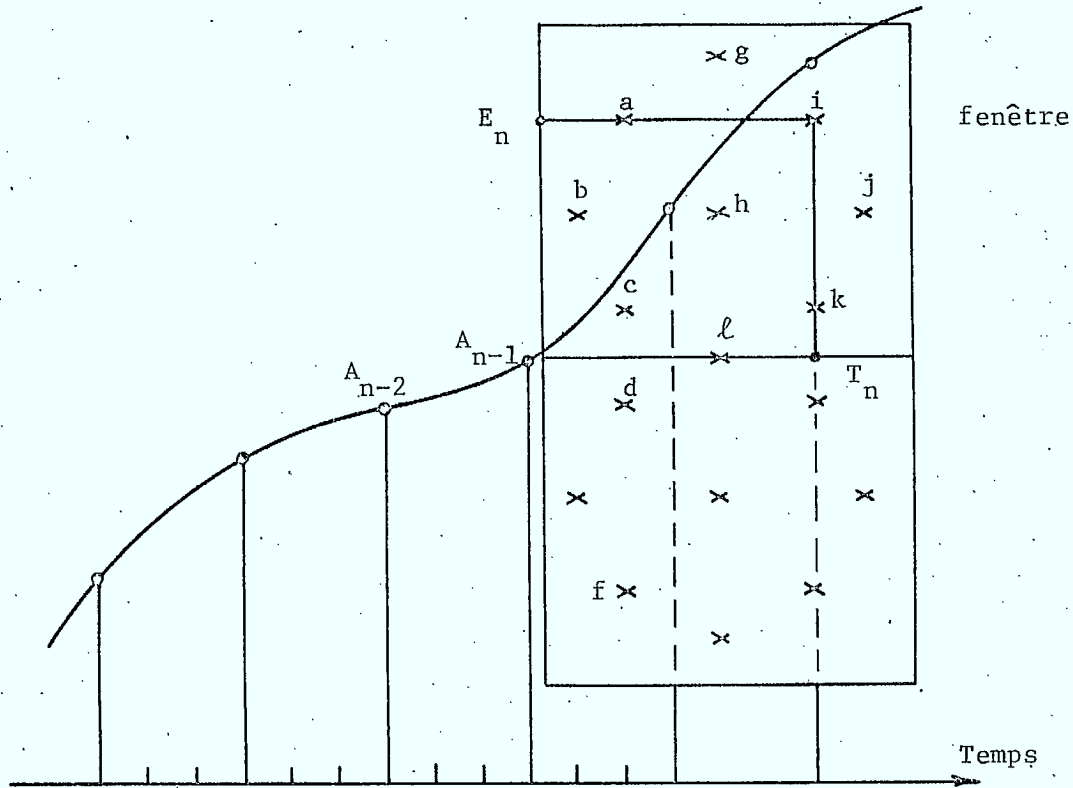


Fig. 6.3 Codage spatio-temporel.

- définit un couple (E_n, T_n) ,
- permet de reconstituer un échantillon A_n , à l'instant T_n variable, calculé par la relation:

$$A_n = A_{n-1} + E_n\{T_n\}$$

où $E_n\{T_n\}$ représente un point de la fonction E_n à l'instant T_n dont l'indice i transmis est l'identificateur. La fenêtre, ensuite, sera décalée d'une durée T_n et ses dimensions éventuellement seront recalculées.

1.2 Choix du représentant dans la fenêtre

Parmi tous les cas possibles (2^m si m bits sont utilisés pour coder les $i, j, k...$), une sélection (par codage en arbre ou autre algorithme) de plusieurs éléments acceptables peut être envisagé. Parmi ceux-ci le choix du meilleur représentant (sur la figure "i" choisi par rapport à "j" et "k") pourrait être effectué par la définition de:

- Un critère d'erreur quadratique (calcul du rapport Signal sur Bruit S/N) entre le véritable $S(t)$ d'origine et le signal reconstitué avec le choix de "i", de "j" ou de "k", en fonction de la recherche d'un maximum du (S/N).
- Un processus de calcul du (S/N) sur des signaux, "sur-échantillonnés" et "filtrés" de façon à comparer, point à point, sur n'importe quel intervalle de temps, les deux courbes:

original $S(t)$ sur-échantillonné et filtré,

reconstitué $S^*(t)$ sur échantillonné et filtré;

et ce, dans la fenêtre sur l'intervalle de temps T_n seulement affecté à chaque indice transmis.

1.3 Rôle du filtrage

Le filtre numérique proposé permet de simuler l'action habituelle du filtre de sortie après le décodage numérique - analogique. En effet, avec un sur-échantillonnage, on peut appliquer une interpolation par blocage illustrée à la figure 6.4.

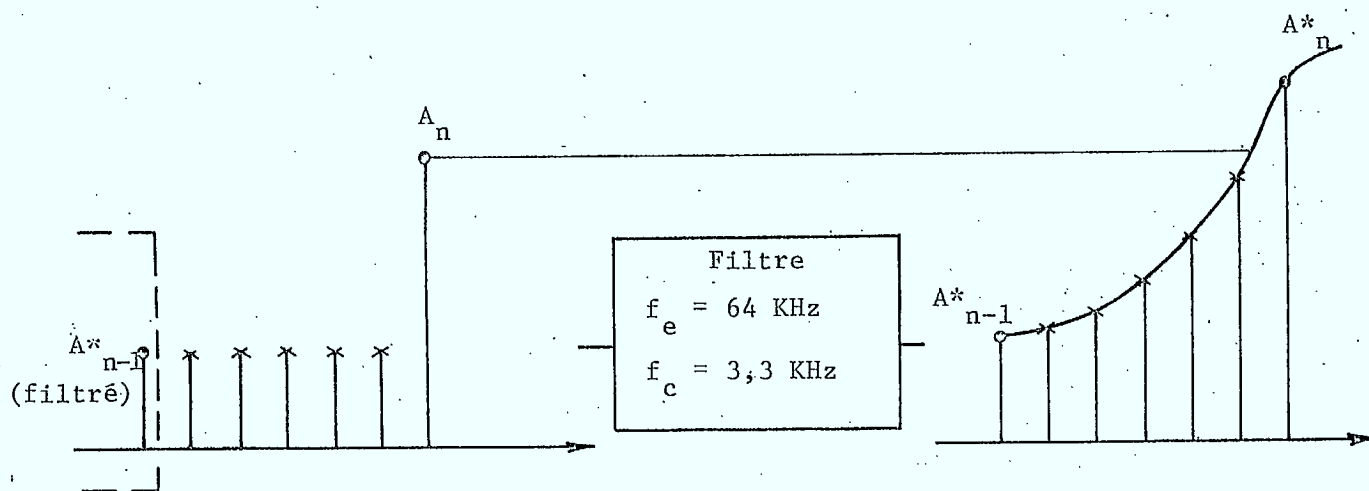


Fig. 6.4 Principe de sur-échantillonnage par blocage.

A partir de A_{n-1}^* (supposé déjà filtré), puis avec interpolation, par blocage, le signal obtenu A_n^* en sortie du filtre sera légèrement différent de A_n . Ainsi la formule (1), en pratique sera réécrite:

$$A_n^* = \Psi \{ A_{n-1}^* + E_n(T_n) \}$$

où $\Psi \{ \}$ représente l'opération globale de "sur-échantillonnage et de filtrage". Par exemple ce peut être un filtre de fréquence de coupure $f_c = 4 \text{ KHz}$, mais fonctionnant à un rythme (d'échantillonnage de base) de 64 KHz , ce qui en définit les coefficients...

2. Application du principe - Essais de codage

Dans les essais préliminaires effectués, le fichier d'origine demeure à 8 KHz et il n'y a pas de sur-échantillonnage. L'objectif de ces es-

sais de codage consiste à réduire le débit moyen en imposant une fenêtre de largeur variable entre 1 et 8 périodes d'échantillonnage à 8 KHz.

Dans la recherche des représentants, un critère de mesure (rapport S/N limité au calcul non pas de la puissance mais seulement aux amplitudes relatives), est pondéré par un facteur associé au temps. Ce procédé vise à choisir, à qualité sensiblement égale, le représentant le plus éloigné dans le temps, de façon à réduire au maximum le débit numérique moyen.

2.1 Rôle du filtrage

Il est apparu que le choix du filtre choisi pour restituer le signal intermédiaire, à tous les intervalles de temps à l'intérieur de la fenêtre, est très critique, (voir figure 6.3).

En effet chaque représentant possible, dans la fenêtre, donne une forme d'onde filtrée qui est comparée à la forme d'onde du signal original de parole lui-même filtré. Ces deux opérations de filtrage sont imposées pour tenir compte du retard du filtre numérique et de rotations de phase inévitables sur les signaux.

Si le filtre est d'ordre élevé, le nombre de calcul mis en jeu est considérable puisqu'il faut analyser chaque courbe obtenue pour chaque représentant dans la fenêtre.

Le blocage, à l'intérieur de la fenêtre, des échantillons au niveau constant précédemment codé/décodé (figure 6.4), s'avère peu approprié à la technique proposée. Il a été remplacé par une interpolation linéaire entre les points A_{n-1}^* et A_n . Ces échantillons intermédiaires sont alors filtrés.

On a représenté sur les figures suivantes (figures 6.5 à 6.8) quelques représentations graphiques du principe proposé.

On constate que l'effet du filtrage conduit à considérer que la position de chaque échantillon dans la fenêtre d'origine est elle-même filtrée par le filtre choisi. La fenêtre filtrée est donc globalement déplacée par rapport à la courbe originale $S(t)$ de la parole (elle-même filtrée). Si le filtre choisi perturbe trop la position de ces échantillons de la fenêtre, on peut se placer au niveau du codeur dans des conditions de codage très mauvaises, car:

- Les échantillons (filtrés) déplacés sont tous hors de la trajectoire du signal original,
- L'algorithme d'adaptation des dimensions de la fenêtre (positions relatives des échantillons entre eux) ne peut suivre les variations du signal d'origine.

Des conditions divergentes peuvent alors apparaître.

2.2 Résultats du codage préliminaire

Avec un filtre d'ordre 2, limité à une fréquence de coupure haute de 3300 Hz, le principe proposé donne quelques résultats significatifs.

Toutefois l'état de l'étude en cours, s'il a permis de vérifier la validité du procédé envisagé, ne permet pas la conclusion définitive sur cette technique.

Les problèmes inhérents au principe et le développement de techniques de calculs plus efficaces (codage en arbre) s'imposent pour prolonger efficacement cette analyse.

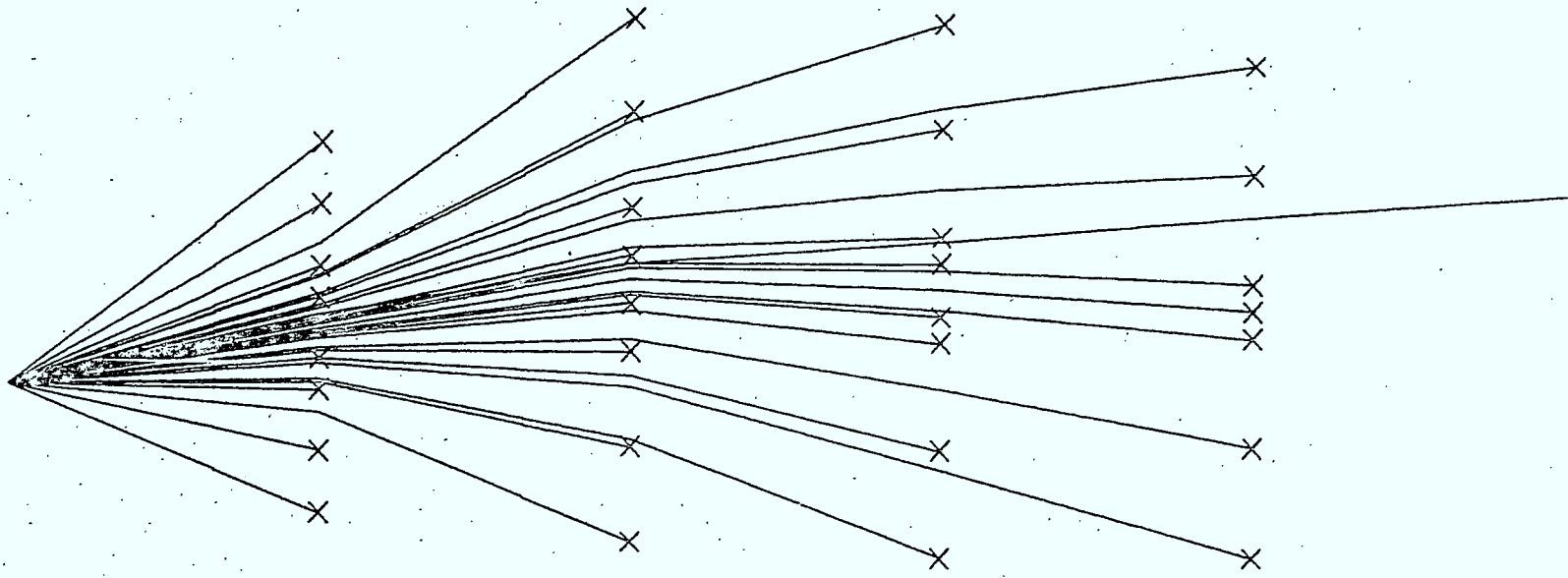


Fig. 6.5 Exemple d'effet de filtrage de la position des échantillons (de la fenêtre reconstituée) en sortie du filtre du codeur.

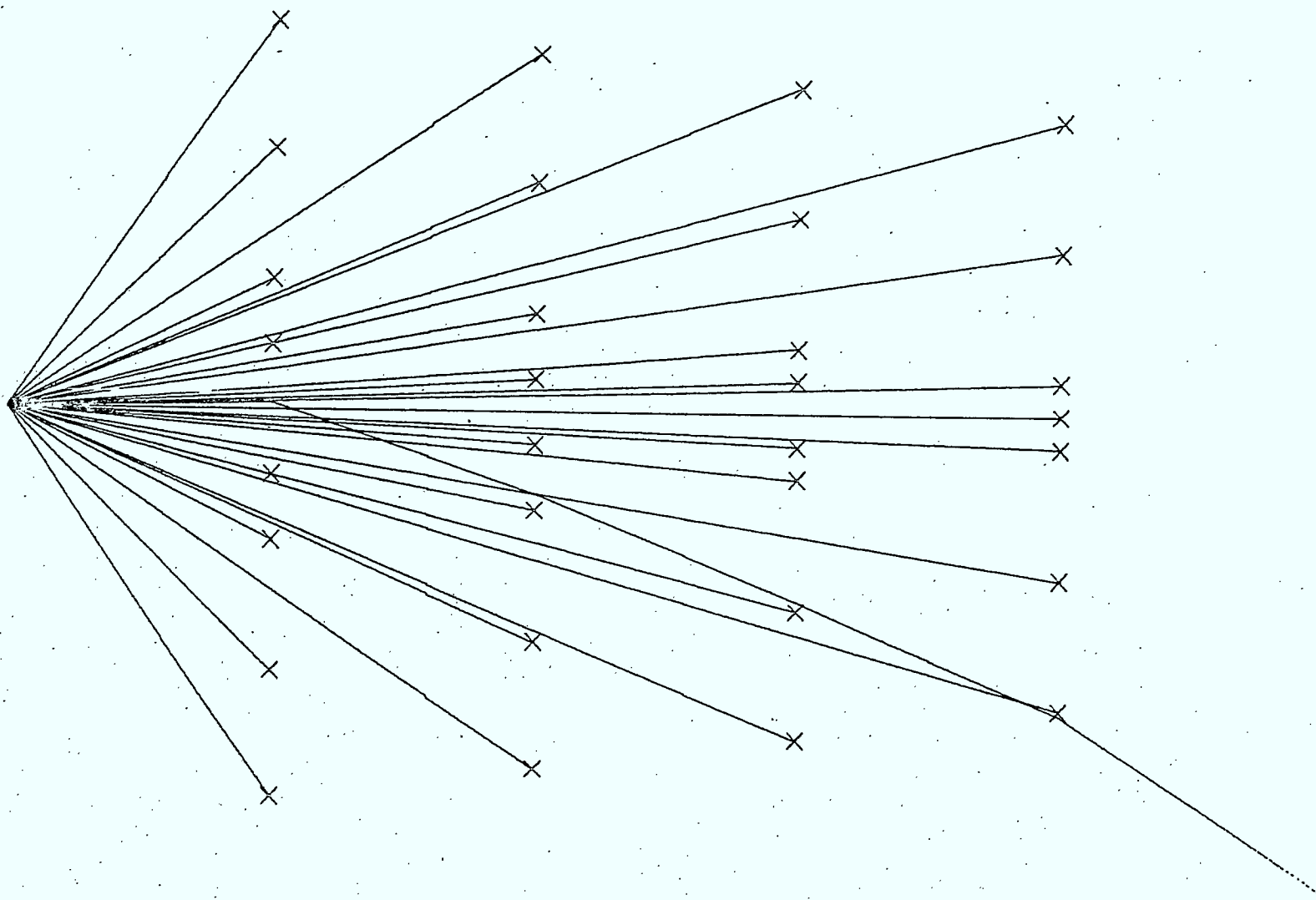


Fig. 6.6 Exemple d'effet de filtrage de la position des échantillons (de la fenêtre reconstituée) en sortie du filtre du codeur.

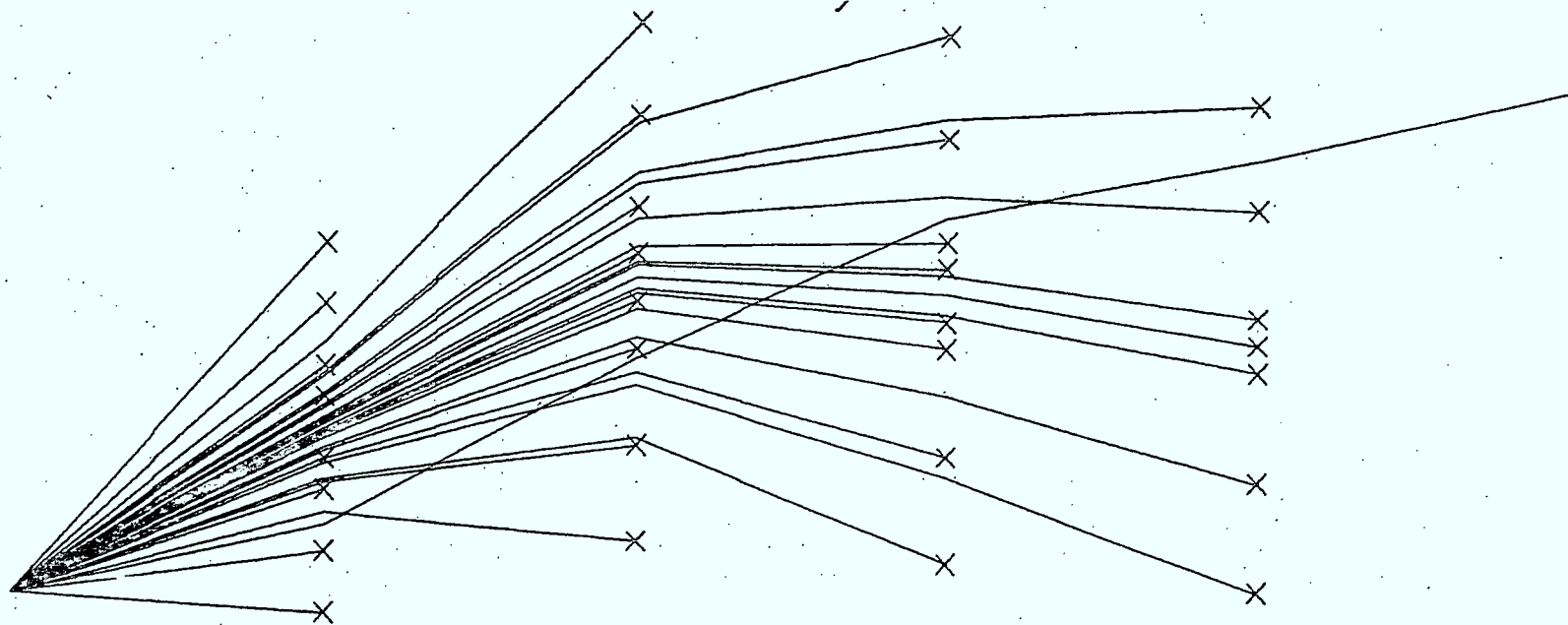


Fig. 6.7 Exemple d'effet de filtrage de la position des échantillons (de la fenêtre reconstituée) en sortie du filtre du codeur.

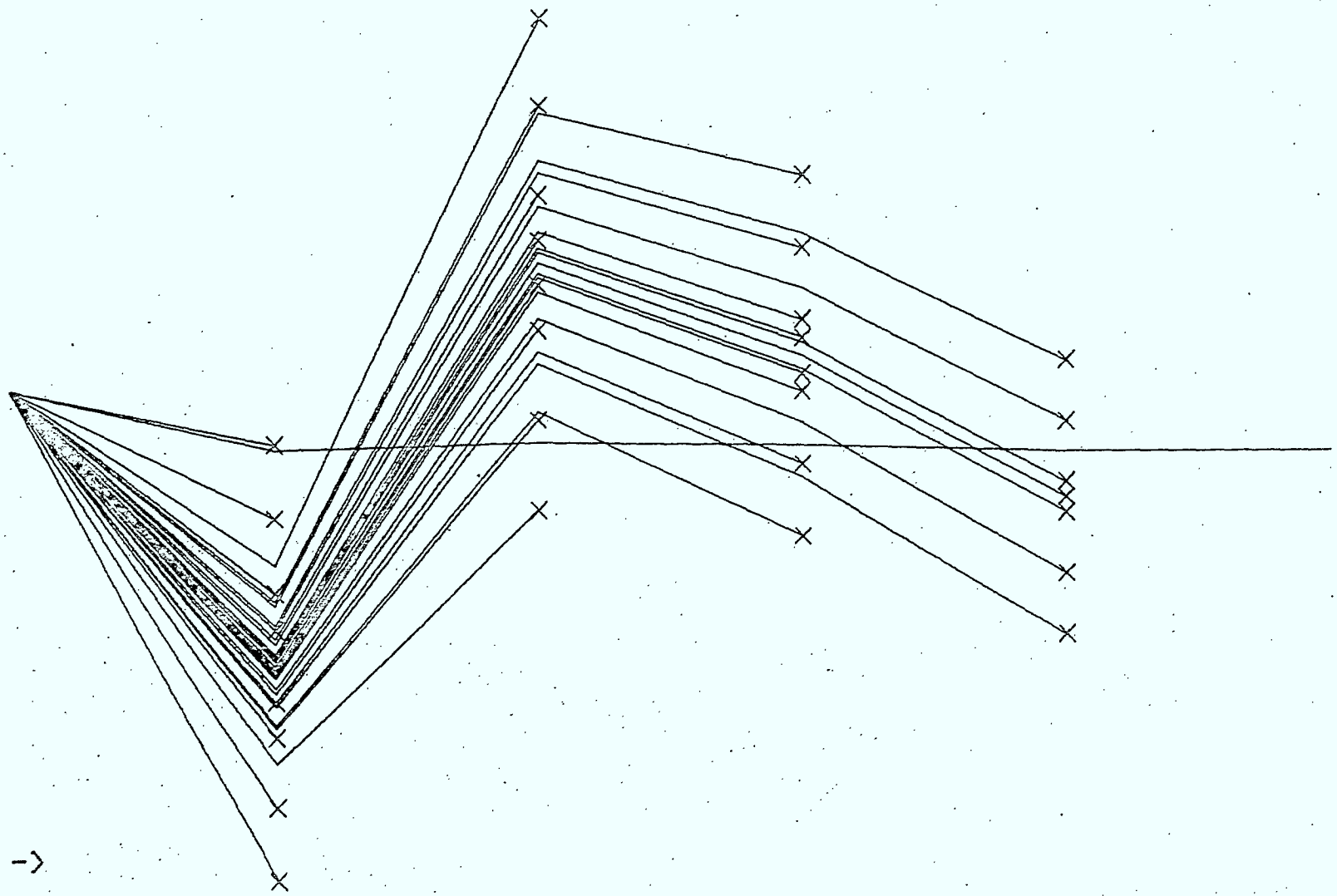


Fig. 6.8 Exemple d'effet de filtrage de la position des échantillons (de la fenêtre reconstituée) en sortie du filtre du codeur.

CONCLUSION GENERALE

Selon l'approche envisagée, codage et transmission avec réduction du débit à 32 Kbit/s (4 bits par échantillon codé) ou transmission analogique, le principe de codage en "temps élastique" semble donner des résultats prometteurs:

- Codage numérique:

Avec un débit de 32 Kbit/s et un retard (dû à la mémoire) de 32 millisecondes, les enregistrements présentés prouvent que la qualité du codage obtenu équivaut au moins à celle d'un PCM 6 bits et pour la plupart des phrases au PCM 8 bits.

Une comparaison avec les codages classiques à 32 Kbit/s (codage par modulation Delta, codage différentiel adaptatif ADPCM, codage par "sous-bandes"...) s'imposerait avant de donner des conclusions définitives. En tout cas le volume et la complexité de calcul (en longueur binaire par exemple) sont considérablement plus réduits dans le principe proposé que dans les codages ADPCM où des coefficients adaptatifs (de 2 à 4...) sont calculés par des algorithmes du gradient...

- La quantification optimale ne semble pas essentielle tant au niveau objectif (rapport S/N) que subjectif (tests d'écoute informels).

Selon l'intérêt présenté pour les codages dits "à 32 Kbit/s", il nous semblerait que cette technique mériterait un développement plus approfondi par une recherche plus optimale des paramètres de l'algorithme et un montage en simulation de tests d'évaluation comparée de qualité avec des techniques de codage déjà connues (modulation Delta, ADPCM, etc...)

- Transmission analogique

Le traitement totalement numérique reste d'une grande simplicité de principe, ce qui autorise certains avantages pour le codage proposé.

L'intégration au codeur et au récepteur d'organes hybrides (analogiques et numériques) ne pose pas de problèmes majeurs et les circuits spécialisés de traitement de signal (NEC, Texax Inst.) se prêteront aisément à l'implantation d'algorithmes.

Le problème spécifique de re-synchronisation semble possible d'après la solution simulée et l'influence du bruit, présent en communications radio-mobile, n'affecte que peu le rendement de l'algorithme.

A cette étape du travail, il apparaît que si l'intérêt présenté par la forme d'onde $\phi_n(t)$ (angle de codage qui ne présente que peu de modulation d'amplitude) devient un paramètre essentiel d'une liaison radio-mobile, modulation MF (FM) ou BLU (SSB), il serait utile de prolonger les études de simulation en cours par:

- une génération complète du système de transmission simulé par un transfert numérique/analogique, une modulation appropriée et un filtrage de canal avec bruit analogique: le signal ainsi obtenu sur bande magnétique sera à nouveau numérisé.
- un traitement numérique complet simulant le décodeur, à condition d'avoir prévu au codeur le principe spécialisé de traitement des silences et du bruit de fond pour l'efficacité, à la réception, de la re-synchronisation numérique.

ANNEXE AU RAPPORT - CASSETTE ENREGISTREE

La cassette jointe à ce rapport présente quelques résultats de codage en "temps élastique" et illustre l'effet de la réduction de débit numérique autorisé par ce système avec un maintien d'une bonne qualité de parole codée.

Il faut noter que l'enregistrement a été fait avec une courbe de correction "DOLBY" affectant la réponse en fréquence à la lecture (sur un appareil non doté de cette correction).

Quelques enregistrements #1 à 6 présentent une comparaison de phrases avec la référence MIC ou PCM à divers débits.

Deux autres enregistrements présentent:

- d'une part, l'effet de la réduction de taille de la mémoire (de 1024, 512 et 256 échantillons).
- d'autre part, un codage en présence de bruit plus ou moins élevé (bruit blanc additif) où le rapport signal à bruit est ajusté de 30 dB à 3 dB seulement.

Enregistrement #6 - Phrases à 6,4 KHz, filtrées 300-3000 Hz

Phrase 1 PCM 4 bits puis codage élastique 6 bits
4 bits.

Phrase 2 Même présentation

Effet de la taille mémoire

Enregistrement #7 - Phrases à 8 KHz, filtrées 300-3300 Hz

La même phrase répétée 3 fois, soit:

1^{ère} fois Mémoire 1024 mots (128 ms de retard)
2^{ème} fois Mémoire 512 mots (64 ms de retard)
3^{ème} fois Mémoire 256 mots (32 ms de retard).

On notera que les rapports S/N varient peu

S/N	25.3	25.4	23.1
S/N _g	29.1	29.5	29.1

quand la taille mémoire est réduite, ce qui limitera le délai de transmission à qualité égale

Effet du bruit de fond

Enregistrement #8 - Phrases à 8 KHz, filtrées 300-3300 Hz

Un bruit blanc additif est superposé à la parole avec des rapports S/N globaux de:

30dB, 20dB, 10dB, 5dB et 3dB

et présentés successivement sur la même phrase..

L'original bruité précède la phrase codée/décodée

On notera, sans traitement de bruit particulier, que le codage en temps élastique n'est pas perturbé par le bruit même important. Il contribue même à "assourdir" légèrement le bruit d'origine.

