

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT : D'ELECTRONIQUE

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

PROJET DE FIN D'ETUDES

SUJET

**Determination de la Fréquence
Fondamentale du Signal de
Parole et Etude de l'onde
Glottique par Filtrage Inverse**

Proposé par :

Melle M. GUERTI

Etudié par :

Mr M. ALAOUCHICHE
Mr M.N. HADDOUCHE

Dirigé par :

Melle M. GUERTI

Promotion : Janvier 1988



قُلْ إِنْ هَدَيْتِي وَسُئِلْتِي
وَمُحْيَيْتِي وَمَمَاتِي لِلَّهِ رَبِّ
الْعَالَمِينَ ﴿١٦٢﴾ لَا شَرِيكَ لَهُ وَبِذَلِكَ
أُمِرْتُ وَأَنَا أَوَّلُ الْمُسْلِمِينَ ﴿١٦٣﴾
سُورَةُ الْإِنْعَامِ

* D E D I C A C E S *

A ma mère, à mon père:

Pour tous les sacrifices qu'ils ont fait pour moi.

A mes frères et soeurs.

A mes oncles et tantes.

A tous les frères qui me sont chers.

A tous ceux que j'aime: En particulier ma femme, L. GUENANE

Je dédie ce modeste travail.

- MOHAMED -

---ooOoo---

A ma mère, à mon père

A mes frères et soeurs

A tous ceux qui me sont chers: En particulier, S. ADJAOUT

Je dédie ce modeste travail.

- MUSTAPHA NABIL -

---ooOoo---

* R E M E R C I E M E N T S *

المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

Nous tenons à remercier vivement notre promoteur M^{elle} M. GUERTI pour son aide, ses conseils et les intéressantes suggestions qu'elle nous a prodigué tout au long de la réalisation de ce travail.

Nous sommes très reconnaissant à toutes les personnes qui ont participé de près ou de loin à l'élaboration de ce mémoire.

Que tous les professeurs, qui ont contribué à notre formation, trouvent ici l'expression de notre profonde gratitude.

—ooOoo—

	Pages
Introduction Générale	1
Chapitre I - Notions de phonétique	3
I-1 - Introduction	4
I-2 - Description de l'appareil phonatoire humain	4
I-2-1 - Les poumons	4
I-2-2 - La trachée artère	4
I-2-3 - Le conduit vocal	4
I-3 - Fonctionnement acoustique	5
I-4 - Caractéristiques acoustiques du signal de parole	6
I-5 - Production du son	7
I-5-1 - Définition d'un son	7
I-5-2 - Les différents types de sons	7
I-6 - Classification des différents sons	8
I-6-1 - Les voyelles	8
I-6-2 - Les consonnes	8
I-7 - Conclusion	9
Chapitre II - Elements de traitement du signal	10
II-1 - Introduction	11
II-2 - Transformation de Fourier (TF)	11
II-2-1 - Définition de différentes sortes de TF	11
II-2-2 - La transformée de fourier rapide	12
II-3 - Echantillonnage, quantification et codage	12
II-3-1 - Théorème d'échantillonnage	13
II-3-2 - Quantification	13
II-3-3 - Codage	13
II-4 - Convolution et corrélation	14
II-4-1 - Convolution	14
II-4-2 - Corrélation	14
II-5 - Fenêtrage	15
II-5-1 - Le modèle de Hanning	15
II-5-2 - Le modèle de Hamming	15
II-5-3 - Caractéristiques et critères de choix	16
II-6 - Le filtrage	16
II-6-1 - Définition	16
II-6-2 - Les filtres analogiques	17
II-6-3 - Les filtres numériques	17
II-7 - Conclusion	20

Chapitre III - Les différentes méthodes d'analyse de la parole	21
III-1 - Introduction	22
III-2 - L'analyse spectrale	22
III-2-1 - L'analyse analogique	22
III-2-2 - L'analyse numérique	23
III-3 - L'analyse temporelle	23
III-3-1 - La méthode de passage par zéro	24
III-3-2 - La méthode de la fonction d'autocorrélation	24
III-3-3 - La méthode de la prédiction linéaire	24
III-4 - La méthode d'interpolation linéaire	29
III-5 - Conclusion	30
Chapitre IV - Détermination de la fréquence fondamentale (F_0)	31
IV-1 - Introduction	32
IV-2 - Le filtrage passe-bas	32
IV-3 - Détection de F_0 par la fonction d'autocorrélation	33
IV-3-1 - Autocorrélation modifiée	33
IV-3-2 - Méthode de SONDHI	34
IV-3-3 - Méthode de DUBNOWSKI	37
IV-4 - Technique de traitement parallèle	39
IV-4-1 - Principe général de la méthode	39
IV-4-2 - Fonctionnement des estimateurs de pitch	41
IV-4-3 - Estimation F_0 et décision	41
IV-5 - Observations	43
IV-5-1 - La méthode d'autocorrélation	43
IV-5-2 - La méthode d'autocorrélation modifiée	43
IV-5-3 - La méthode de traitement parallèle	43
IV-6 - Conclusion	44
Chapitre V - Analyse de l'onde glottique par filtrage inverse	45
V-1 - Introduction	46
V-1-1 - La méthode du masque de ROTHENBERG	46
V-1-2 - Méthode du tube sans réflexion	47
V-1-3 - Caractéristiques générales de l'onde glottique	48
V-2 - Représentation physiologique et modèle	49
V-2-1 - Modélisations linéaire du système de production	49
V-2-2 - Schéma physiologique et modèle	50
V-2-3 - Analyse	51
V-3 - Application à une voyelle	58
V-3-1 - Prétraitement " manuel "	58
V-3-2 - "traitement " automatique "	60
V-4 - Conclusion	70
Conclusion générale	71
Bibliographie	73
Annexe: Programmation	74



INTRODUCTION GENERALE

L'intérêt porté à la parole n'est pas nouveau. Depuis très longtemps, l'homme s'est intéressé à ce phénomène qui lui permet de s'exprimer, de communiquer ou tout simplement de vivre en société.

Au XVIII^{ème} siècle, le rêve tant caressé par l'homme de faire parler la machine se réalise enfin. Il suffit d'appuyer sur les touches correspondantes et vous serez dispensés de plusieurs tâches (calcul, tracé des courbes mathématiques, enregistrement, programmation, etc...).

En tant que phénomène physique, la parole est le signal acoustique produit par le système vocal. Elle nécessite des processus de production et de perception permettant de transmettre des informations d'un groupe d'individus à un autre sous forme acoustique.

Bien que l'étude des mécanismes de sa production et analyse ait fait l'objet de nombreux travaux, ceux-ci ne sont pas encore tous élucidés jusqu'à ce jour.

Depuis quelques années, les recherches fondamentales sur le fonctionnement de l'appareil phonatoire humain se sont intensifiées, stimulées par le champ des applications possibles : transmission de la parole à faible débit, synthèse et reconnaissance automatiques de la parole, aide aux handicapés, traduction automatique et apprentissage d'une langue, etc...

La reconnaissance du système de production de la parole s'améliore depuis quelques années grâce à l'emploi de techniques récentes permettant une analyse dynamique sur les plans acoustiques et articulatoires.

Notre projet de fin d'études comprend deux axes principaux :

- La détermination du pitch " F_0 " qui est un problème important dans le traitement automatique de la parole : il ne s'agit pas de reconnaître ce qui est dit, mais d'identifier la personne qui le dit.
- L'étude de l'onde glottique par filtrage inverse en vue de connaître sa forme et en déduire ses caractéristiques générales pour pouvoir la modéliser et l'analyser.

Pour atteindre ces objectifs, le travail a été mené de la manière suivante :

Le premier chapitre, traite l'anatomie de l'appareil phonatoire humain et son fonctionnement acoustique ainsi que la classification des différents sons du langage.

Le second est consacré au rappel de quelques éléments de base du traitement de signal indispensable par la suite.

Au troisième, nous verrons quelques techniques d'analyse du signal de parole en assistant surtout sur les méthodes qui seront utilisées dans ce projet.

Dans le quatrième, nous exploitons tout ce qui a été décrit précédemment pour déterminer " F_0 ", par les méthodes d'autocorrélation et de traitement parallèle.

Dans le dernier chapitre, nous présenterons l'analyse de l'onde glottique par filtrage inverse en utilisant la technique de covariance dans l'intervalle de la fermeture de la glotte.

Enfin nous présenterons, en annexe, les programmes écrits en Basic sur l'Olivetti M24 de l'E.N.P.A., correspondants :

- à la détermination de " F_0 " par la méthode d'autocorrélation.
- au calcul de " F_0 " par la méthode d'autocorrélation modifiée.
- à la détection de " F_0 " par la méthode de traitement parallèle.

-----oooOooo-----

CHAPITRE I

NOTIONS DE PHONETIQUE

-----ooOoo-----

I - 1 - INTRODUCTION

Nous étudierons brièvement dans ce chapitre l'anatomie des organes contribuant à la production de la parole et de leur fonctionnement acoustique.

Ce qui va nous permettre de comprendre la formation des sons élémentaires dans le but d'estimer la fréquence fondamentale du signal de parole et d'analyser l'onde glottique par filtrage inverse.

I - 2 - DESCRIPTION DE L'APPAREIL PHONATOIRE HUMAIN

I - 2 - 1 - LES POUMONS

Mis à part leur fonctionnement de respiration, ils génèrent l'air nécessaire à la production des sons par expiration produisant le flux qui sera conduit par la trachée artère.

I - 2 - 2 - LA TRACHEE ARTERE

C'est une cavité quasi-cylindrique supportant à son extrémité supérieure "le larynx" qui joue le rôle de respiration, de nutrition et de phonation. Ce dernier supporte à sa base deux muscles : "les cordes vocales" susceptibles de vibrer. L'espace circonscrit par celles-ci est "la glotte". La fréquence de vibration des cordes vocales est appelée fréquence laryngienne ou fréquence fondamentale (F_0) ou encore "Pitch" en anglais.

I - 2 - 3 - LE CONDUIT VOCAL

Il se compose de deux parties :

- Le conduit buccal qui est formé d'une cavité pharyngo-buccale de forme et de volume variables dûs à la mobilité de la langue et du maxillaire.
- Le conduit nasal de longueur 12 cm environ et composé de deux fosses fixes communiquant en parallèle avec la cavité buccale à l'aide du voile du palais ou "velum" (fig-1).

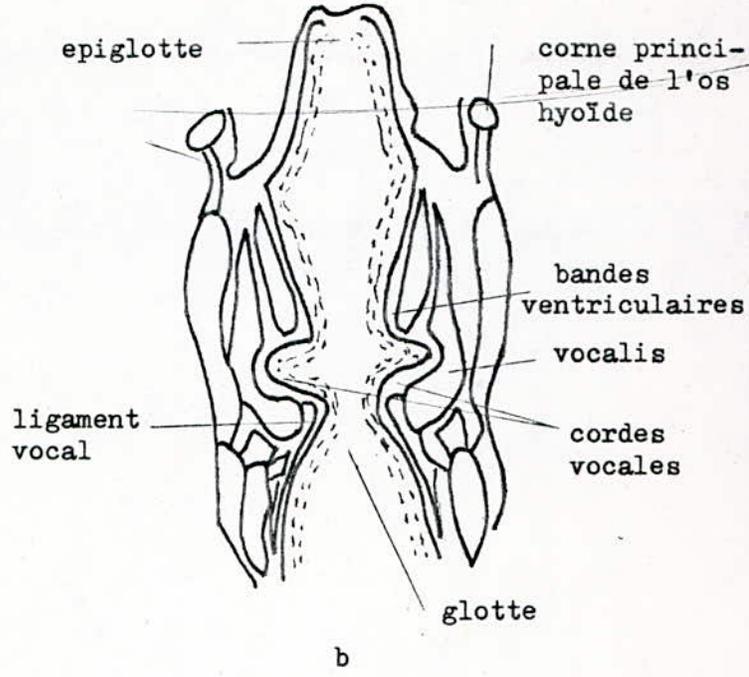
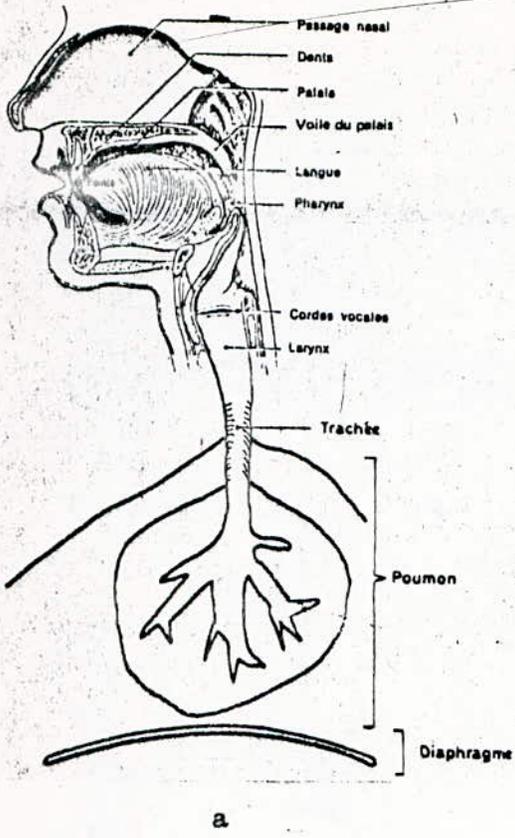


Figure 1 : a -- Appareil phonatoire humain
 b - Configuration interne du larynx

I - 3 - FONCTIONNEMENT ACOUSTIQUE

Le système acoustique peut être décrit comme suit :
 "Un générateur" qui est représenté par les poumons. Il alimente "l'excitateur" caractérisé par les cordes vocales. Celui-ci produit des signaux acoustiques devant être filtrés par "le résonateur" ou conduit vocal qui est composé de quatre cavités (fig-2).

Le générateur crée un flux d'air, celui-ci est mis en vibration par l'excitateur qui sera ensuite modifié par la fonction de transfert du résonateur.

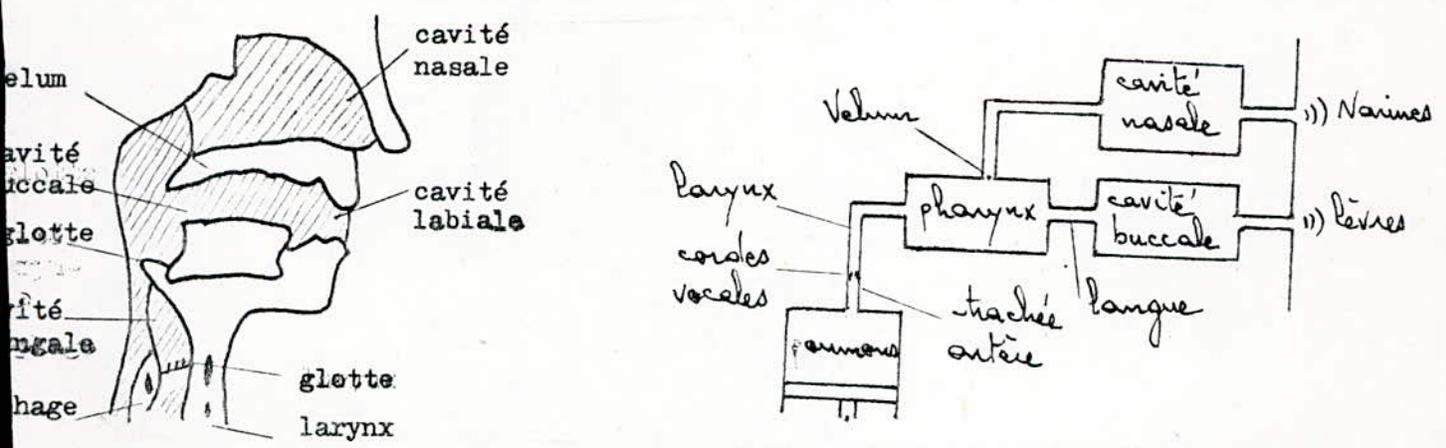


Figure-2 : L'appareil phonatoire en tant que système acoustique.

Dans la figure-2a, le conduit vocal se compose essentiellement des cavités pharyngale et buccale. Le phénomène physique "parole" résulte de l'excitation de ce système par l'ensemble poumons-cordes vocales pour les voyelles, et par un courant d'air agissant sur les parois du conduit pour certaines consonnes.

La cavité nasale intervient également lors de la production des consonnes du type "m" ou "n" et des voyelles nasales du type "on";

La figure-2b représente les trois cavités couplées qui constituent l'appareil vocal humain. La cavité nasale, indéformable, est couplée au pharynx par le vélum, la position de la langue règle le volume et le couplage des deux autres cavités. Chaque cavité, prise isolément, résonne à des fréquences déterminées par sa forme et son volume : le couplage des deux cavités fait apparaître des fréquences du système déplacées par rapport aux fréquences propres de chacune d'elles.

I - 4 - CARACTERISTIQUES ACOUSTIQUES DU SIGNAL DE PAROLE

- Le signal glottique a un spectre de raies périodiques, sa fréquence "Fo" varie de 70 à 150 Hz pour les hommes et de 150 à 300 Hz pour les femmes et au delà de 300 Hz pour les enfants.
- L'onde glottique a une forme approximativement triangulaire, le spectre de ce signal est par conséquent riche en harmoniques et présente une pente moyenne de : -12 dB/Octave. (cf. Ch V § 1-2).
- Le spectre transmis dans la parole s'étend de 50 Hz à 3 KHz pour les sons voisés et jusqu'à 6 KHz pour les sons non voisés.
- La fonction de transfert du conduit vocal présente un certain nombre de pôles appelés "formants" : c'est la concentration des harmoniques renforcés dans certaines bandes de fréquence particulières à chaque voyelle. Elle présente des zéros appelés "anti-formants" (figure-3).

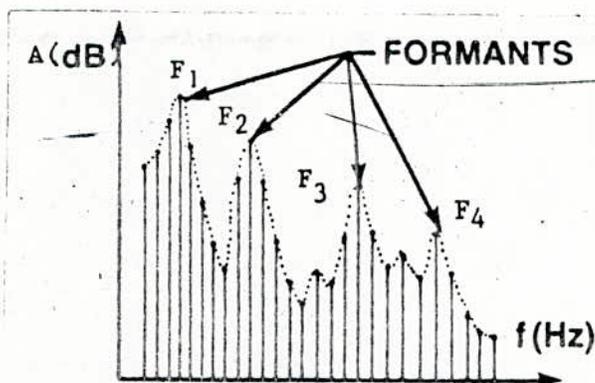


Figure-3 : Spectre vocalique présentant quatre résonances formantiques.

I - 5 - PRODUCTION DU SON

I - 5 - 1 - DEFINITION D'UN SON

Un son est une variation de pression du milieu ambiant (Air, eau...) pouvant être perçu par l'oreille. Dans la phonétique, un son est un phénomène physique qui résulte de l'excitation des cavités de l'ensemble vocal sous l'action d'un flux d'air qui agit sur les cordes vocales, celles-ci provoquent une vibration du milieu matériel dans laquelle elle peut se propager pour exciter notre ouïe.

I - 5 - 2 - LES DIFFERENTS TYPES DE SONS

L'appareil vocal présente deux catégories de sons :

- Voisés ou sonores :

La production de ce type de sons est illustrée par la figure-4.

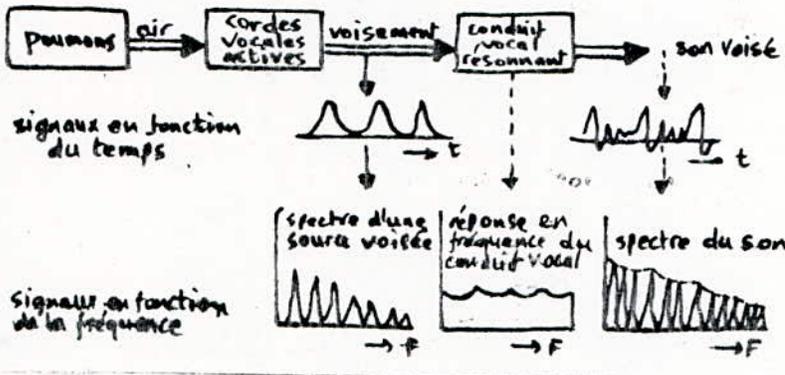


Figure-4

- Non voisés ou sourds :

Il sont dûs à un passage d'un flux d'air turbulent agissant sur la cavité buccale qui perturbe son passage (fermeture partielle de la bouche, contact de la langue avec les dents etc...) (fig-5).

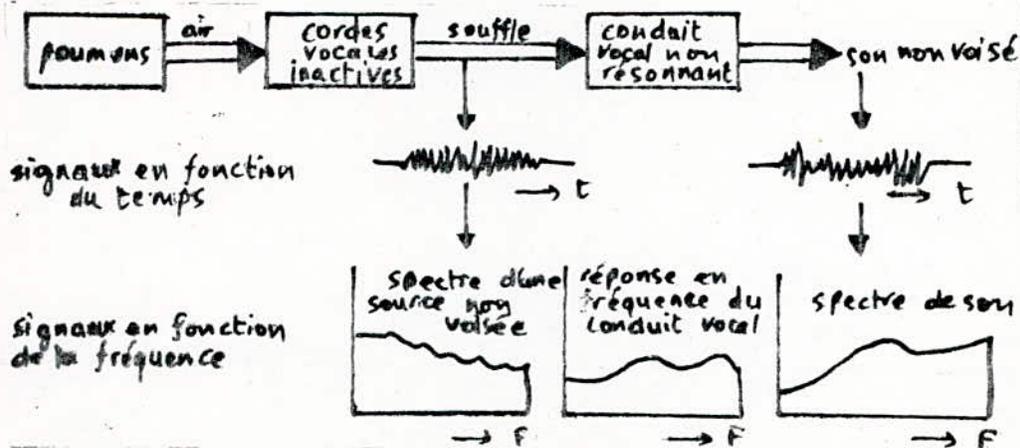


Figure 5 : Production d'un son non voisé.....

I - 6 - CLASSIFICATION DES DIFFERENTS SONS

Elle est basée sur les caractères voisé-non voisé, oral-nasal, le mode et le lieu d'articulation, par exemple :

En se relevant à l'horizontale, le velum interdit toute communication entre les cavités orale et nasale, on a dans ce cas un "SON ORAL" par contre, si le velum est abaissé, on obtient un "SON NASAL".

En phonétique, les sons du langage se divisent en deux classes :

Les voyelles et les consonnes.

I - 6 - 1 - LES VOYELLES

Elles sont caractérisées par le passage libre de l'air. La source d'excitation étant la vibration des cordes vocales, on distingue :

- Les voyelles orales, exemple : |a| , |u| , |i| , ...
- Les voyelles nasales, exemple : |ã| , |õ| , |ẽ| , ...

I - 6 - 2 - LES CONSONNES

Elles sont produites lors d'une constriction ou fermeture du passage de l'air. On distingue :

- Les consonnes fréquatives ou constrictives : Elles sont produites par le passage continu de l'air à travers un retrécissement du conduit vocal. Celles-ci peuvent être voisées (|v| , |z|) ou non voisées (|f| , |s|).
- Les consonnes plosives ou occlusives : Pour produire une plosive, il faut procéder à une fermeture plus tenue en un point particulier du conduit vocal, puis à un relâchement brusque. Elles peuvent être aussi voisées (|b| , |d| , |g|) ou non (|p| , |t| , |k|).
- Les consonnes nasales : Elles sont obtenues par une fermeture partielle à l'avant du conduit buccal avec l'abaissement du velum faisant, du conduit nasal, le canal de transmission. ex : |n| , |m| .

.../...

- Les semi-voyelles : [w| , |y| , |j| sont des transitions rapides du conduit vocal, ce qui les apparente aux consonnes, mais celles-ci continuent à fonctionner en mode résonnant, ce qui les apparente aux voyelles, d'où leur nom.
- Les consonnes liquides : [l| , |r| sont dues à une obstruction du flux d'air dans la bouche qui n'est pas suffisante pour créer une friction.

I - 7 - CONCLUSION

Ces notions sont à la base de toute étude portant sur le traitement automatique de la parole, sur le plan phonétique et articulatoire.

CHAPITRE II

ELEMENTS DE TRAITEMENT DU SIGNAL

---ooOoo---

II - 1 - INTRODUCTION

L'intérêt du traitement de signal n'est plus à démontrer. Utilisé sous les formes les plus diverses, il est, en effet, indispensable à la manipulation de l'information afin de la transmettre et de l'identifier avec une efficacité constamment accrue, en sorte qu'aujourd'hui, le traitement du signal est devenu une discipline scientifique à part entière, d'un niveau élevé et dont les bases, les méthodes et les possibilités ne peuvent plus être ignorées de l'ingénieur et du physicien.

Pour ces raisons, nous avons commencé par donner un très bref aperçu sur les éléments principaux du traitement afin de faciliter notre étude par la suite.

II - 2 - TRANSFORMATION DE FOURIER (TF)

Un signal quelconque a deux représentations possibles :

- Une représentation temporelle, $y = f(t)$
- Une représentation fréquentielle, $Y = F(f)$

Celles-ci sont reliées entre elles par la TF.

II - 2 - 1 - DEFINITION DES DIFFERENTES SORTES DE TF

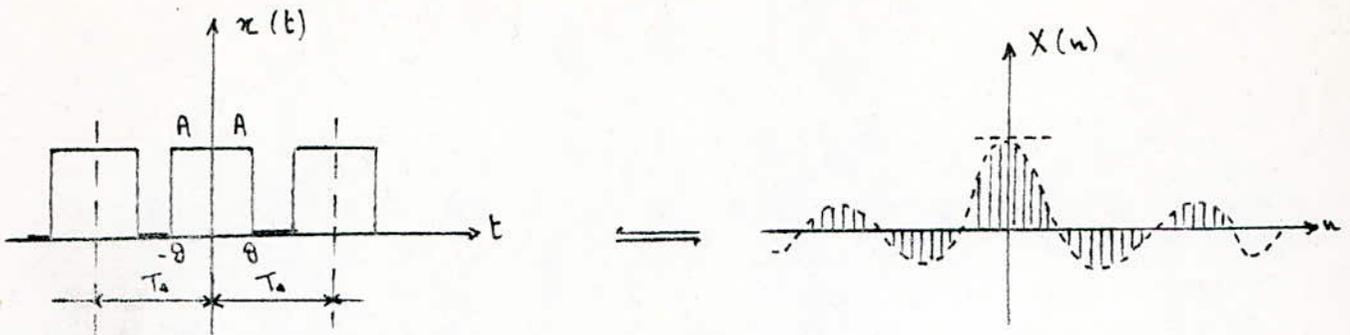
La transformation de fourier directe (TFD) fournit le spectre d'un signal défini dans le domaine temporel, tandis que la transformation de fourier inverse (TFI) permet de revenir du spectre au spectre initial.

II-2-1-a - Pour les signaux périodiques à temps continu

- La TFD : $X(n) = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \cdot \exp(-j2\pi nft) dt$ (1)

- La TFI : $x(t) = \sum_{n=-\infty}^{+\infty} X(n) \cdot \exp(j2\pi nft)$ (2)

Exemple : Figure ci-après montre le spectre d'une onde rectangulaire, périodique ; les spectres sont apériodiques et discrets.

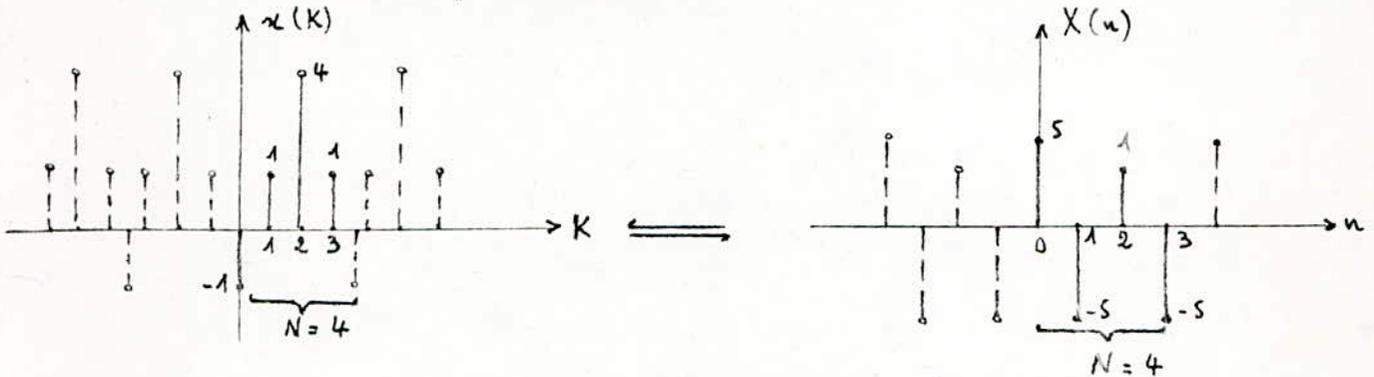


II-2-1-b - Pour les signaux périodiques à temps discret

- La TFD : $X(n) = \sum_k x(k) \cdot \exp(-j2\pi nk/N)$ (3)

- La TFI : $x(k) = \frac{1}{N} \sum_n X(n) \cdot \exp(j2\pi nk/N)$ (4)

Exemple : La figure, ci-dessous, montre que les spectres sont périodiques et à fréquence discrete.



II - 2 - 2 - LA TRANSFORMEE DE FOURIER RAPIDE (TFR)

L'emploi de la TFD comporte une sévère limitation due à la capacité du calculateur. Une TFR est alors une organisation méthodique du calcul des TFD ramenant le nombre des opérations à effectuer de "N²" à une valeur de l'ordre de N.Log₂ N. Cette organisation consiste à utiliser certains algorithmes de calcul (Cooley-Sand...) et à relever les symétries ou les ressemblances qui existent à l'intérieur même de la matrice W (J.MAX, 1981):

$$W = \exp(-j2\pi k/N)$$

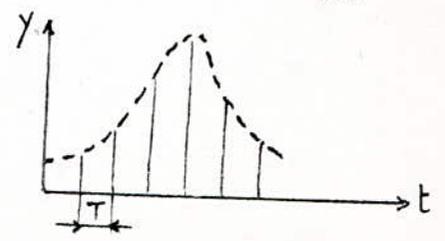
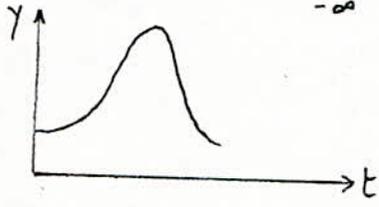
II - 3 - ECHANTILLONNAGE, QUANTIFICATION ET CODAGE

Dans de très nombreux cas, on ne traite pas directement les signaux analogiques (donc continus) fournis par les capteurs de mesure, mais on les échantillonne à une fréquence "F_e = 1/T_e", c'est à dire que ces signaux sont observés non pas d'une manière continue mais à certains instants seulement. Ces échantillons seront quantifiés, codés et stockés dans la mémoire des calculateurs en vue d'utilisation ultérieure.

II - 3 - 1 - THEOREME D'ECHANTILLONNAGE

L'échantillonnage consiste à représenter une fonction du temps $Y(t)$ par ses valeurs $Y(nT)$ à des instants multiples d'une durée "Te" appelée période d'échantillonnage. Echantillonner une fonction $Y(t)$ revient à multiplier cette fonction par une suite d'impulsions de dirac (peigne de dirac) de fréquence "Fe".

$$Y_e(t) = Y(t) \cdot \sum_{-\infty}^{+\infty} \delta(t-k/Fe) \quad (5)$$



Pour pouvoir reconstituer un signal à partir de ces échantillons, il faut que la fréquence d'échantillonnage soit supérieure ou égale à deux fois la plus haute fréquence contenue dans le spectre (Théorème de Shanon).

II - 3 - 2 - QUANTIFICATION

C'est l'approximation de chaque valeur du signal $Y(t)$ par un multiple entier d'une quantité élémentaire "q" appelée échelon de quantification.

Si "q" est constante quelque soit le signal, la quantification est dite uniforme. Cette opération revient à faire passer le signal dans un organe qui possède une caractéristique en marche d'escalier appelé quantificateur "Q". La manière dont l'approximation est faite définit le centrage de cette caractéristique. Toute valeur du signal comprise entre $(\frac{n-1}{2})q$ et $(\frac{n+1}{2})q$ est arrondie à nq (fig-6).

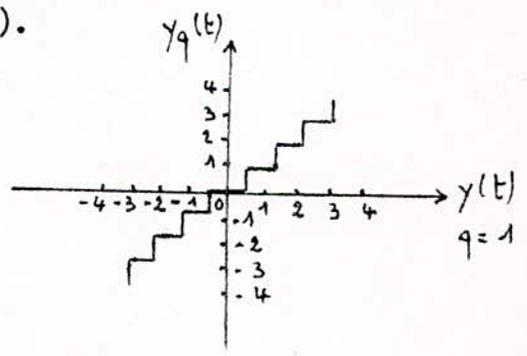
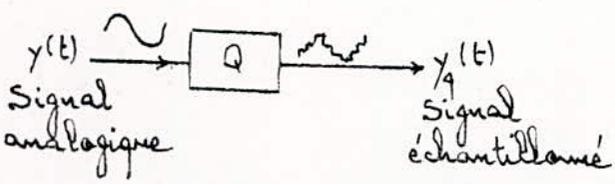


Figure -6 : Opération de quantification

II - 3 - 3 - CODAGE

Le signal échantillonné et quantifié en amplitude est représenté par une suite de nombres binaires. Si chaque nombre compte N bits, le nombre maximum d'amplitudes quantifiées qu'il est possible de distinguer s'élève à 2^N .

II - 4 - CONVOLUTION ET CORRELATION

Les notions de convolution et de corrélation occupent une place importante dans le traitement de signal, de ce fait il est indispensable de les connaître et de savoir utiliser leurs propriétés.

II - 4 - 1 - CONVOLUTION

La convolution est définie à partir de la réponse impulsionnelle par une intégrale dans le cas des signaux analogiques. Ainsi, si "e(t)" désigne le signal à filtrer, "h" la réponse impulsionnelle du filtre, le signal filtré "S(t)" est donné par :

$$S(t) = \int_{-\infty}^t e(\tau) \cdot h(t - \tau) d\tau \quad (6)$$

qui s'écrit symboliquement :

$$S(t) = e(t) * h(t) \quad (7)$$

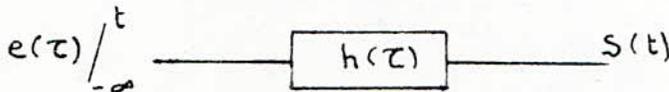


Figure-7 : Opération de convolution

On peut passer d'un produit de convolution à un produit simple (Théorème de Plancherel) par les relations suivantes :

$$e(t) * h(t) = E(f) \cdot H(f) \quad (8)$$

$$e(t) \cdot h(t) = E(f) * H(f) \quad (9)$$

II - 4 - 2 - CORRELATION

C'est une opération qui permet d'évaluer la ressemblance d'un signal avec un autre. Pour cela, on décale leurs échelles de temps d'un intervalle "τ" et on multiplie l'un par l'autre. L'autocorrélation procède des mêmes opérations, mais exprime la ressemblance d'un signal avec lui-même.

Considérons deux signaux X(t) et Y(t), leur fonction d'intercorrélation est définie par :

$$C_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) \cdot y(t - \tau) dt \quad (10)$$

En particulier, pour un même signal X(t), sa fonction d'autocorrélation est définie par :

$$C_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) \cdot X(t - \tau) dt \quad (11)$$

Les fonctions de corrélation jouissent de multiples propriétés. Nous citerons parmi elles la plus importante :

La densité spectrale est la (TF) de la fonction de corrélation :

$$S_{xx}(f) = TF \ C_{xx}(\tau) ; S_{xy}(f) = TF \ C_{xy}(\tau) \quad (12)$$

II - 5 - FENETRAGE

Pour étudier un signal sur un intervalle de durée finie " ϑ ", il suffit de multiplier ce signal par une fenêtre définie sur cette même durée. Donc le rôle principal de la fenêtre est de limiter la durée d'un signal, l'exemple le plus simple est la fenêtre rectangulaire définie par :

$$\phi(t) = \begin{cases} 1 & t_0 - \vartheta/2 < t < t_0 + \vartheta/2 \\ 0 & \text{ailleurs} \end{cases} \quad (13)$$

Cette fenêtre n'est pas très utilisée vu qu'elle possède dans le domaine fréquentiel des lobes latéraux relativement amples de période $\frac{1}{\vartheta}$ qui amènent des distorsions gênantes, ceci est connu sous le nom de phénomènes de "Gibbs". Cet effet peut être atténué avec l'utilisation d'autres types de fenêtres. Les modèles les plus connus sont les fenêtres de Hanning et de Hamming, etc...

II - 5-1 - LE MODELE DE HANNING

J-Van Hann a proposé ce modèle qui est défini dans le domaine temporel par la relation :

$$\psi(t) = \frac{1}{2} \left(1 + \cos \frac{2t}{\vartheta} \right) \quad 0 \leq t \leq \vartheta \quad (14)$$

La (TF) de cette dernière est de la forme :

$$\phi(f) = \frac{\frac{1}{2} \sin(\pi f \vartheta)}{\sin \pi f} - \frac{\frac{1}{4} \sin(\pi(f + \frac{1}{\vartheta})\vartheta)}{\sin \pi(f + \frac{1}{\vartheta})} + \frac{\frac{1}{4} \sin(\pi(f - \frac{1}{\vartheta})\vartheta)}{\sin \pi(f - \frac{1}{\vartheta})} \quad (15)$$

Les variations des fonctions ci-dessus dans le domaine temporel et fréquentiel sont données par la figure 8.

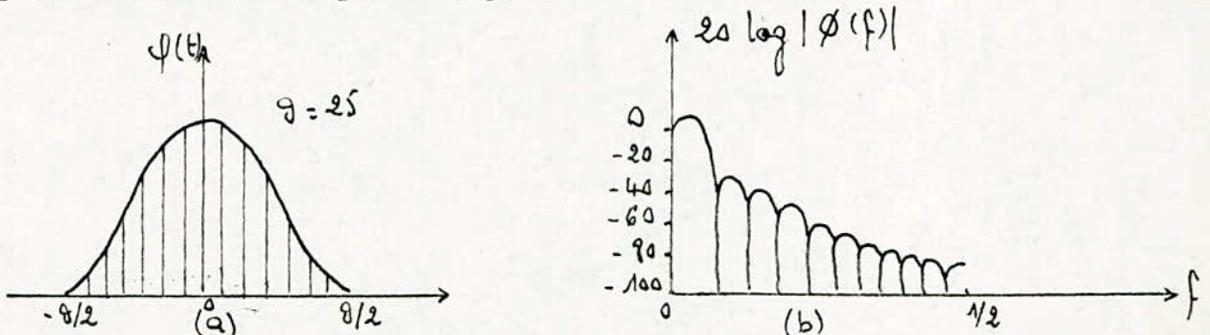


Figure-8 : Variation de la fonction de Hanning : a - Domaine temporel
b - Domaine fréquentiel

II - 5 - 2 - LE MODELE DE HAMMING

Sa fonction est donnée dans le domaine temporel sous une forme généralisée :

$$\psi(t) = \begin{cases} \alpha(1-\alpha) \cos(2\pi t/\vartheta) & \text{pour } |t| \leq \vartheta/2 \\ 0 & \text{ailleurs} \end{cases} \quad (16)$$

appelée fenêtre de Hamming généralisée.

- Pour $\alpha = \frac{1}{2}$, on obtient la fenêtre de Hanning.
- Pour $\alpha = 0,54$, on obtient la fenêtre de Hamming.

Sa TF est donnée par :

$$\phi(f) = 0,54 \frac{\sin \pi f \theta}{\sin \pi f} + \frac{0,23 \sin \pi \theta (f - 1/\theta)}{\sin \pi (f - 1/\theta)} + \frac{0,23 \sin \pi \theta (f + 1/\theta)}{\sin \pi (f + 1/\theta)} \quad (17)$$

Les variations des fonctions dans le domaine temporel et fréquentiel sont données par la figure-9.

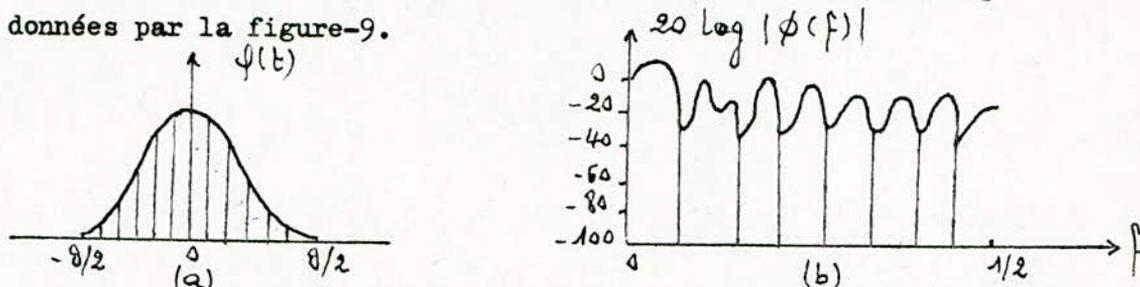


Figure-9 : Variation de la fonction de Hamming : a - Domaine temporel
b - Domaine fréquentiel

II - 5 - 3 - CARACTERISTIQUES ET CRITERES DE CHOIX

Une fenêtre est caractérisée par deux paramètres principaux :

- La longueur de base du pic central.
- Le rapport des amplitudes du premier lobe secondaire et du pic central.

Ce rapport est exprimé en décibels :

$$\lambda = 20 \text{ Log}_{10} (\phi(f)/\phi(0)) \quad (18)$$

Le choix de la fenêtre de pondération est un problème délicat, il dépend de plusieurs paramètres, nous citerons parmi les plus importants : la position, la forme et la durée.

II - 6 - LE FILTRAGE

La transmission intégrale d'une information est très onéreuse. Aussi est-il toujours intéressant de chercher à extraire la partie pertinente de l'information et à éliminer la partie non désirable (bruit). Ceci peut être réalisé grâce à un circuit appelé "filtre".

II - 6 - 1 - DEFINITION

Etant donné un phénomène physique dont le spectre s'étend sur une intervalle de fréquence quelconque, le filtre a pour rôle de renforcer certaines composantes du signal tout en affaiblissant les autres.

II - 6 - 2 - LES FILTRES ANALOGIQUES

Il existe essentiellement trois catégories de ces filtres :
passe-bas, passe-haut et passe-bande.

Dans chaque catégorie, le filtre est caractérisé par sa courbe de réponse ou encore par sa fonction de transfert (fig. 10).

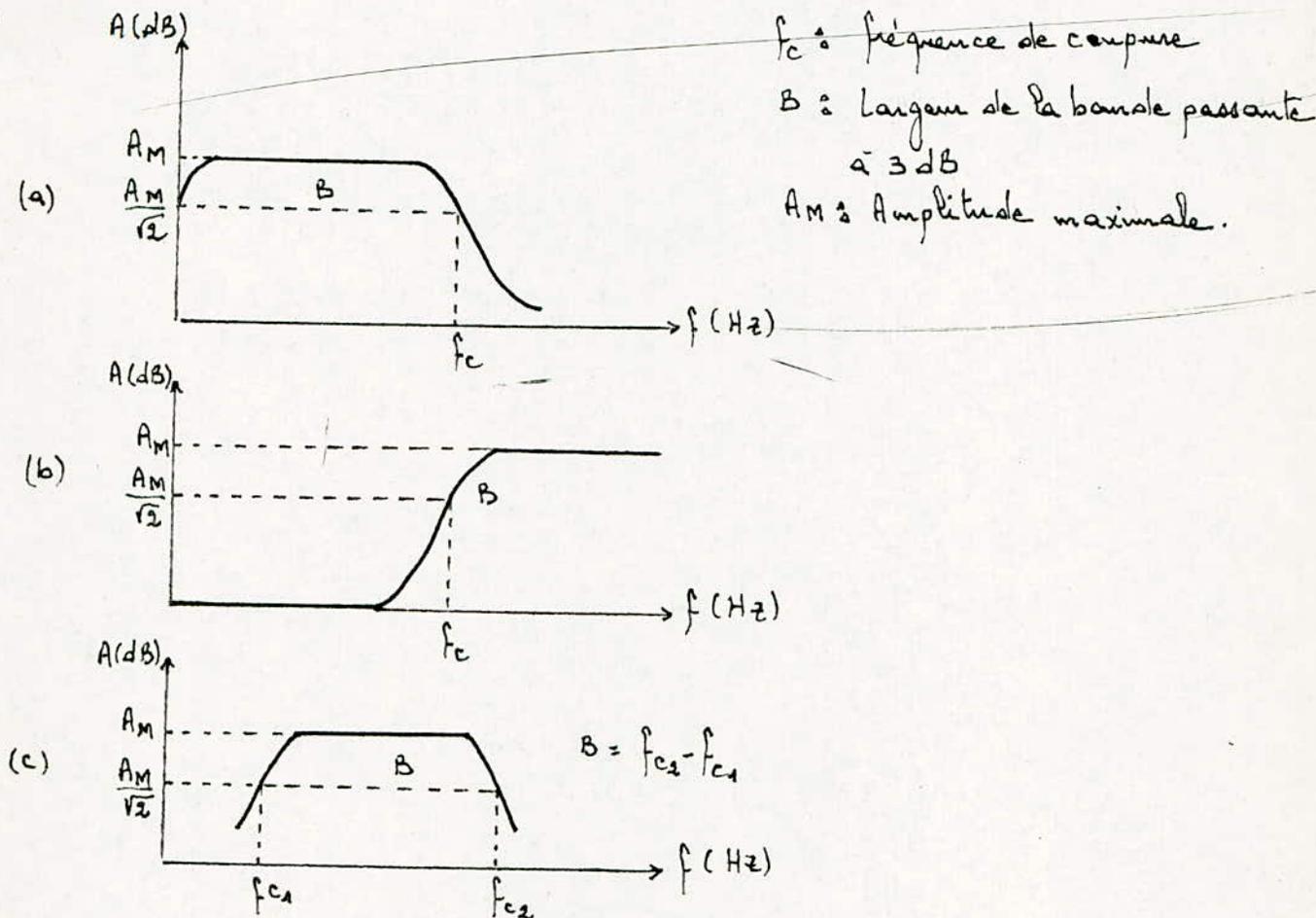


Figure 10 : Courbe de réponse du filtre : a - Passe-bas
b - Passe-haut
c - Passe-bande

II - 6 - 3 - LES FILTRES NUMERIQUES

1 - Définition

Les filtres numériques se rapportent à un système dans lequel un signal échantillonné et numérisé sous forme d'une suite de nombres $X(nT)$ ou plus simplement X_n , (T est la période d'échantillonnage et n le rang de l'échantillon) est transformé en une suite de nombres Y_n qui représente alors le signal de sortie. Dans le cas d'un filtre linéaire, la relation la plus générale entre les suites X_n et Y_n est de la forme :

$$Y_n = \sum_{K=1}^M a_K \cdot Y_{n-K} + \sum_{K=0}^M b_K \cdot X_{n-K} \quad (19)$$

On dit qu'on a un filtre récuratif quand l'un au moins des coefficients "a_K" est non nul.

2 - Les filtres non récuratifs

Un filtre non récuratif est un filtre dont chaque échantillon de sortie est une combinaison linéaire des échantillons présents et passés d'entrée. Les suites d'entrée X(n) et de sortie Y(n) sont reliées par une équation du type suivant qui constitue la relation de définition :

$$Y_n = \sum_{i=0}^{N-1} a_i \cdot X_{n-i} \quad (20)$$

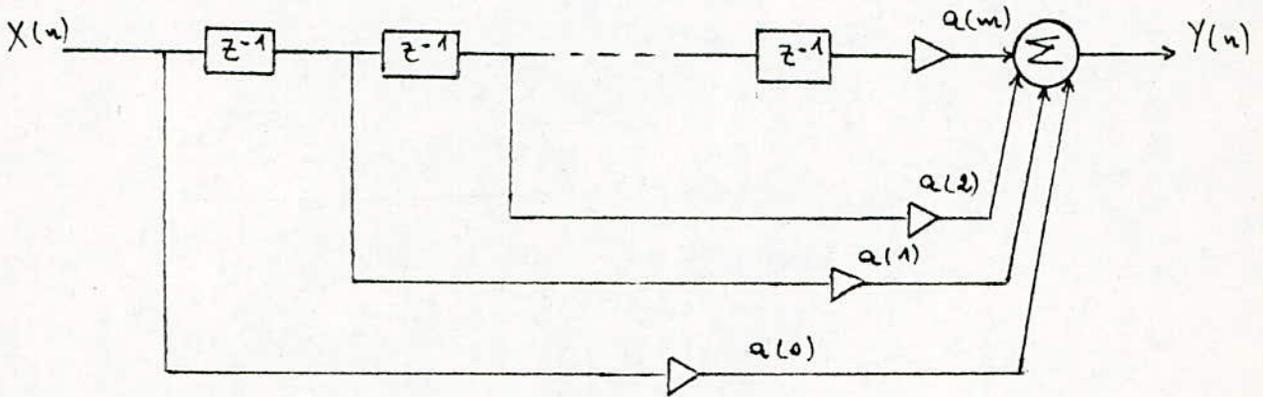


Figure 11 : Réalisation des filtres non récuratifs.

3 - Les filtres récuratifs

Un filtre récuratif est un filtre dont chaque échantillon de sortie est une combinaison linéaire des échantillons présents et passés d'entrée, ainsi que des échantillons passés de sortie.

$$Y_n = \sum_{l=0}^L a_l \cdot X_{n-l} - \sum_{k=1}^K b_k \cdot Y_{n-k} \quad (21)$$

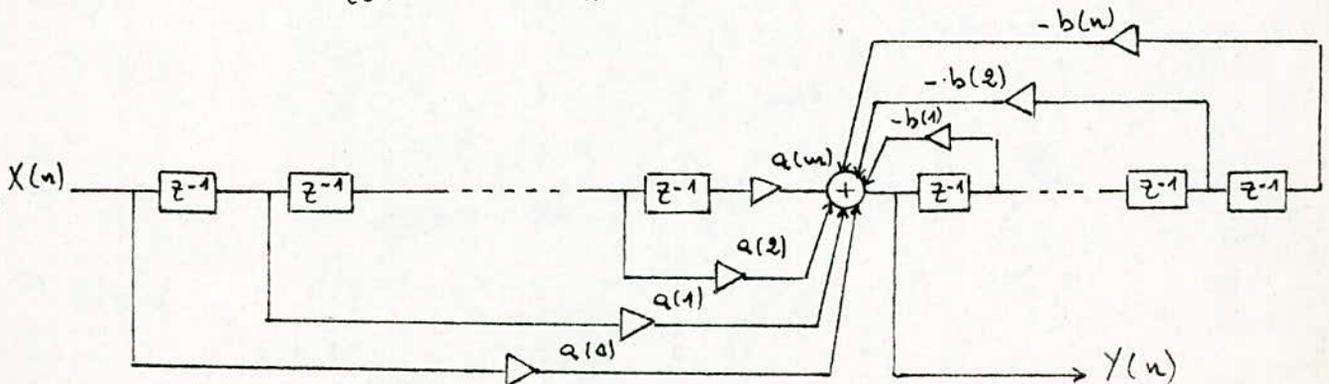


Figure 12 : Réalisation des filtres récuratifs, forme directe.

4 - Le filtre en treillis

La fonction de transfert du modèle de production du signal de la parole est :

$$H(Z) = \frac{1}{A(Z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^i} \quad (22)$$

Le filtre $A(Z)$ peut être donné par une structure en treillis (fig. 16).

Les équations de ce titre sont :

$$f_i(j) = f_{i-1}(j) + K_i b_{i-1}(j-1) \quad (23)$$

$$b_i(j) = b_{i-1}(j-1) + K_i f_{i-1}(j) \quad (24)$$

$$f_0(j) = b_0(j) = S(j) \quad (25)$$

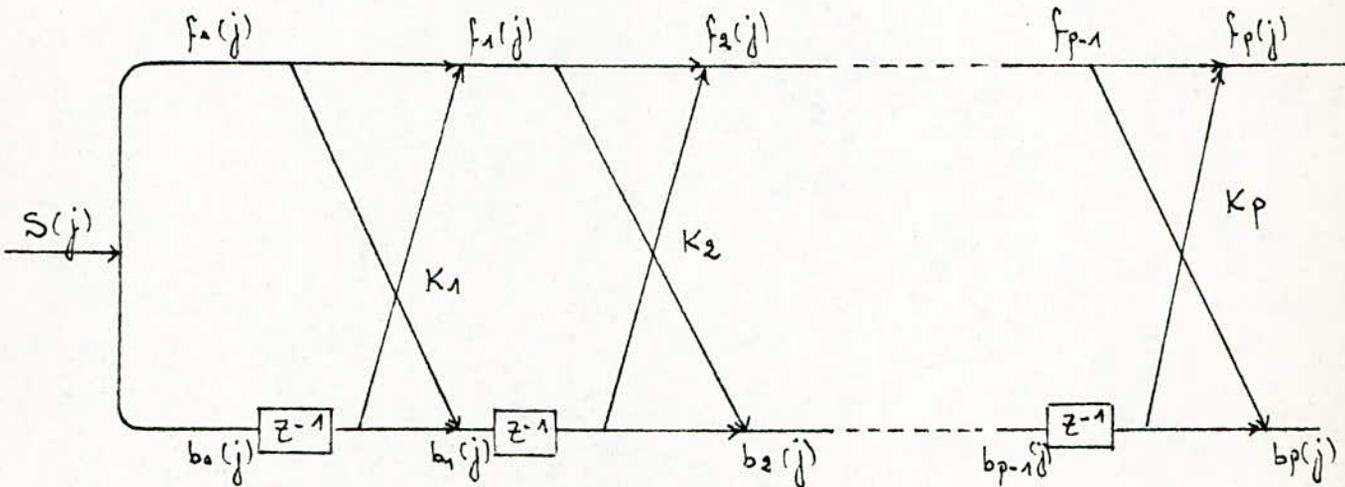


Figure 13 : Filtre inverse en treillis.

Les équations (23) et (24) expriment respectivement l'erreur précédente de prédiction à l'ordre "i" et l'erreur suivante de prédiction pour le même ordre. L'équation (25) exprime l'utilisation d'un prédicteur d'ordre zéro (ou la non utilisation de prédicteur).

5 - Fonction de transfert

Soient $x(k)$ et $y(k)$, les échantillons d'entrée et de sortie d'un filtre, soient $X(Z)$ et $Y(Z)$ leurs transformées en Z respectives. On appelle fonction de transfert d'un filtre numérique, le rapport des transformées $Y(Z)$ et $X(Z)$ soit :

$$H(Z) = Y(Z)/X(Z)$$

- Pour un filtre non récursif : $H(Z) = \sum_{i=0}^{N-1} a_i Z^{-i} \quad (26)$

- Pour un filtre récursif : $H(Z) = \sum_{l=0}^L a_l Z^{-l} / (1 + \sum_{k=1}^K b_k Z^{-k}) \quad (27)$

On appelle "Zéros" de la fonction de transfert les racines de l'équation $Y(Z) = 0$, et "Pôles" les racines de l'équation $X(Z) = 0$.

II - 7 - CONCLUSION

Les fonctions fenêtres étudiées nous permettent d'atténuer les lobes secondaires, seulement de manière différente.

En mesurant λ , il apparait que pour ($\vartheta = 9$), la fenêtre de Hanning conduit à $\lambda = 32$ dB et la fenêtre de Hamming donne $\lambda = -90$ dB. Ceci montre l'énorme atténuation des lobes secondaires, en passant de la fenêtre de Hanning ($\alpha = 1/2$) à celle de Hamming ($\alpha = 0,54$). En bref, les éléments décrits précédemment sont nécessaires pour l'analyse des signaux. Pratiquement on les utilise pour obtenir le signal convenable à l'application voulue et d'extraire les paramètres pertinents.

CHAPITRE III

LES DIFFERENTES METHODES D'ANALYSE

DE LA PAROLE

--ooOoo--

III - 1 - INTRODUCTION

L'étape d'analyse consiste à traiter le signal de parole dans le domaine temporel ou fréquentiel et en tirer des paramètres pertinents de ce signal. Dans tout traitement de la parole, on procède d'abord par une analyse du signal suivant certaines méthodes que nous décrirons dans ce chapitre. Le choix de la méthode se base surtout sur le type d'application et sur le coût.

III - 2 - L'ANALYSE SPECTRALE

Son objectif est la connaissance de la composition spectrale du signal de parole. Les méthodes utilisées dans ce cas sont :

III - 2 - 1 - L'ANALYSE ANALOGIQUE

Elle se base essentiellement sur :

- L'analyse de Fourier qui permet d'obtenir le spectre du signal de la parole en utilisant la transformée de Fourier (cf. Ch II § 2).

- L'analyse instrumentale dans laquelle on distingue :

- . Les analyseurs à filtres en parallèle.

Ils permettent une analyse pratiquement instantanée du signal de la parole tout en appliquant le signal $S(t)$ simultanément à un grand nombre de filtres fixés en fréquence et montés en parallèle. Chacun d'eux analysera la bande de fréquence correspondant à sa propre bande passante. Parmi ces systèmes, on peut citer l'analyseur du vocodeur à canaux (voice coder), constitué d'un banc de dix filtres couvrant la bande de 0 à 3 KHz.

- . Les analyseurs à filtre unique.

Le principe de base de ce genre de système est l'exploration de toute l'étendue du spectre utile du signal de la parole, à l'aide d'un filtre unique.

L'exemple-type de cette catégorie d'analyseurs est le sonographe constituant l'outil d'analyse du synthétiseur à formants.

III-2-2 - L'ANALYSE NUMERIQUE

La numérisation du signal et sa mise dans la mémoire d'ordinateur (fig. 14) faciliteront considérablement son traitement ultérieur.

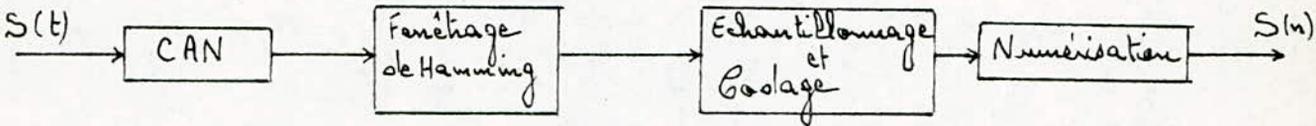


Figure 14 : Schéma du principe d'analyse numérique.

Il existe deux méthodes d'analyse numérique :

- a - L'ANALYSE PAR TFD : (cf. ch II § 2-1)
- b - L'ANALYSE CEPSTRALE

Le principe de cette méthode est de réaliser une séparation entre le spectre du signal glottale et la réponse en fréquence du conduit vocal.

L'hypothèse de base est que : le spectre de la parole est le produit du spectre de la source par celui des caractéristiques de résonance du conduit vocal. Dans ces conditions, si S(t) est le signal d'un son voisé, on a :

$$|S(\omega)|^2 = |G(\omega)|^2 |C(\omega)|^2 \quad (28)$$

- S(ω) : spectre d'amplitude de S(t)
- G(ω) : spectre d'amplitude du signal glottal
- C(ω) : réponse en fréquence du conduit vocal.

Les produits étant plus difficiles à manipuler que les additions, on transforme l'expression du spectre en somme de deux termes. Pour cela, on prend son logarithme :

$$\text{Log} (|S(\omega)|)^2 = \text{Log} (|G(\omega)|)^2 + \text{Log} (|C(\omega)|)^2 \quad (29)$$

Cette méthode permet la détermination du fondamental et l'extraction des formants.

III - 3 - L'ANALYSE TEMPORELLE

Elle nous permet d'étudier certains phénomènes qui sont mieux caractérisés par leur évolution dans le temps que par le spectre, c'est le cas, par exemple, de la fermeture brusque du conduit vocal lors de la production d'une plosive.

Les principales méthodes de cette analyse sont :

III - 3 - 1 - LA METHODE DE PASSAGE PAR ZERO

Le signal $S(t)$ prend la valeur zéro (ou change de signe) à des instants dont la répartition dans le temps est liée à certaines caractéristiques spectrales de $S(t)$. Cependant l'information relative à l'amplitude est perdue car on ne s'intéresse qu'à son signe. Cette méthode est utilisée pour la mesure de la fréquence de mélodie après filtrage passe-bas. Elle présente une grande simplicité de mise en oeuvre.

III - 3 - 2 - LA METHODE DE LA FONCTION D'AUTOCORRELATION

La fonction d'autocorrélation est une des méthodes de décalage, elle est utilisée pour la mesure de F_0 (cf. ch IV § 3). Elle est définie par :

$$Q(m) = \sum_{n=-\infty}^{+\infty} x(n) \cdot x(n+m) \quad (30)$$

$x(n)$: signal numérique
 m : décalage

Il est aisé de constater que :

- Si $x(n)$ est périodique de période T alors $Q(n)$ le sera aussi, la réciproque est vraie.
- Pour $m = 0$ l'énergie du signal $Q(n)$ est maximale. Elle présentera un pic à chaque position $m = nT$ ($n = 0, 1, 2, \dots$).
- Si $x(n)$ est non périodique, sa fonction d'autocorrélation $Q(m)$ ne possède aucun pic en dehors de celui pour $m = 0$.

Pour mettre en oeuvre cette méthode, le signal est découpé en fenêtres temporelles d'une longueur variant de 10 à 20 ms. Théoriquement, la fenêtre doit être courte pour que le paramètre à mesurer puisse être considéré comme suffisamment constant et long pour qu'il soit mesurable.

III - 3 - 3 - LA METHODE DE LA PREDICTION LINEAIRE (LPC)

C'est une technique d'analyse de la parole considérée comme étant l'une des plus puissantes. Elle a été appliquée pour la première fois en 1967 par F. Itakura et S. Saito puis développée par de nombreux chercheurs (J-D. Markel et A-H. Gray, 1976). Cette méthode est considérée à la fois comme temporelle et spectrale. Son importance est due à ce :

- Qu'elle soit fondée sur un modèle simple (filtre numérique récursif ne possédant que des pôles) de production de la parole, constituant une bonne approximation du système phonatoire.

- Qu'elle permette une estimation très précise des paramètres de la parole tels que formant, pitch, spectre, etc...
- Que le calcul de ces paramètres soit relativement rapide.

1 - Principe de la méthode LPC

Elle repose sur le fait qu'un échantillon de parole peut être approximativement prédit par une combinaison linéaire des échantillons précédents pondérés par des coefficients.

Soit : $S(n)$: l'échantillon du signal de parole original
 $\hat{S}(n)$: l'échantillon prédit de $S(n)$

On aura :

$$\hat{S}(n) = \sum_{k=1}^P a(k) \cdot S(n - k) \quad (31)$$

Où, P est l'ordre du prédicteur et $a(k)$ les coefficients du prédicteur.

Connaissant les coefficients de prédiction, on calcule le signal de manière approchée par récurrence et ceci tant qu'il reste stable (intervalle de temps suffisamment faible).

En appliquant à l'équation (31) la transformation en "Z", on aura :

$$\hat{S}(Z) = \sum_{k=1}^P a(k) \cdot Z^{-k} \cdot S(Z) \quad (32)$$

D'où :

$$\hat{S}(Z) = F(Z) \cdot S(Z) \quad (33)$$

Avec $F(Z)$: le filtre linéaire de prédiction.

D'autre part, l'erreur de prédiction entre le signal original et le signal prédit est donnée par :

$$e(n) = S(n) - \hat{S}(n) \quad (34)$$

$$E(Z) = S(Z) - \hat{S}(Z) \text{ (par la T.Z.)} \quad (35)$$

Les équations (32) et (35) donnent :

$$E(Z) = S(Z) \left[1 - \sum_{k=1}^P a(k) \cdot Z^{-k} \right] = S(Z) [1 - F(Z)] \quad (36)$$

On pose : $A(Z) = 1 - F(Z)$, on aura :

$$E(Z) = S(Z) \cdot A(Z) \quad (37)$$

Ainsi on obtient le modèle de prédiction linéaire (fig. 15)



Figure 15 : a - Représentation du filtre prédicteur $F(Z)$.

b - Représentation équivalente du filtre inverse.

On se base sur le principe que le conduit vocal filtrant le signal d'excitation "U(n)" peut être assimilé à un filtre récursif. La prédiction linéaire consiste donc à exprimer la valeur du n^{ème} échantillon en fonction des "p" échantillons précédents et de l'entrée "U(nT)" du système, si le signal est échantillonné avec une période d'échantillonnage "T", cette valeur est donnée par :

$$\hat{S}(nT) = \sum_{k=1}^p a(k).S(nT - kT) + GU(nT) \quad (38)$$

Où G est un facteur de gain.

On obtient ainsi le modèle de production de parole dans le domaine temporel (fig. 19), en appliquant la transformée en Z à l'équation (38) on obtient :

$$S(Z) = \sum_{k=1}^p a(k).S(Z)z^{-k} + GU(Z) \quad (39)$$

d'où

$$H(Z) = \frac{S(Z)}{U(Z)} = \frac{G}{1 - \sum_{k=1}^p a(k).z^{-k}} \quad (40)$$

d'où le modèle de production de la parole dans le domaine fréquentiel.

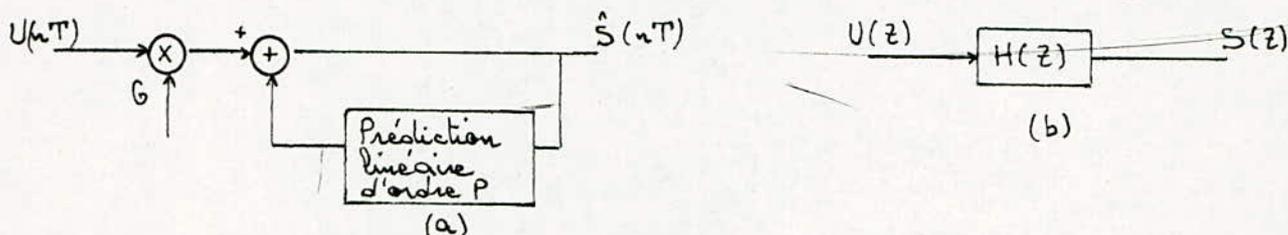


Figure 16 : Modèle de production de la parole :

a - Domaine temporel. b - Domaine fréquentiel

Le modèle utilisé dans la méthode LPC est fondé sur les trois hypothèses suivantes :

- La source vocale et le conduit vocal peuvent être considérés comme constants pour les intervalles de temps courts.
- Pendant ces intervalles de temps, on suppose que l'on approche la fonction de transfert du conduit vocal à un filtre récursif.
- Le signal de parole de synthèse est constitué par le signal de sortie d'un filtre dont les coefficients a(k) avec 1 k p sont renouvelés.

D'autre part, la prédiction linéaire permet d'estimer des paramètres pertinents du signal, pour cela deux critères sont utilisés :

a) Le critère (L₁) de la valeur absolue de l'erreur minimale

Il consiste en la minimisation de la quantité :

$$L_1 = \sum_n \left| S(n) - \sum_{k=1}^p a(k).S(n-k) \right| \quad (41)$$

b) Le critère (L₂) des moindres carrés

Ce critère est le plus utilisé, il consiste en la minimisation de l'erreur quadratique totale, à savoir que :

$$e(n) = S(n) - \hat{S}(n) = S(n) - \sum_{k=1}^p a(k) \cdot S(n-k) \quad (42)$$

Soit E l'erreur quadratique totale :

$$E = \sum_n e^2(n) = \sum_n \left[S(n) - \sum_{k=1}^p a(k) \cdot S(n-k) \right]^2 \quad (43)$$

On cherche à minimiser "E" par rapport à chaque coefficient prédicteur "a(k)", pour cela on procède par la dérivée partielle suivante :

$$\frac{\partial E}{\partial a(i)} = 0 \text{ pour } 1 \leq k \leq p, 1 \leq i \leq p \text{ d'où :}$$

$$2 \sum_n \left[S(n) - \sum_{k=1}^p a(k) \cdot S(n-k) \right] S(n-i) = 0 \implies \sum_n \sum_{k=1}^p a(k) \cdot S(n-k) \cdot S(n-i) = \sum_n S(n) S(n-i) \quad (44)$$

L'équation (44) peut s'écrire encore :

$$\sum_{k=1}^p a(k) \sum_n S(n-k) S(n-i) = \sum_n S(n) S(n-i) \quad (45)$$

L'équation (45) constitue un système de "p" équations à "p" inconnues.

2 - Les principales méthodes de la LPC

Pour la résolution du système linéaire d'équations précédentes, on utilise deux méthodes :

a - La méthode de covariance

Elle consiste à trouver les "p" coefficients a(k) en posant :

$$C(i,k) = \sum_n S(n-k) \cdot S(n-i) \quad (46)$$

L'équation (45) devient :

$$\sum_{k=1}^p a(k) \cdot C(i,k) = C(i,0) \text{ avec } 1 \leq i \leq p, 1 \leq k \leq p$$

Supposant qu'on a une séquence de N échantillons de parole :

$$S(n) = [S(0), S(1), \dots, S(N-1)]$$

La méthode de covariance se résume comme suit :

$$a(k) \cdot C(i,k) = C(i,0) \quad (47)$$

$$\text{avec } C(i,k) = \sum_{n=0}^{N-1} S(n-i) \cdot S(n-k)$$

L'équation (47) s'écrit sous forme matricielle :

$$\begin{bmatrix} C(1,1) & C(1,2) & \dots & C(1,p) \\ C(2,1) & C(2,2) & \dots & C(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ C(p,1) & C(p,2) & \dots & C(p,p) \end{bmatrix} \times \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = \begin{bmatrix} C(1,0) \\ C(2,0) \\ \vdots \\ C(p,0) \end{bmatrix} \quad (48)$$

qui devient sous forme condensée :

$$[C_p] \cdot [A_p] = [C_0]$$

C_p : Matrice de covariance

A_p : Vecteur colonne de composantes $a(k)$

C_0 : Vecteur colonne de composantes $C(i,0)$.

La structure de C_p est symétrique et définie positive.

b - La méthode d'autocorrélation

Cette méthode est fondée sur les hypothèses suivantes :

- Le signal est nul à l'extérieur de l'intervalle considéré. Ceci est réalisé à l'aide d'un fenêtrage temporel de largeur "N".
- Pour tout le temps ($n \in]-\infty, +\infty[$), chaque échantillon est prédit par les "p" échantillons qui le précèdent.
- La minimisation de l'erreur quadratique entre le signal fenêtré et le modèle se fait de $-\infty$ à $+\infty$

Posons : $m = n - i$

L'équation (45) devient :

$$\sum_{k=1}^p a(k) \sum_m S(m) \cdot S(m + |i - k|) = \sum_m S(m + |i|) \cdot S(m) \quad (49)$$

En général, l'autocorrélation est donnée par :

$$C(i,k) = \sum_{n=-\infty}^{+\infty} S(n - i) \cdot S(n - k) \quad (50)$$

Puisque le signal est limité par une fenêtre

$$S(m) = 0 \text{ pour } m \notin [0, N - 1]$$

L'équation (50) devient :

$$C(i,k) = \sum_{m=0}^{N-1-|i-k|} S(m) \cdot S(m + |i - k|) \quad (51)$$

D'autre part, on définit les coefficients de corrélation par :

$$R(i - k) = \sum_{m=0}^{N-1-|i-k|} S(m) \cdot S(m + |i - k|), \text{ d'où :} \quad (52)$$

$$R(i) = \sum_{m=0}^{N-1-|i|} S(m) \cdot S(m + |i|)$$

La résolution du système (49) donne :

$$R(i) = \sum_{k=1}^p a(k) \cdot R(|i - k|) \quad 1 \leq i \leq p \quad (53)$$

L'équation (53) se traduit par :

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \times \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \quad (54)$$

Nous remarquons que la matrice carrée (R_p) est symétrique et définie positive. Les éléments situés sur ses diagonales parallèles à la diagonale principale sont égaux : c'est la matrice de "TOEPLITZ".

III - 4 - LA METHODE D'INTERPOLATION LINEAIRE

Elle consiste à calculer la valeur d'un échantillon " S_n " à partir d'échantillons se trouvant de part et d'autre de lui-même.

Soit : $S(n)$: l'échantillon du signal de parole original

$\hat{S}(n)$: l'échantillon prédit par interpolation linéaire

On aura :

$$\hat{S}(n) = - \sum_{k=1}^{M'} a_k \cdot S_{n-k} - \sum_{l=1}^{M'} b_l \cdot S_{n+l} \quad (55)$$

D'où :

$$\hat{S}(n) = - \sum_{k=1}^{M'} a_k \cdot S_{n-k} + b_k \cdot S_{n+k} \quad (56)$$

L'erreur de prédiction s'écrit alors :

$$e(n) = S(n) - \hat{S}(n) = S_n + \sum_{k=1}^{M'} a_k \cdot S_{n-k} + b_k \cdot S_{n+k} \quad (57)$$

L'erreur quadratique moyenne sera donnée par :

$$E(n) = [e(n)]^2 = \left[S_n + \sum_{k=1}^{M'} a_k \cdot S_{n-k} + b_k \cdot S_{n+k} \right]^2 \quad (58)$$

La minimisation s'obtient lorsque : $\frac{\partial E}{\partial a_j} = 0, \frac{\partial E}{\partial b_j} = 0, j=1, \dots, M'$

$$a) \frac{\partial E}{\partial a_j} = \left(S_n + \sum_{k=1}^{M'} a_k S_{n-k} + b_k S_{n+k} \right) S_{n-j} \quad (59)$$

$$\frac{\partial E}{\partial a_j} = C_{aj} + \left[\sum_{k=1}^{M'} a_k C_{k,j} + b_k C_{-k,j} \right] = 0 \quad (60)$$

$$b) \frac{\partial E}{\partial b_j} = \left(S_n + \sum_{k=1}^{M'} a_k S_{n-k} + b_k S_{n+k} \right) S_{n+j} \quad (61)$$

$$\frac{\partial E}{\partial b_j} = C_{aj} + \left[\sum_{k=1}^{M'} a_k C_{k,-j} + b_k C_{-k,-j} \right] = 0 \quad (62)$$

Pour optimiser les coefficients a_k et b_k , on ne s'intéresse qu'à la méthode de covariance dont le coefficient de covariance est :

$$C_{ij} = \sum_{n=M'+1}^{N-M'} S_{n-i} \cdot S_{n-j} \quad (63)$$

En injectant " C_{ij} " dans les équations (60) et (62), on aboutit au système matriciel suivant :

$$\begin{bmatrix} C_{01} \\ C_{02} \\ \vdots \\ C_{0L} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1L} \\ C_{21} \\ \vdots \\ C_{L1} & \dots & C_{LL} \end{bmatrix} \begin{bmatrix} C_{1,-1} & \dots & C_{1,-M} \\ \vdots \\ C_{M,-1} & \dots & C_{M,-M} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_L \end{bmatrix}$$

$$\begin{bmatrix} C_{0,-1} \\ \vdots \\ C_{0,-L} \end{bmatrix} = \begin{bmatrix} C_{-1,1} & \dots & C_{-1,M} \\ \vdots \\ C_{-M,1} & \dots & C_{-M,M} \end{bmatrix} \begin{bmatrix} C_{-1,-1} & \dots & C_{-1,-M} \\ \vdots \\ C_{-M,-1} & \dots & C_{-M,-M} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_L \end{bmatrix}$$

Sait :

$$\begin{cases} |C_{0i}| = |C_{ij}| |C_{i,-j}| \\ |C_{0,-i}| = |C_{-i,j}| |C_{-i,-j}| \end{cases} \quad (64)$$

III - 5 - CONCLUSION

L'analyse par LPC donne de bons résultats pour le calcul des paramètres de la parole. La méthode de covariance est généralement utilisée sans aucune restriction, le seul problème réside en la stabilité du modèle. Par contre, la méthode d'autocorrélation donne une stabilité du modèle mais ne donne pas des résultats précis à cause de la troncature du signal original.

L'interpolation linéaire n'est pas plus performante que la LPC sur le plan de la minimisation de l'erreur quadratique moyenne, et par là même, sur la modélisation.

CHAPITRE IV

DETERMINATION DE LA FREQUENCE FONDAMENTALE

-----ooOoo-----

IV - 1 - INTRODUCTION

En analyse de la parole, la détection de F_0 est l'un des problèmes les plus complexes. Ayant fait l'objet d'un grand nombre d'études et de réalisations, la détection de F_0 consiste à :

- Déterminer la fonction de voisement.
- Estimer F_0 dans le cas où la séquence est voisée.

Les méthodes utilisées pour la détection de F_0 sont nombreuses et se classent suivant les techniques d'analyse utilisées.

Dans ce chapitre, nous nous sommes particulièrement intéressés aux techniques suivantes :

- La fonction d'autocorrélation.
- Le traitement parallèle.

IV - 2 - LE FILTRAGE PASSE-BAS

Puisque F_0 est située dans les basses fréquences et varie entre 50 et 300 Hz (selon les individus) le signal analysé est d'abord filtré à travers un filtre passe-bas afin de réduire partiellement l'effet des harmoniques et des formants sur les résultats de la détection.

La fréquence de coupure est généralement fixée à 800 ou 900 Hz.

Dans notre étude, nous avons opté pour un filtre numérique de Tchebyshev d'ordre 3. Sa fréquence de coupure varie entre 0,8 et 1 KHz.

Sa formulation temporelle est donnée par :

$$Y(n) = a_1 S(n) + a_2 Y(n-1) \quad (65)$$

$$X(n) = a_3 Y(n) + a_4 X(n-1) + a_5 X(n-2) \quad (66)$$

avec : $a_1 = 1 - e^{-\alpha_1 T}$

$$a_2 = e^{-\alpha_1 T}$$

$$a_3 = 1 - 2e^{-\alpha_2 T} \cos \beta_2 T + e^{-2\alpha_2 T}$$

$$a_4 = 2e^{-\alpha_2 T} \cos \beta_2 T$$

$$a_5 = -e^{-2\alpha_2 T}$$

$$\alpha_1 = (0,3572) 2\pi f_c$$

$$\alpha_2 = (0,1786) \pi f_c$$

$$\beta_2 = (0,8938) \pi f_c$$

$$T = \frac{1}{F_e}, \quad F_e : \text{fréquence d'échantillonnage.}$$

$$Y(n) = 0 \quad n < 0, \quad X(n) = 0 \quad n > 0 \quad -33-$$

$S(n)$ et $Y(n)$ sont respectivement l'entrée et la sortie du filtre.

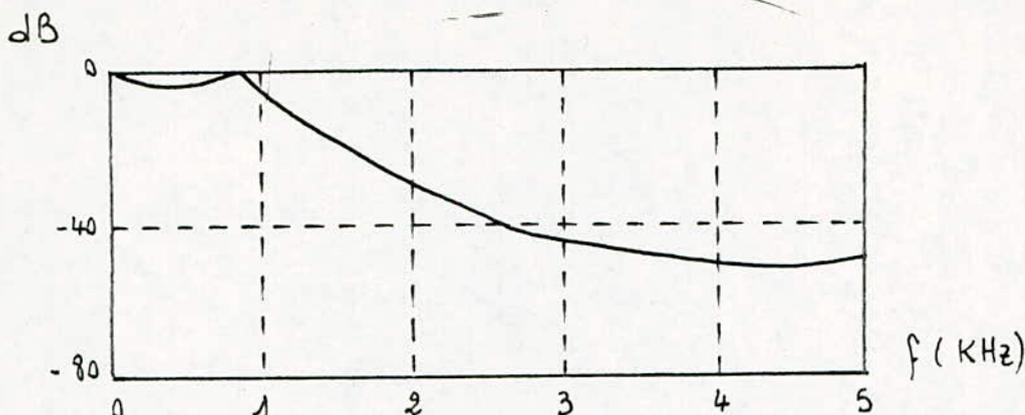


Figure 17 : Réponse fréquentielle du filtre de Tchebyshev d'ordre 3.

IV - 3 - DETECTION DE F_0 PAR LA FONCTION D'AUTOCORRELATION

Pour estimer la fréquence d'un signal périodique, il suffit de rechercher les positions des pics de sa fonction d'autocorrélation.

En appliquant cette méthode au signal de parole, des tests ont révélé son insuffisance pour constituer un bon détecteur de pitch. En effet :

- La non-stationnarité du signal de parole, sa connaissance sur une durée limitée impose le fenêtrage, ce qui a pour effet d'atténuer le pic de la fonction d'autocorrélation à court terme et de réduire leur nombre.
- La constitution formantique des sons voisés fait apparaître dans la fonction d'autocorrélation des pics additionnels dûs aux formants. Il est pratiquement impossible de les distinguer des pics relatifs au pitch. L'estimation de F_0 est ainsi erronée.
- Temps de calcul relativement long.

Pour remédier à ces inconvénients, les recherches ont abouti à une version modifiée de la méthode d'autocorrélation.

La mesure de la fonction d'autocorrélation à long terme d'un signal non stationnaire comme donné par l'équation (30) n'est pas significative. Ainsi il est préférable de définir une fonction d'autocorrélation à court terme qui opère sur des segments courts du signal, tel que :

$$Q(m) = \frac{1}{N} \sum_{n=0}^P X(n) \cdot X(n+m) \quad (67)$$

IV - 3 - 1 - AUTOCORRELATION MODIFIEE

Le signal est d'abord filtré à travers un passe-bas (900 Hz), ce qui élimine l'effet du second et des hauts formants, mais préserve

le premier à cause de sa proximité de F_0 . Une transformation non linéaire (NL) est ensuite appliquée au signal filtré : le temps de calcul est réduit et NL1 et NL2 peuvent être identiques. (fig. 18).

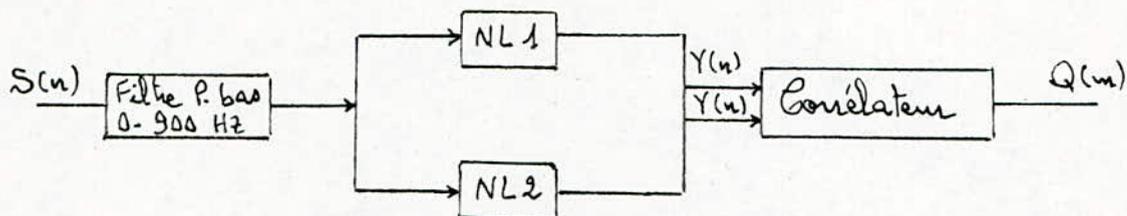


Figure 18 : Schéma synoptique du calcul de l'autocorrélation modifiée.

Dans notre étude, nous nous sommes intéressés à l'application de deux d'entre elles.

IV - 3 - 2 - AUTOCORRELATION MODIFIEE AVEC ECETAGE
ET COMPRESSION (Méthode de Sondhi)

Cette méthode a été introduite par M.M Sondhi en 1968. A chaque fenêtre d'analyse, le maximum absolu (CL) du signal filtré est calculé et toute portion comprise entre $\pm CL$ est éliminée (écrêtage par le centre). Les échantillons restants sont décalés vers le centre (compression). Ils sont supposés retenir la périodicité du signal dans le cas où elle existe.

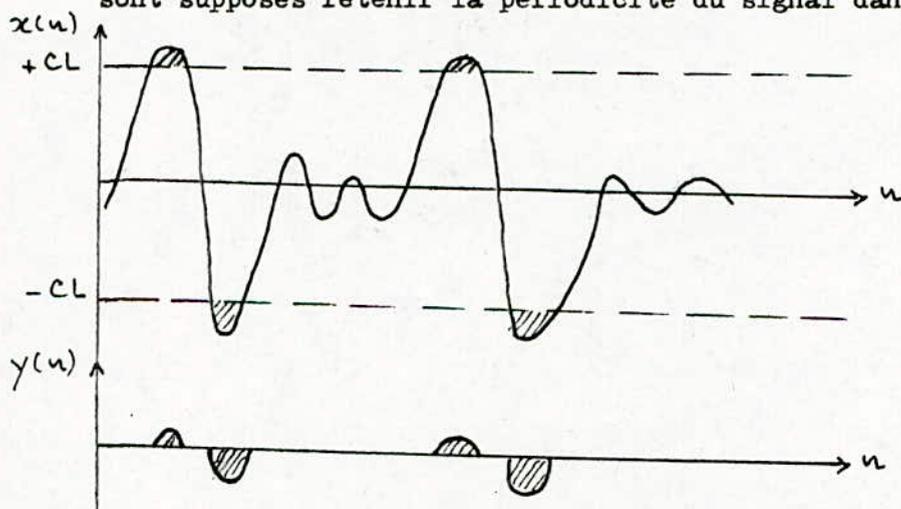


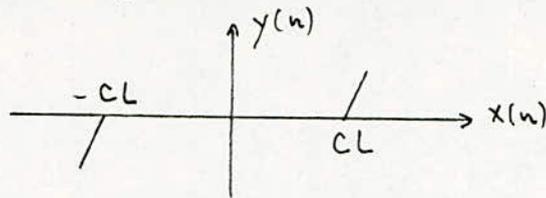
Figure 19 : Procédé d'écrtage et de compression.

Plus tard, ce procédé a été formulé par Rabiner sous le nom de non-linéarité CLC (compressed center clipper).

$$Y(n) = CLC X(n) = \begin{cases} X(n) - CL & X(n) \gg CL \\ 0 & |X(n)| < CL \\ X(n) + CL & X(n) \ll -CL \end{cases} \quad (68)$$

CL : est appelé seuil d'écrtage.

La caractéristiques entrée-sortie de cette non-linéarité est :



La structure formantique du signal est plus ou moins affectée selon l'ordre de grandeur du coefficient CL. Dans ses expériences, Sondhi a opté pour CL = 30 %.

L'autocorrélation du signal vocal ainsi modifiée (Y(n)) permet l'extraction de la fonction de voisement et l'estimation du fondamental.

$$Q(m) = \frac{1}{N} \sum_{n=0}^{N-1} Y(n) \cdot Y(n+m) \quad (69)$$

La détermination de la fonction de voisement peut être dans ce cas basée sur des critères énergétiques :

- Pour les sons voisés, une grande partie de l'énergie du signal est transportée par la fréquence fondamentale, elle est donc concentrée dans les basses fréquences.
- Pour les sons non voisés, l'énergie est plutôt concentrée vers les hautes fréquences.

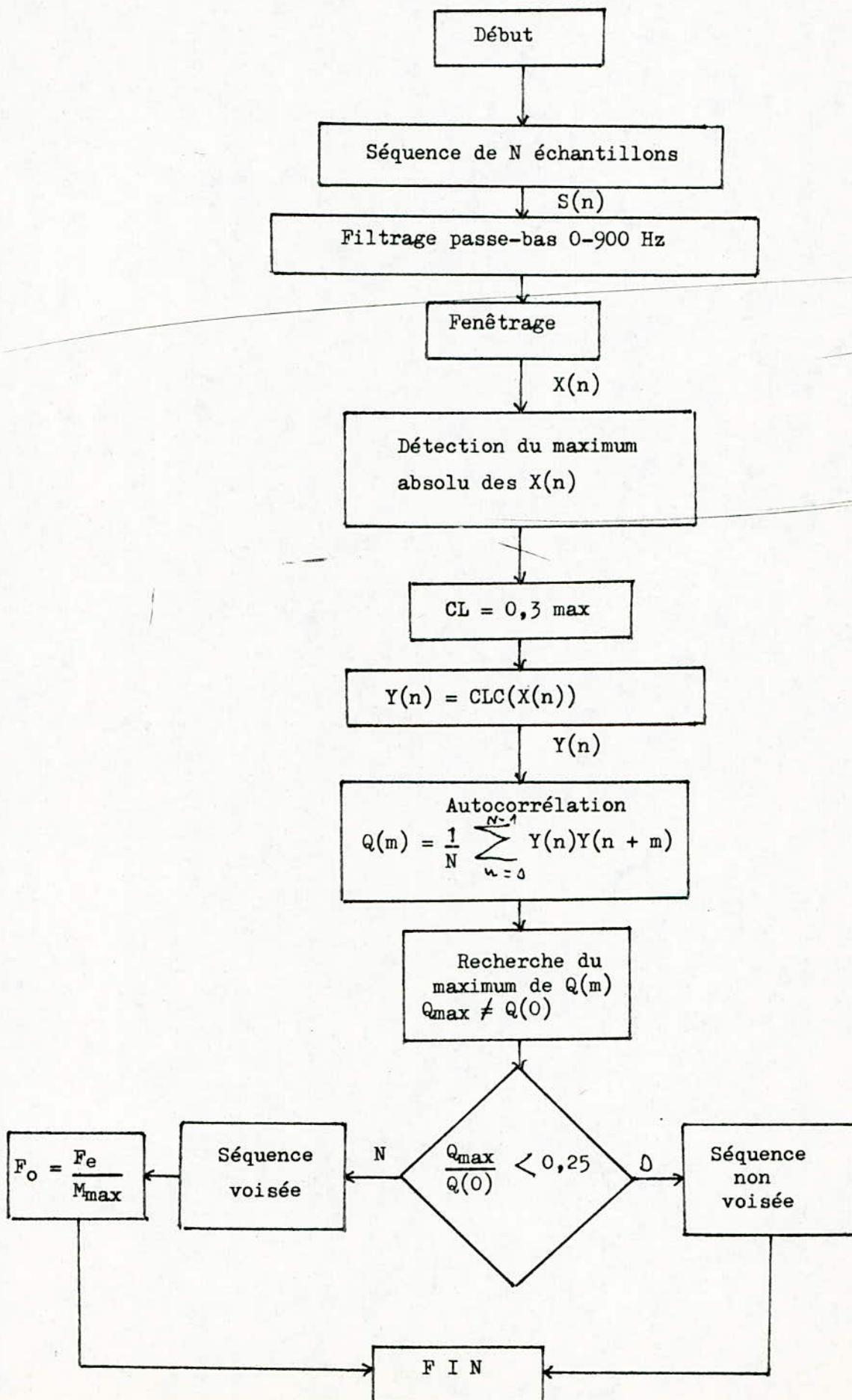
Dans cette méthode, nous comparons le rapport $Q_{max}/Q(0)$ à un seuil préalablement fixé, tel que Q_{max} différent de $Q(0)$, où $Q(0)$ est le plus grand des pics de la fonction d'autocorrélation (grande énergie).

Cependant, une méthode plus simple consiste à chercher le plus grand pic (Q_{max}) de l'autocorrélation, en dehors de $Q(0)$. Après plusieurs expériences de Sondhi, un seuil de 0,25 pourrait donner la meilleure estimation de F_0 :

- Si $\frac{Q_{max}}{Q(0)} < 0,25$: le signal est non voisé
- Si $\frac{Q_{max}}{Q(0)} \geq 0,25$: le signal est voisé, F_0 est donnée par l'abscisse de ce pic : M_{max}

alors : $F_0 = \frac{F_e}{M_{max}}$, F_e : fréquence d'échantillonge.

Organigramme : Détection de F_0 par la méthode de M.M. Sondhi.

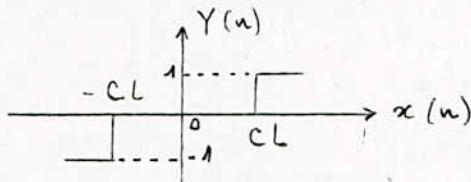


IV - 3 - 3 - METHODE D'AUTOCORRELATION MODIFIEE AVEC
DOUBLE ECRETAGE : (Méthode de DUBNOWSKI)

Cette méthode a été introduite par DUBNOWSKI en 1976. Elle consiste en l'application au signal filtré, d'une non linéarité définie par :

$$Y(n) = \text{Sgn } X(n) = \begin{cases} 1 & \text{si } X(n) \geq CL \\ 0 & \text{si } |X(n)| < CL \\ -1 & \text{si } |X(n)| \leq -CL \end{cases} \quad (70)$$

La caractéristique entrée-sortie de cette non linéarité est illustrée par :



Le signal est ainsi codé sur trois niveaux 1, 0 et - 1, ce qui s'implifie considérablement le calcul de sa fonction d'autocorrélation. Le seuil d'écrêtage "CL" est égal à 0,68 M. Où M représente le plus petit des maximums absolus MAX₁ et MAX₃ calculés respectivement sur la première et la dernière position de la fenêtre d'analyse.

Le spectre se trouve aplati, ce qui réduit fortement l'influence du premier formant sur les pics de la fonction d'autocorrélation Q(m). La décision de voisement est déterminée par la comparaison de Q(0) au plus grand pic Q_{max}(m).

$$\frac{Q_{\max}}{Q(0)} < 0,25 : \text{ la séquence est non voisée.}$$

$$\frac{Q_{\max}}{Q(0)} \geq 0,25 : \text{ la séquence est voisée.}$$

F₀ est calculée comme suit : $F_0 = \frac{F_e}{M_{\max}}$

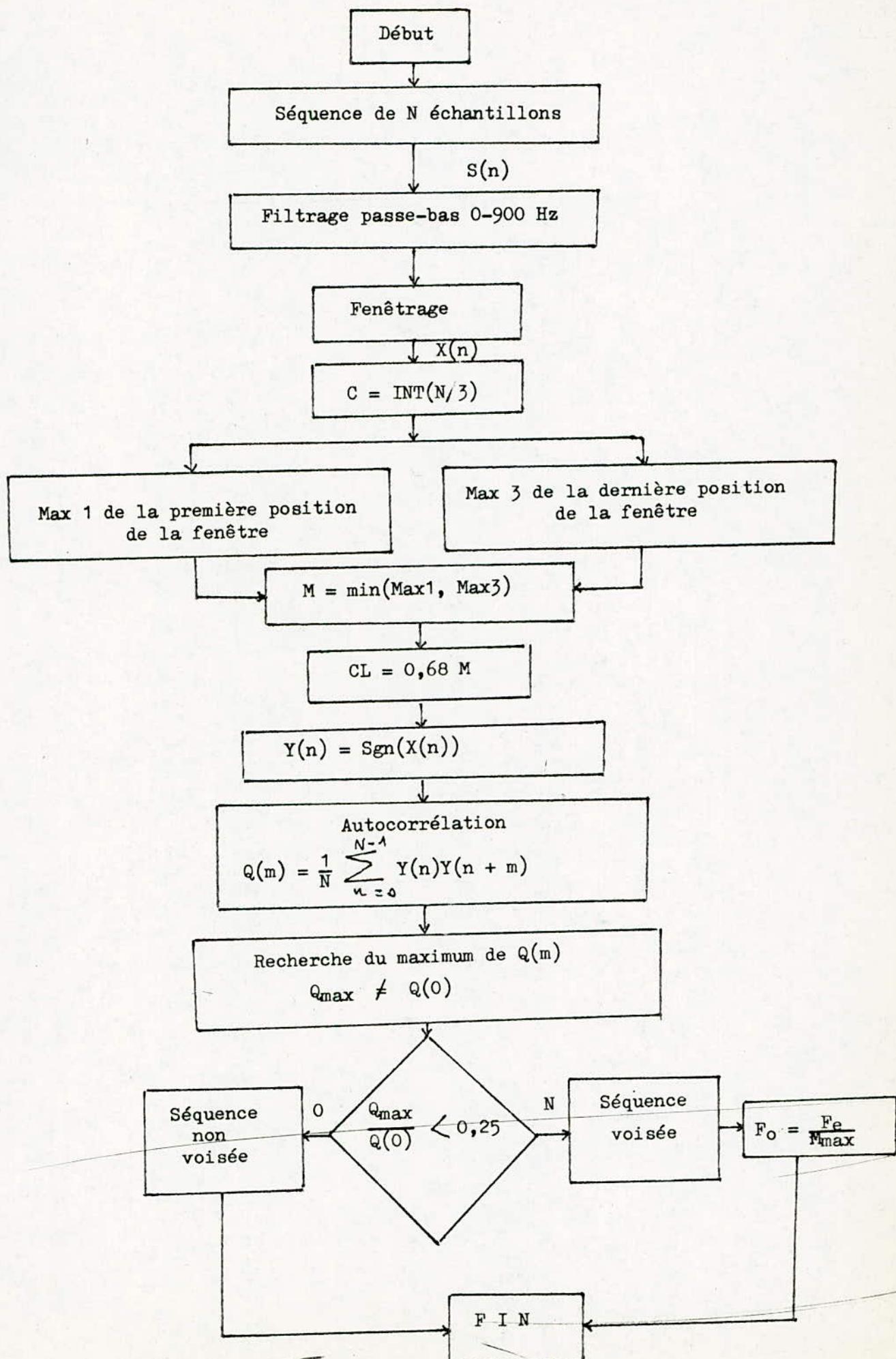
M_{max} : Abscisse de Q_{max}(m)

L'utilisation de la transformation "signe" a l'avantage de constituer un outil simple et efficace de détection du pitch en temps réel. En effet, l'autocorrélation étant réduite à :

$$Q(m) = \frac{1}{N} \sum_{n=0}^{N-1} \text{Sgn } X(n) \cdot \text{Sgn } X(n+m) \quad (71)$$

Dans ces méthodes, "m" sera pris égal au $\frac{2}{3}$ du nombre d'échantillons N.

Organigramme : Détection de F_0 par la méthode de Dubnowski



IV - 4 - TECHNIQUE DE TRAITEMENT PARALLELE POUR L'ESTIMATION DU PITCH

Cette méthode élaborée par Rabiner et Gold, 1969, se distingue par le fait que l'estimation de " F_0 " est effectuée sur le signal vocal dans le domaine temporel.

IV - 4 - 1 - PRINCIPE GENERAL DE LA METHODE

L'estimation de " F_0 " se fait en quatre étapes (fig. 20).

- a - Filtrage passe-bas du signal vocal : Le signal est filtré par un passe-bas (900 Hz), le même que celui décrit en(IV-2). Il ne subsistera que la région du premier formant à cause de sa proximité de " F_0 ".
- b - Génération de six trains d'impulsions m_i ($i = 1, 2, 3, \dots, 6$), caractérisant les variations du signal vocal (repérage des pics et vallées), comme suit :
 - m_1 : Amplitudes des pics
 - m_2 : Différence d'amplitude, pic-vallée précédente.
 - m_3 : Différence d'amplitude de deux pics successifs (si celle-ci est négative alors $m_3 = 0$).
 - m_4 : Amplitude des vallées.
 - m_5 : Différence d'amplitude, vallée-pic précédent.
 - m_6 : Différence d'amplitude de deux vallées successives (si celle-ci est négative alors $m_6 = 0$). (fig. 21)
- c - Les six trains d'impulsions " m_i " sont les entrées respectives de six détecteurs de pitch identiques (estimateurs de F_0) indépendants.
- d - Calcul final de " F_0 " (si elle est présente), en se basant sur la comparaison des résultats donnés par les estimateurs.

Toutes les amplitudes seront prises ensuite en valeur absolue, ainsi tous les trains d'impulsions générés par le bloc (2) seront à impulsions positives.

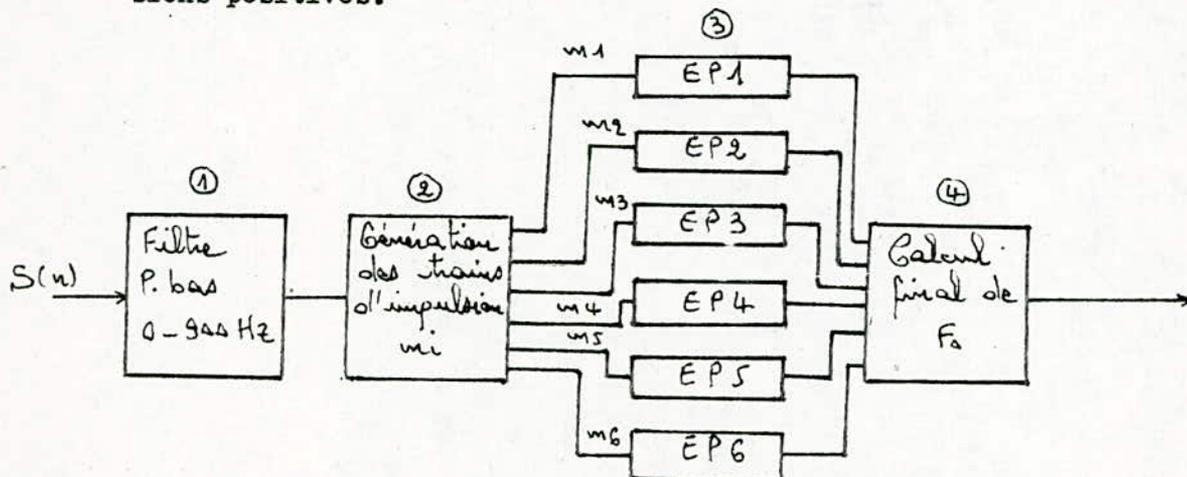


Figure 20 : Schéma synoptique du détecteur de pitch à traitement parallèle.

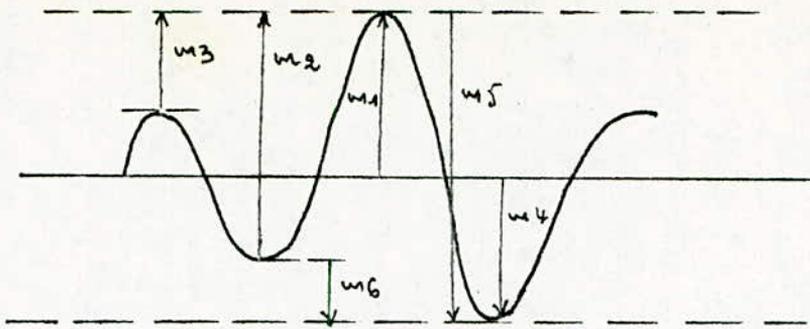
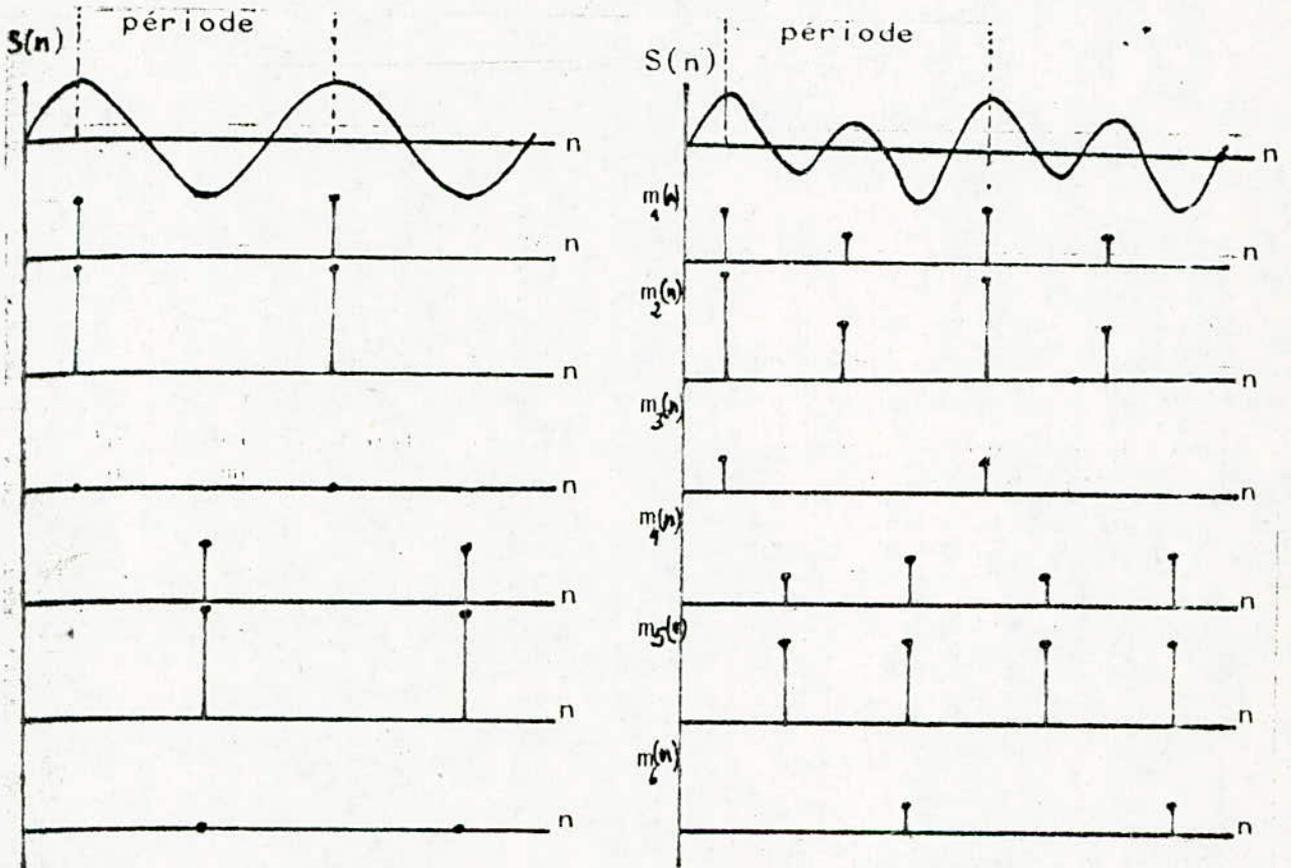


Figure 21 : Impulsions caractérisant les variations du signal.

Le choix de ce type de mesures est justifié en considérant les deux cas extrêmes de signaux qui peuvent se présenter (fig. 22).



a) présence du fondamental uniquement.

b) présence du fondamental et d'un second harmonique.

Figure 22 : Cas extrêmes de signaux

On constate que lorsqu'il y a présence du fondamental seul (fig. 22-a) les mesures m_3 et m_6 sont nulles. Le temps séparant deux impulsions dans (m_1, m_2, m_4, m_5) donne $T_0 = \frac{1}{2F_0}$.

Dans le cas b où il y a présence du fondamental et d'un second harmonique, m_3 et m_6 seront les plus susceptibles de donner une mesure exacte de F_0 .

Quoique la mesure de "F₀" sur les trains d'impulsions m₁, m₂, m₄ et m₅ est imprécise, il est démontré ultérieurement comment l'estimation finale de "F₀" a une grande probabilité d'être correcte. A noter que les impulsions m₁, m₂ et m₃ sont générées à chaque détection de pics positifs, tandis que les impulsions m₄, m₅ et m₆ sont générées à chaque détection de pics négatifs.

IV - 4 - 2 - FONCTIONNEMENT DES ESTIMATEURS DE PITCH

Les trains d'impulsions m_i (i = 1, ..., 6) sont appliqués respectivement aux six estimateurs. A ce niveau, un tri des impulsions est effectué, ne seront prises en compte que les impulsions dépassant un certain seuil S variable dans le temps :

$$S(t) = A e^{-\frac{t-t_0}{\tau}} \quad (72)$$

Après chaque détection d'impulsion d'amplitude "A", il s'écoule un temps t₀ pendant lequel aucune impulsion ne sera prise en compte.

t₀ sera de l'ordre de 2 à 3 ms.

S(t) décroît exponentiellement en fonction du temps avec une constante de temps prise fixe : τ = 5 ms (fig. 23).

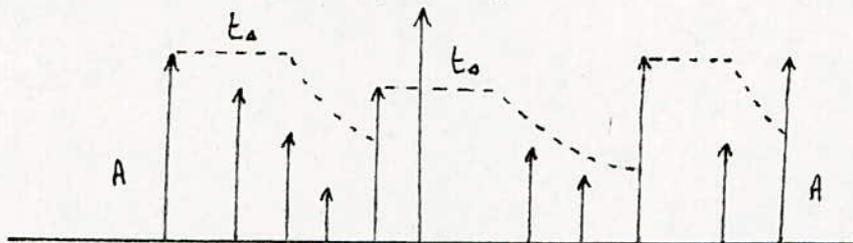


Figure 23 : Tri des impulsions.

Un maximum de quatre impulsions est recueilli puis généré par chaque estimateur (fig. 24)

IV - 4 - 3 - ESTIMATION DE F₀ ET DECISION

Une matrice [6 x 6] est formée, chaque colonne représente les différentes mesures de "F₀", effectuées sur le train d'impulsions généré par un détecteur. La détermination des éléments P_{ij} de cette matrice est illustrée par la figure ci-dessous :

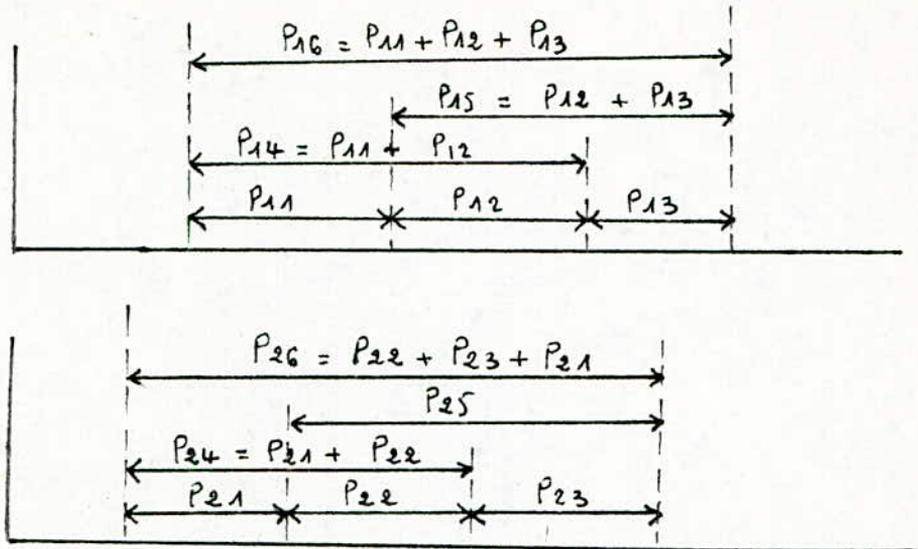


Figure 24 : Formation de la matrice des P_{ij}

Estimateur n°	1	2	3	4	5	6
	P_{11}	P_{61}

	P_{16}	P_{66}

Chaque élément de la matrice est susceptible d'être une mesure de " T_0 " (si elle existe).

- Il arrive parfois à cause d'une forte perturbation du formant proche de " F_0 " que les éléments P_{ij} ($i=1, \dots, 6$ et $j=1, 2, 3$) indiquent la valeur du second harmonique et dans ce cas la valeur de " F_0 " est plutôt approchée par les éléments P_{ij} ($i=1, \dots, 6$ et $j=4, 5, 6$).
- L'élément qui aura compté le plus grand nombre de "coincidences" se verra être l'estimation finale de " F_0 ".
- Une coïncidence est réalisée lorsque la différence entre l'élément "candidat" et l'élément à comparer est inférieure ou égale à un certain seuil.

$$\begin{cases} |P_{i1} - P_{kj}| \leq \beta \\ \beta = 30 \text{ Hz} \quad 1 \leq k \leq 6 \end{cases}$$

Si P_{i1} comptabilisera un nombre supérieur ou égal à 5 alors $P_{i1} = T_0$, dans le cas contraire, une décision de non voisement est prise.

IV - 5 - OBSERVATIONS

IV - 5 - 1 - LA METHODE D'AUTOCORRELATION

Cette méthode présente les inconvénients suivants :

- Temps de calcul important par rapport aux autres méthodes.
- L'influence des formants introduit des erreurs sur le calcul du maximum de la fonction d'autocorrélation sur une fenêtre.

Le seul avantage de cette méthode reste sans doute la simplicité de mise en oeuvre.

IV - 5 - 2 - LA METHODE D'AUTOCORRELATION MODIFIEE

Nous avons utilisé deux types de non linéarités :

- a - La non linéarité CLC, l'effet de celle-ci est illustré par la figure 25.

On constate :

- Une accentuation du pic de la fonction d'autocorrélation due au fondamental, cette accentuation dépend du seuil d'écrêtage.
- Un aplatissement du spectre, une homogénéisation en quelque sorte, les amplitudes des formants, annihilant de ce fait leurs effets sur la détermination de "F₀". D'après Rabiner, cet aplatissement du spectre est souhaitable pour une bonne appréciation de "F₀".
- Temps de calcul moins élevé que pour la méthode d'autocorrélation.
- Cette méthode exige un bon choix du seuil d'écrêtage, dans notre étude on l'a pris égal à 30 % (opté par Sondhi).

- b - La non linéarité Sgn, les résultats obtenus par cette méthode sont illustrés par la figure 26, celle-ci présente les mêmes avantages que la précédente avec un temps de calcul moindre.

IV - 5 - 3 - LA METHODE DE TRAITEMENT PARALLELE

Cette méthode est caractérisée par sa difficulté relative de mise en oeuvre mais présente un avantage certain quant au temps de calcul. Pour améliorer cet algorithme deux modifications peuvent être introduites :

- 1 - La première a pour but d'accélérer l'exécution et de limiter la valeur de "F₀" à 220 Hz (donc la voix masculine est la plus convenable pour l'analyse).
- 2 - La deuxième modification conduit à une réalisation "hardware" appréciable, plus simple que celle de l'algorithme original et donne une fréquence fondamentale de la voix au dessous de 300 Hz.

Cet algorithme ainsi que ces deux modifications ont été déjà trouvés précis et efficaces dans les tests sur la parole réelle et synthétique.

La vitesse de calcul de cet algorithme (et ces modifications) est plus rapide que celle des techniques utilisant la fonction d'autocorrélation.

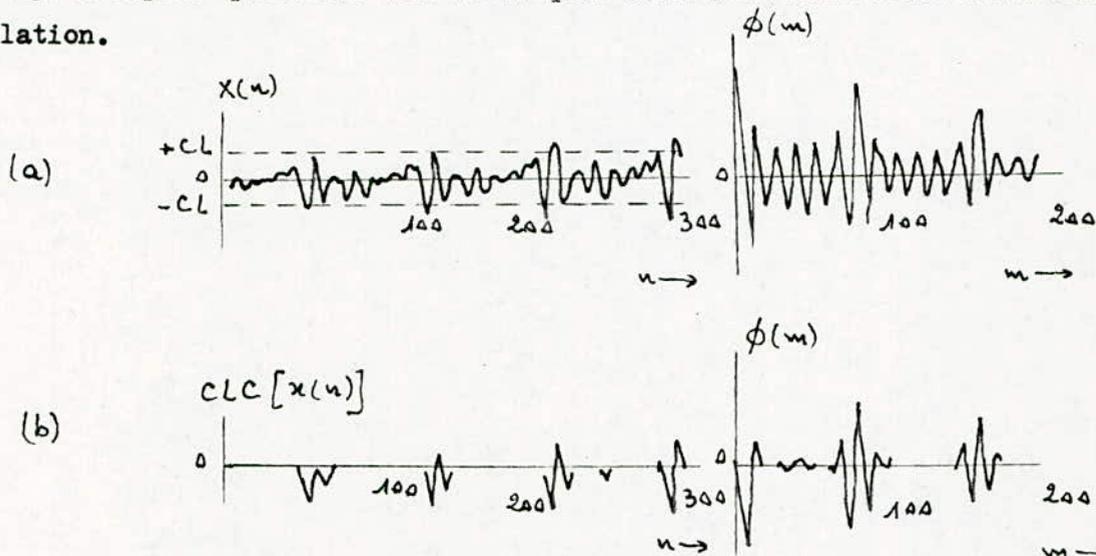


Figure 25 : Effet de la non linéarité CLC sur l'autocorrélation.

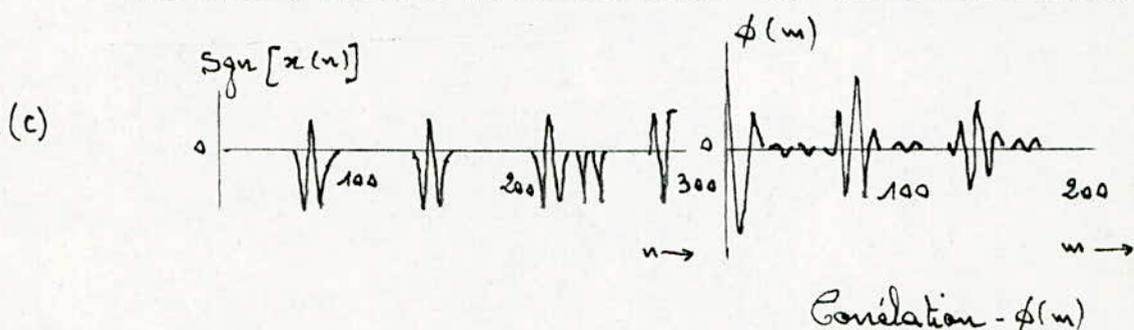


Figure 26 : Effet de la non linéarité Sgn sur l'autocorrélation.

IV - 6 - CONCLUSION

Nous avons constaté qu'une bonne estimation de " F_0 " exige :

- L'utilisation d'une fenêtre convenable.
- La durée de la fenêtre d'analyse contient au moins 2 à 3 périodes du fondamental, prendre par exemple pour un locuteur mâle, une fenêtre de 20 ms ($F_0 = 150$ Hz en moyenne).
- Dans le cas de l'estimation par la méthode d'autocorrélation, le nombre d'échantillons de la fonction d'autocorrélation doit être égal au moins à $\frac{2}{3} N$ (N est le nombre d'échantillons par fenêtre).

CHAPITRE V

ANALYSE DE L'ONDE GLOTTIQUE

PAR FILTRAGE INVERSE

—oooOooo—

V - 1 - INTRODUCTION

La connaissance du signal glottique U_g est très importante dans l'analyse et la synthèse etc, de la parole. Ce chapitre est consacré à la connaissance de la forme de l'onde glottique en vue de la modéliser et de l'analyser par filtrage inverse. Pour cela, plusieurs expériences ont été réalisées. Parmi les plus importantes, nous citerons les suivantes :

V - 1 - 1 - LA METHODE DU MASQUE DE ROTHENBERG (1972)

Cette technique est très sensible car elle nécessite la connaissance préalable du filtre inverse du conduit vocal (figure 27).

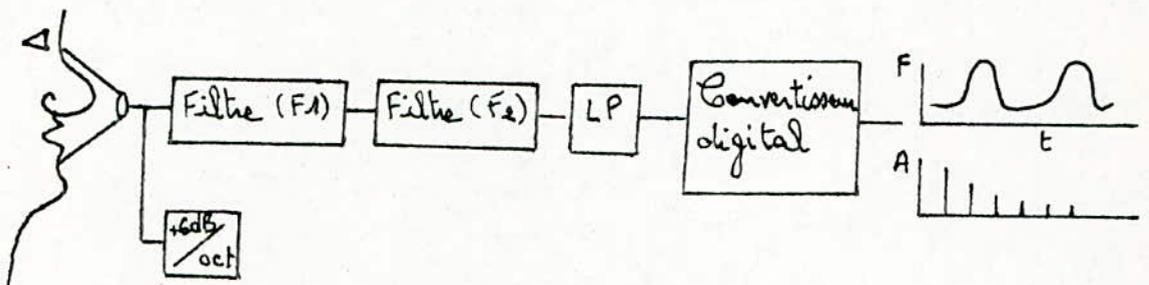


Figure 27 : Mesure de l'onde glottique par filtrage inverse.

La vitesse de l'air aux lèvres est enregistrée au moyen d'un masque spécial (pneumotograph).

La figure 28 représente l'onde glottique obtenue à l'aide de ce système. La pression subglottique P_g est calculée par extrapolation à partir de la pression intraorale (supraglottique).

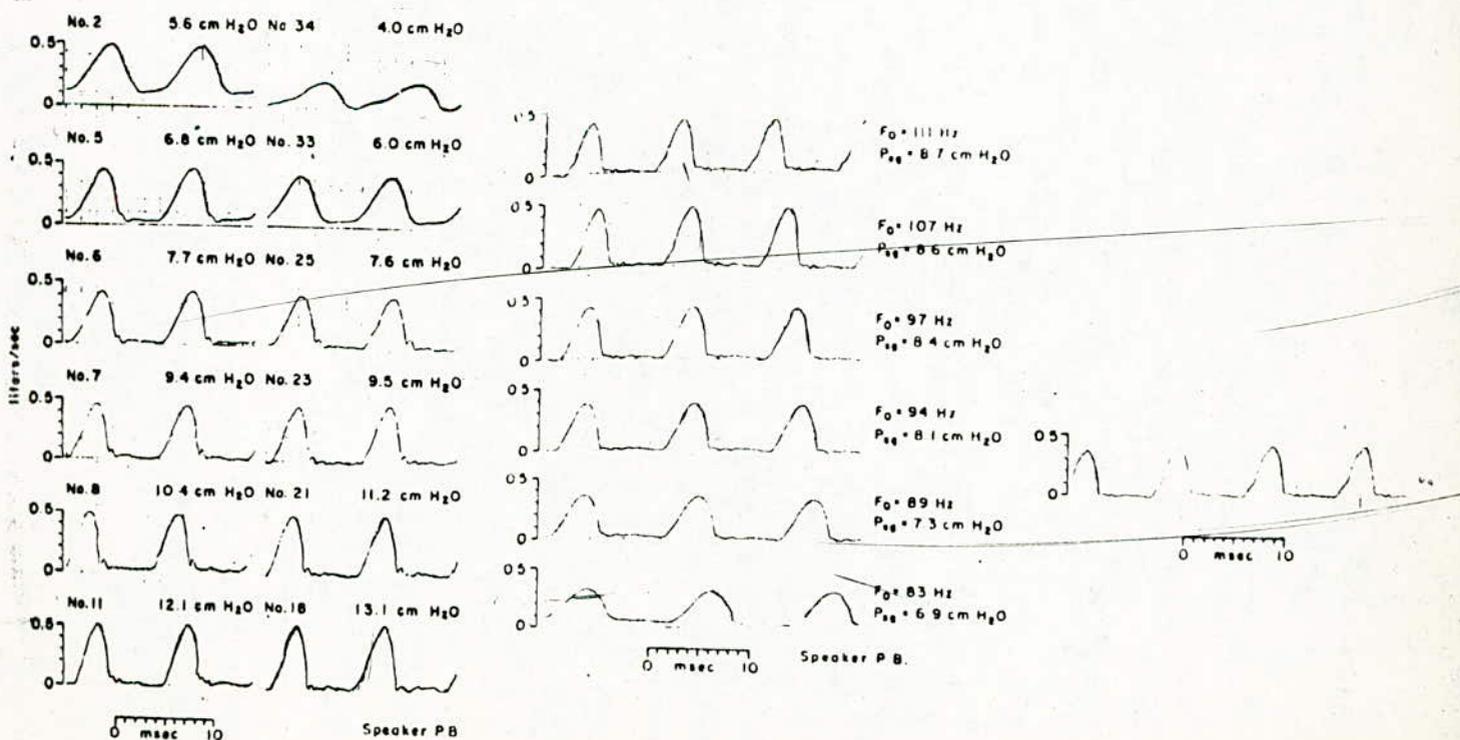


Figure 28 : Forme de l'onde glottique obtenue par la méthode du masque de Rothenberg.

On notera l'augmentation de " F_0 " et de l'intensité avec la pression subglottique ainsi que la période de fermeture de la glotte à laquelle correspond le conduit vocal en oscillation libre.

V - 1 - 2 - L'ADAPTATION D'IMPEDANCE (Méthode du tube sans réflexion)

Cette approche simple par sa conception a été proposée par Shondi en 1975 puis utilisée par Fisher en 1976. Son principe est le suivant :

Si le conduit vocal était parfaitement uniforme et sans réflexion, le signal d'excitation pourrait être récupéré aux lèvres sous sa forme originale. Cette condition peut être approchée par l'émission d'une voyelle neutre et l'utilisation d'un long tube (environ 2 m de long) bourré à son extrémité de mousse synthétique afin d'éviter toute réflexion, un capteur est placé dans ce tube afin d'enregistrer la forme de U_g . (figure 29).

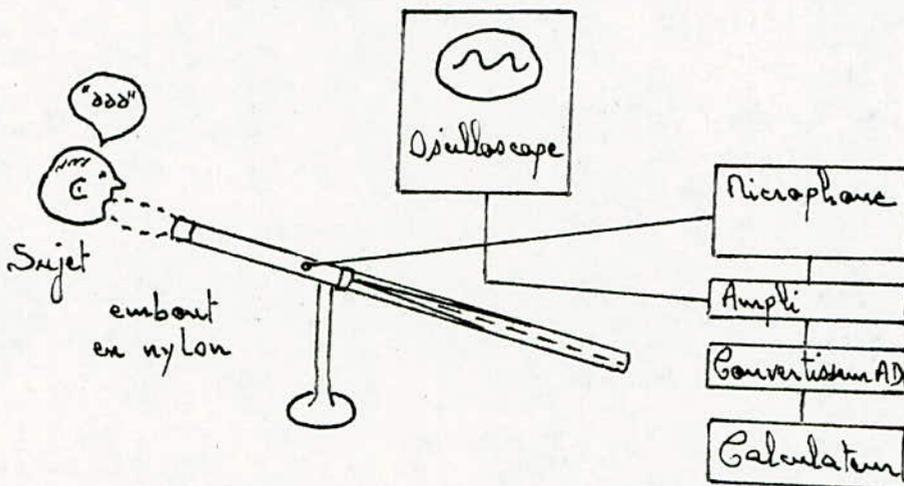
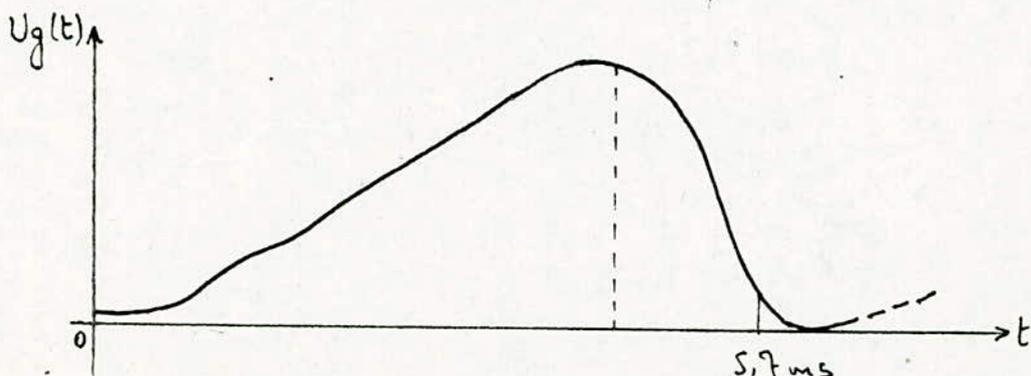


Figure 29 : Mesure de l'onde glottique par adaptation d'impédance.

La forme d'onde ainsi obtenue lors des premières expériences est la suivante :



Cette méthode ne peut être employée pour l'émission d'une voyelle quelconque, car alors le conduit vocal prendra une forme incompatible avec la structure du tuyau prolongateur (désadaptation d'impédance).

V - 1 - 3 - CARACTERISTIQUES GENERALES DE L'ONDE GLOTTIQUE

A partir des observations précédentes, nous pouvons tirer les conclusions suivantes :

- Le conduit vocal est excité à la fermeture de la glotte, le front descendant est plus rapide que le front montant.
- Pour les faibles pressions, la fermeture n'est pas toujours complète, il subsiste un débit résiduel.
- Quand la pression augmente, la valeur maximale du débit augmente mais le signal glottique devient plus étroit (F_0 augmente).
- Le signal glottique présente de petites ondulations attribuables à l'interaction acoustiques avec le conduit vocal (Guérin et Boe, 1979).

Sur un plan qualitatif, les différentes méthodes d'extraction permettent de considérer que l'onde glottique possède la forme suivante (fig. 30) :

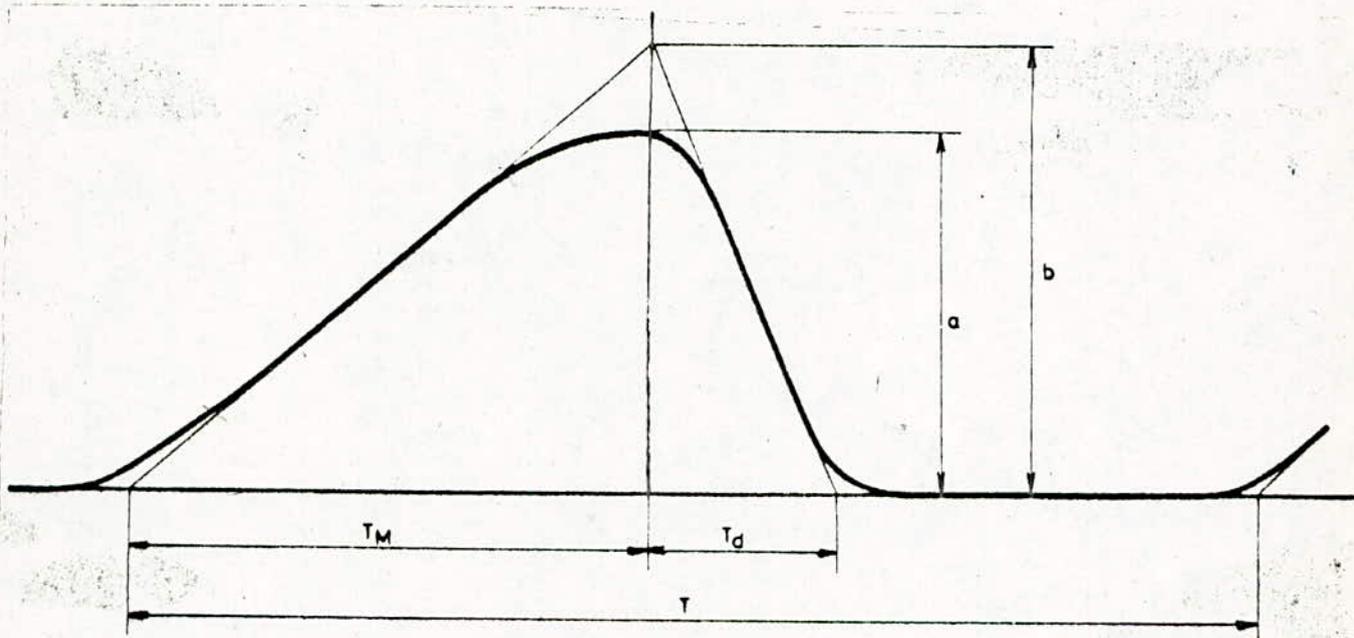


Figure 30 : Caractéristique "estimée" de l'onde glottique.

On définit :

- Le temps de montée et de descente T_M et T_D
- La période du signal T
- Le taux d'ouverture $T.O = \frac{T_M + T_D}{T}$
- Le facteur de crête $\frac{a}{b}$

Les valeurs relevées pour ces divers paramètres varient dans les limites suivantes (notons les différences notables dues à l'emploi de la méthode) :

$$0,64 < T_0 < 0,89 \quad (\text{Zemlin, 1959, FANT, 1960})$$

$$0,4 < T_0 < 0,83 \quad (\text{Lindqvist, 1965})$$

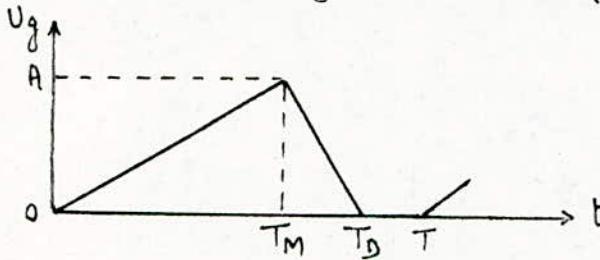
$$0,37 < T_0 < 1 \quad (\text{Sonesson, 1960})$$

$$0,45 < T_0 < 0,75$$

$$0,7 < \frac{a}{b} < 0,9 \quad \left. \vphantom{0,7 < \frac{a}{b} < 0,9} \right\} (\text{Lindqvist, 1970})$$

$$\frac{T_M}{T_D} > 2$$

Les approximations qui ont été utilisées comme signal glottique avaient la forme triangulaire suivante (figure 31) :



A est l'amplitude du signal
a est la pente de l'ouverture
b est la pente de la fermeture

Figure 31 : Forme triangulaire du signal glottal.

V - 2 - REPRESENTATION PHYSIOLOGIQUE ET MODELE

V - 2 - 1 - MODELISATION LINEAIRE DU SYSTEME DE PRODUCTION

Les études de FANT en 1960 et FLANAGAN en 1965 ont montré que, dans le cas des sons sonores sans nasalité, la fonction de transfert du conduit vocal peut être représentée par un modèle ne possédant que des pôles. En 1971, ATAL a montré que les sons voisés sont produits par une excitation glottique dans la transformée en Z peut être approchée par :

$$G(Z) = \frac{1}{(1 - e^{-cT} z^{-1})^2} \quad (73)$$

où e^{-cT} est un pôle réel de $G(Z)$ situé dans le cercle unité.

Le conduit vocal possède une fonction de transfert de la forme

$$V(Z) = \frac{1}{\prod_{i=1}^k [1 - 2e^{-c_i T} \cos(b_i T) z^{-1} + e^{-2c_i T} z^{-2}]} \quad (74)$$

T est la période d'échantillonnage, k représente le nombre de résonances définies par le modèle.

Le rayonnement aux lèvres peut alors être assimilé à celui d'une source sphérique. La pression "P" à une distance r de la bouche est donnée par :

$$p(r) = \frac{j\omega_0 \rho V_0}{4\pi r} e^{-j\frac{\omega r}{v}} \quad (75)$$

Où ρ est la densité de l'air, ω la pulsation, v la vitesse de propagation des ondes et v_0 la vitesse de l'air au niveau des lèvres. Le rapport pression au capteur/vitesse de l'air aux lèvres est alors proportionnel à un terme de la forme "1-Z⁻¹" que l'on appellera R(Z), qui est un terme relatif au rayonnement. La combinaison de ces trois facteurs G(Z), V(Z) et R(Z) permet de constituer la fonction de transfert H(Z) possédant la forme suivante :

$$H(Z) = G(Z) \cdot V(Z) \cdot R(Z) = \frac{1 - z^{-1}}{(1 - e^{-cT} z^{-1})^2 \left[\prod_{i=1}^K (1 - 2e^{-c_i T} \cos(b_i T) z^{-1} + e^{-2c_i T} z^{-2}) \right]} \quad (76)$$

La valeur de cT étant générale très petite devant "1" et le numérateur disparaît alors en même temps que (1 - e^{-cT} z⁻¹)

Dans le cas des sons voisés, on obtient alors :

$$H(Z) = \frac{S(Z)}{E(Z)} = \frac{1}{(1 - e^{-cT} z^{-1}) \left[\prod_{i=1}^K (\dots) \right]} = \frac{1}{A(Z)} \quad (77)$$

et A(Z) peut se mettre sous la forme :

$$A(Z) = \sum_{k=0}^M a_k \cdot Z^{-k} \quad \text{avec } a_0 = 1$$

et E(Z) = A(Z) \cdot S(Z) est l'équation du filtrage inverse.

V - 2 - 2 - SCHEMA PHYSIOLOGIQUE ET MODELE

La représentation classique du signal phonatoire est la suivante :

(figure 32)

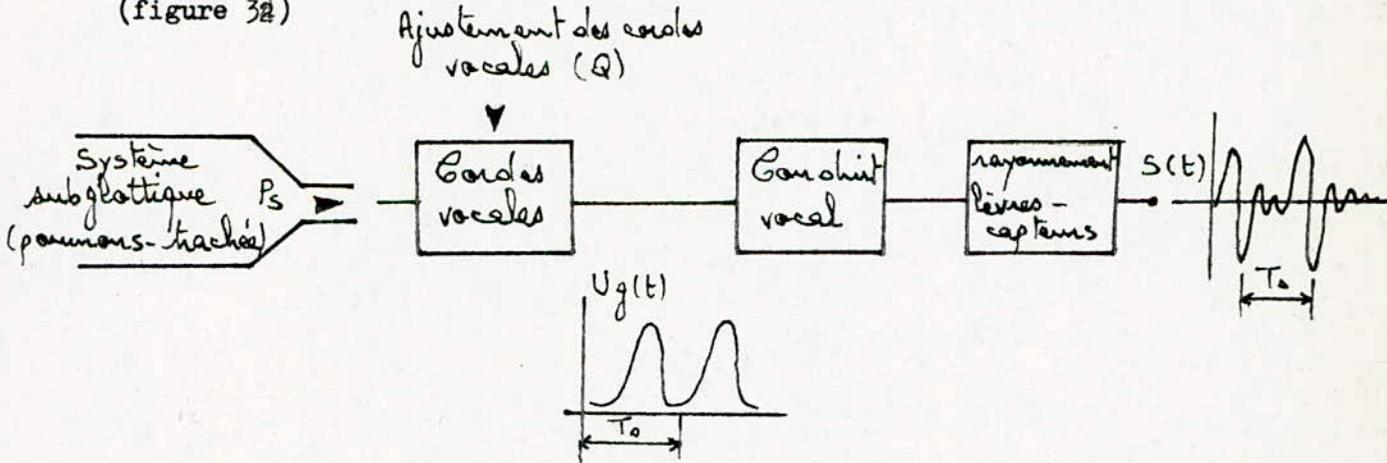


Figure 32 : Représentation physiologique du système phonatoire.

La représentation linéaire du système phonatoire est une modélisation globale du système car le filtre "1/A(Z)" prend en compte, non seulement le conduit vocal, mais également la source et le rayonnement : il y a équivalence entre la branche supérieure et inférieure de la figure 33.

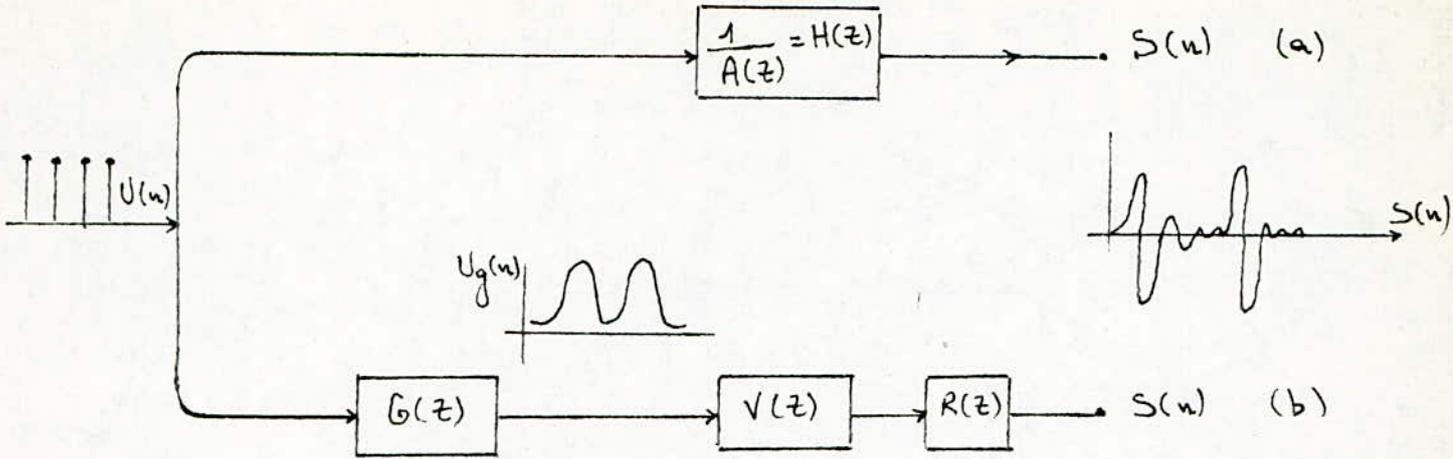
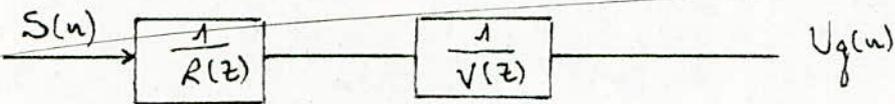


Figure 33 : Modèle linéaire de production.

Ce modèle n'est validé que si les sons produits sont voisés, les fréquences sont inférieures à 5 KHz. A partir du schéma précédent, on peut déduire le filtrage inverse correspondant.



Le filtre $V(Z)$ peut être représenté par l'expression suivante :

$$V(Z) = \frac{1}{\sum_{i=0}^K -v_i Z^{-i}} \quad (78)$$

K est un nombre entier pair, il correspond au nombre de " $\frac{k}{2}$ " formants du modèle. Connaissant l'expression de ces deux filtres, l'expression de U_g est la suivante :

$$U_g(Z) = \frac{S(Z)}{V(Z) \cdot R(Z)} \quad (79)$$

Le problème consiste donc à calculer le filtre $V(Z)$, comment éliminer l'influence de la source et du rayonnement sur la fonction de transfert du filtre $1/A(Z)$?

V - 2 - 3 - ANALYSE

Supposons que l'onde glottique $U_g(t)$ possède la forme indiquée dans la figure 34.

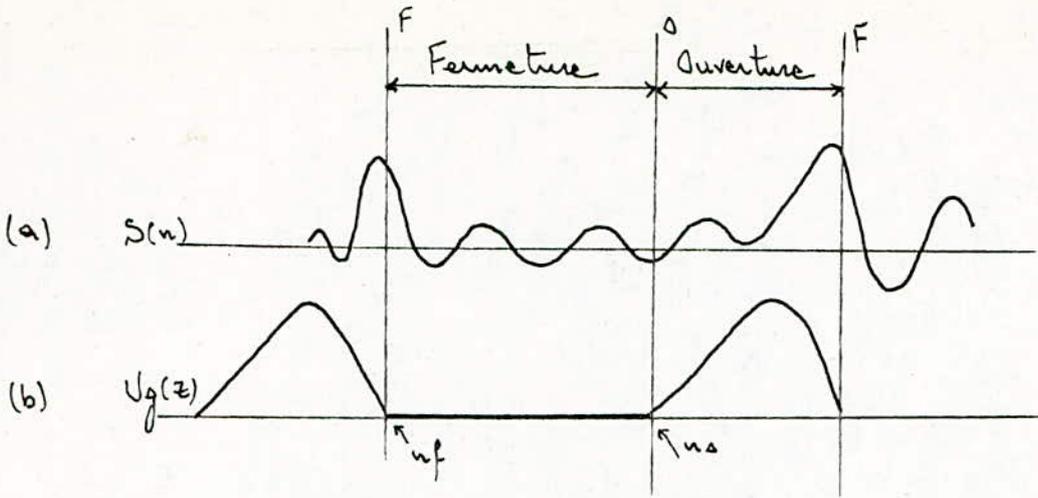


Figure 34 : Forme de $S(t)$ et de $U_g(t)$.

Soit n_f l'échantillon correspondant à l'instant de fermeture de la glotte et n_o l'échantillon correspondant à l'ouverture :

$$U_g(n) = 0 \quad \text{pour } n_f < n < n_o$$

Pendant la fermeture de la glotte, la fonction de transfert $G(Z)$ sera éliminée ainsi que $R(Z)$ car elle possède un zéro à la fréquence $f = 0$. Dans ce cas, le modèle de la figure 33 devient :

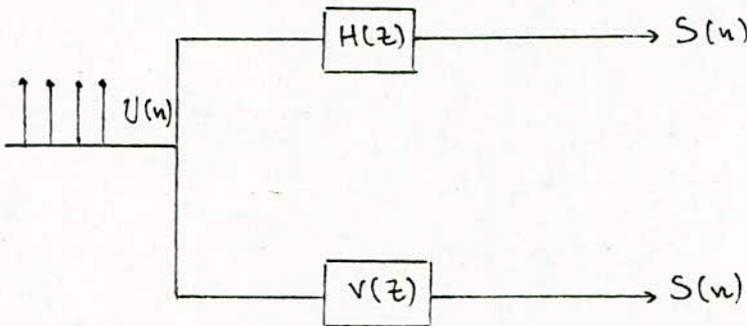


Figure 35 : Modèle linéaire équivalent.

Calculer le filtre $V(Z)$ revient à trouver le filtre $A(Z)$. Le calcul de ce dernier va être effectué par deux méthodes :

a - Méthode de la prédiction linéaire

Sachant que la fonction de transfert du conduit vocal peut être simulée à celle d'un filtre numérique ne possédant que des pôles, le modèle autorégressif paraît donner une meilleure modélisation du conduit vocal (fig. 36).

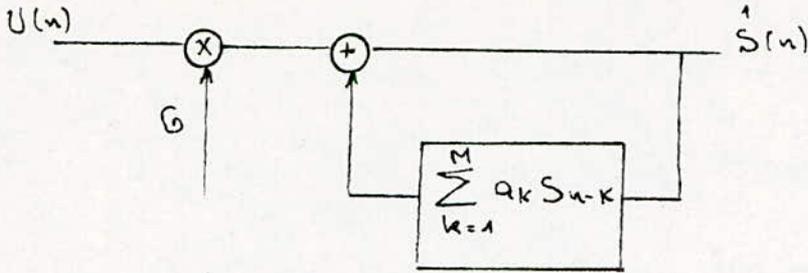


Figure 36 : Modèle autorégressif dans le cas de la LPC.

Ce modèle permet le calcul d'un échantillon à partir d'une entrée et d'une combinaison linéaire des échantillons de sortie précédents.

$$\hat{s}(n) = - \sum_{k=1}^M a_k \cdot s(n-k) + G \cdot u(n) \quad (80)$$

G est un facteur de gain, le signe "-" a été choisi arbitrairement.

Les propriétés générales de ce modèle sont d'une part, une mise en oeuvre relativement simple et d'autre part, une bonne approximation du filtre récursif (conduit vocal).

La transformée en Z de l'expression précédente est :

$$\hat{S}(Z) = - \sum_{k=1}^M a_k \cdot S(Z) \cdot Z^{-k} + G \cdot U(Z) \quad (81)$$

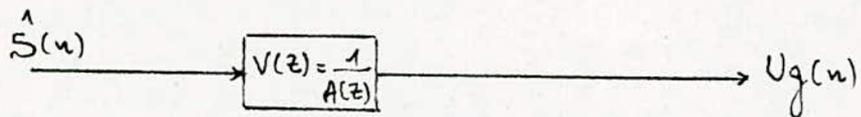
d'où :

$$\frac{\hat{S}(Z)}{U(Z)} = \frac{G}{1 + \sum_{k=1}^M a_k \cdot Z^{-k}} = \frac{G}{\sum_{k=0}^M a_k \cdot Z^{-k}} \quad \text{avec } a_0 = 1 \quad (82)$$

Au terme G près, ceci est l'expression du terme H(Z). on définit de la même façon le filtre inverse :

$$A(Z) = \sum_{k=0}^M a_k \cdot Z^{-k}$$

D'après ce qui précède, la détermination de U_g se résume comme suit :



- Optimisation des coefficients a_k : l'erreur quadratique moyenne sera minimale dans l'intervalle $[n_f, n_0]$ (cf. ch III § 3-3).

b - Méthode d'interpolation linéaire

Le modèle autorégressif est illustré par la figure 37.

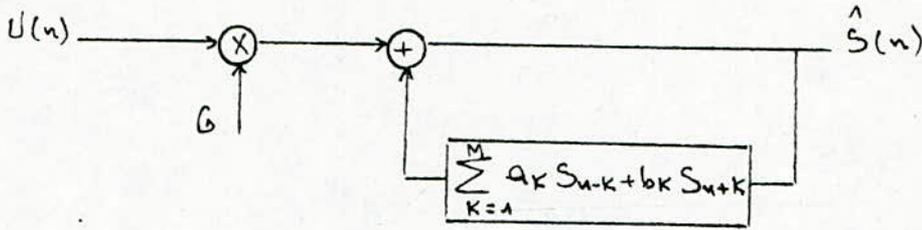


Figure 37 : Modèle autorégressif sans le cas de l'interpolation

Le signal prédit sera :

$$\hat{S}(n) = - \sum_{k=1}^L (a_k S_{n-k} + b_k S_{n+k}) + G U(n) \quad (83)$$

La transformée en Z de s(n) est :

$$\hat{S}(Z) = - \sum_{k=1}^L (a_k z^{-k} + b_k) S(z) + G U(z) \quad (84)$$

d'où :

$$\frac{\hat{S}(Z)}{U(Z)} = \frac{G}{1 + \sum_{k=1}^L (a_k z^{-k} + b_k)} = \frac{G}{A_L(z)} \quad (85)$$

avec

$$A_L(z) = 1 + a_1 z^{-1} + \dots + a_L z^{-L} + b_1 z^1 + \dots + b_L z^L$$

En multipliant par Z^{-L} et en positionnant les termes, on obtient :

$$Z^{-L} \cdot A_L(z) = b_L + b_{L-1} z^{-1} + \dots + b_1 z^{-L+1} + a_0 z^{-L} + a_1 z^{-L+1} + \dots + a_L$$

avec : $a_0 = 1$

La fonction de transfert est alors :

$$\frac{\hat{S}(Z)}{U(Z)} = \frac{G z^{-L}}{b_L + b_{L-1} z^{-1} + \dots + a_0 z^{-L} + \dots + a_L z^{-2L}} = \frac{G z^{-L}}{\sum_{k=0}^{2L} C_k z^{-k}} \quad (86)$$

avec

$$k = 0, \dots, L-1$$

$$C_k = b_{L-k}$$

$$k = L, \dots, 2L$$

$$C_k = a_{k-L} \text{ avec } a_0 = 1$$

En comparant cette dernière avec la fonction de transfert calculée dans le cas de la prédiction linéaire, on voit que l'ordre du filtre inverse est le même dans les deux cas. L'apparition du terme " Z^{-L} " n'a pas d'incidence sur le module du spectre car $|Z^{-L}| = 1$. Ce terme exprime le retard entre l'échantillon prédit \hat{S}_n et le dernier échantillon de la trame temporelle considérée S_{n+L} .

L'erreur quadratique moyenne sera minimisée dans l'intervalle $[n_f, n_o]$ (cf. ch III § 4).

Le problème consiste maintenant à déterminer l'intervalle de la fermeture de la glotte.

Ainsi, la recherche manuelle qui consiste à déterminer la séquence du signal de parole possédant un maximum d'oscillations libres amorties, ce qui n'est pas très précis, se double maintenant de la détermination automatique à l'aide d'un algorithme de calcul, des positions éventuelles de la fenêtre d'analyse pour lesquelles l'erreur quadratique moyenne est minimale. Cette méthode permet de déterminer avec précision les échantillons n_f et n_o marquant respectivement la fermeture et l'ouverture de la glotte.

Afin de s'assurer que les résultats de l'analyse ne sont pas fonction d'un gain quelconque, l'erreur quadratique moyenne sera normalisée, soit $E_N = \frac{E}{E_0}$, E_0 représente l'énergie du signal.

Par conséquent, les coefficients du filtre $A(Z)$ correspondant aux valeurs minimums de E_N représente la fonction transfert du conduit vocal $V(Z)$ sans l'influence de la glotte ni des lèvres.

- Procédure : Il faut préciser que le résultat d'un filtrage inverse, quel que soit la méthode employée (cf. ch V § 2-3 a et b), dépend beaucoup d'un certain nombre de facteurs liés à la qualité de la parole, tels que :

- Le bruit ambiant
- Les basses fréquences dues aux techniques d'enregistrement
- La distorsion due à l'enregistrement et à la bande.

L'enregistrement d'un corpus se fera de préférence en chambre sourde, à l'aide d'un magnétophone à bande. La solution qui consiste à numériser directement la parole sans passer par l'intermédiaire d'un support magnétique est cependant souhaitable afin d'éliminer les phénomènes de distorsion, d'enregistrement et de lecture.

Pour éliminer en partie les deux premières causes, il est effectué un filtrage passe-haut ($F_c = 60 \text{ Hz}$).

Le filtre passe-bas ($F_c = 3,6 \text{ KHz}$) permet de respecter les hypothèses du modèle linéaire tout en évitant un éventuel recouvrement du spectre dû à l'échantillonnage à 8 KHz .

La suite de l'analyse s'effectue alors sous forme numérique (fig. 38).

La nécessité de se placer dans l'intervalle de fermeture de la glotte afin d'obtenir les caractéristiques du conduit vocal seul, demande l'emploi de l'analyse par la méthode de covariance. Celle-ci est appliquée sur une fenêtre de M points à chaque pas de calcul, la fenêtre est décalée d'un échantillon.

La valeur minimum de E_N correspond à une suite de coefficients de prédiction, a_k , $k = 1, M$, permettant de définir le filtre inverse $A(z)$ qui est implanté sous la forme d'un filtre en treillis (cf. ch. II § 6-3).

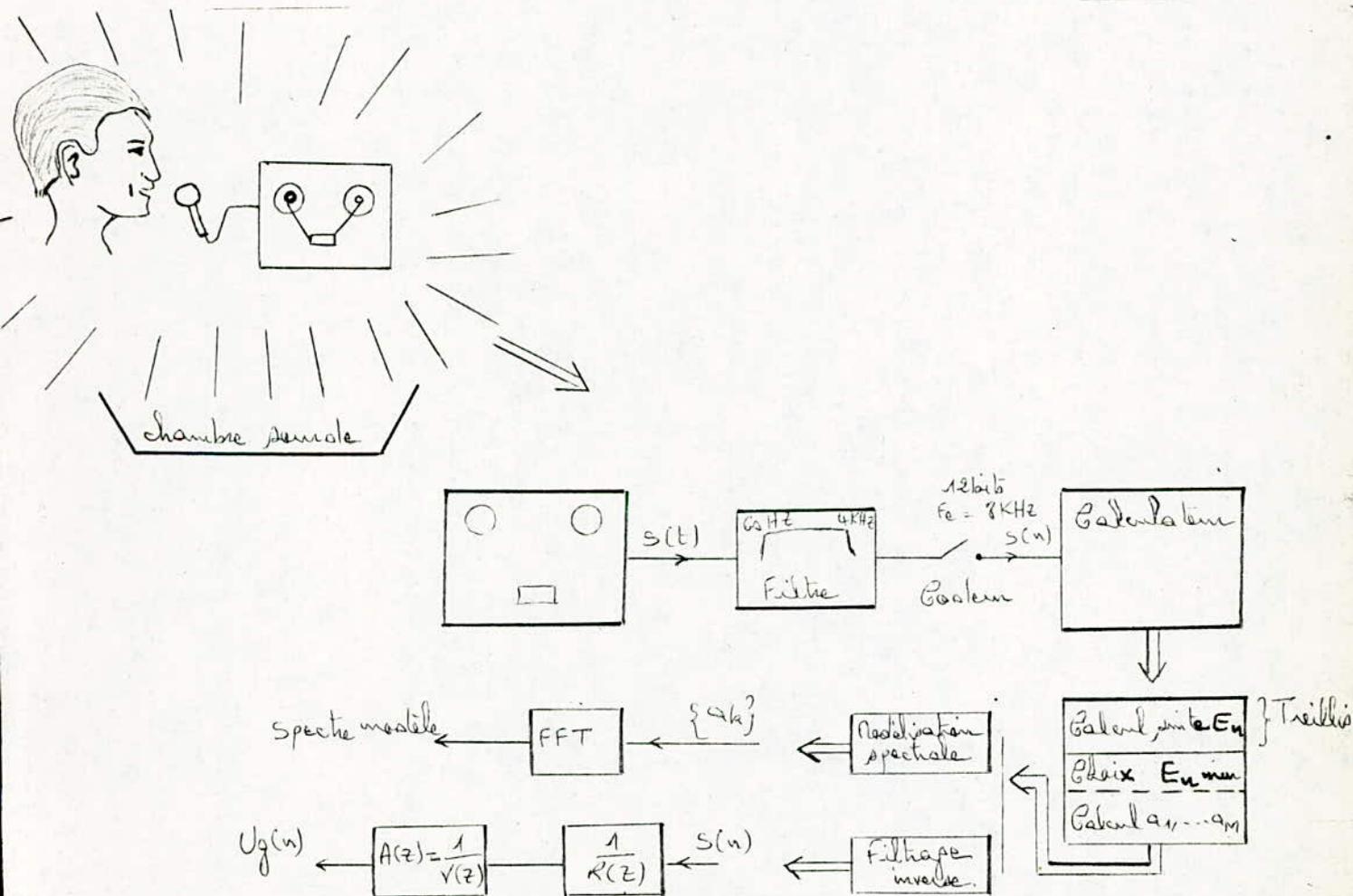


Figure 38 : Technique d'analyse par prédiction linéaire du signal de parole.

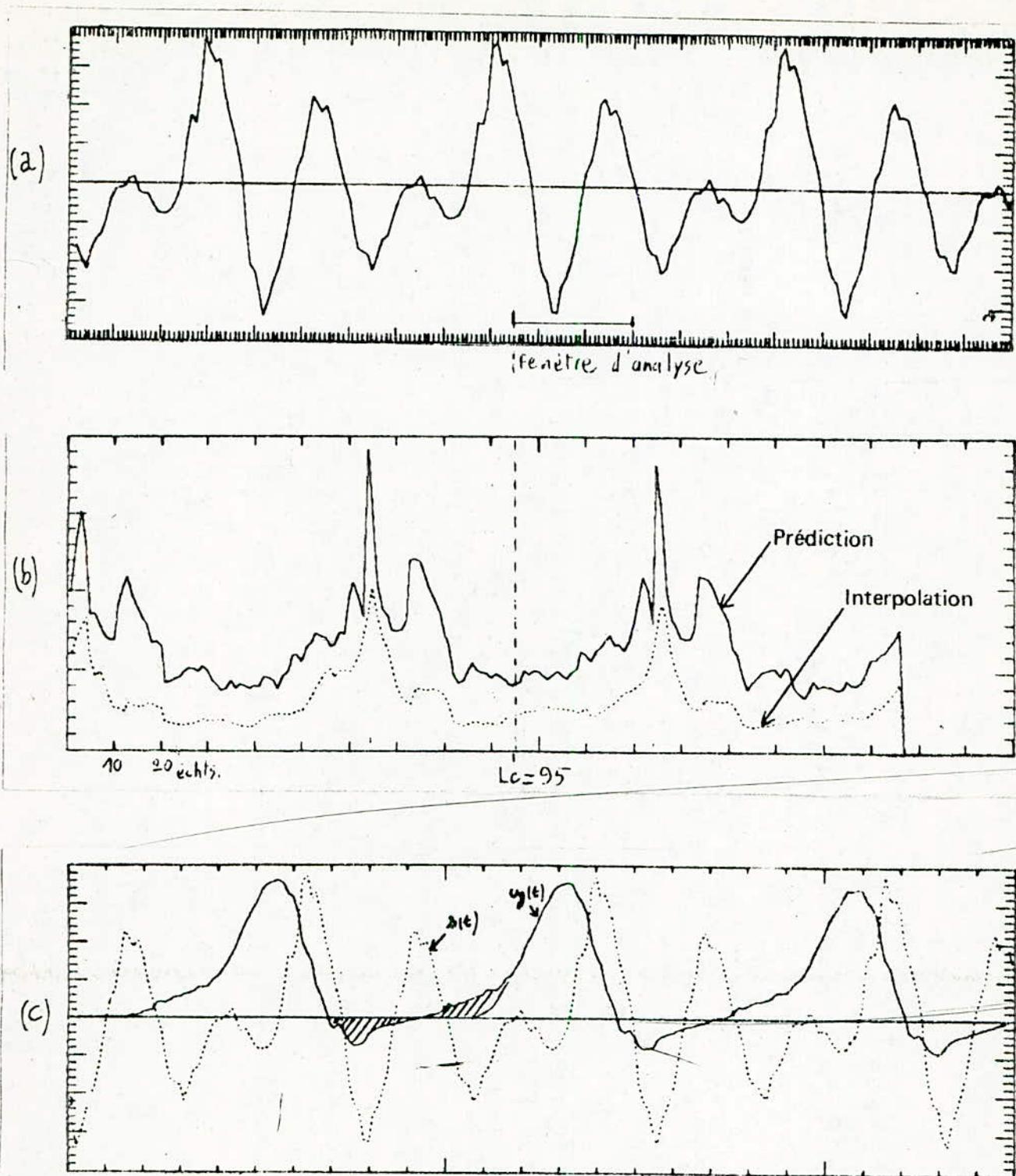


Figure 39 : a - Signal acoustique/e/
b - Erreur quadratique moyenne E_N
c - Onde glottique obtenue par filtrage inverse

V - 3 - APPLICATION A UNE VOYELLE

V - 3 - 1 - PRETRAITEMENT " MANUEL "

La voyelle étudiée est un /e/ (fig. 39a)

Le spectre de puissance de ce signal est donné par la figure-40
Les fréquences de résonance des trois premiers formants se situent à 380, 2120 et 2780 Hz.

Une première partie du traitement consiste à imposer les valeurs des paramètres N et M .

N: Nombre des chantillons inclus dans la fenêtre d'analyse.

M: Nombre de coefficients de prédiction du filtre A (Z).

N est fixé à 25 à partir de l'observation des oscillations du signal de parole, et M = 8.

La figure-39b représente l'erreur quadratique moyenne en fonction de la position de la fenêtre.

Cas de la prédiction linéaire: L'échantillon correspondant à l'erreur quadratique moyenne est $L_c = 95$, elle est calculée sur la fenêtre $[L_c, L_c + N - 1]$ appartenant à l'intervalle de fermeture de la glotte.

La figure-39c représente l'allure de $U_g(t)$ déduite par filtrage inverse.

Deux causes peuvent être à l'origine de la déformation de U_g par rapport à sa forme " idéale " qui est schématiquement représentée en enlevant la partie hachurée.

- Si le signal contient des zéro, ce modèle ne les prend pas en compte à moins d'augmenter le nombre de coefficients prédicteurs.

- Le bruit peut perturber le signal $S(t)$.

La figure-41 ~~représente la fonction de transfert~~ $V(Z)$ du conduit vocal seul.

L' " idéal " en matière de modélisation par covariance dans l'intervalle de fermeture de la glotte est d'obtenir une pente de la fonction de transfert $V(Z)$ aussi plate que possible et une onde $U_g(t)$ compatible avec les résultats déjà obtenus (cf.chV § 13)

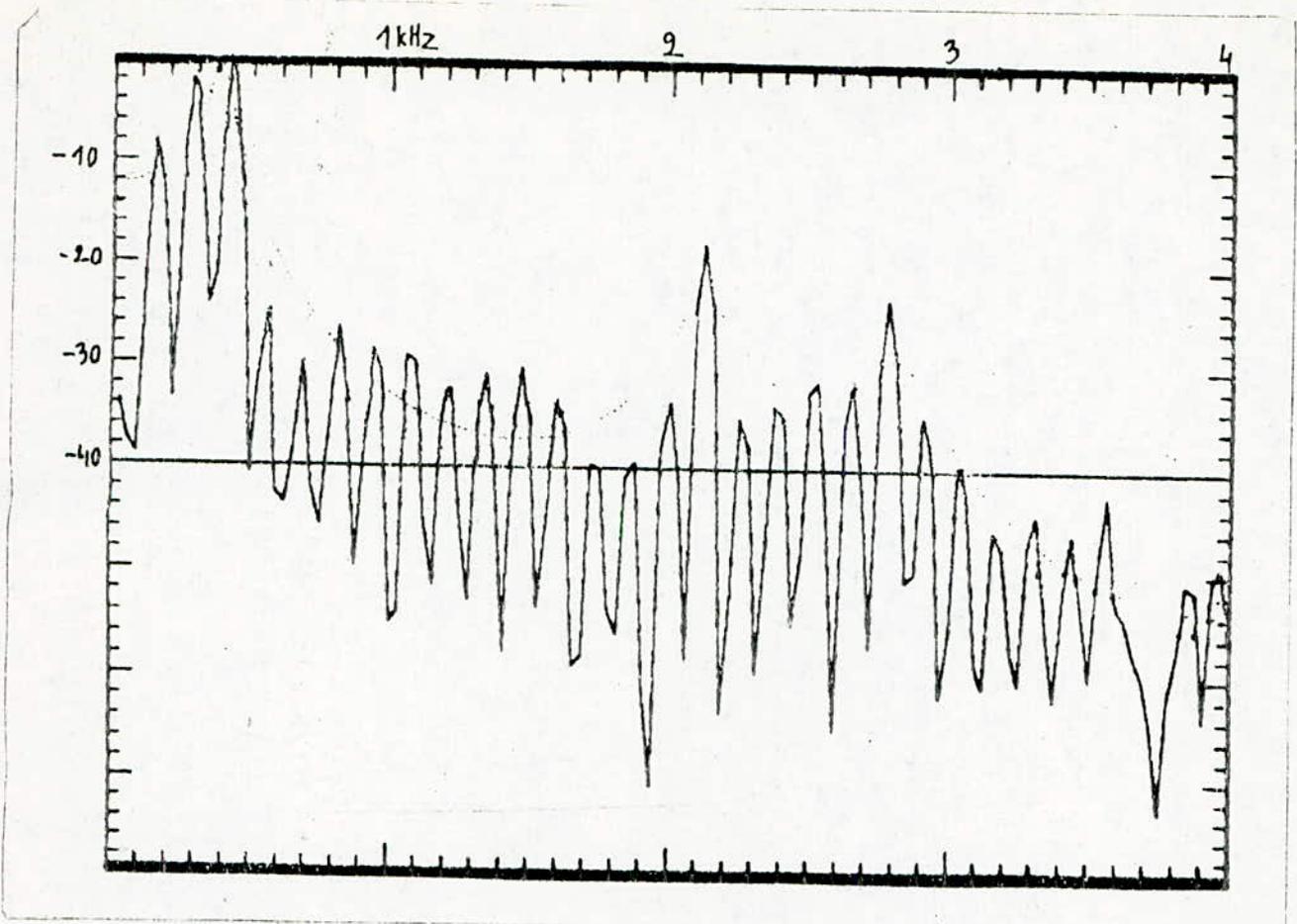


Figure 40 : Spectre de puissance d'un /e/

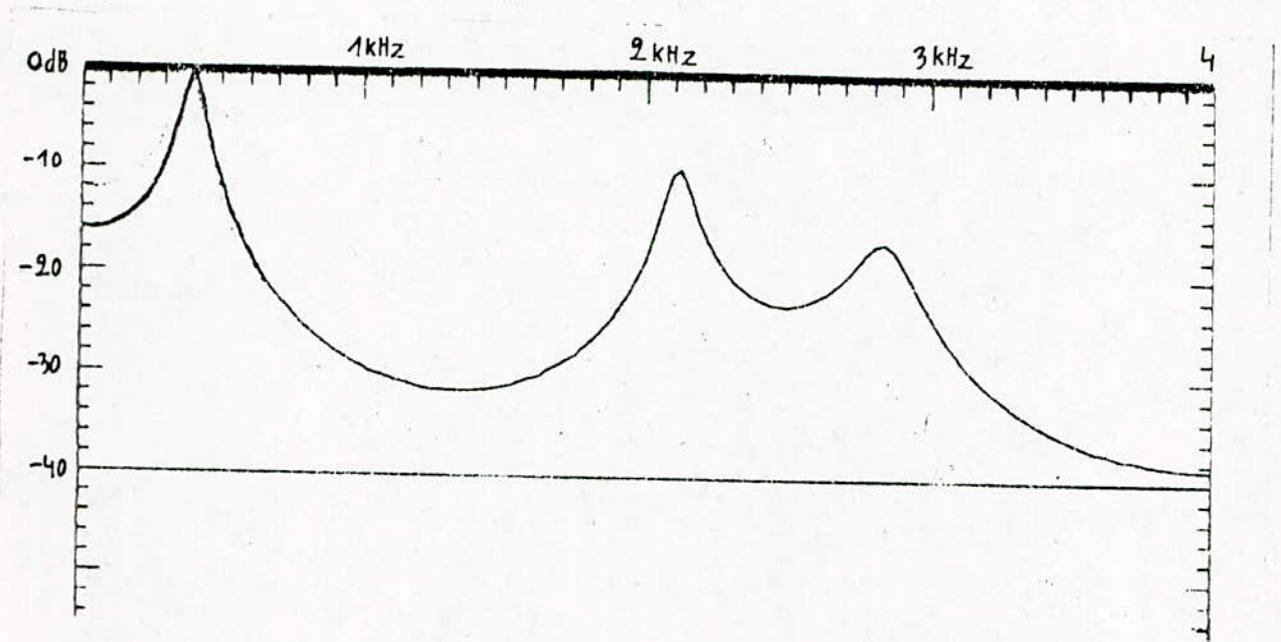


Figure 41 : Fonction de transfert du conduit vocal seul

Notons que ces deux critères sont liés car l'obtention de $U_g(t)$ est conditionnée par la modélisation préalable du conduit vocal.

L'objectif du paragraphe suivant est de déterminer automatiquement les valeurs de M et N ainsi que la position optimale de la fenêtre d'analyse en fonction des critères précédents.

V - 3 - 2 - TRAITEMENT " AUTOMATIQUE "

a - Cas de la prédiction linéaire.

Les figures 42 à 45 nous montrent les effets du déplacement d'une fenêtre d'analyse.

Les paramètres N et M sont fixés deux cas sont considérés:

$N = 25, M = 8$ et $N = 25, M = 10$. La paramètre L_c qui détermine le déplacement de la fenêtre varie de 65 à 120 par pas de 5 ou de 10.

- OBSERVATIONS

Les figures 42 et 44 qui représentent la forme temporelle de l'onde glottique $U_g(t)$ montrent clairement que la valeur de L_c optimum, déterminant l'onde U_g la plus " réaliste ", correspond à $L_c = 95$, soit effectivement le minimum de E_N .

Avant et après cette position de la fenêtre d'analyse, on assiste à une dégradation du signal $U_g(t)$.

Les figures 43a et 45 représentant la forme de la fonction de transfert du filtre $1/A(Z)$ confirment ce choix, le spectre étant le plus plat pour $L_c = 95$.

La modélisation est meilleure dans le cas où $M = 10$, le spectre est plus plat et la forme de U_g est sensiblement meilleure, notons que l'erreur quadratique moyenne E_u est moindre; ceci ne signifie nullement comme le montre la figure 43b qu'il faille adopter une modélisation, caractérisée par les paramètres N et M, fournissant une erreur quadratique moyenne inférieure à celle fournie par une modélisation caractérisée par d'autres paramètres. Seule la valeur de l'erreur quadratique moyenne minimale pour une modélisation donnée permet de définir l'intervalle de fermeture de la glotte et de calculer le filtre $V(Z)$.

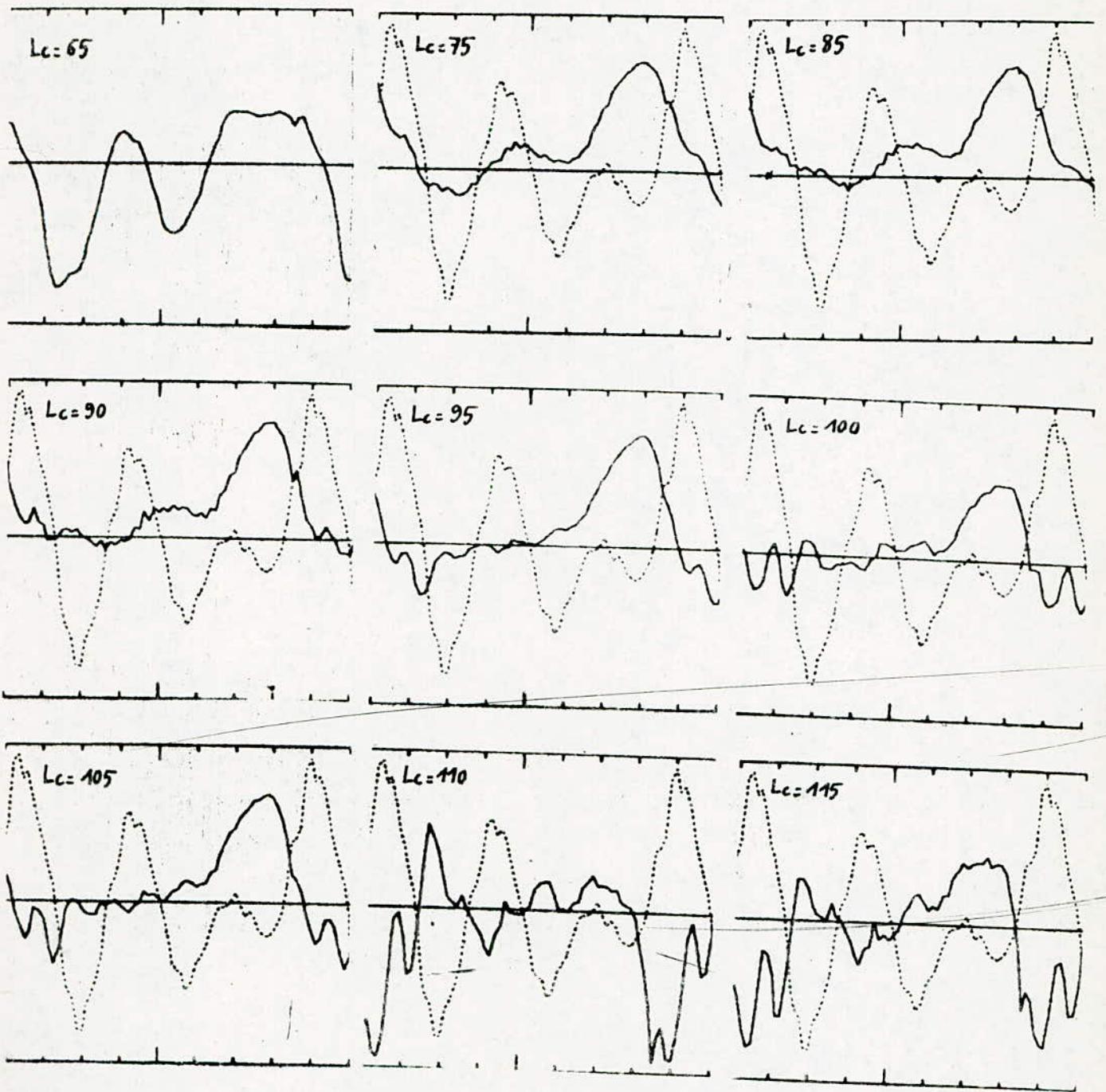


Figure 42 : $U_g(t)$; Glissement d'une fenêtre de 25 points, $M = 80$

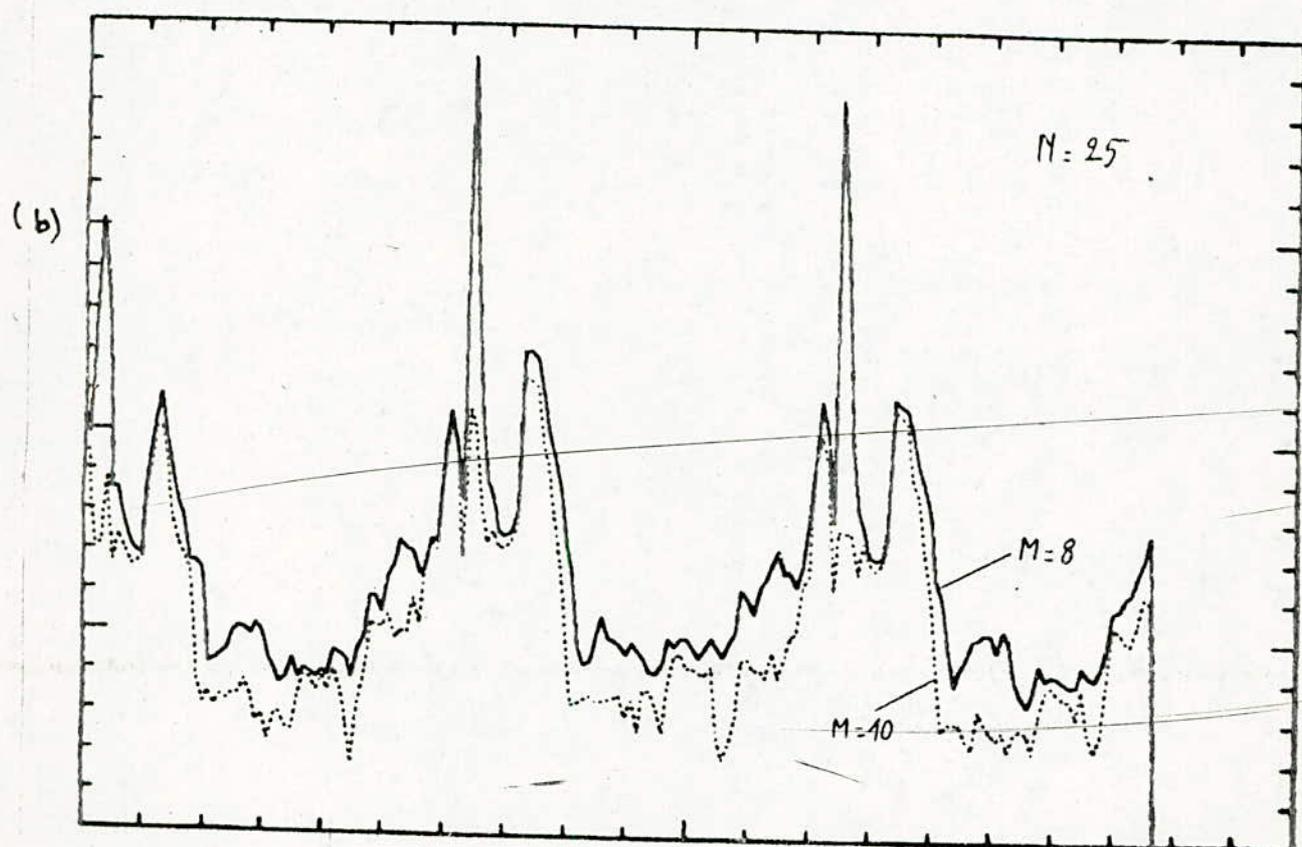
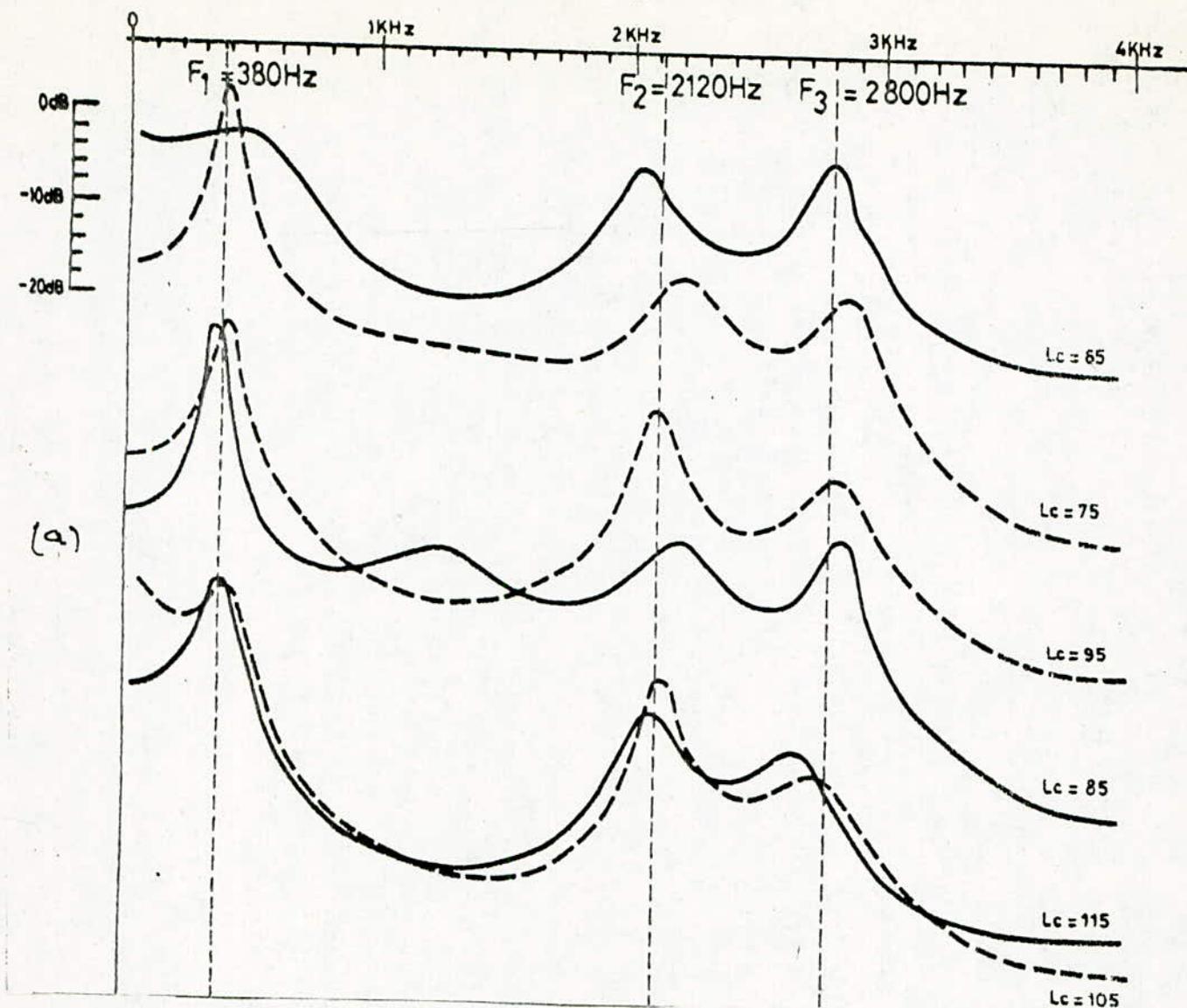


Figure 43 : a - Fonction de transfert du conduit vocal, $N = 25$, $M = 8$
 b - Erreur quadratique moyenne pour $M = 8$ et $M = 10$

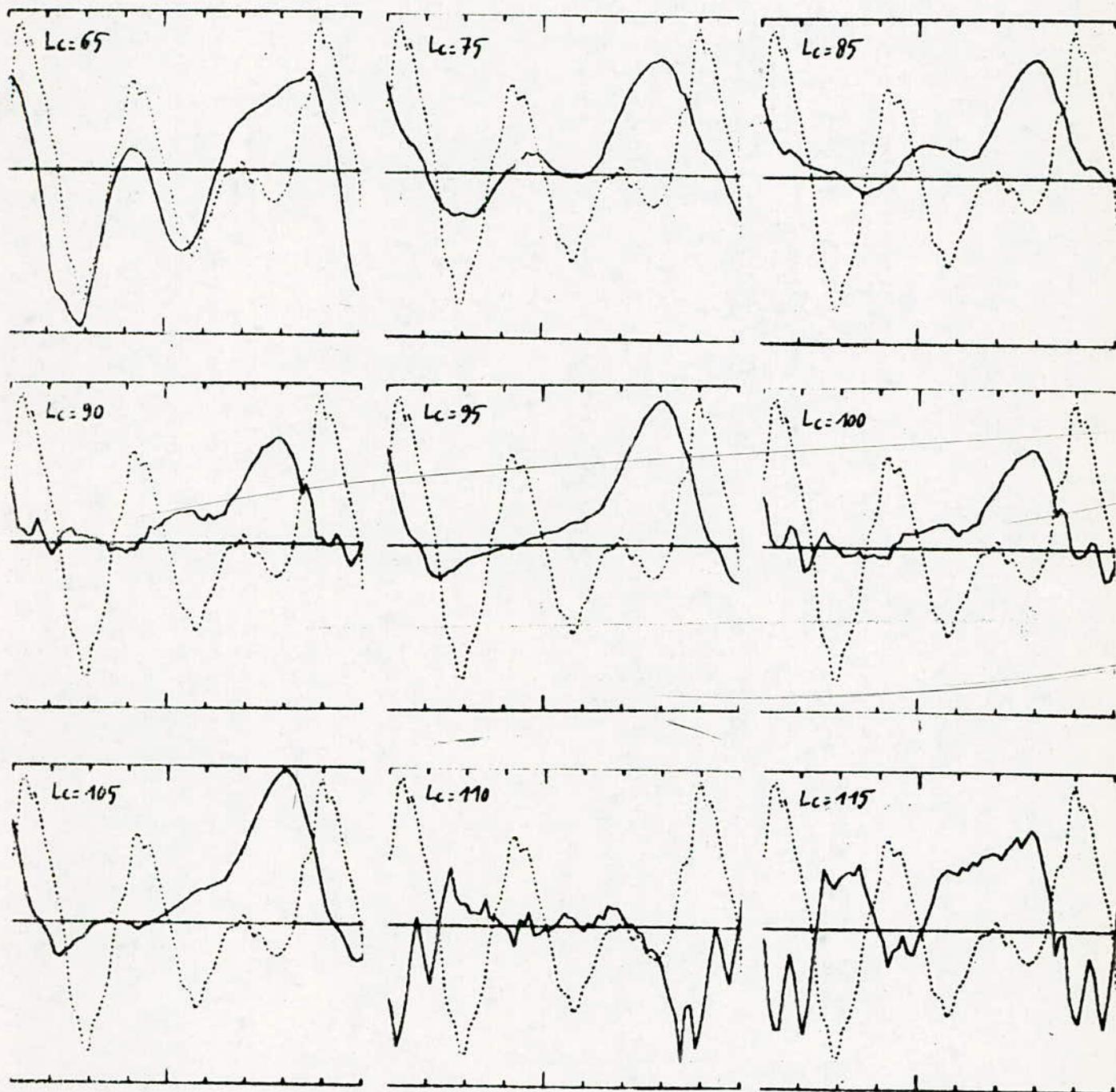


Figure 44 : $U_g(t)$; Glissement d'une fenêtre de 25 points, $M = 10$

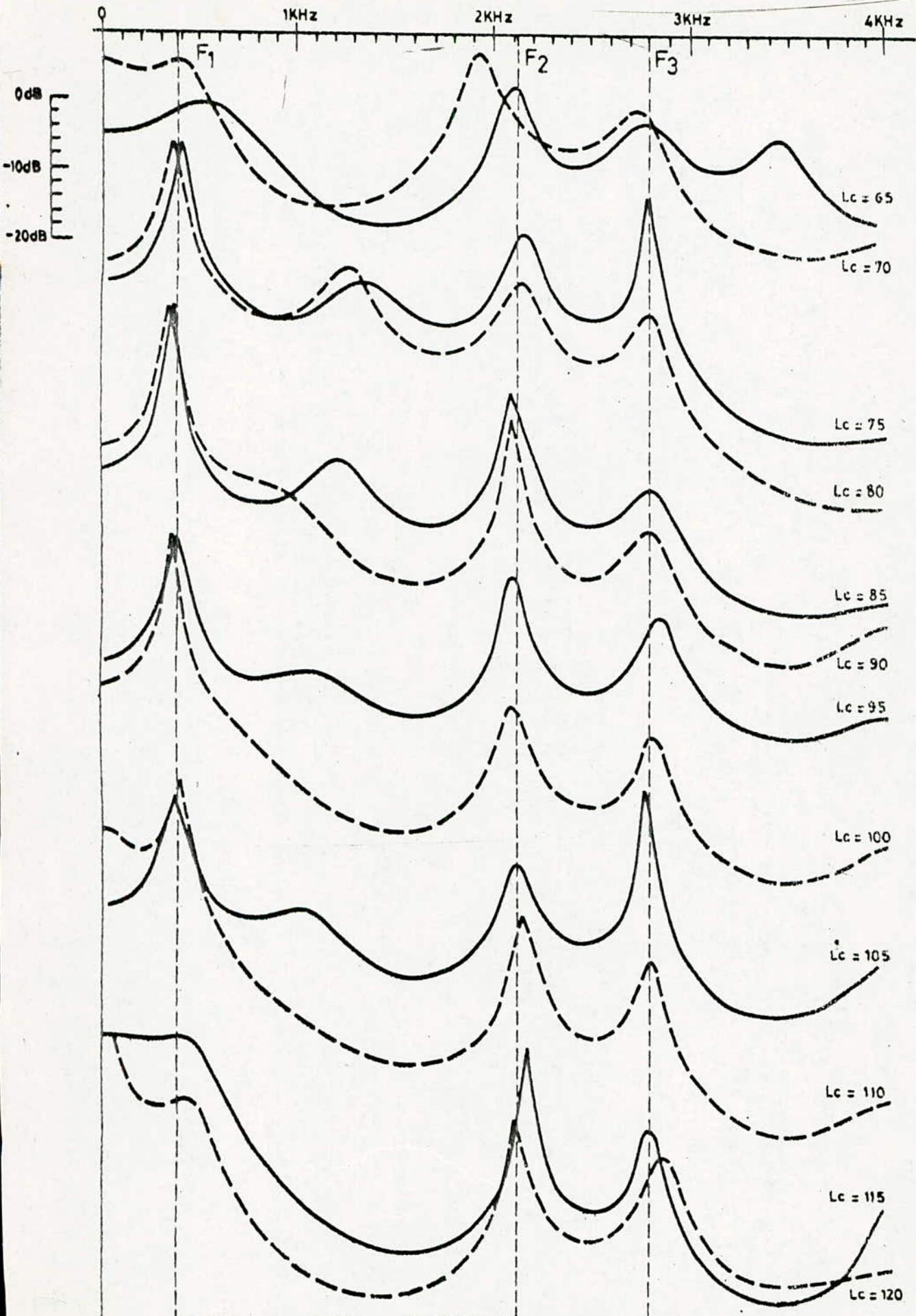


Figure 45 : Fonction de transfert du conduit vocal, $N = 25$, $M = 10$

b - Cas de l'interpolation linéaire.

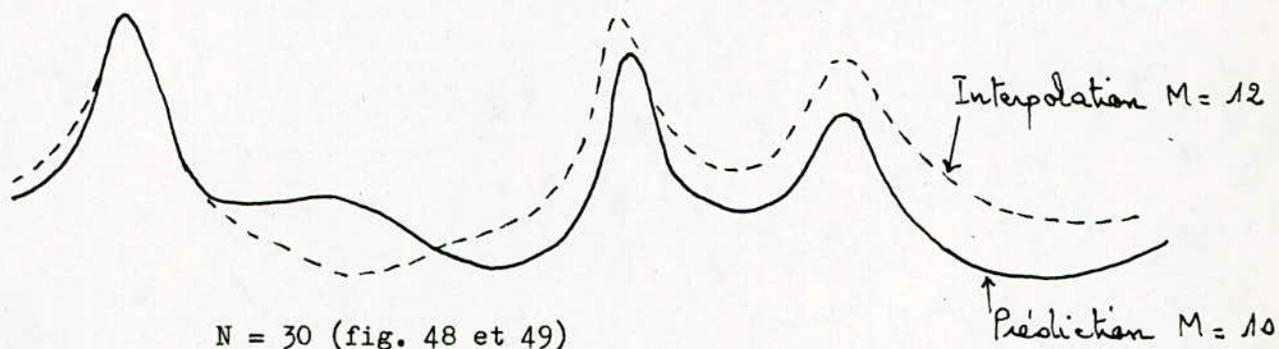
Nous avons déterminé l'emplacement optimal de la fenêtre d'analyse. Il s'agit maintenant de faire varier N et M dans une gamme plus large afin de découvrir leurs valeurs optimales. Ainsi en comparant les résultats obtenus par la méthode d'interpolation avec ceux obtenus à l'aide de la méthode de prédiction, nous observons :

Les figures 46 et 48 représentent les modélisations de U_g pour différentes valeurs des paramètres N et M , l'emplacement de la fenêtre étant fixe.

$N = 25$ (fig. 46 et 47)

La méthode de prédiction donne les valeurs réalistes de U_g pour $M = 6, 8, 10$ et 12 (fig. 46). Le spectre modèle, quant à lui, ne devient représentatif qu'à partir de $M = 8$ et la réalisation la plus plate correspond à $M = 10$ (fig. 47).

La méthode d'interpolation ne donne un résultat acceptable que pour $M = 12$, tant pour U_g que pour le spectre modèle. En effet, ce dernier est plus plat que celui obtenu par prédiction linéaire correspondant à $M = 10$.



La méthode de prédiction donne des résultats encore acceptables bien que la fenêtre d'analyse commence à pénétrer dans l'intervalle d'ouverture de la glotte (fig. 48). La méthode d'interpolation ne donne des résultats que pour $M = 6$ (fig. 49), mais alors les résonances du conduit vocal ne sont plus modélisées avec assez de coefficients (fig. 49).

$N = 35$ (fig. 48 et 49)

Aucune des deux méthodes ne permet d'obtenir une modélisation acceptable, la fenêtre d'analyse étant plus large que l'intervalle de fermeture de la glotte.

$N = 20$ (fig. 48 et 49)

Ici également les deux méthodes ne donnent aucun résultat, car la portion du signal analysé n'est pas assez grande et donc pas assez représentatif du signal $S(t)$ durant cette période d'oscillation libre du conduit vocal.

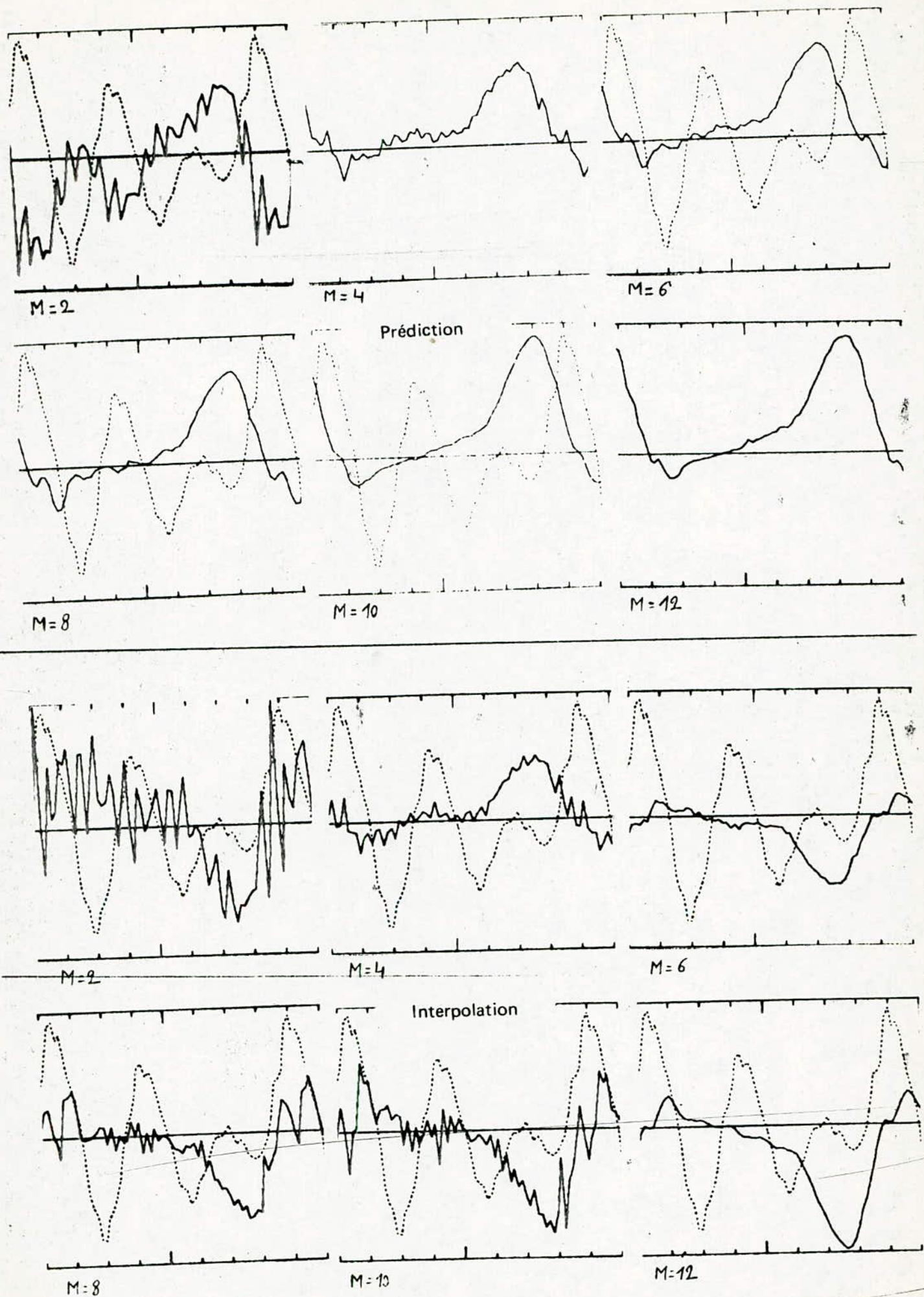


Figure 46 : $U_g(t)$; Fenêtre fixe, $N = 25$, M varie de 2 à 12

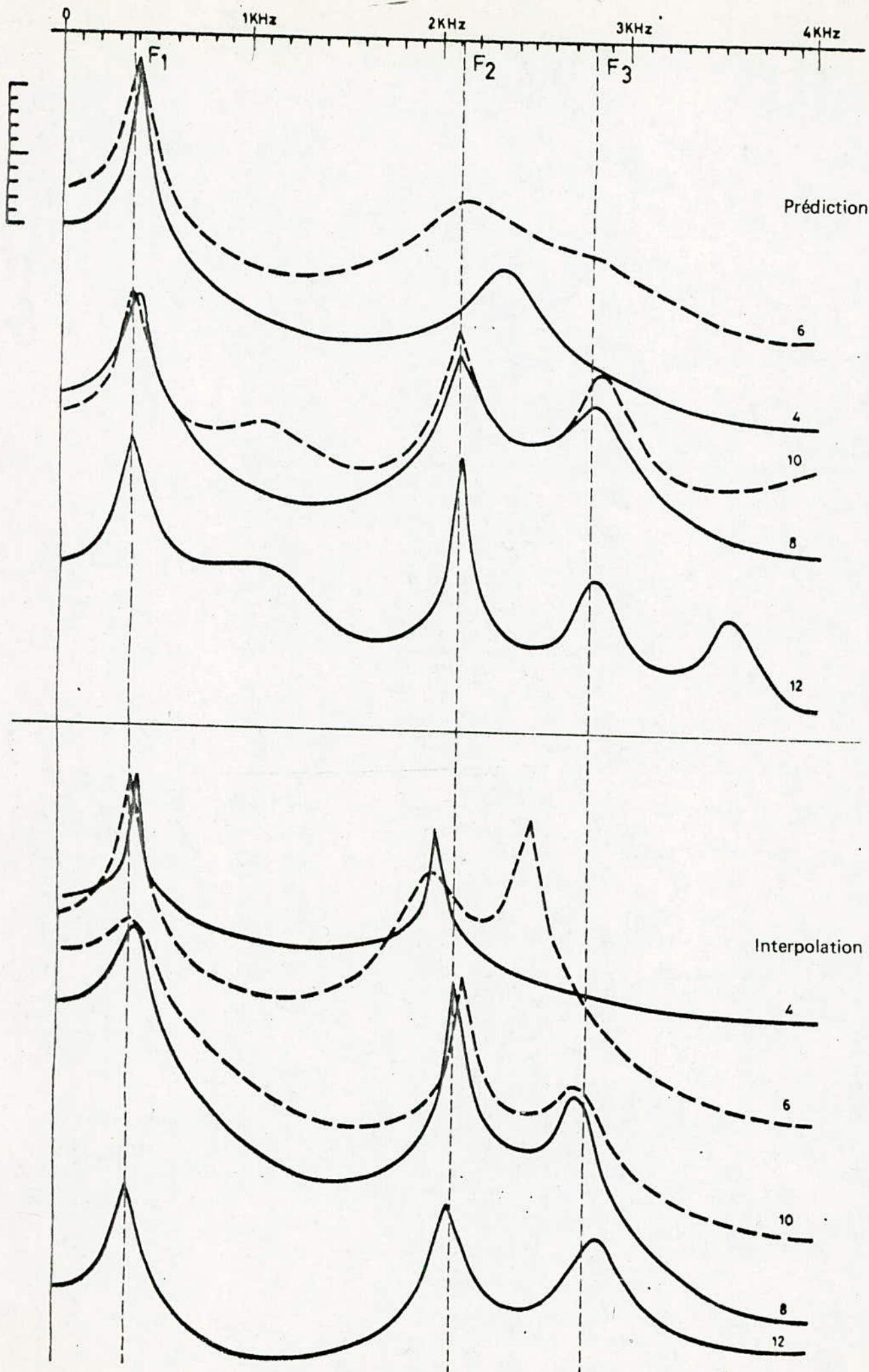
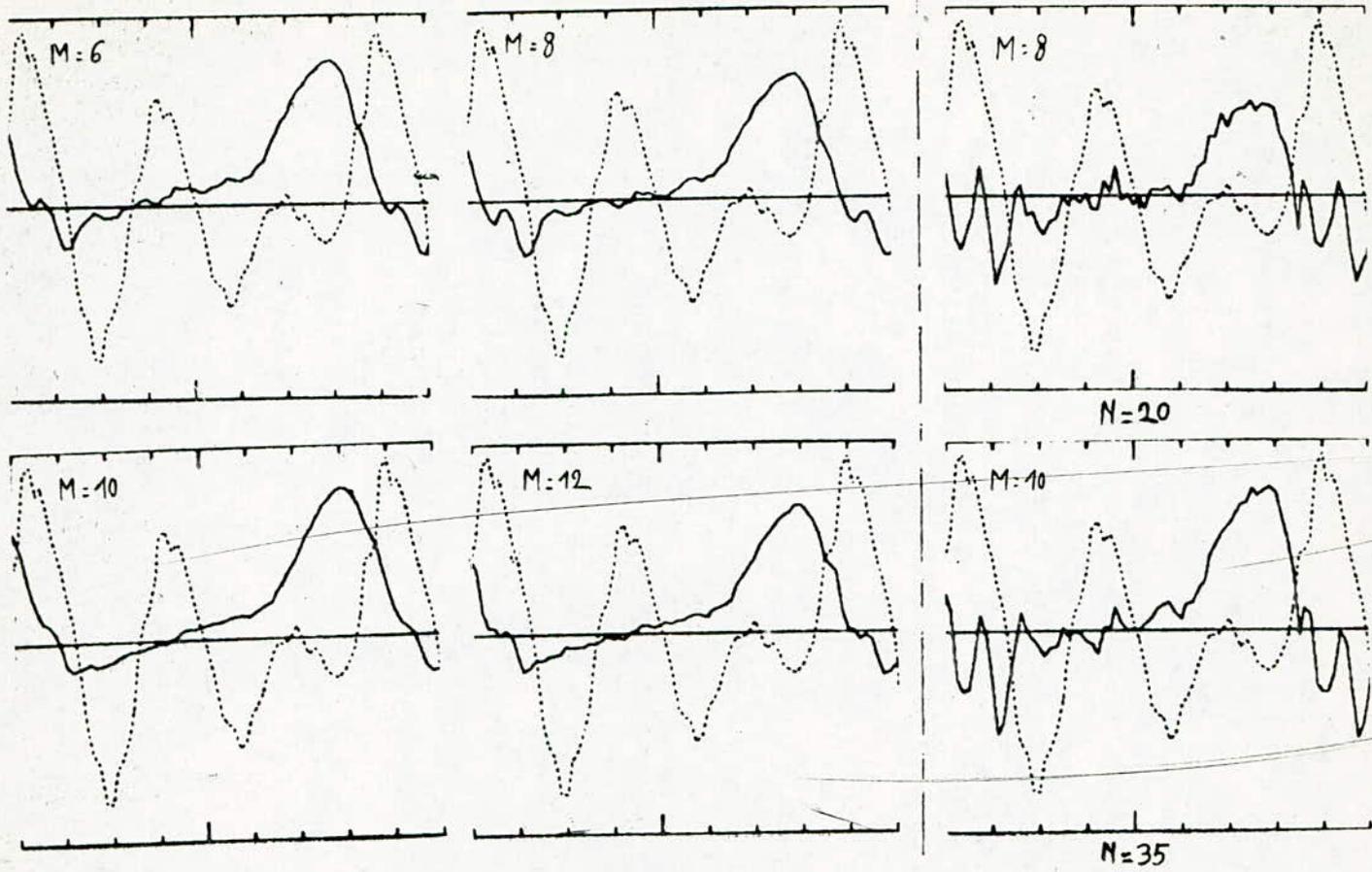


Figure 47 : Fonction de transfert du conduit vocal, $N = 25$, $M = 4, 12$

Prédiction



Interpolation

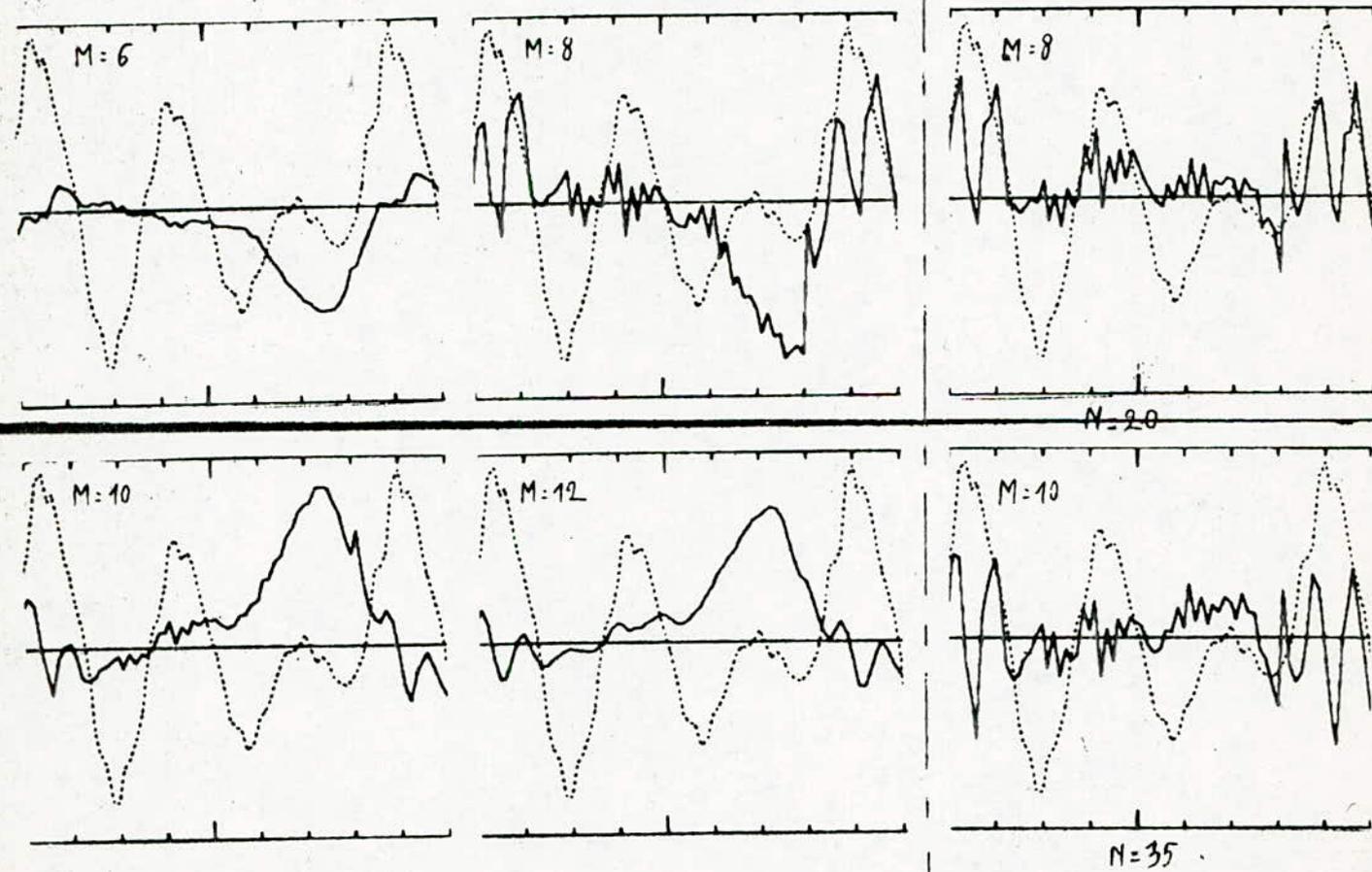


Figure 48 : $U_g(t)$; Fenêtre fixe, $N = 30$, M varie de 6 à 12

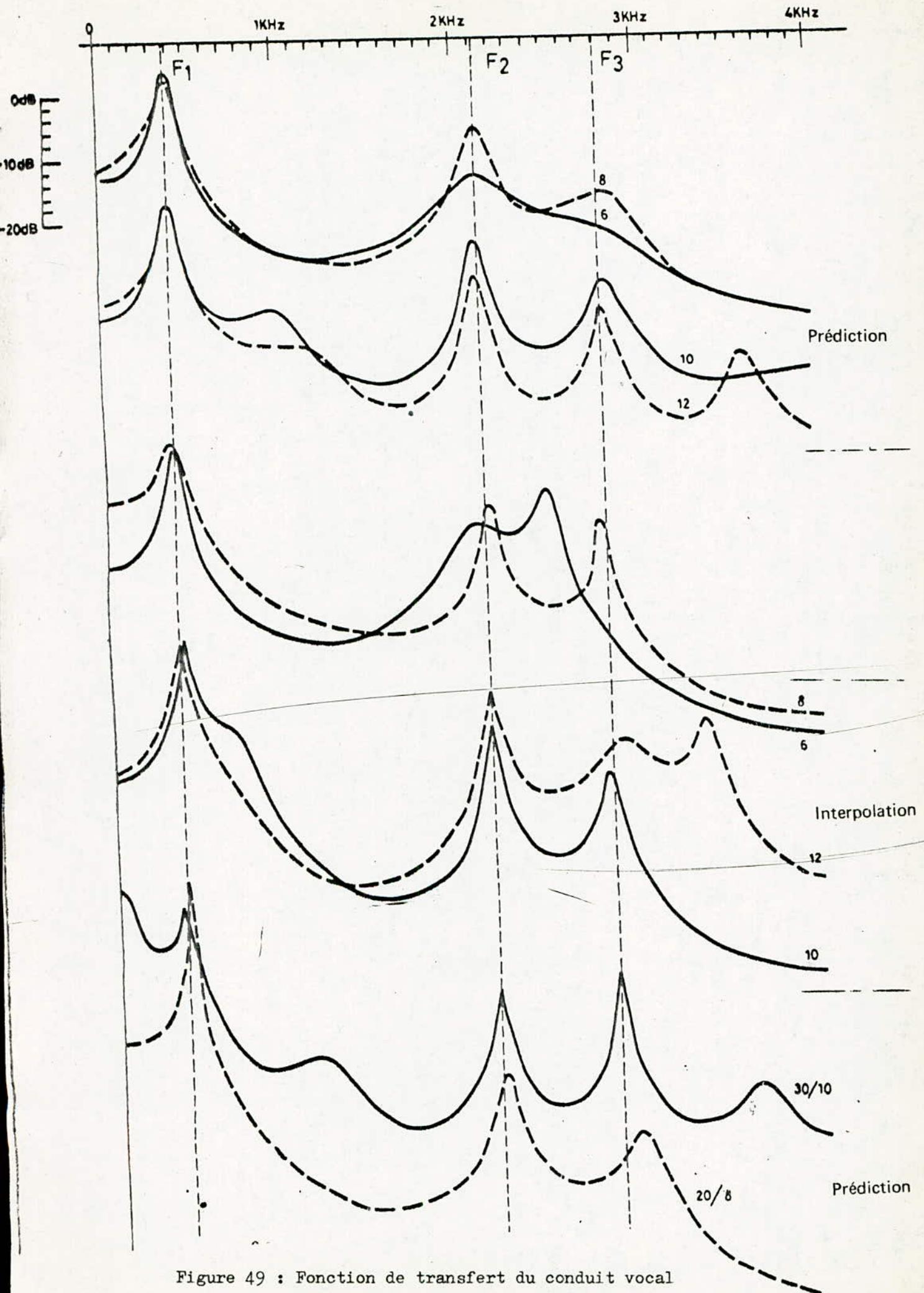


Figure 49 : Fonction de transfert du conduit vocal

V - 4 - CONCLUSION

En faisant varier simultanément les 3 paramètres L_c , N et M , il est possible de déterminer leurs valeurs optimales, permettant alors le calcul de l'onde de débit et de la fonction de transfert du conduit vocal seul jugées les plus réalistes.

A partir de ces observations nous pouvons conclure que la méthode d'interpolation qui donne une erreur quadratique moyenne inférieure à celle obtenue par prédiction (fig. 39b), est beaucoup plus dépendante des valeurs des différents paramètres que sont:

- N , la longueur de la fenêtre d'analyse.
- M , le nombre de coefficients du filtre.
- L_c , l'emplacement de la fenêtre d'analyse.

Aussi pour extraire l'onde glottique U_g , l'analyse par prédiction linéaire qui paraît plus faible dans le cas de la covariance.

CONCLUSION GENERALE

Dans ce projet, nous avons étudié quelques aspects du traitement du signal intervenant dans l'analyse de la parole, en mettant l'accent sur l'estimation de la fréquence fondamentale qui est un paramètre pertinent du signal de parole ainsi que sur l'analyse de l'onde glottique par filtrage inverse.

Pour déterminer cette fréquence, trois méthodes ont été étudiées en détail, à savoir :

La méthode d'autocorrélation simple et modifiée et la technique de traitement parallèle. Les avantages et les inconvénients d'application propres à chacune d'elles ont été mentionnés. Nous avons opté pour la méthode de traitement parallèle. Les raisons qui ont motivé notre choix sont les suivantes :

- L'algorithme ainsi que ses deux modifications déjà cités sont plus précis et efficaces dans les tests sur la parole réelle et synthétique (B. Gold et L-R. Rabiner, 1969).
- La vitesse de calcul est plus rapide que celle des techniques utilisant la fonction d'autocorrélation.

En ce qui concerne la détermination de l'onde glottique à partir du signal de parole, la méthode de filtrage inverse adoptée utilise la technique de covariance dans l'intervalle de fermeture de la glotte. Les résultats fournis sont cohérents avec ceux observés par ailleurs. Cependant, ils dépendent fortement de la qualité du signal acoustique utilisé et de la nature de la voyelle étudiée. Parfois, deux signaux représentatifs de deux voyelles identiques peuvent ne pas avoir la même forme temporelle et conduiront à deux résultats différents. Or, si la méthode de covariance dans l'intervalle de fermeture de la glotte est théoriquement la meilleure pour ce type d'utilisation, elle reste très sensible à ce phénomène.

Pour les contraintes, vu la non disponibilité du matériel nécessaire pour réaliser nos expériences, tels que l'audiomètre, l'analyseur de spectre, etc... Nous étions amenés à procéder d'une manière approchée sur des exemples de signaux simples pour pouvoir les utiliser dans nos programmes et mener à bien notre travail.

Néanmoins, notre étude pourrait permettre d'effectuer des mesures sur les valeurs des impédances d'entrée et de sortie du conduit vocal, qui est une étude importante sur l'interaction source - conduit vocal (Guérin et Boe, 1979). Cela peut être l'objet d'une recherche ultérieure bien approfondie.

B I B L I O G R A P H I E

- BELLANGER. M, 1981, " Traitement numérique du signal " Ed. Masson.
- FANT. G, 1979, " Glottal source and excitation analysis " STL, QPRS, Vol. 1, pp 85-100.
- FLANAGAN. J-L, 1958, " Somme properties od glottal sound source " JSHR, Vol. 1,pp 99-116.
- GOLG. B and RABINER. L-R, Aug 1969, " Parallel processing techniques for estimating pitch periods of speech in the time domain " JASA, Vol. 42, pp 442-448.
- GUERIN. B, 1978, " Etude du fonctionnement de la source vocale, simulation d'un modèle ", Thèse Doctorat d'Etat, Université de Grenoble.
- GUERIN. B, BOE. L-J, 1979, " Influence du couplage accoustique avec le conduit vocal ", Séminaire larynx et parole, Grenoble, pp 103-144.
- GUIBERT. J, 1979, " La parole compréhension et synthèse par les ordinateurs " Ed. Presses universitaires de France (PUF).
- HAVET. P, 1980, " Contribution à l'étude de la source vocale: Détermination des paramètres de commande d'un modèle de source vocale et étude de l'onde glottique par filtrage inverse ", Thèse Docteur-Ingénieur, Université de Rennes.
- KUNT. M, 1981, " Traitement numérique des signaux " Ed. Dunod.
- LIFERMAN. J, 1980, " Les méthodes rapides de transformation du signal et application aux mesures physiques ", Tome 1, Ed. Masson.
- MALMBERG.B, " La phonétique, que sais-je? " Ed. PUF.
- MAX. J, 1981, " Méthodes et techniques de traitement du signal et applications aux mesures physiques ", Tome 1, Ed. Masson.
- ROTHENBERG. M, 1972, " A new inverse-filtering technique for deriving the glottal air-flow wave form during voicing ", JASA, Vol. 53, pp 1632-1645.
- SCHAFER. R-W and RABINER. L-R, 1978 " Digital processing of speech signals " séries Editor.
- SONDHI. MM, June 1968, " New Méthods of pitch extraction ", IEEE, trans, Audio electroacoust, Vol. AU-16, pp 262-266.
- SONDHI. MM, 1975, " Measurement of the glottal wave form ", JASA, Vol. 57, pp 228-234.

ANNEXE : PROGRAMMATION

---oooOooo---

```

10 ****** METHODE D'AUTOCORRELATION *****
20 *x(i):signal filtre passe bas et fenetre
30 *o:nombre d'echantillons de la fonction d'autocorrelation
40 *FE:frequence d'echantillonnage de X(I)
50 FOR M=0 TO O-1
60 Q(M)=0
70 FOR I=M TO N-1
80 Q(M)=Q(M)+X(I)*X(I-M)
90 NEXT I
100 Q(M)=Q(M)/N
110 NEXT M
120 FOR I=1 TO O-1
130 ****** Recherche du 1er minimum de q(i) *****
140 IF Q(I+1)>Q(I) THEN Q(I)=MIN:S=I+1
150 NEXT I
160 ****** Recherche du 1er maximum de q(i) *****
170 I=S
180 FOR J=O-1 TO S STEP -1
190 IF Q(I)<Q(J) THEN I=I+1:GOTO 190
200 IF I=J THEN MAX Q(I):IMAX=I ELSE NEXT J
210 ****** Decision de voisement *****
220 *v=1 séquence voisée
230 *v=0 séquence non voisée
240 IF (MAX/Q(O))>=.25 THEN FO=FE/IMAX:V=1
250 IF (MAX/Q(O))<.25 THEN FO=0:V=0
260 END

```

```

10 ****** FILTRAGE PASSE-BAS *****
20 *FC:frequence de coupure /KHZ
30 *T: periode d'echantillonnage /ms
40 FC=.8
50 T=1/FE
60 B1=(.3572)*2*3.141592*FC
70 B2=(.1786)*3.141592*FC
80 B3=(.8938)*3.141592*FC
90 A1=1-EXP(-(B1*T))
100 A2=EXP(-(B1*T))
110 A3=1-(2*EXP(-(B2*T))*COS(B3*T))+EXP(-(2*B2*T))
120 A4=2*EXP(-(B2*T))*COS(B3*T)
130 A5=-EXP(-(2*B2*T))
140 **** initialisation ***
150 U(0)=A1*S(0):X(0)=A3*U(0)
160 U(1)=A1*S(1)+A2*U(0):X(1)=A3*U(1)+A4*X(0)
170 U(2)=A1*S(2)+A2*U(1):X(2)=A3*U(2)+A4*X(1)+A5*X(0)
180 **** X(I):echantillon filters ***
190 FOR I=1 TO 2-1
200 U(I)=A1*S(I)+A2*U(I-1)
210 X(I)=A3*U(I)+A4*X(I-1)+A5*X(I-2)
220 NEXT I
230 END

```

```

0 ***** PITCH PAR L'AUTOCORRELATION MODIFIEE *****
0
0 ***** Calcul du seuil d'ecretage C *****
0 N1=INT(N/3+.5)
0 L=0:H=N1-1:Q=1
0 I=L
0 FOR J=L TO H
0 C=ABS(X(I))-ABS(X(J))
0 IF C<0 THEN I=I+1:GOTO 80
00 IF I=J THEN M(Q)=ABS(X(I)) ELSE NEXT J
10 Q=Q+1:L=N-N1-2:Q=N-1
20 IF Q<3 THEN 60
30 IF M(1)<M(2) THEN C=.64*M(1) ELSE C=.64*M(2)
40 ***** Codage et autocorrelation *****
50 FOR I=1 TO N-1
60 IF X(I)>=C THEN Y(I)=1
70 IF ABS(X(I))<C THEN Y(I)=0
80 IF X(I)<=-C THEN Y(I)=-1
90 NEXT I
100 Q(M)=0
110 FOR M=0 TO Q-1
120 Q(M)=0
130 FOR I=M TO N-1
140 Q(M)=Q(M)+Y(I)*Y(I-M)
150 NEXT I:Q(M)=Q(M)/N
160 NEXT M
170 ***** Recherche du 1er min de Q(i) *****
180 FOR I=1 TO Q-1
190 IF Q(I+1)>Q(I) THEN MIN=Q(I):S=I+1 ELSE NEXT I
200 ***** Recherche du max de Q(i) *****
210 I=S
220 FOR J=S TO Q-1
230 IF Q(I)<Q(J) THEN I=I+1:GOTO 330
240 IF I=J THEN MAX=Q(I):IMAX=I ELSE NEXT J
250 ***** Decision de voisement *****
260 *v=1 séquence voisée
270 *v=0 séquence non voisée
280 IF MAX/Q(0)>=.25 THEN FO=FE/IMAX:V=1
290 IF MAX/Q(0)<.25 THEN FO=0:V=0
300 END

```

```

0 ***** Methode de sondhi *****
0 *x(i) signal filtre passe-bas
0 ***** Calcul du seuil d'ecretage CL *****
0 I=0
0 FOR J=0 TO N-1
0 IF ABS(X(I))<ABS(X(J)) THEN I=I+1:GOTO 60
0 IF I=J THEN M=ABS(X(I)) ELSE NEXT J
0 CL=M*.3
0 ***** Codage du signal et autocorrelation *****
00 FOR I=0 TO N-1
10 IF ABS(X(I))<CL THEN Y(I)=0
20 IF X(I)>=CL THEN Y(I)=X(I)-CL
30 IF X(I)<=-CL THEN Y(I)=X(I)+CL
40 NEXT I
50 FOR M=0 TO Q-1
60 Q(M)=0
70 FOR I=M TO N-1
80 Q(M)=Q(M)+Y(I)*Y(I-M)
90 NEXT I
100 Q(M)=Q(M)/N
110 NEXT M
120 ***** Recherche du minimum de Q(i) ***
130 FOR I=0 TO Q-1
140 IF Q(I+1)>Q(I) THEN Q(I)=MIN:S=I+1 ELSE NEXT I
150 ***** Recherche du maximum de Q(i) ***
160 I=S
170 FOR J=0-1 TO S STEP -1
180 IF Q(I)<Q(J) THEN I=I+1:GOTO 280
190 IF I=J THEN MAX=Q(I):IMAX=I ELSE NEXT J
200 ***** Decision de voisement *****
210 *v=1 séquence voisée
220 *v=0 séquence non voisée
230 IF MAX/Q(0)>=.25 THEN FO=FE/IMAX:V=1
240 IF MAX/Q(0)<.25 THEN FO=0:V=0
250 END

```

```

*****DETECTION DU PITCH PAR LA METHODE PARALLELE*****
s(i):signal filtre passe bas
DIM M(400),N(6,400),V(100),PI(100),P(7,7),FA(7),B(7)
I=1:L=1:J=1
FOR K=3 TO N-1.....m1.....
IF (S(K-3)<S(K))*S(K)*(S(K+3)<S(K))*S(K-1)<S(K))*S(K+1)<S(K))*S(K-2)<S(K))*S(K+
<S(K)) THEN N(1,K)=ABS(S(K)):M(I)=S(K):PI(L)=K:I=I+1:L=L+1
IF (S(K-3)>S(K))*S(K+3)>S(K))*S(K-2)>S(K))*S(K-1)>S(K))*S(K+2)>S(K))*S(K
1)>S(K)) THEN N(4,K)=ABS(S(K)):M(I)=S(K):V(J)=K:I=I+1:J=J+1
NEXT K:IG=I
IF M(1)>M(2) THEN OC=1 ELSE OC=0
IF OC=1 THEN L=2 ELSE L=1.....m2.....
FOR I=2 TO IG STEP 2
IF PI(L)=0 THEN 200
N(2,PI(L))=ABS(M(I+OC)-M(I+OC-1))
L=L+1
NEXT I
L=1.....m3.....
FOR I=2 TO IG STEP 2
N(3,PI(L))=M(I-OC+2)-M(I-OC)
IF N(3,PI(L))<0 THEN N(3,PI(L))=0
L=L+1
NEXT I
IF OC=1 THEN L=1 ELSE L=2.....m5.....
FOR I=2 TO IG-1 STEP 2
IF V(L)=0 THEN 440
N(5,V(L))=ABS(M(I-OC+1)-M(I-OC))
L=L+1
NEXT I
IF OC=1 THEN L=2 ELSE L=1.....m6.....
FOR I=2 TO IG STEP 2
IF V(L)=0 THEN 410
N(6,V(L))=(M(I+OC)-M(I+OC+1))
IF N(6,V(L))<0 THEN N(6,V(L))=0
L=L+1:NEXT I
.....selection des impulsions.....
FOR J=1 TO 6
U=1:I=1
IF N(J,I)><>0 THEN B(U)=I:L=I:M=N(J,I):GOTO 660
I=I+1
IF I>N-1 THEN END ELSE 440
FOR I=L TO N-30
IF N(J,I+30)>=M*EXP(-(I+30)/80) THEN 490 ELSE 510
M=N(J,I+30):U=U+1:B(U)=I+30
IF I>3 THEN 520
NEXT I
..... Formation de la matrice .....
RA=B(2)-B(1)
P(J,1)=F/RA:PRINT P(J,1)
P(J,2)=F/(B(3)-B(2)):PRINT P(J,2)
P(J,3)=F/(B(4)-B(3))
P(J,4)=P(J,1)+P(J,2)
P(J,5)=P(J,3)+P(J,2)
P(J,6)=P(J,1)+P(J,2)+P(J,3)
NEXT J:STOP
FOR J=1 TO 6
FOR I=1 TO 6
NEXT I
NEXT J
..... Nombre de coïncidences .....
mar:marge(Hz)
fa(i):nbre de coïncidences
p(i,1):candidats
MAR=30
FOR I=1 TO 6
FA(I)=0
FOR J=1 TO 6
FOR K=1 TO 6
IF (K=I)AND(J=1) THEN 760
IF ABS(P(I,1)-P(K,J))<=MAR THEN FA(I)=FA(I)+1
NEXT K
NEXT J
NEXT I
..... DECISION -CALCUL DE FO .....
J+1
FOR I=6 TO 1 STEP -1
IF FA(J)<FA(I) THEN J=J+1:GOTO 820
IF J=I THEN MAX=FA(I) ELSE NEXT I
IF MAX>=5 THEN FO=P(I,1):V=1 ELSE FO=0:V=0
*****

```