الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE





وزارة التعليم و البحث العلمي
MINISTERE DE L'ENSEIGNEMENT ET DE LA RECHERCHE SCIENTIFIQUE

# ECOLE NATIONALE POLYTECHNIQUE

## DEPARTEMENT d'ELECTRONIQUE

# PROJET DE FIN D'ETUDES

**SUJET**

# DETECTION DE LA FREQUENCE FONDAMENTALE A L'AIDE DE LA TECHNIQUE DE PREDICTION LINEAIRE

**Proposé par :**
M^lle M. GUERTI

**Etudié par :**
Malika SADAT
Djida YAHLALI

**Dirigé par :**
M^lle M. GUERTI

**PROMOTION** : JUIN 85

E.N.P. 10, Avenue Hacen Badi – EL-HARRACH – ALGER

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم و البحث العلمي
MINISTERE DE L'ENSEIGNEMENT ET DE LA RECHERCHE SCIENTIFIQUE

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT d'ELECTRONIQUE

# PROJET DE FIN D'ETUDES

**SUJET**

# DETECTION DE LA FREQUENCE FONDAMENTALE A L'AIDE DE LA TECHNIQUE DE PREDICTION LINEAIRE

**Proposé par :**
M^lle M. GUERTI

**Etudié par :**
Malika SADAT
Djida YAHLALI

**Dirigé par :**
M^lle M. GUERTI

**PROMOTION :** JUIN 85

E.N.P. 10, Avenue Hacen Badi – EL-HARRACH – ALGER

***DEDICACES***

A MON MARI ET A MON FILS SID-ALI.

A MA GRAND-MERE FATIMA.

A MES PARENTS ET A MES BEAUX-PARENTS.

A MES TANTES,ONCLES ET LEURS ENFANTS.

A MES FRERES ET SOEURS.

A MES AMIES:ZOUBIDA,SALEHA,SAIDA,MALIKA,DJAMILA ET NACERA.

A TOUS CEUX QUI ME SONT CHERS.

-DJIDA-

A MON MARI ET A MON FILS YACINE.

A MES PARENTS ET A MES BEAUX-PARENTS.

A MES AMIES.

-MALIKA-

** REMERCIEMENTS **

Nous remercions notre promoteur Melle M. GUERTI de nous avoir suivies et conseillées durant tout notre travail.

Nous adressons aussi nos remerciements à M. A. HADJ-SALAH, Directeur de l'Institut de Linguistique et de Phonétique de l'Université d'Alger de nous avoir autorisées l'utilisation du calculateur TEKTRONIX 4052.

Que M. R. WOROSZCZUK enseignant à l'I.L.P et ses collègues trouvent ici le témoignage de notre reconnaissance pour l'aide qu'ils nous ont apportée en programmation.

La frappe de ce polycopié a été soigneusement faite par Melle Z. AIT-MEKIDECHE qu'elle trouve dans ces lignes nos vifs remerciements.

Nous finirons en remerciant tous les enseignants qui ont contribué à notre formation.

S O M M A I R E .

DEUXIEME PARTIE

# I N T R O D U C T I O N

La parole est un signal extrêmement dense. Il vehicule à la fois l'information relative au contenu du message et celle relative au locuteur. Il contient des données sur l'accent, le rythme et l'intonation de ce dernier. L'analyse dont le but est d'extraire du signal vocal les paramètres de base (les formants, la fréquence fondamentale le spectre etc...) est donc necessaire afin de réduire son débit tout en conservant son intelligibilité.

Plusieurs méthodes analogiques et numériques d'analyse ont été décrites. Elles peuvent servir dans les domaines de reconnaissance, synthèse et transmission de la parole dont les buts consécutifs sont:

- Communiquer avec les machines en leur permettant de reconnaitre le langage parlé.

-" Faire parler les ordinateurs"c'est à dire créer de la parole synthétique.

-Permettre la communication parlée à distance .

La fréquence fondamentale "$F_o$" (appelée pitch) est égale à la fréquence d'ouverture et de fermeture de la glotte( espace circonscrit par les cordes vocales). Sa valeur est determinee par la pression sous glottique ainsi que par la masse et la tension des cordes vocales.

Le parametre "$F_o$" joue un rôle trés important car il permet de savoir si un son est voisé ou non. Dans la synthèse de la parole par exemple il contribue au naturel de la voix synthétique. Ainsi des méthodes de plus en plus perfectionnées sont proposées par les chercheurs dans le domaine de la parole. Le premier detecteur de pitch a été proposé par GRUTZMACHER et LOTTERMOSER (1937). Son principe reposait sur un circuit de base de t temps linéaire, commandé par des impulsions se répetant à la fréquence du fondamental.

L'une des techniques la plus repandue et la plus récente pour l'éstimation des parametres de base du signal de la parole est le codage préditif linéaire (L.P.C).L'algorithme de la technique simplifiée du filtre inverse,S.I.F.T) décrit par MARKEL (1972),est une méthode efficace et sûre d'extraction du "pitch", basée sur le principe du L.P.C. Elle fait appel au calcul de la fonction d'autocorrélation de la fonction d'erreur obtenue par filtrage inverse du signal de parole. Le filtre inverse permet d'éliminer la réponse du conduit vocal de ce dernier, en affaiblissant ses résonances. Le signal d'erreur présente ainsi des impulsions qui correspondent au fondamental. La mesure du temps séparant deux de ces impulsions successives donne la valeur de sa période dont l'inverse est égal au "pitch".

L'importance de cette technique réside dans sa capacité de fournir des estimations trés précises et sa relative rapiditéd'exécution des algorithmes( grâce aux progrés de la technologie des calculateurs).

Notre travail est composé de deux parties. La premiére est consacrée à l'étude théorique du sujet et la deuxième à l'élaboration des programmes.

Dans la premiére partie nous distinguons quatre chapitres Dans le premier chapitre nous décrivons l'appareil phonatoire humain ainsi que la classification des différents sons du langage. Puis suit un chapitre dans lequel les différentes méthodes d'analyse de la parole sont développées en détaillant la méthode du codage prédictif linéaire.

Les équations résultant de l'analyse prédictive et leurs solutions, sont exposées dans le chapitre trois. Enfin un dernier chapitre est consacré aux différentes méthodes de detection du "pitch" en insistant sur la technique simplifiée du filtre inverse (S.I.F.T ).

# P R E M I E R E  P A R T I E

# CHAPITRE PREMIER

# P R O D U C T I O N  D E  L A  P A R O L E

Dans ce chapitre, nous décrirons l'appareil phonatoire humain, le phénomène de production de la parole ainsi que les différents sons du langage.

Le signal de la parole est composé d'une suite de sons qui servent de support pour véhiculer l'information. Ces sons résultent des fluctuations rapides de la pression de l'air et leur arrangement est régi par les lois du langage.

## 1.1/ L'appareil phonatoire humain

Les principaux organes composant cet appareil sont: les poumons, les cordes vocales et le conduit vocal . (fig 1.1)

Nous allons examiner chacun de ces sous-ensembles en nous limitant aux points susceptibles de jouer un rôle dans la phonation.

### 1.1.1/ Les poumons

Les poumons jouent le rôle de generateur d'air sous pression constante. Il sont reliés à la source des sons"voisés" qui est le larynx par la trachée artère.

### 1.1 2/ Les cordes vocales

Les cordes vocales sont des muscles dont la longueur, la tension et l'épaisseur determinent le fondamental. L'espace circonscrit par celles-ci est appelé "glotte". Il est possible de rapprocher les cordes vocales les unes des autres, la fermant ainsi.

Pendant la respiration la glotte est ouverte, mais pour la phonation elle est fermée tout le long de la ligne médiane produisant ainsi les sons sonores. Quand la partie inférieure de la glotte est ouverte en laissant passer l'air, nous obtenons la voix chuchotée(FIG 1.2)

Initialement, les cordes vocales sont accolées entre elles, la pression sous-glottique tend à écarter celles-ci en les deformant vers le haut, brusquement elles finissent par s'écarter provoquant le passage de l'air.

Et grâce à leur élasticité et à la chute de pression, elles se referment à nouveau.

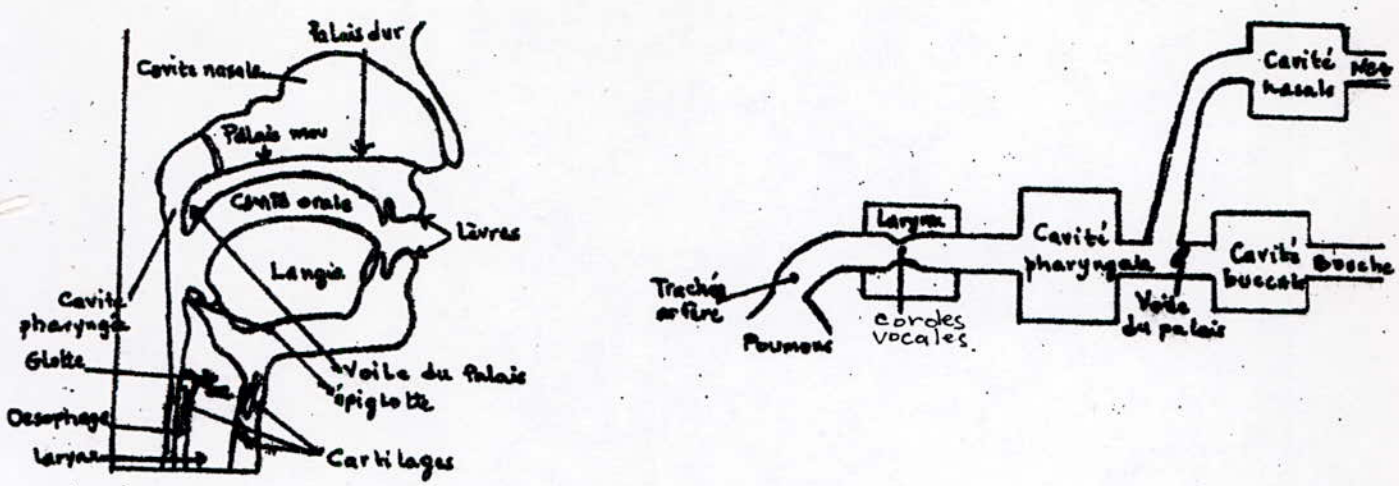
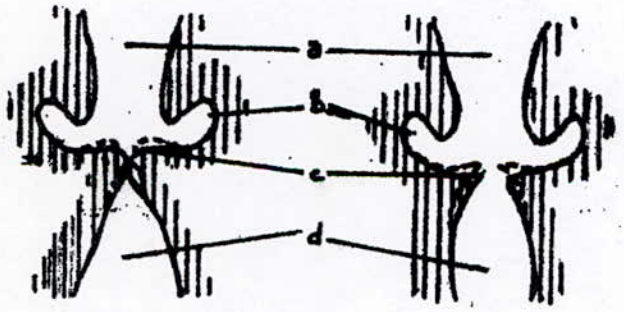



FIG 1-1

a- SYSTEME VOCAL HUMAIN

b- MODELE DE L'APPAREIL PHONATOIRE HUMAIN




a- Cavité Du Pharynx
b- Ventricules De Morgagni
c- Cordes Vocales
d- Cavité De La trachée Artère

FIG 1-2 : COUPE TRANSVERSALE DU LARYNX.

Le cycle recommence produisant ainsi un signal périodique dont la période est celle du fondamental (ou "pitch" en anglais) (fig 1.3)

1.1.3/ Le conduit vocal

Il est formé par le pharynx et les cavités orale et nasale

Le pharynx qui est la connexion entre l'oesophage et la bouche constitue avec la cavité buccale, la cavité "pharyngo-buccale".

Cette dernière commence au niveau de la glotte et se termine aux lèvres. Par contre, la cavité nasale débute au niveau du vélum et s'achève aux narines. Quand le voile du palais est abaissé, le conduit oral présente une dérivation qui est à l'origine des sons nasalisés.

Le conduit vocal joue le rôle d'un filtre. dont la courbe de réponse en fréquence présente des maxima appelés "FORMANTS" ( fig 1.4).

1.2/ Caractères sonore-sourd des sons

L'appareil vocal présente deux modes d'excitation selon la vibration des cordes vocales.

1.2.1/ Les sons sonores

L'excitation du système phonatoire est dûe à la mise en vibration des cordes vocales.

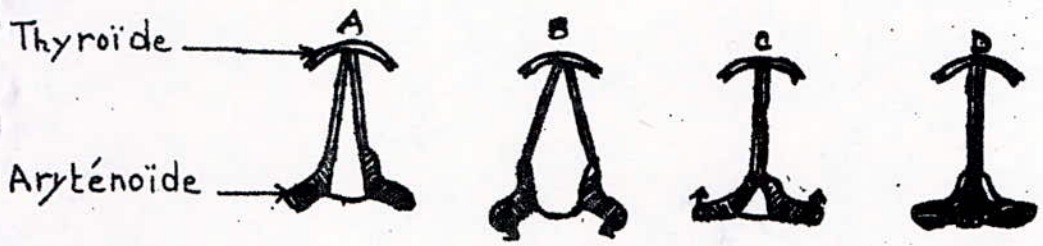
L'onde acoustique ainsi créee a une forme périodique (ou plutot quasi-périodique car il est rare que deux impulsions glottiques soient exactement identiques). La fréquence fondamentale "$F_o$" correspond ainsi à l'ouverture et à la fermeture périodique de la glotte.(fig 1.5-1.6)

1.2.2/ Les sons sourds

Dans ce cas l'excitation du système phonatoire est dûe à un passage turbulent d'air à travers une constriction située en un point du conduit vocal. Il s'agit d'un bruit: phénomène aléatoire et apériodique dont le spectre est relativement uniforme.(fig 1.7).

1.3/ Classification des différents sons

Les sons du langage peuvent être classés selon la vibration des cordes vocales(sonores-sourds), la position du voile du palais selon qu'il est abaissé ou non (oral,nasal) et le lieu d'articulation) (fig. 1.8).



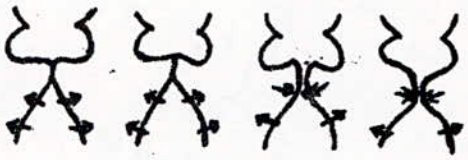

(a): Position de la glotte pendant:

A: La respiration normale.

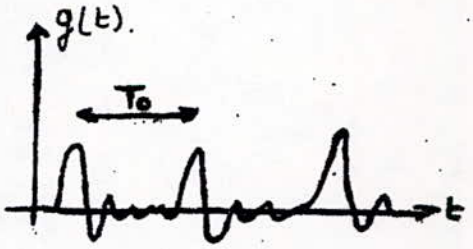
B: La respiration forte.

C: La voix chuchotée.

D: La phonation. (D'apres MALMBERG).



(b): Fonctionnement des cordes vocales.
(D'aprés LIENARD).




(c): Signal glottal filtré par le conduit vocal.

Fig.1.3: Formation du signal vocal.

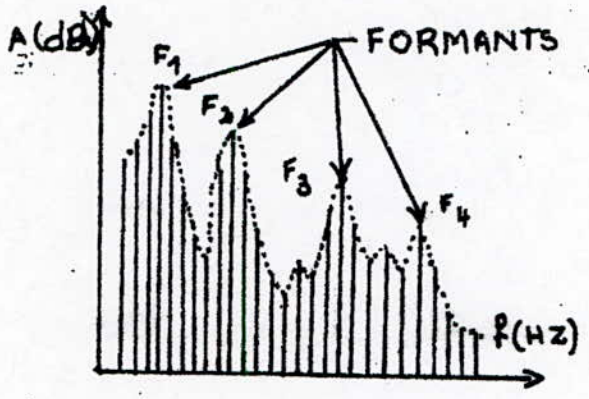



FIG 1-4 Spectre Vocalique Presentant Quatre Resonances Formantiques.

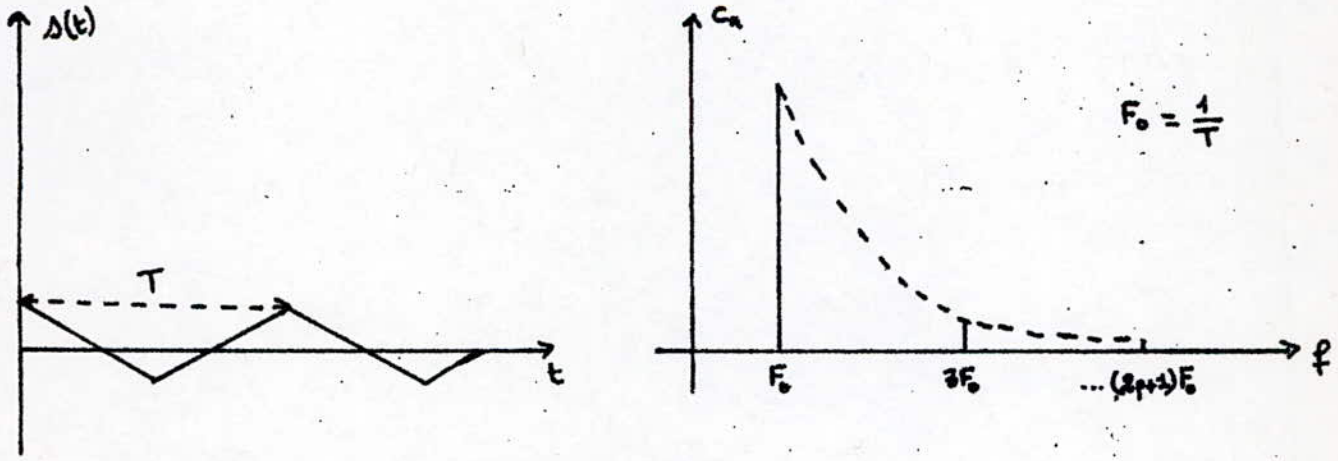



FIG 1-5

a) SIGNAL GLOTTIQUE b) SPECTRE DU SIGNAL GLOTTIQUE.

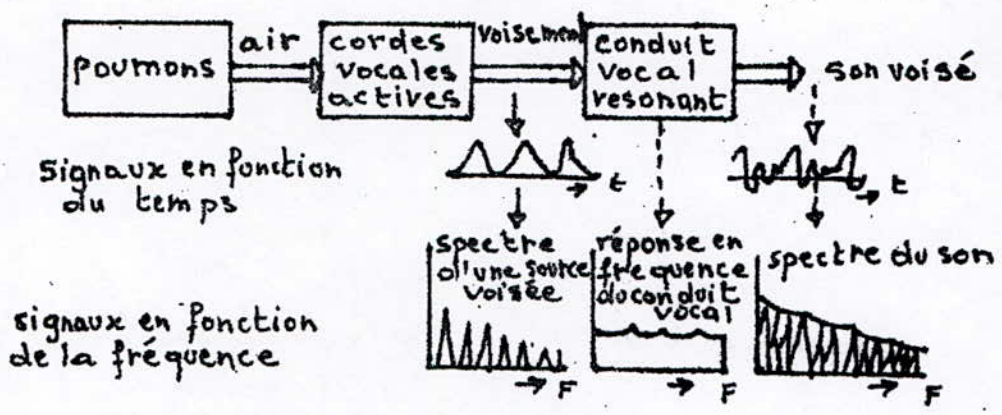



Fig. 1.6: Production d'un son voisé.

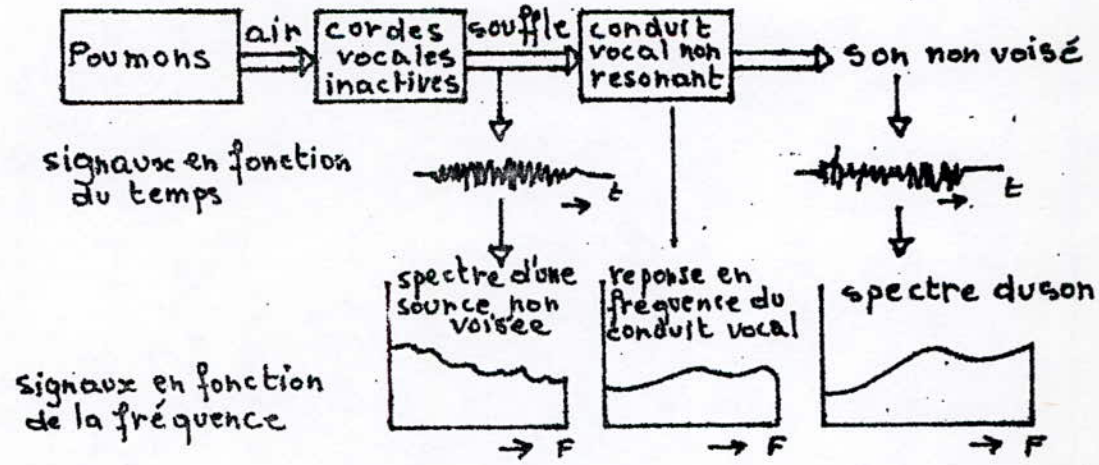



Fig. 1.7: Production d'un son non voisé.

| *Dénominations* | *Symbole phonétique* | *Mot-clé* | *Observations* |
|---|---|---|---|
| *Consonnes fricatives :* | [f] | fameux | |
| (constrictives) | [s] | saucisson | « sourdes », ou non-voisées |
| | [ʃ] | chat | |
| | [v] | vert | |
| | [z] | zèbre | « sonores », ou voisées |
| | [ʒ] | janvier | |
| *Consonnes nasales :* | [m] | menthe | |
| (occlusives nasales) | [n] | Nantes | |
| | [ɲ] | agneau | |
| *Consonnes liquides :* | [l] | salon | souvent appelée « latérale » |
| | [r] | bureau | nombreuses variantes en français ; les principales sont notées [r] , [R] , [ʁ] |
| *Consonnes plosives :* | [p] | pari | |
| (occlusives orales) | [t] | bateau | « sourdes », ou non-voisées |
| | [k] | écart | |
| | [b] | barbare | |
| | [d] | badaud | « sonores », ou voisées |
| | [g] | langue | |
| *Voyelles orales :* | [i] | lit | |
| | [e] | été | |
| | [ɛ] | marais | |
| | [y] | Ursule | |
| | [œ] | peur | son voisin de la voyelle neutre théorique [ə] |
| | [ə] | petit | son voisin de [œ], mais souvent plus court, ou à prononciation facultative. |
| | [ø] | jeu | |
| | [a] | patte | Cette distinction tend à disparaître au profit d'un A moyen |
| | [ɑ] | pâte | |
| | [ɔ] | sol | |
| | [o] | saule | |
| | [u] | bijou | |
| *Voyelles nasales :* | [ɛ̃] | brin | Cette distinction tend à disparaître, notamment à Paris au bénéfice du seul [ɛ̃] |
| | [œ̃] | brun | |
| | [ã] | chant | |
| | [ɔ̃] | bonjour | |
| *Semi-voyelles :* | [j] | paille | appelées quelquefois semi-consonnes |
| | [ɥ] | lui | |
| | [w] | Louis | |

**Figure 1.8 : Classification des sons du français**

**(d'après LIENARD)**

1.3.1/ Les voyelles

On appelle voyelles, des sons produits par le passage libre de l'air lorsque les cordes vocales vibrent. Cet air est modifié par les variations de forme de la cavité bucco-nasale.

On distingue d'aprés la voie d'échappement de l'air, les voyelles orales et nasales.

Il est possible de classer les voyelles , selon la fréquence moyenne de leurs deux premiers formants $F_1$ et $F_2$. P. DELATTRE a suivi cette procédure pour établir le triangle vocalique des voyelles françaises. (fig 1.9).

D'aprés cette classification, on distingue trois sortes de voyelles:

- Les voyelles compactes: Pour celles-ci les formants $F_1$ et $F_2$ sont groupés aux basses fréquences.

Exemple: [U], [O]

-Les voyelles médianes: Les formants $F_1$ et $F_2$ dans ce cas sont plus ou moins écartés.

Exemple: [a], [oe]

- Les voyelles diffuses: les deux formants $F_1$ et $F_2$ dans ce cas sont trés écartés l'un de l'autre, $F_2$ se trouvant dans les hautes fréquences. Exemple: [i], [E].

1.3.2/ Les consonnes

Les consonnes sont des sons produits lors d'une fermeture ou d'un rétrécissement du passage de l'air. Il existe trois sortes selon le mode d'excitation du système phonatoire.

a. Les fricatives

Le retrécissement du conduit oral entraîne l'émission d'un bruit. On est ainsi en présence des fricatives.

-Elles sont dites voisées, quand elles sont produites par l'association du bruit et de la vibration des cordes vocales.

Exemple : [Z], [V].

$F_1$ : Premier formant.

$F_2$ : Deuxieme formant.

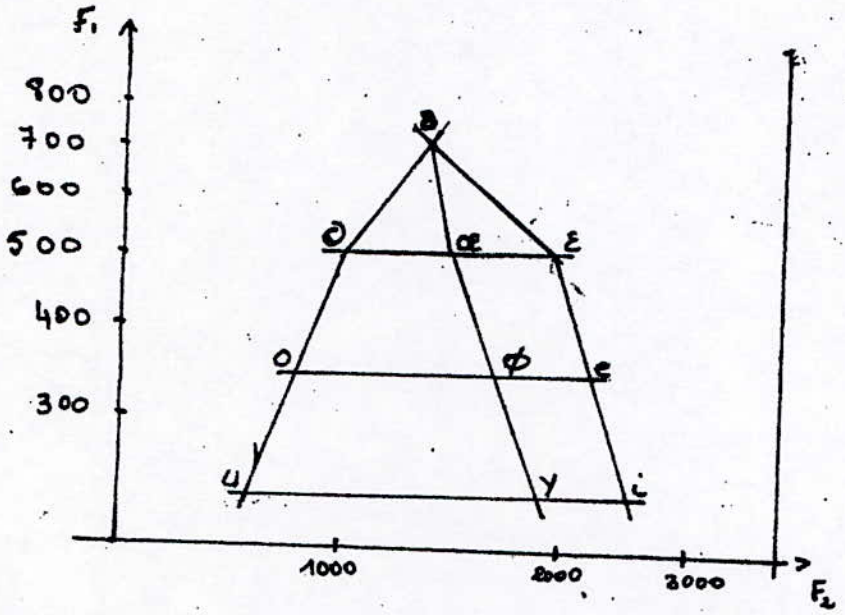



Fig. (1.9) : TRIANGLE VOCALIQUE DE P. DELATTRE (1948). D'APRES EMERIT

Par contre pour les non voisées, elles sont produites par du bruit en certains points de constriction du conduit vocal :

Exemple· [s], [ʃ], [j]

b/ Les occlusives

Une fermeture complète en un point particulier du conduit vocal suivie d'une ouverture brusque produit les occlusives. Celles-ci sont dites voisées quand il ya la contribution des cordes vocales.

exemple· [b], [d], [g].

Elles sont non voisées sinon.

Exemple· [p], [t], [k].

c/ Les nasales

Elles correspondent à une fermeture partielle à l'avant du conduit buccal, l'abaissement du voile du palais fait du conduit nasal la seule voie de sortie du son.

Exemple [m], [n].

d/ Les semi-voyelles et les liquides

Les transitions rapides du conduit vocal et sa continuité en fonctionnement en mode resonant font que les semi-voyelles sont apparentées aux consonnes et aux voyelles en même temps.

Exemple· [w] [y].

Quand d'autres phénomènes interviennent on a les liquides.

Exemple: [l], [r].

# CHAPITRE DEUX

# L'ANALYSE DE LA PAROLE

Dans ce chapitre nous citons les principales méthodes d'analyse de la parole et le principe de base du codage prédictif. Pour être analysé le signal vocal doit être échantillonné car le calculateur numérique a besoin d'un certain temps pour effectuer les opérations arithmétiques ou logiques préscrites par le programme. Il ne saurait traiter de façon continue l'information qu'il reçoit. Ainsi nous donnons un bref aperçu sur les operations d'échantionnage et de filtrage qui sont nécessaires dans l'analyse du signal de la parole.

## 2.1/ Echantillonnage

L'opération d'échantillonnage consiste à représenter un signal en fonction du temps S(t) par ses valeurs S(nT) à des instants multiples entiers d'unedurée T, appelée période d'echantillonnage.

Soit U(t) la distribution de masses unitaires aux points de l'axe réel, multiples entiers de la période T.(distributions de Dirac)

$$U(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT) \qquad (2.1)$$

Soit U(f) le spectre de U(t)

$$U(f) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} \delta\left(f - \frac{n}{T}\right) \qquad (2.2)$$

Il est constitué de raies d'amplitude $\frac{1}{T}$ aux fréquences qui sont des multiples entiers de la fréquence d'échantillonnage $F_e = \frac{1}{T}$

La suite des valeurs du signal S(nT) correspond au produit de l'ensemble des signaux élémentaires qui constituent U(t) par le signal S(t). L'opération d'échantillonnage affecte le spectre du signal échantillonné. Elle introduit une périodicité du spectre dans l'espace des fréquences. (fig 2.1).

$$S_e(f) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} S\left(f - \frac{n}{T}\right) \qquad (2.3)$$

Où $S_e(f)$ est le spectre du signal échantillonné et $\frac{1}{T}$ le spectre du signal S(f) le spectre du signal S(t).

Théorème d'échantillonnage ( théorème de SHANNON)

Si le signal continu S(t) dont le spectre de fréquence est borné et échantillonné à une fréquence $F_e$ au moins deux fois égale à la fréquence maximale $F_c$ de son spectre ( $F_e \geq 2F_c$ ), il n'y a pas perte d'information lors de l'opération d'échantiollonnage. Il est possible de reconstituer le signal S(t) sans deformations. (fig. 2.2).

## 2.2/ Filtres numériques

### 2.2.1/ Définition

Un filtre numérique F est un algorithme de calcul par lequel une séquence de nombres $\{x(n)\}$ dite séquence d'entrée est transformée en une séquence de nombres $\{y(n)\}$ dite séquence de sortie.

$$\{y(n)\} = F\{x(n)\} \qquad (2.4).$$

Son unité de calcul est munie des opérations suivantes:

- l'addition
- multiplication
- retard

Ils ont été développés et étudiés dans le but de pouvoir simuler les filtres analogiques sur ordinateurs.

Dans le cas des filtres linéaires la relation entre les séquences x(n) et y(n) est de la forme:

$$y(n) = \sum_{k}^{M} a_k\, y(n-k) + \sum_{k=0}^{M} b_k\, x(n-k) \qquad (2.5)$$

avec M un nombre entier quelconque

Les coefficients $a_k$ et $b_k$ vont servir à la construction du filtre numérique. Lorsque l'un au moins des coefficients $a_k$ est non nul, on obtient un filtre recursif.

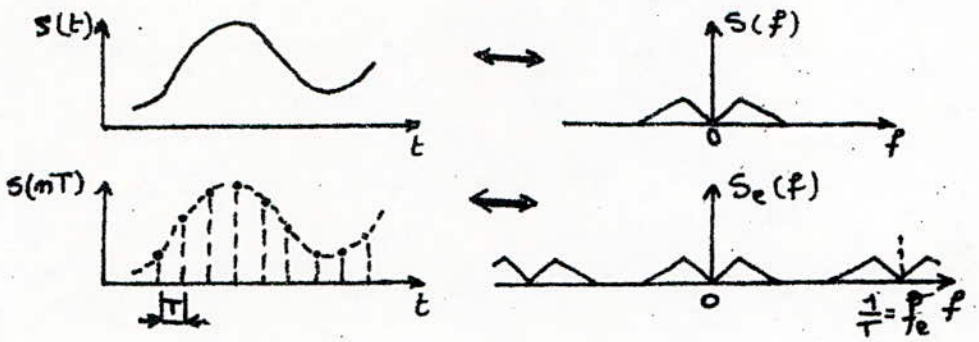
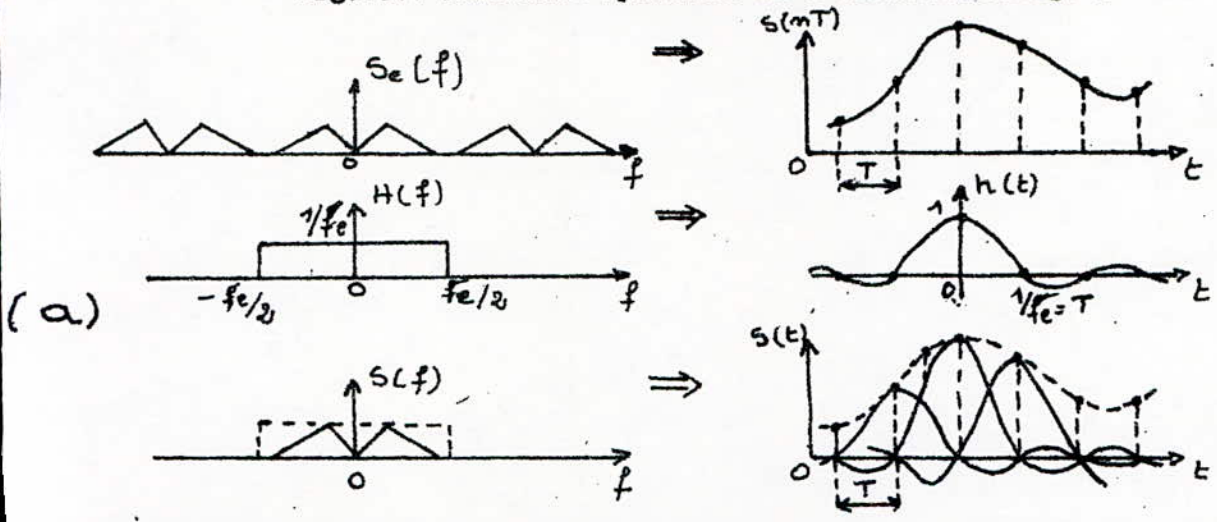



Fig.2.1: Incidence spectrale de l'échantillonnage .




Pour restituer le signal d'origine il faut supprimer la periodicité introduite par l'échantillonnage ,c'est à dire éliminer les bandes images ,operation qui peut être réalisée à l'aide d'un filtre passe bas dont la fonction de transfert H(f) vaut 1/Fe jusqu'à la fréquence Fe/2 et 0 aux fréquences supérieures. En sortie d'un tel filtre apparait un signal continu qu'il est possible d'exprimer en fonction des valeurs S(nT). Si le spectre d'origine contient des composantes aux fréquences supérieures ou égales à Fe/2 les bandes images chevauchent la bande de base. On dit qu'il ya repliement de bande.

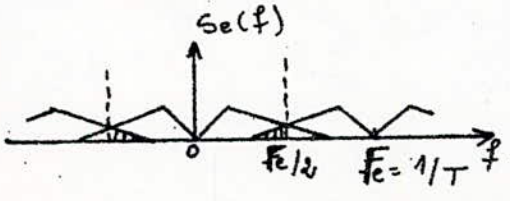



Fig. 2.2 : (a) Reconstitution du signal après échantillonnage
(b) Repliement de bande.

2.22/ Fonction de transfert d'un filtre numérique

Par définition on appelle fonction de transfert H(Z) d'un filtre numérique le rapport suivant:

$$H(Z) = \frac{Y(Z)}{X(Z)} = \frac{N(Z)}{D(Z)} \qquad (2.6)$$

avec Y(Z) et X(Z) les transformées en Z respectivement de y(n) et x(n).

On appelle zéros de la fonctions de transfert les racines de l'équation N(Z) =0 et pôles, les racines de l'équation D(Z) = 0

Les filtres non recursifs n'ont que des zéros ce qui assure leur stabilité, par contre les filtres récursifs ont des pôles et des zéros

L'analyse de la parole consiste à extraire du signal vocal un nombre réduit de parametres pertinents pepresentant les caractères de la parole. Le contenu du signal analysé est codé puis stocké en mémoire dans l'ordinateur afin qu'il puisse produire de la parole artificielle.

Parmi les méthodes d'analyse de la parole, on trouve essentiellement les méthodes classiques de traitement ( transformée de Fourier, corrélation)

L'application de ces méthodes souffre cependant de sérieuses limitations à cause du caractère non stationnaire et pseudo-périodique de la parole.

2.3 L'analyse spectrale

Le principe de l'analyse spectrale est basée sur les transformations de Fourier et le filtrage. Parmi ces méthodes nous pouvons citer:

2.3.1/ L'analyse par la synthèse

Cette méthode suppose la connaissance préalable du spectre réel du signal à analyser. Son principe est le suivant : on se donne un certain nombre de parametres caracterisant le conduit vocal,à partir de ces derniers on déduit un spectre qui est comparé au spectre réel du signal analysé.

La selection du modèle s'adaptant le mieux au spectre réel se fait à l'aide du critère des moindres carrés par exemple. Cette méthode présente l'inconvenient d'être trop longue car il faut recommencer à chaque fois le processus de comparaison. Si le modèle de prédiction est bien choisi, le processus doit converger et permettre d'atteindre une valeur minimale de l'erreur entre le signal original et le signal de synthèse.

2.3.2/ L'analyse cepstrale

Le principe de cette méthode est de réaliser une séparation entre la source d'éxcitation et la réponse du conduit vocal.

Soit un signal voisé dont le spectre en amplitude est S(f) son spectre en puissance $/S(f)/^2$ peut se mettre sous forme:

$$/S(f)/^2 = /G(f)/^2 \times /C(f)/^2 \qquad (2.7)$$

avec: G(f): spectre de la source glottale

C(f): réponse en fréquence du conduit vocal.

en prenant le logarithme de cette expression on obtient:

$$\log\left\{/S(f)/^2\right\} = \log\left\{/G(f)/^2 \times /C(f)/^2\right\}$$
$$= \log/G(f)/^2 + \log /C(f)/^2 \qquad (2.8)$$

prenons la transformée de Fourier des deux membres de l'égalité:

$$TF\left\{\log /S(f)/^2\right\} = TF\left\{\log/G(f)/^2\right\} + TF\left\{\log /C(f)/^2\right\} \qquad (2.9)$$

en elevant au carré TF $\log/S(f)/^2$ nous obtenons le spectre de puissance du logarithme du spectre de puissance. C'est le cepstre

L'axe horizontal du cepstre porte une grandeur qui possède la dimension d'un temps: on l'appelle "quéfrence".

L'avantage de cette méthode est qu'elle permet une bonne séparation source/ conduit vocal mais elle a l'inconvenient d'être trop longue car elle comporte deux transformées de Fourier.

L'analyse cepstrale est utilisée pour la détection du pitch.

2.4 L'analyse temporelle

C'est une technique qui permet d'analyser les aspects temporels du signal de parole. En effet certains évenements tels que la fermeture brusque du conduit vocal lors de la production d'un plosive sont mieux caractérisés par l'évolution temporelle du signal que par son spectre.

nous distinguons parmi ces méthodes:

2.4.1/ La méthode des passages par zéros du signal

Cette méthode permet la localisation des fréquences des premiers formants. Le signal de parole S(t) s'annule ou change de signe à des instants dont la répartition dans le temps est liée à certaines caractéristiques spectrales de S(t). L'information relative à l'amplitude du signal est perdue, puisque l'on ne s'interesse qu'à son signe.

Cette méthode ne permet donc pas la determination d'un modèle du conduit vocal. Son avantage réside dans le fait qu'elle est simple et trés rapide.

2.4.2/ La méthode d'autocorrélation

Une analyse dans le domaine temporel consiste à chercher les lois de periodicité du signal S(t). elle s'effectue par le calcul de la fonction d'autocorrelation:

$$g(\tau) = \int_{-\infty}^{+\infty} S(t)S(t-\tau)\,dt \qquad (2.10)$$

où l'on compare la valeur de la fonction à l'instant t, à la valeur de cette même fonction à l'instant " $t+\tau$ ". Le traitement de cette fonction par une transformée de Fourier conduit à la connaissance de la densité spectrale de S(t).

En effet on a:

$$g(\tau) = \int_{-\infty}^{+\infty} P(f)\, e^{+2\pi jf}\,df \qquad (2.11)$$

où:

$$P(f) = \int_{-\infty}^{+\infty} g(\tau)\, e^{-2\pi jf}\,d\tau \qquad (2.12)$$

P(f) étant la densité spectrale de puissance de S(t): $P(f) = /S(f)/^2$

La fonction d'autocorrélation peut être calculée à l'aide de procédés numériques à partir du signal échantillon soit en calculateurs, soit à l'aide de materiel spécialisé appelé "autocorrélateur".

EEle peut être aussi directement à partir du signal analogique S(t) en utilisant des lignes à retards, des modulateurs et des sommateurs.

Cette méthode est souple mais elle necessite le recours à des ordinateurs pour la réalisation de calculs qui sont souvent longs et qui occupent un espace mémoire trop grand.

2.5/ L'analyse prédictive

La méthode d'analyse par prédiction linéaire est considérée aussi bien comme une méthode temporelle que comme une méthode spectrale, de ce fait elle est exposée aux même limitations que les autres méthodes d'analyse.

Son choix est dû à l'utilisation d'un filtre numérique ne possédant que des pôles.Elle est ainsi fondée sur un modèle simple de production de la parole, constituant une bonne approximation du système phonatoire.

Le principe de cette méthode était connu depuis longtemps par les mathématiciens sous la forme de l'approximation d'une fonction par un polynôme.

Elle a été développée en 1966 par F.ITAKURA et S. SAITO, mais ce n'est qu'en 1972 que les concepts et algorithmes nécessaires furent détaillés par MARKEL.

2.5.1/ Principe du codage prédictif linéaire (L.P.C)

Le principe du codage prédictif linéaire est fondé sur l'hypothèse selon laquelle un échantillon du signal de parole S(nT) ou plus simplement $S_n$ (T est la période d'échantillonnage et n un nombre entier) est prédit approximativement par une somme pondérée linéairement, d'un certain nombre d'échantillons le précédant immédiatement.

Le signal prédit s'écrit:

$$\hat{S}_n = \sum_{k=1}^{p} \alpha_k S_{n-k} \qquad 1 \leq K \leq P \qquad (2.13)$$

p est l'ordre du prédicteur et $\alpha_k$ un ensemble de coefficients réels appelés coefficients du prédicteur.

Cette hypothèse est justifiée par le fait que physiologiquement la forme du conduit vocal, n'évolue pas rapidement (lors de la production des voyelles par exemple, le passage libre de l'air fait que le conduit vocal se déforme trés peu). Il en résulte que le spectre à court terme du signal de la parole évolue lentement. Ceci permet de considérer ce signal stationnaire sur des intervalles de temps de l'ordre de 10 à 25 ms et les parametres " $\alpha_k$" sont constants sur ces intervalles de temps . La non stationnarité globale se manifeste par le calcul des coefficients du prédicteur tous les 10 à 25 ms.

Ces coefficients vont servir à la construction d'un filtre linéaire ne possédant que des pôles et qui sert à modéliser le conduit vocal.

2.5.2/ Application: Modèle de production de la parole

Le signal de parole étant le résultat de l'action de filtrage du conduit vocal sur un signal de source( ce filtrage se traduit par les formants), il est tout à fait naturel de modéliser l'effet du conduit vocal par un filtre linéaire H(z) ne contenant que des pôles dont la fonction de transfert est de la forme:

$$H(Z) = \frac{G}{1 - \sum_{k=1}^{p} a_k Z^{-k}} \qquad 1 \leq K \leq P \qquad (2.14)$$

avec: G: gain du filtre

$a_k$: ensemble de coefficients réels

p: nombre de pôles (formants)

Remarque :

La nature particulière du filtre " H(Z)" n'est pas restrictive.

En effet, FANT (1960) et FLANAGAN (1965) ont démontré que dans les cas des sons sonores, la fonction de transfert du conduit vocal n'a que des pôles. Dans le cas des sons nasalisés, elle comprend de surcroît des zéros. Mais chaque zéro peut être remplacé par un ou plusieurs pôles permettant ainsi d'atteindre son effet. Le nombre de pôles dépend de la précision de representation requise.

Cette approximation n'a pas de conséquence sur le plan perceptif du fait que l'oreille est bien plus sensible à la localisation d'un maximum d'énergie dans l'echelle des fréquences(pôle) qu'à celle d'un minimum (zéro).

Suivant la nature du son émis, nous aurons le modèle de production de la figure (2.2)

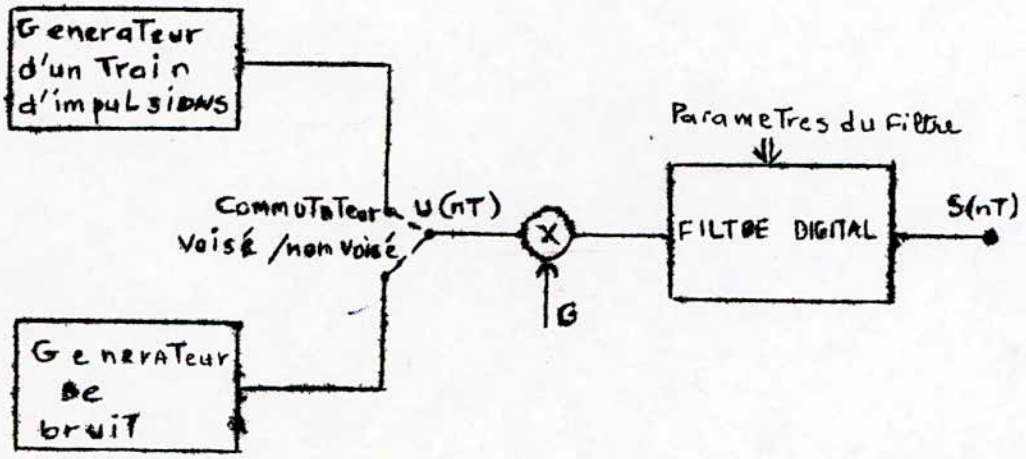



Fig 2.2: diagramme simplifié des blocs du modèle de production de la parole.

Il suffit donc d'appliquer à l'entrée de ce système une suite d'impulsions à périodes du fondamental, ou une séquence de bruit blanc pour obtenir à la sortie un signal équivalent au signal de la parole.

$$S(Z) = H(z) \cdot U(Z) \qquad (2.15)$$

U(Z) et S(Z) représentent respectivement les transformées en Z du signal d'excitation u(nT) et celui de sortie S(nT).

On a :

$$\sum_{n=0}^{\infty} S(nT) Z^{-n} = \frac{G}{1 - \sum_{K=1}^{P} a_K Z^{-K}} \sum_{n=0}^{\infty} U(nT) Z^{-n} \qquad (2.16)$$

ou encore :

$$\sum_{n=0}^{\infty} \left[ S(nT) - \sum_{K=1}^{P} a_K S((n-K)T) \right] Z^{-n} = G \sum_{n=0}^{\infty} U(nT) Z^{-n} \qquad (2.17)$$

d'où

$$S(nT) = \sum_{K=1}^{P} a_K S(n-K)T + G\, U(nT) \qquad (2.18)$$

Or, nous avons vu précédemment que le signal prédit de la parole se met sous la forme (2.13).

L'erreur de prédiction étant définie par:

$$e_n = S_n - \hat{S}_n = S_n - \sum_{K=1}^{P} \alpha_K S_{n-K} \qquad (2.19)$$

En lui appliquant la transformée en Z nous obtenons:

$$E(Z) = S(Z) \left[ 1 - \sum_{K=1}^{P} \alpha_K Z^{-K} \right] \qquad (2.20)$$

ou $E(Z) = S(Z).D(Z)$

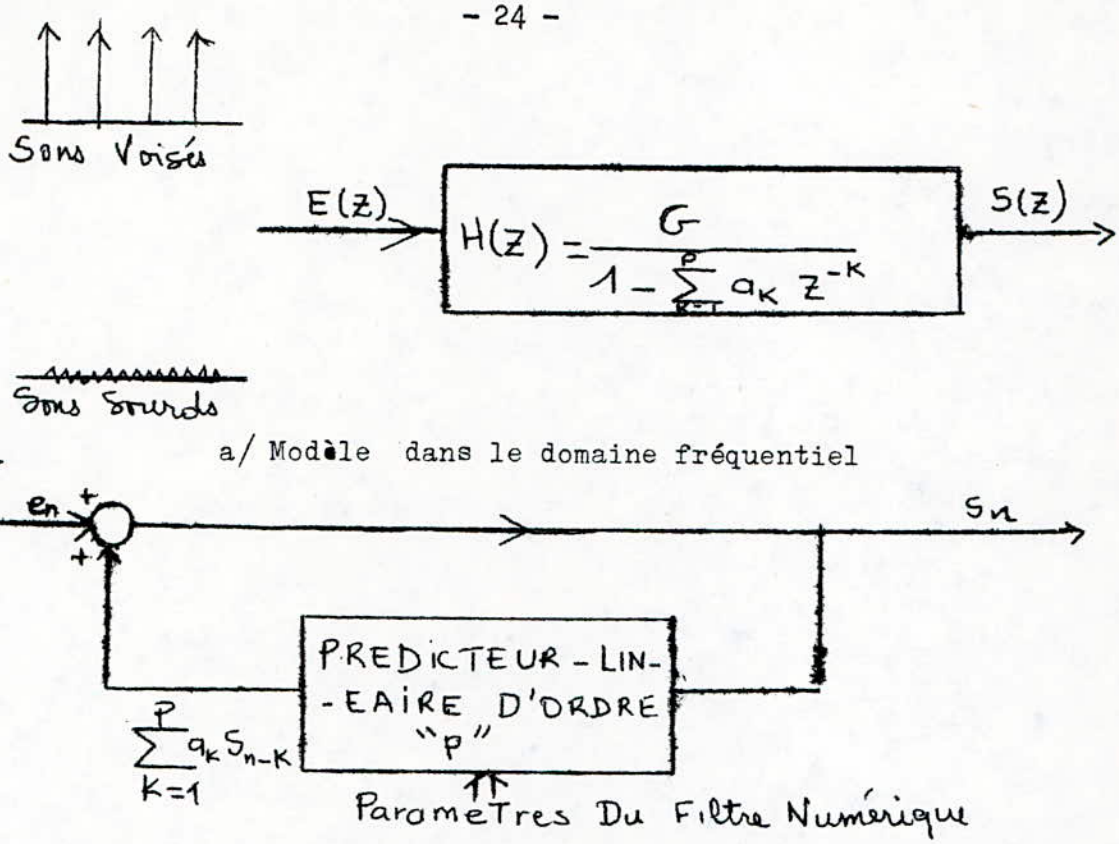
avec: $D(Z) = 1 - \sum_{K=1}^{P} \alpha_K Z^{-K} \qquad 1 \leqslant K \leqslant P \quad (2.21)$

où E(Z) et S(Z) sont respectivement les transformées en Z de "$e_n$" et "$S_n$". En comparant (2.13) et (2.18), nous remarquons que si le signal de parole obéit au modèle de la figure (22) et si $\alpha_K = a_k$ alors: $e_n = GU_n \qquad (2.22)$

Dans ce cas le filtre de l'erreur de prédiction défini dans l'équation (2.20) sera un filtre inverse pour le système de production de la parole défini dans l'équation (2.14) d'où:

$$H(Z) = \frac{G}{D(Z)} = \frac{G}{1 - \sum_{K=1}^{P} a_K Z^{-K}} \qquad (2.23)$$

La figure (2.3) représente le modèle de production de la parole dans les domaines temporel et fréquentiel.




a/ Modèle dans le domaine fréquentiel

b/ Modèle dans le domaine temporel

Fig 2.3 : Modèle de la production de la parole.

Le modèle de production de la parole de la figure (2.3) dispose d'un commutateur qui selectionne le mode d'excitation. Deux générateurs font fonction de source d'excitation, l'un emettant des impulsions d'amplitude et depériodicité variable pour les sons sonores: l'autre un bruit blanc pour les sons sourds.

Le filtre récursif se décompose endeux opérations:

- Une estimation " $\hat{S}_n$", obtenue à l'aide d'un filtre linéaire d'estimation à P" retards.
- Une addition ( $\hat{S}_n + G\,U_n$) où $G\,U_n$ est la valeur de la sourced'excitation à l'instant " nT".

La production de la parole suppose que l'on alimente le synthétiseur toutes les 10ms à 25ms par un ensemble de données:

- les parametres "$a_k$" du filtre ou du prédicteur, au nombre de P.
- Le facteur de gain G.
- La période du fondamental.
- L'indicateur de voisement ou non qui commande le commutateur.

2.5.3/ Calcul des coefficients $a_k$

Le problème de base de l'analyse prédictive est la détermination des coefficients "$a_k$" de telle sorte que la relation (2.13) soit optimale.

Le critèere d'optimisation utilisé est arbitraire. Le but est de minimiser l'energie de l'erreur de prédiction dans un intervalle quelconque . Le critère des moindres carrés est le plus utilisé car il conduit à un système d'équations linéaires faciles à résoudre théoriquement.

L'interet de cette approche provient du fait que si le signal est généré par l'équation (2.18) avec des coefficients constants dans le temps et excité par des impulsion uniques aussi bien que par un bruit stationnaire, alors on pourrait voir que les coefficients prédicteurs qui résultent de la minimisation de l'erreur quadratique de la prédiction sont identiques aux coefficients de (2.13) ( $\alpha_k = a_k$ $\quad 1 \leq k \leq p$)

Soit E l'erreur quadratique totale:

$$E = \sum_n e_n^2 = \sum_n \left(s_n - \sum_{k=1}^{p} a_k s_{n-k}\right)^2 \qquad (2.24)$$

Pour que E soit minimale, il faut que pour chaque $a_k$ ( k=1,...p) la dérivée partielle correspondante de E soit nulle.

$$\frac{\partial E}{\partial a_k} = 0 \qquad 1 \leq k \leq p \qquad (2.29)$$

d'où l'obtention du système d'équations:

$$\sum_n s_n s_{n-i} = \sum_n \sum_{k=1}^{p} a_k s_{n-k} s_{n-i} \qquad (2.26)$$

Les coefficients "$a_k$" sont les "p" inconnus de ces "p" équations linéaires (2.14) appelés "équations normales".

Plusieurs méthodes de résolutions du système (2.26) sont disponibles parmi lesquelles. on trouve la méthode de covariance et d'autocorrélation.

2.5.4/ Determination du nombre de coefficients "$a_k$"

Nous complétons notre précedente étude sur le modèle de l'appareil vocal par un aspect du filtre récursif. Il s'agit du nombre "p" de coefficients $a_k$ (qui est aussi égal au nombre de pôles du filtre modèle) nécessaire pour avoir une approximation valable.

En admettant que le spectre du signal vocal puisse être décrit à l'aide de cinq formants (trois formants suffisent pour décrire le spectre du signal vocal, les deux autres contibuent à caractériser la voix) ceux-ci seront situés dans une bande de 5KHz. En tenant compte du théorème d'échantillonnage de SHANNON ($F_e \geq 2F_c$) la fréquence d'échantillonnage du signal de parole sera: $F_e$ = 10KHz ($T_e$=0,1ms).

La vitesse du son étant de 340m/s, et la longueur moyenne du conduit vocal est égal à 17cm, alors le temps mis par l'onde sonore pour se propager depuis la glotte jusqu'aux lèvres est de 0,5ms. En tenant compte de la reflexion et de la transmission du son on doit évaluer le temps d'un aller-retour du son, soit T=1ms ,ce temps est équivalent au nombre d'échantillons contenus dans la mémoire du prédicteur.

Et comme la période d'échantillonnage est $T_e$= 0,1ms la valeur correspondante de p ($p = \frac{T}{T_e}$) sera égale à 10. La contribution de la glotte et des lèvres étant équivalente à une paire de pôles réels cela porte le nombre de coefficients à " 12".

# CHAPITRE TROIS

## EQUATIONS DU CODAGE PREDICTIF LINEAIRE ET LEUR SOLUTION

### 3.1/ Les méthodes essentielles d'analyse par prédiction linéaire.

Nous avons vu dans le chapitre précédent que les coefficients " "$a_k$" du prédicteur sont obtenus par la résolution d'un système de "p" équations linéaires à "p" inconnues (équation 2.14)

$$\sum_{n} \sum_{K=1}^{p} a_K S_{n-K} S_{n-i} = \sum_{n} S_n S_{n-i} \qquad \begin{matrix} 1 \leq i \leq p \\ 1 \leq K \leq p \end{matrix}$$

Si nous avons défini: $C(i,k) = \sum_{n} S_{n-K} S_{n-i}$ (3.1)

alors l'équation (2.14) peut s'écrire d'une manière plus compacte:

$$\sum_{K=1}^{p} a_K C(i,k) = C(i,0) \qquad 1 \leq i \leq p \quad (3.2)$$

Pour chercher les coefficients prédicteurs optimaux nous devons en premier lieu calculer les quantités: C(i,k) avec $0 \leq k \leq p$ et $1 \leq i \leq p$. Puis il faut résoudre l'équation (3.2) pour obtenir les coefficients "$a_k$". Jusqu'ici il n'a pas été indiqué explicitement les bornes de sommation dans (2.14),(3.1)et (3.2). Nous allons voir deux méthodes de l'analyse par prédiction linéaire qui définiront ces bornes.

Nous citons au passage une troisième méthode qui n'est pas appliquée au signal de la parole car l'intervalle de prédiction est limité par l'ordre des équations linéaires à résoudre, il s'agit de la méthode exacte.

#### 3.1.1/La méthode exacte

Les hypothèses sur lesquelles est fondée la méthode exacte sont les suivantes:

a- le signal est défini exactement pour"2p" échantillons consécutifs.

b- un échantillon de signal de parole peut être prédit exactement à partir des "p" échantillons précédents.

c- l'hypothèse (b) est valable pour les"p"échantillons consécutifs qui suivent:

L'ensemble de ces hypothèses est représenté par les équations suivantes:

$$\sum_{K=1}^{p} a_K S_{n-K} = S_n \qquad n = p, p+1, \ldots 2p-1 \quad (3.3).$$

La méthode exacte suppose que l'erreur e(nT) est identiquement nulle à chaque instant. Cela entraîne que dans le modèle de production de la parole u(nT) est nulle pour tout n. Donc cette méthode implique qu'il n'y a pas d'impulsions du fondamental pendant l'intervalle de temps correspondant aux "2p" échantillons de parole nécéssaires à l'analyse. Par conséquent l'utlisation de cette méthode ne s'applique pas à notre modèle.

3.1.2/La méthode de covariance

Elle suppose le signal non stationnaire à l'interieur de l'intervalle d'analyse, elle tient compte de la variation spectrale dûe à un décalage du signal à l'interieur de cet intervalle. Elle fait les suppositions suivantes:

a- le signal est défini pour "N+p" échantillons consécutifs avec N entier.
b- un échantillon du signal de parole peut être prédit à l'aide des "p" échantillons précédents.
c- l'hypothèse (b) est valable pour les N échantillons consécutifs.
d- l'écart quadratique total entre le signal original et sa valeur prédite est minimisé pour l'ensemble des N échantillons consécutifs.

On en déduit les formulations suivantes:

$$E_n = \sum_{n=p}^{N-1} e_n^2 \qquad (3.4)$$

alors C(i,k) devient:

$$C(i,k) = \sum_{n=p}^{N-1} S_{n-i} S_{n-k} \qquad \begin{matrix} 0 \leq K \leq p \\ 1 \leq i \leq p \end{matrix} \qquad (3.5)$$

Si nous changeons l'indice de sommation, nous pouvons exprimer C(i,k) comme:

$$C(i,k) = \sum_{n=p-i}^{N-1-i} S_n S_{n+i-k} . \qquad (3.6)$$

Cette approche est similaire à celle fondée sur la fonction d'autocorrélation. Elle fournit une fonction qui n'est pas une fonction d'autocorrélation réelle, mais plutôt une comparaison entre deux segments de signaux de parole de longueur finie, similaires mais non identiques.

Le système d'équation à résoudre s'écrit sous la forme;

$$\sum_{K=1}^{P} a_K C(i,K) = C(i,0) \qquad \begin{matrix} 1 \leq i \leq P \\ 0 \leq K \leq P \end{matrix} \qquad (3.7)$$

La forme matricielle de cette équation est la suivante:

$$\begin{bmatrix} c(1,1) & c(1,2) & \ldots . c(1,p) \\ c(2,1) & c(2,2) & \ldots . . c(2,p) \\ & & \\ c(p,1) & c(p,2) & \ldots . . c(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \\ a_p \end{bmatrix} = \begin{bmatrix} c(1,0) \\ c(2,0) \\ \\ c(p,0) \end{bmatrix} \qquad (3.8)$$

comme $c(i,k) = c(k,i)$ la matrice pxp de covariance C est symétrique. Cette équation matricielle peut être résolue efficacement avec l'algorithme de Cholesky.

3.1.3 La méthode d'autocorrelation

Cette méthode considère le signal stationnaire dans un intervalle de temps. Elle suppose que le spectre à court terme du signal est invariant dans la trame considérée. Ceci est réalisé par un fenêtrage temporel préalable du signal. Les suppositions de la méthode d'autocorrelation sont les suivantes:

a- le signal est nul à l'exterieur de l'intervalle $0 \leq n \leq N-1$

Cette condition est réalisée par une fenêtre (cf l'annexe)

Soit $S(n)$ un échantillon du signal de parole et $W(n)$ une fenêtre de longueur finie, identiquement nulle à l'exterieur de l'intervalle $0 \leq n \leq N-1$. Le signal $S_N(n)$ "vu" par la fenêtre s'écrit/:

$$S_N(n) = S(n) \cdot W(n) \qquad (3.8)$$

Le choix de la fenêtre est compliqué. Il depend du son à analyser.

Ainsi pour les sons voisés il est nécéssaire d'examiner l'influence de la largeur et de la position de la fenêtre d'analyse. L'utilisation des fenêtres de Hamming ou Hanning est souvent satisfaisante (fig 3.1)

b- chaque échantillon numérique peut être prédit par ses "p" échantillons précédents et ceci pour tout le temps ($n \in ]-\infty, +\infty[$)

Remarque:

Si $S_n$ est different de zéro seulement pour $0 \leqslant n \leqslant N-1$, alors l'erreur de prédiction correspondante e(n), pour le prédicteur d'ordre "p", sera différente de zéro dans l'intervalle $0 \leqslant n \leqslant N+p-1$ .

Ainsi pour ce cas "$E_n$" est exprimé comme suit :

$$E_n = \sum_{n=-\infty}^{+\infty} e_n^2 = \sum_{n=0}^{N+p-1} e_n^2 \qquad (3.9)$$

Les bornes dans l'expression de C(i,k) dans (3.1) sont identiques à celles de l'équation (3.9) on a donc:

$$C(i,k) = \sum_{n=0}^{N+p-1} S_{n-i} S_{n-k} \qquad \begin{matrix} 0 \leqslant k \leqslant p \\ 1 \leqslant i \leqslant p \end{matrix} \qquad (3.10)$$

Ceci peut être exprimé comme suit:

$$C(i,k) = \sum_{n=0}^{N-1-(i-k)} S_n S_{n+i-k} \qquad (3.11)$$

Nous remarquons que C(i,k) est identique à la fonction d'auto-correlation définie pour les intervalles de temps trés courts évalué pour (i-k). C'est à dire:

$$C(i,k) = R(i-k) \qquad (3.12)$$

où $R(l) = \sum_{n=0}^{N-1-l} S_n S_{n+l}$

puisque R(l) est une fonction paire $R(-l) = R(l)$ il vient alors:

$$R(l) = \sum_{n=0}^{N-1-|l|} S_n S_{n+|l|} \qquad (3.13)$$

et $C(i,k) = R(|i-k|) \qquad 1 \leqslant i \leqslant p \quad ; \quad 0 \leqslant k \leqslant p \qquad (3.14)$

Par conséquent l'équation (3.2) s'écrit:

$$R(i) = \sum_{k=1}^{p} a_k R(|i-k|) \qquad 1 \leqslant i \leqslant p \; ; \; 0 \leqslant k \leqslant p \qquad (3.15)$$

* Fenêtre de HAMMING

$$W(m)=\begin{cases}0,54 + 0,46 \cos (2\pi m)/(N-1) & 0 \leq m \leq N-1 \\ 0 & \text{Partout ailleurs}\end{cases}$$

* Fenêtre de HANNING

$$W(m)=\begin{cases}0,5 + 0,5 \cos (2\pi m)/(N-1) & 0 \leq m \leq N-1 \\ 0 & \text{Partout ailleurs}\end{cases}$$

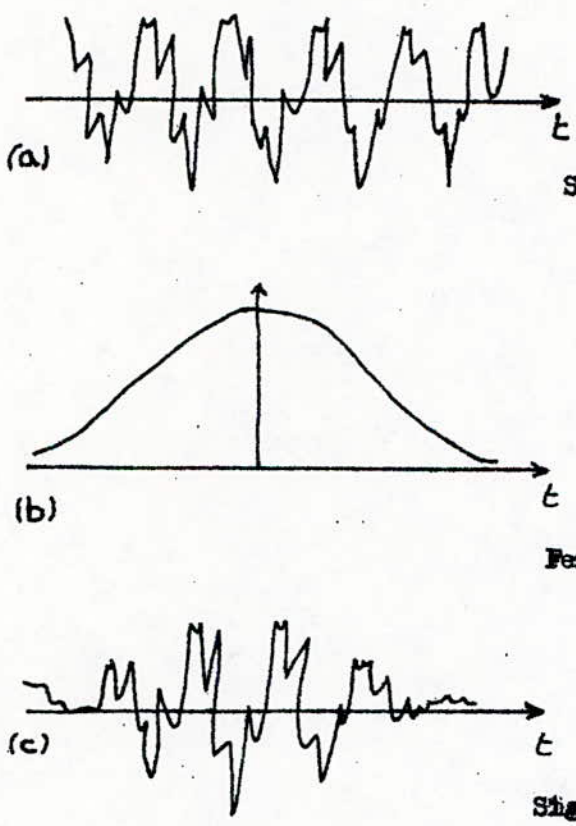



Fig.3.1:Effet de la fenêtre de HAMMING sur le signal

L'ensemble des équations données par la relation (3.15) peut être exprimé dans une forme matricielle comme suit:

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(p-1) \\ R(1) & R(0) & \cdots & R(p-2) \\ R(2) & R(1) & \cdots & R(p-3) \\ & & & \\ R(p-1) & R(p-2) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \\ R(p) \end{bmatrix} \qquad (3.16)$$

La matrice carrée "R p " des valeurs d'autocorrelation est symétrique et définie positive. De plus les éléments situés de part et d'autre de la diagonale sont égaux. Il suffit alors de "p" valeurs d'autocorrelation pour sa définition complète. C'est une matrice de "Toeplitz" et l'algorithme de DURBIN permet de résoudrecette équation matricielle d'une façon réccurente.

3.1.4/Erreur quadratique totale minimale

Nous avons:

$$E_n = \sum_n e_n^2 = \sum_n \left(S_n - \sum_{K=1}^{p} a_K S_{n-K}\right)^2$$

$$E_n = \sum_n \left(S_n^2 - 2 S_n \sum_{K=1}^{p} a_K S_{n-K} + \sum_{K=1}^{p}\sum_{i=1}^{p} a_K a_i S_{n-K} S_{n-i}\right)$$

$$E_n = \sum_n S_n^2 - 2\sum_{K=1}^{p} a_K \sum_n S_n S_{n-K} + \sum_{K=1}^{p} a_K \sum_{i=1}^{p} a_i \sum_n S_{n-K} S_{n-i}$$

Or d'aprés l'équation (2.14) nous avons la relation:

$$\sum_{K=1}^{p} a_K \sum_n S_{n-K} S_{n-i} = \sum_n S_n S_{n-i}$$

de même:

$$\sum_{i=1}^{p} a_i \sum_n S_{n-i} S_{n-K} = \sum_n S_n S_{n-K}$$

En remplaçant cesdeux relations dans l'expression de "$E_n$" on a :

$$E_p = \sum_n S_n^2 - \sum_{K=1}^{p} a_K \sum_n S_n S_{n-K} \qquad (3.17)$$

a/Méthode de covariance

Dans la méthode de covariance nous avions les relations suivantes:

$$C(i,k) = \sum_{n=p}^{N-1} S_{n-i}\, S_{n-k}$$

$$C(0,0) = \sum_{n=p}^{N-1} S_n^2$$

$$C(0,k) = \sum_{n=p}^{N-1} S_n\, S_{n-K}$$

En remplaçant ces équations dans l'expression de "$E_p$" nous obtenons l'erreur quadratique totale minimale qui s'écrit:

$$E_p = C(0,0) - \sum_{k=1}^{p} a_k\, C(0,k) \qquad (3.18)$$

b-Méthode d'autocorrelation

Dans ce cas nous avons la relation suivante:

$$R(k) = \sum_{n=0}^{N-1-|k|} S_n\, S_{n+/k/}$$

d'où $R(0) = \sum_{n=0}^{N-1} S_n^2$

En remplaçant ces expressions dans celles de $E_p$ nous aurons:

$$E_p = R(0) - \sum_{k=1}^{p} a_k\, R(k) \qquad (3.19)$$

c- Erreur normalisée

L'erreur normalisée $V_p$ est définie comme suit:

$$V_p = \frac{E_p}{R_o} \qquad (3.20)$$

$$\text{d'où } V_p = 1 - \sum_{k=1}^{p} a_k\, r(k) \qquad (3.21)$$

avec $r(k) = \dfrac{R(k)}{R_o}$ $\forall k$

L'erreur normalisée $V_p$ est définie comme étant le rapport de l'energie representée par les échantillons de l'erreur de prédiction à l'energie representée par les échantillons du signal de parole pour la sequence considérée.

Les coefficients"$r_k$" sont appelés coefficients normalisés de la fonction d'autocorrelation.

3.1.5/ Calcul du facteur de gain G

La determination du facteur de gain G de la fonction de transfert H(Z) repose sur le critère suivant: l'energie totale continue dans la séquence de signal de synthèse doit être égale à l'energie totale de la sequence correspondante du signal analysé.

En supposant que le signal est analysé sur une période du fondamental, à la synthèse il n'y aura qu'une seule impulsion d'excitation du filtre pendant l'intervalle de temps correspondant. Par conséquent le critère précédent peut être reformulé comme suit: l'energie totale contenue dans la réponse impulsionnelle de H(Z) doit être égale à l'energie totale du signal analysé.

Le modèle du signal est donné par:

$$\hat{S}_n = \sum_{K=1}^{P} a_K \hat{S}_{n-K} + G\,\delta_n \qquad (3.22)$$

$$\delta_n = \begin{cases} 1 & n=0 \\ 0 & n\neq 0 \end{cases}$$

$\delta_n$ est une impulsion de Dirac

D'aprés ces deux équations nous pouvons écrire:

$\hat{S}_n = 0$ ; pour $n<0$

$\hat{S}_o = G$ ; pour $n=0$

et $\hat{S}_n = \sum_{K=1}^{P} a_K \hat{S}_{n-K}$ ; pour $n>0$

Par définition l'autocorrelation du signal prédit est:

$$\hat{R}_i = \sum_{n=-\infty}^{+\infty} \hat{S}_n \hat{S}_{n+i} \qquad \forall\, i \qquad (3.23)$$

pour i =0

$$\hat{R}_o = \sum_{n=0}^{\infty} \hat{S}_n^2 = \hat{S}_o^2 + \sum_{n=1}^{\infty} \hat{S}_n \sum_{k=1}^{P} a_k \hat{S}_{n-k} \qquad (3.24)$$

$$\hat{R}_o = G^2 + \sum_{K=1}^{P} a_k \left( \sum_{n=1-K}^{\infty} \hat{S}_m \hat{S}_{m+k} \right) \qquad m= n-k \qquad (3.25)$$

Du fait que $\hat{S}_m = 0$ pour $m<0$ il vient:

$$\hat{R}_o = G^2 + \sum_{K=1}^{P} a_k \sum_{m=0}^{\infty} \hat{S}_m \hat{S}_{m+k} \qquad (3.26)$$

$$\hat{R}_o = G^2 + \sum_{k=1}^{P} a_k \hat{R}_k \quad \text{d'où} \quad G^2 = \hat{R}_o - \sum_{k=1}^{P} a_k \hat{R}_k \qquad (3.27)$$

Et pour $i > 1$

$$\hat{R}_i = \sum_{n=0}^{\infty} \hat{S}_n \hat{s}_{n+i} \qquad (3.28)$$

$$\hat{R}_i = \sum_{n=0}^{\infty} \hat{S}_n \sum_{k=1}^{p} a_k \hat{S}_{n+i-k} = \sum_{k=1}^{p} a_k \sum_{n=0}^{\infty} \hat{S}_n \hat{S}_{n+i-k} \qquad (3.29)$$

$$\hat{R}_i = \sum_{k=1}^{p} a_k \hat{R}_{|i-k|} \qquad 1 < i < \infty \qquad (3.30)$$

L'équation (3.15) est identique à ( 3.30) pour $1 \leqslant k \leqslant p$.

La fonction d'autocorrelation obeit par conséquent à la même équation matricielle de la minimisation de l'erreur quadratique de prédiction à partir de l'autocorrelation du signal $R_i$. De ce fait les fonctions d'autocorrelation $\hat{R}_i$ et $R_i$ doivent satisfaire l'équation suivante: $\hat{R}i = cR_i \qquad o \leqslant i \leqslant p \qquad (3.31)$

C constante à determiner

Afin de conserver l'égalité de l'energie de $S_n$ et $\hat{S}_n$ il faut que $\hat{R}_o = R_o$ et par conséquent $C=1$ d'où $\hat{R}_i = R_i \qquad 0 \leqslant i \leqslant p \qquad (3.32)$

On peut calculer le gain G de telle sorte que $\hat{R}_o$ soit égale à $R_o$ en utilisant l'erreur quadratique minimale.

D'aprés l'équation (3.19) $E_p = R_o - \sum_{k=1}^{p} a_k R_k$

Plus le signal est "prédictible" plus cette difference est minimale. Ainsi pour que $\hat{R}_o$ soit égale à $R_o$, nous avons :

$$E_p = G^2 \qquad (3.33)$$

En fonction de l'erreur normalisée $V_p$ nous avons :

$V_p = \frac{\epsilon_p}{R_o}$ d'où $G^2 = V_p R_o \qquad (3.34)$

3.1.6/ Stabilité du prédicteur

Les méthodes décrites precedemment ont pour but de determiner les valeurs optimales des parametres $a_k$ de la fonction de transfert H(Z) ne possedant que des pôles. Or il est important de savoir si le filtre dont la fonction de transfert calculée est H(Z), est stable ou non .

Par définition H(Z) est stable si tous les pôles sont situés à l'interieur du cercle unité du plan Z. ces pôles sont les racines du polynôme du dénominateur de cette fonction de transfert c'est à dire:

$$1-\sum_{K\geq 1}^{p} a_k\, Z^{-k} = 0$$

Les instabilités sont généralement le fait du premier formant oui a une largeur de bande faible et dont l'energie est la plus élevée.

Compte tenu du fait que les méthodes de prédiction linéaire tendent à accentuer les pointes spectrales du signal analysé, le pôle representant le premier formant se situera dans le voisinage immédiat du cercle unité du plan Z et la moindre erreur de calcul peut suffire pour le deplacer hors du cercle unité. D'autant plus pour les voix de femmes pour lesquelles le premier formant se confond avec le fondamental.

ATAL et HANAVER (1971) ont décrit un algorithme qui permet de detecter l'instabilité. Dans ce cas ils proposent le calcul des racines, la detection de(s) pôle(s) de module supérieur à l'unité, et la division de ce(s) pôle(s) par son (leur) module(s). Cette solution est assez compliquée car elle fait appel à un calcul de zéros complexes d'un polynôme.

Il existe une solution plus simple oui consiste en une detection puis correction d'instabilité. Cette correction consiste à "augmenter le rayon" du cercle unité pour englober les pôles se situant à l'exterieur. Elle est réalisée de la manière suivante:

$$a'_k = a_k \exp(-ck) = a_k \left(\frac{1}{\alpha}\right)^k \qquad (3.35)$$

(ELMALAWANY 1975)

c: est une constante positive

$\alpha = 1 + \delta$ avec $\delta$ : incrementation du rayon du cercle unité

3.1.7/ Comparaison des méthodes

POUr la méthode d'autocorrelation le signal est supposé stationnaire à l'interieur d'une trame. La non stationnarité globale se manifeste par le fait que les coefficients de prédiction changent d'une trame à l'autre.

Alors que la méthode de covariance assume que le signal est encore non stationnaire dans une trame. Théoriquement cette hypothèse de non stationnarité est plus réaliste , mais cette méthode n'assure pas la stabilité du modèle. Un algorithme de detection de la stabilité du filtre est indispensable. Par contre la méthode d'autocorrelation peut garantir la stabilité. Par conséquent les coefficients "$a_k$" calculés par cette méthode peuvent être directement utilisés en synthèse.

La comparaison des méthodes d'autocorrelation et de covariance n'a de signification que par rapport à une application bien determinée . Chaque méthode présente des avantages et des inconvenients liés à des contraintes d'application.

| | Méthodes de prediction linéaire | |
|---|---|---|
| propriété | autocorrelation | covariance |
| fenêtre | necessaire | non |
| stabilité | garantie théoriquement | non garantie |
| efficacité | efficace | efficace |

Tableau 3.1. Comparaison des deux méthodes de prédiction linéaire

3.2/ Solution des équations du codge prédictif linéaire

Pour chacune des deux méthodes dévelloppés précedemment nous pouvons calculer les coefficients $a_k$ en resolvant un système de "p" équations à "p" inconnues du type AX =b. Nous pouvons faire appel à une variété de techniques pour la résolution d'un tel système d'équations linéaires. Les méthodes itératives du type Gauss-Seidel et Jacobi bien que performantes, ont l'inconvenient d'avoir un temps de calcul indeterminé . Il existe aussi des méthodes directes telles que la triangularisation de Gauss et celle de Gauss- Jordan .

Ces dernières utilisant un grand nombre d'opérations, il faut donc chercher d'autres techniques plus rapides et utiliant moins d'opérations donc diminution de la place mémoire.

Grâce à la propriété particulière des coefficients de la matrice "A", il est possible de résoudre ces équations de manière beaucoup plus efficace qu'il ne l'est en general. Dans ce qui suit nous developpons deux méthodes de resolution rapides.

3.2.1/ Méthode de la decompositionde Cholesky (ou de la racine carrée)

Pour la méthode de covariance, l'ensemble des équations à résoudre est de la forme.

$$\sum_{K=1}^{P} a_K C(i,K) = C(i,0) \qquad \begin{matrix} 1 \leq i \leq P \\ 0 \leq K \leq P \end{matrix} \quad (3.2)$$

La notation matricielle est donc:

$$Ca = b \qquad (3.36)$$

où: C: matrice des élements $c(i,k)$

a: matrice colonne des elements $a_k$

b: matrice colonne des éléments $c(i,0)$

Comme C est une matrice symétrique definie positive le système d'équations ( 3.2) peut être résolu d'une manière efficace. Cette méthode de résolution est appelée décomposition de Cholesky. Lors de sa programmation sur calculateur elle occupe un espace mémoire plus petit que celui occupé avec les autres méthodes.

Dans cecas la matrice C peut s'écrire:

$$C = vD\,v^T \qquad (3.37)$$

où v est une matrice triangulaire inferieure dont les éléments de la $^{ie}$ diagonale principale sont égaux à l'unité, et D une matrice diagonale. $V^T$ est la matrice transposée de V.

l'équation (3.37) s'écrit donc:

$$\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{p1} \\ C_{21} & C_{22} & \cdots & C_{p2} \\ \vdots & & & \\ C_{p1} & C_{p2} & \cdots & C_{pp} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ V_{21} & 1 & \cdots & 0 \\ \vdots & & & \\ V_{p1} & V_{p2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & d_p \end{bmatrix} \begin{bmatrix} 1 & V_{21} & \cdots & V_{p1} \\ 0 & 1 & \cdots & V_{p2} \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad (3.38)$$

A partir de l'equation (3.24) nous obtenons le système d'équations suivant:

$$C_{ij} = \sum_{K=1}^{j} V_{ik}\, d_k V_{jk} \qquad 1 \leq j \leq i-1 \qquad (3.39)$$

Les élements de la matrice V et D sont obtenus en resolvant le système (3.39) on a:

$$C_{ij} = \sum_{K=1}^{j-1} V_{ik}\, d_k\, V_{jk} + V_{ij}\, d_j\, V_{jj} \qquad (3.40)$$

or $V_{jj} = 1$ d'où:

$$V_{ij}\, d_j = C_{ij} - \sum_{K=1}^{j-1} V_{ik}\, d_k\, V_{jk} \qquad 1 \leq j \leq i-1 \qquad (3.41)$$

Pour le calcul des éléments de la diagonale de D on a:

$$C_{ii} = \sum_{K=1}^{i} V_{ik}\, d_k\, V_{ik} \qquad (3.42)$$

$$C_{ii} = \sum_{K=1}^{i-1} (V_{ik})^2\, d_k$$

$$C_{ii} = \sum_{K=1}^{i-1} (V_{ik})^2 d_k + (Vii)^2 d_i$$

or $V_{ii} = 1$ d'où:

$$d_i = C_{ii} - \sum_{K=1}^{i-1} (V_{ik})^2\, d_k \qquad i \geq 2 \qquad (3.43)$$

avec la condition initiale:

$$d_1 = C_{11} \qquad (3.44)$$

A partir de l'équation (3.41) et de la connaissance de $d_1$ nous pouvons calculer les éléments $V_{i,1}$ et de l'équation (3.43) nous determinons $d_2$.

Ainsi connaissant les éléments des matrices Vet D, nous pouvons calculer récursivement les éléments de la matrice, colonne a.

A partir de (3.36) et (3.37) nous déduisons que:

$$V D v^T a = b \qquad (3.45)$$

posons: $D\, V^T a = Y \qquad (3.46)$

d'où: $V\, Y = b \qquad (3.47)$

De l'équation (3.46) nous avons:

$$V^T a = D^{-1} Y \qquad (3.48)$$

Et à partir de l'équation (3.47) et de la matrice V nous pouvons déduire:

$$Y_i = C(i,0) - \sum_{j=1}^{i-1} V_{ij} Y_j \qquad 2 \leq i \leq p \qquad (3.49)$$

pour i =1 $\quad Y_1 = C_{10}$

L'équation (3.48) conduit à la relation récursive suivante:

$$a_i = \frac{Y_i}{d_i} - \sum_{j=i+1}^{p} V_{ji}\, a_j \qquad 1 \leq i \leq p-1 \qquad (3.50)$$

Et pour i= p, nous obtenons la condition initiale suivante:

$$a_p = y_p / d_p \qquad (3.51)$$

les coefficients $a_k$ sont ainsi determinés en utilisant les équations (3.41),(3.43), (3.49),(3.50)

Erreur ouadratique totale de prediction

l'erreur ouadratique totale pour la méthode de covariance s'écrit:

$$E_p = C(0,0) - \sum_{K=1}^{p} a_K\, C(0,K) \qquad \text{(équation 3.18)}$$

Elle s'écrit sous forma matricielle:

$$E = C(0,0) - a^T b \qquad (3.52)$$

Nous allons cher cher l'erreur quadratique en fonction des éléments des matrices Y et D. L'equation (3.48) peut s'écrire sous la forme suivante:

$$(V^T a)^T = (D^{-1} Y)^T \qquad (3.53)$$

puisoue V est une matrice triangulaire inferieure et D une matrice diagonale nous pouvons écrire:

$$a^T V = Y^T (D^{-1})^T = Y^T D^{-1} \qquad (3.54)$$

d'où la relation suivante:

$$a^T = Y^T D^{-1} V^{-1} \qquad (3.55)$$

Nous remplaçons $a^T$ par sa valeur dans l'équation (3.52)

$$E = C(0,0) - Y^T D^{-1} V^{-1} b \qquad (3.56)$$

en remplaçant b par VY nous aurons :

$$E = C(0,0) - Y^T D^{-1} V^{-1} VY$$

$$E = C(0,0) - Y^T D^{-1} Y \qquad (3.57)$$

à partir de l'équation précédente nous pouvons déduire que :

$$E_p = C(0,0) - \sum_{K=1}^{P} \frac{Y_K^2}{d_K} \qquad (3.58)$$

3.2.2/ Méthodes ou algorithme de Durbin

Pour la méthode d'autocorrélation l'équation matricielle à résoudre pour les coefficients du prédicteur est de la forme :

$$\sum_{K=1}^{P} a_K R(|i-K|) = R_i \quad 1 \leq i \leq p \quad \text{équation (3.15)}$$

En exploitant la nature de Toeplitz de la matrice d'autocorrélation plusieurs procédures récursives ont été concues pour résoudre ce système d'équations mais la plus efficace est la méthode récursive de Durbin. Cette dernièce est très rapide car elle utilise "P(P+1)" opérations et occupe une place mémoire réduite.

Cet algorithme s'énonce comme suit :

$$\begin{cases} E(0) = R(0) \\ K_i = \left[ R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} \cdot R(i-j) \right] / E(i-1) \quad 1 \leq i \leq p \\ a_i^{(i)} = K_i \\ a_j^{(i)} = a_j^{(i-1)} - K_i \cdot a_{i-j}^{(i-1)} \quad 1 \leq j \leq i-1 \\ E(i) = (1-K_i^2)\, E(i-1) \end{cases} \qquad (3.59).$$

Ces équations sont résolues récursivement pour i = 1, 2, ...p et la solution finale est donnée comme : $a_j = a_j^{(p)} \quad 1 \leq j \leq p \qquad (3.60).$

Nous remarquons que dans la résolution du système d'équations pour obtenir les coefficients $a_j$ d'un prédicteur d'ordre p, on calcule tous les coefficients de tous les prédicteurs d'ordre inférieur à p.

Les coefficients $K_i$ qui apparaissent dans l'algorithme de Durbin sont appelés " coefficients de correlation partielle" ou " coefficients de reflexion". Ces coefficients sont trés importants pour l'analyse predictive du signal vocal.

$$/ K_i / < 1 \qquad (3.61)$$

Cette relation est une condition nécéssaire et suffisante pour la stabilité du filtre modèle.

Il est interessant de comparer les quantités d'opérations nécéssaires à la résolutuion des équations (tableau 3.2)

Tableau 3.2: Comparaison du nombre d'operations nécessaires dans chaque méthode. ( D'aprés Makhoul, 1972)

| | Taille mémoire | Nombre d'opérations |
|---|---|---|
| Triangularisation de Gauss | $p^2$ | $p(2p^2+6p-2)/6$ |
| Méthode de la racine carrée (décomposition de Cholesky) | $p(p+1)/2$ | $p(p^2+6p/11)/6$ |
| Méthode d'autocorrélation (algorithme de Durbin) | $2p$ | $p(p+1)$ |

# CHAPITRE QUATRE

# D E T E C T I O N  D E  L A  F R E Q U E N C E  F O N D A M E N T A L E

Introduction

Dans le domaine de la parole on a toujours besoin de savoir si le sighal est voisé on non. La synthèse ainsi que la reconnaissance de la parole font appel à la connaissance du fondamental.

La frequence "$F_o$" se situe pour les femmes dans la gamme (150Hz à 300Hz), pour les enfants, elle est supérieure à 400Hz et pour les hommes, elle se trouve dans l'intervalle (70Hz à 150Hz).

Cette difference est dûe à l'anatomie des cordes vocales, car ces dernières sont plus minces chez les femmes et les enfants que chez les hommes. En général tous les locuteurs hommes out femmes ont une période du fondamental "$T_o$" qui s'inscrit dans l'intervalle 2,5 à 20ms, soit une fréquence fondamentale inscrite dans l'intervalle 50 à 400Hz (ELMALLAWANY, 1975).

L'extraction de " $F_o$ " est importante dans l'analyse du signal de la parole. Mais ceci est un problème du fait de la quasi-périodicité du signal glottique (forme de l'onde d'excitation non constante.) ( cf chapitre 1 ).

La fréquence fondamentale ou " pitch " varie non seulement d'un locuteur à un autre mais aussi d'une élocution à une autre pour un même locuteur ce qui prouve plus la difficulté de sa determination.

De nombreuses méthodes ont été proposées dans le but de detecter le pitch . Nous en décrivons quelques unes.

## 4.1/ Méthode d'intercorrelation avec une fonction peigne

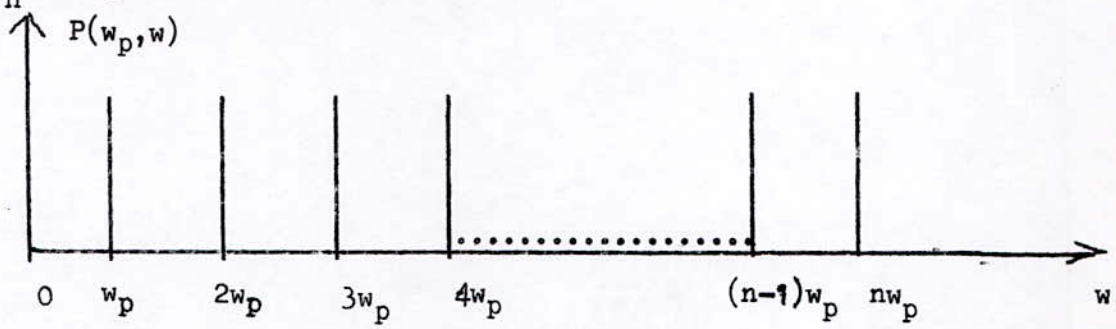
Cette méthode est basée sur la recherche d'une structure harmonique dans le spectre d'amplitude . (P. MARTIN, 1981)

Le principe est le suivant: On effectue le calcul de l'intercorrelation entre le spectre d'amplitude / F(w)/ d'un son voisé et une fonction " peigne" p(wp,w). avec "w" définissant la frequence et "$w_p$" la distance entre deux dents respectives du peigne.

La fonction d'intercorrelation s'écrit:

$$P(w_p,w) = \sum_n An\ \delta(nw_p - w) \qquad (4.1)$$

avec $A_n$ : amplitude de la raie d'ordre n.




Fonction peigne aux dents d'amplitude constante.

Le maximum de la fonction d'intercorrelation est obtenu lorsque la distance entre deux dents consécutives est égale à la période du signal à analyser, c'est à dire lorsque $wp = w_o$. Cette méthode consiste en une recherche des maxima du spectre situés à des fréquences harmoniques les uns des autres.

4.2/Méthode cepstrale

Une seconde analyse de Fourier du signal de la parole mène à la detection de sa periodicité.

Comme nous le savons ( cf, chap 2), le spectre instantané du signal de parole est consideré comme le produit du spectre instantané du signal source et de la fonction de transfert du conduit vocal.

En utilisant l'echelle logarithmique ce produit devient une addition.

Ainsi le signal source et la fonction de transfert du conduit vocal sont séparées. La transformation inverse de Fourier fournit le cepstrum.

Pour les sons voisés les oscillations rapides et periodiques du spectre donnent lieu à une raie d'abscisse éloignée donnant "$T_o$" (fig 4.2, (a)).

Pour les sons non voisés les oscillations lentes et aperiodiques donnent lieu à une courbe étalée (fig 4.2, (b)).

4.3/ Determination de "$F_o$" à partir de ses harmoniques.

Le principe consiste à detecter deux harmoniques. La frequence fondamentale dans ce cas correspond au plus grand commun diviseur des fréquences de ces harmoniques.

exemple: $f_1 = 770Hz$ $f_2 = 1220Hz$

le plus grand diviseur commun est :

$F_c = 110Hz$

$F_1$ et $F_2$ correspondent au septième et onzième harmonique

4.4/Detection de " $F_o$ " dans l'analyse prédictive

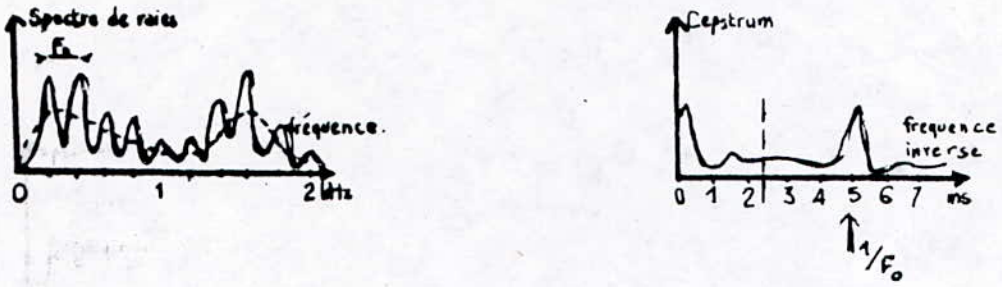
L'analyse prédictive apporte un appui important pour la resolution du problème de détection du fondamental de la voix. Elle donne des résultats satisfaisants. Nous pouvons à partir de la fonction décrivant l'erreur entre le signal réel et le signal prédit obtenir l'information recherchée.

4.4.1/ Le principe de la méthode

Pour extraire " $F_o$ " il faut élaborer un signal permettant de determiner facilement les instants d'excitation du conduit vocal par le signal glottique, de calculer " $F_o$ ". On obtient ce signal par la méthode dite du filtrage inverse " .

On applique les échantillons S(nT) du signal de parole à l'entrée du filtre inverse décrit precedemment et dont l'equation polynômiale est la suivante:

$$D(Z) = 1 - \sum_{K=1}^{P} a_k Z^{-k} \qquad (4.2)$$




(a) : Cepstre d'un son voisé


(b) : Cepstre d'un son non voisé

Fig. 4.2 : Détermination du fondamental à l'aide du cepstre.

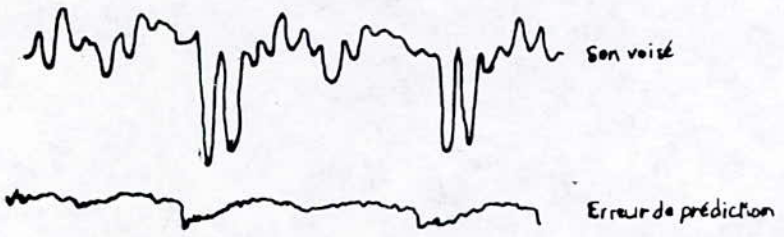



Fig. 4.3 : Echantillon d'un son voisé et erreur de prédiction correspondante. L'erreur est petite sauf au moment où apparaissent les impulsions du fondamental. Il est ainsi possible de mesurer la hauteur du son, en examinant le signal d'erreur, indirectement à travers sa fonction d'autocorrélation.

Le filtre D(Z) est au facteur de gain " G" prés, le filtre inverse du signal prédit qui définit la fonction de transfert approchée du système vocal. Donc en appliquant le signal S(nT) à l'entrée de ce filtre inverse le signal e(nT) obtenu à la sortie sera proche du signal d'excitation du système vocal $\dot{u}(nT)$.

nous avons: $e(nT) = G\,u(nT)$ (4.3)

Ainsi ce signal d'erreur comporte des impulsions du fondamental. La mesure du temps séparant deux impulsions successives donne la valeur de la periode " $T_o$ " dont l'inverse est égal à "$F_o$" (fig 4.3).

La detection de voisement dans cette technique se fait en deux étapes:

- Detecter le voisement
- Mesurer la periode dans le cas d'une sequence sonore.

4.4.2/ Decision " voisée / non voisée "

deux critères simples ont été établis pour determiner si la sequence analysée est voisée ou non. Ces critères sont fondés sur le fait que les sons non voisés ont une grande partie de leur energie concentrée dans les hautes frequences, le nombre de passages par zéro du signal de parole est plus élevé dans ce cas . Ainsi le premier critère consiste à determiner le nombre de passages par zéro du signal et dont la valeur doit être inferieure à un certain seuil (2 à 3ms) ( ELMALAWAN , 1975) pour que la sequence soit voisée.

Le deuxième critère consiste à établir le rapport entre la longueur de la trajectoire du signal de parole dans la séquence et la somme des valeurs absolues des échantillons.

le critère est exprimé par :

$$C = \frac{\sum_{n=1}^{N-1} /S_{n+1} - S_n /}{\sum_{n=1}^{N} /S_n/} \quad \begin{cases} > 90\ \% \longrightarrow \text{non voisé} \\ \leq 90\ \% \longrightarrow \text{voisé} \end{cases} \qquad (4.4)$$

Ces critéres étant trop simples , ils ne tiennent pas compte de toutes les configurations possibles. Il en résulte une fausse décision, d'où il a fallu avoir recours au critère de l'erreur normalisée $V_p$ .

- Erreur normalisée $V_p$

L'erreur normalisée " $V_p$ " a été définie comme suit:

$$V_p = 1 - \sum_{k=1}^{p} a_k r_k \qquad \text{où} \quad r_k = \frac{R_k}{R_o}$$

$$V_p = \frac{E_p}{R_o}$$

où $E_p$ : erreur de prediction

$R_o$ fonction d'autocorrelation

Il a été constaté qu'en general les sons non voisés possèdent une erreur normalisée plus importante que les sons voisés, ce qui suggère oue " $V_p$ " peut servir de detecteur de voisement.

(MAkhoul , (1972) a fait une remarque trés importante, il a demontré que " $V_p$ " ne dépend que de la forme du spectre et non du caractère de voisement . En effet d'aprés la formule

$$G(Z) = \frac{1}{(1 - e^{-cT} \times Z^{-1})^2} \qquad (4.5)$$

c: celerité du son , T : période d'échantillonnage.

On remarque que le spectre de la source sonore est inversement proportionnel au carré de la fréquence. Par conséquent pour les sons sonores, le spectre du signal d'excitation présente une décroissance rapide de l'énergie. Donc le maximum d'énergie dans ce cas est concentré vers les basses fréquences et il en découle une erreur normalisée relativement faible. Pour les sons non voisés par contre l'énergie est mieux répartie sur l'ensemble du spectre et l'erreur est par conséquent plus importante. Cependant cette propriété n'est pas toujours vérifiée.

De plus la précision des calculs est réduite et les erreurs résultant de l'arrondi et de la troncature tendent à rendre la valeur de "$V_p$" faible, ce qui conduit à de fausses décisions en faveur du voisement . Ces défauts incitent les chercheurs dans ce domaine à trouver d'autres critères plus sûrs pour detecter le voisement.

4.4.3/ Decimation et interpolation

Dans le traitement numérique du signal de la parole, on a toujours besoin de changer la vitesse d'échantillonnage ( 1/T) du signal discret. Les procédés de réduction et d'augmentation de cette celle-ci s'appellent respectivement " décimation " et " interpolation ".

—Décimation/

Supposons que l'on veuille réduire la vitesse d'échantillonnage par un facteur k. Par conséquent il faut calculer une nouvelle séquence correspondante aux échantillons, prise avec une periode $T' = k\,T$

donc $y(n) = x(n\,T') = x(n\,k\,T)$

$y(n) = x(kn)$

avec $x(n)$ signal d'entrée

$y(n)$ signal de sortie

—interpolation/

Si on veut augmenter la vitesse d'echantiollonnage par un facteur L, on doit calculer une nouvelle séquence correspondante aux échantillons, prise avec une periode $T' = T/L$

donc: $y(n) = x(nT') = x(n\,T/L)$

$y(n) = x(n/L)$ pour $n = 0, \pm L, \pm 2L \ldots$

On doit compléter les échantillons pour toutes les valeurs de n par un procédé d'interpolation.

4.4.4/ Detection du fondamental à l'aide du S.I.F.T

( Technique Simplifiée du Filtre Inverse )

Cette technique a été présentée par MARKEL (1972) pour la detection du voisement et la mesure du fondamental dans le cas d'une séquence sonore.

Le schéma fonctionnel de l'algorithme, dont nous allons décrire le principe est illustré dans la figure (4-4).

Son principe est basé sur le calcul de la fonction d'autocorrélation. Ce pendant le calcul direct de cette fonction à partir du signal de parole pose un problème. En effet le signal de parole est un produit de convulution entre le signal quasi-périodique de la source sonore et la réponse du conduit vocal. Or, les formants ont des largeurs de bande suffisamment étroites pour produire des oscillations d'amplitudes élevées dans la fonction d'autocorrélation. Ainsi il peut y avoir interférences entre ces oscillations et la composante représentant la période du fondamental. Ces interférences sont surtout dues au premier formant $F_1$ et conduisent à une estimation erronée de la période fondamentale.

Avant de calculer l'autocorrelation il faut éliminer la réponse du conduit vocal. La prédiction linéaire est un moyen simple permettant l'affaiblissement des résonances du conduit vocal. Dans le schéma fonctionnel, nous remarquons que le signal de parole est échantillonné à 10 KHZ. Il est ensuite filtré à l'aide d'un filtre passe-bas dont la fréquence de coupure est de 800 HZ. Le signal $\hat{s}_n$ ainsi obtenu ne peut contenir qu'une ou deux résonances du conduit vocal, donc un prédicteur d'ordre "trois" suffit pour les affaiblir.
Les trois coefficients "$a_k$" du prédicteur sont estimés à l'aide de la méthode d'autocorrélation. La convolution entre le signal $\hat{s}_n$ et le filtre inverse permet d'obtenir le signal d'erreur $\hat{e}n$ dans lequel les résonances du conduit vocal et les pics parasites sont éliminés.

Un calcul de la fonction d'autocorrélation de $\hat{e}n$ est effectué pour des retards zéro et de 2,5 à 20 ms qui représentent les valeurs extrêmes que peut prendre la période fondamentale. Sa valeur correspondra au retard de la composante maximale de la fonction d'autocorrélation,L'interpolation est utilisée afin d'obtenir une meilleure estimation du "pitch". Après cela la valeur "Ri" doit satisfaire à une condition pour décider si la séquence analysée est ou non voisée.

On peut estimer que le signal d'erreur pour les sons non voisés est considéré comme un bruit blanc gaussien de valeur moyenne nulle. Ainsi théoriquement quand la séquence de signal analysé devient grande, la fonction d'autocorrélation R(n) tend vers zéro pour n différent de zéro. Et pour une longueur finie de la séquence du signal analysé on peut déterminer un seuil tel que avec une certaine

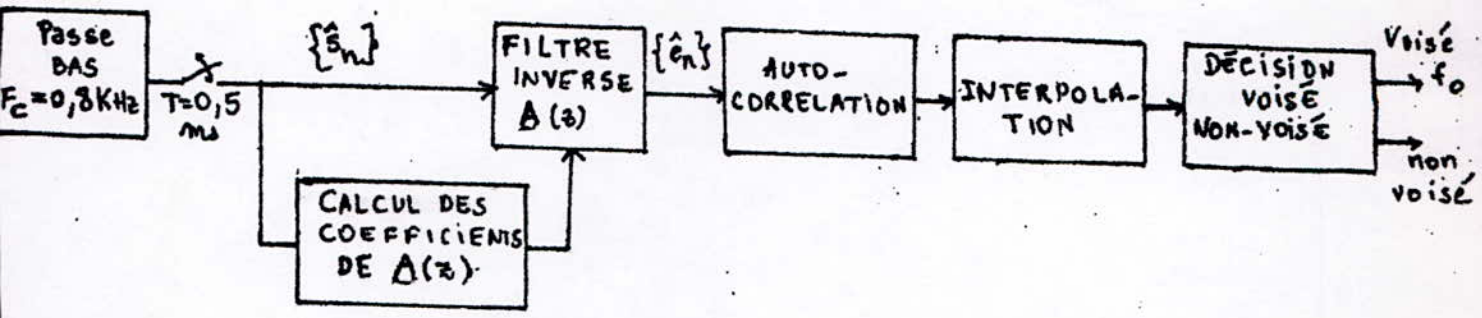



Fig.4.4: Schema fonctionnel du detecteur du fondamental.

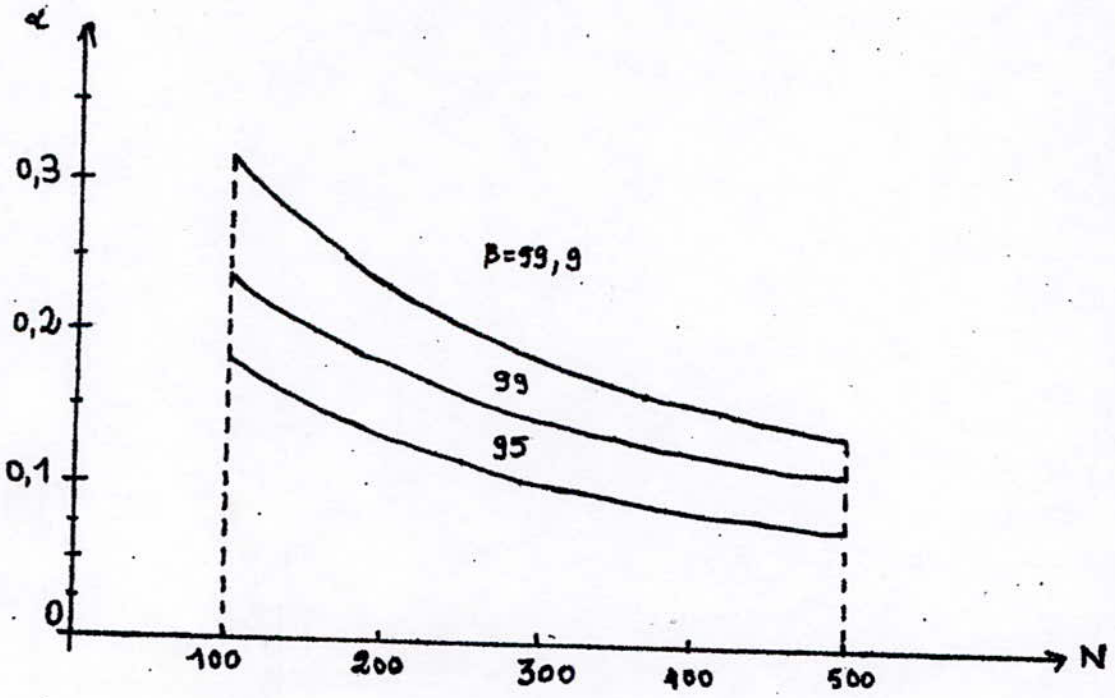



Fig.45: Courbes du seuil $\alpha$ en fonction du nombre N d'échantillons du signal pour differentes valeurs de l'intervalle de confiance $\beta$ .

probabilité et pour un certain intervalle de confiance aucun échantillon de la fonction d'autocorrélation (excepté celui de l'origine) ne dépasse la valeur de ce seuil.

Nous pouvons obtenir ce seuil $\alpha$ qui satisfait à la relation suivante :

$$P_r \ (r \leq \alpha) = 0{,}01\ \beta$$

$\alpha$ est obtenu à partir d'un graphe de la fonction $\alpha = f(N)$ pour des intervalles de confiance $\beta$ différents . (Voir fig. 4.5)
N étant le nombre d'échantillons du signal.

$r = \frac{R(n)}{R(0)}$ , avec R(n) le premier maximum de la fonction d'autocorrélation à partir de R(0), dans la séquence considérée la valeur de r comparée à $\alpha$, nous permet de déduire si la séquence du signal est voisée ou non.

A partir de la fig. 4-5 nous pouvons constater, par exemple que pour $\alpha = 0{,}2$ et $N \geq 250$, la probabilité d'avoir une erreur est inférieure à 0,001.

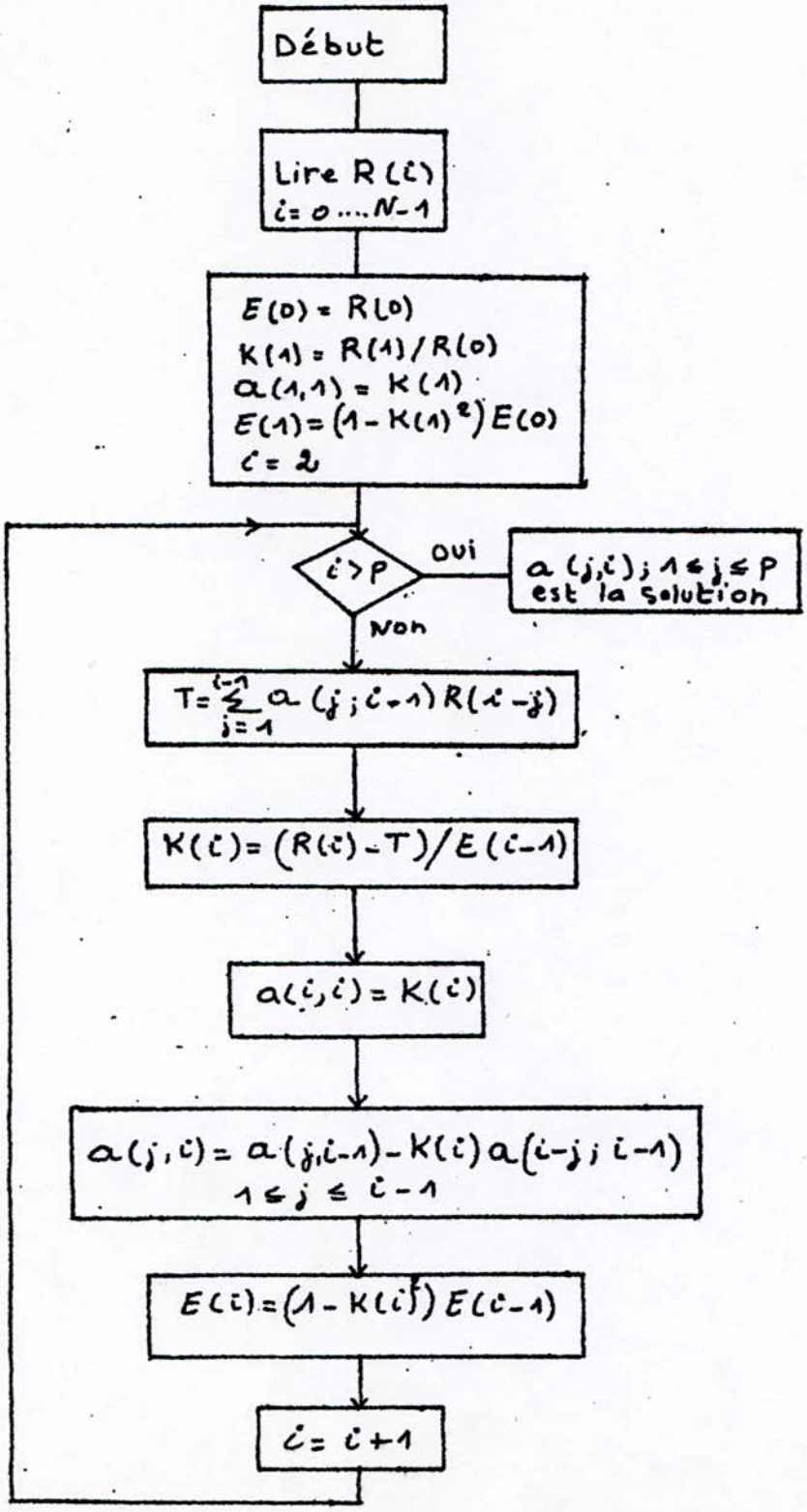
Le signal de parole, dans le cas des sons non-voisés, peut quelquefois être assez différent d'un bruit gaussien, alors on est conduit à relever $\alpha$, pour ne pas commettre d'erreurs dans la décision "voisée-non voisée". Le nombre d'échantillons utulisé pour calculer la fonction d'autocorrélation à court terme a été fixé à 200 . D'après la figure 4-5, on constate que la valeur $\alpha$, correspondant à ce nombre N pour un intervalle de confiance $\beta$ supérieur à 99 % est alors environ 0,2. Pour une longueur de la séquence égale à 80 le seuil $\alpha$ est de l'ordre de 0,35. Dans le cas d'une fin de segment de parole voisée, la période du fondamental varie dans une large proportion. Les périodes du signal sont alors faiblement correlées et la valeur du maximum de la fonction d'autocorrélation qui correspond à une valeur moyenne de la période du fondamental est de ce fait notablement réduite, d'où la nécessité de réduire ce seuil. Cette réduction se fait par une pondération du nombre $\alpha$ par un facteur 0,8.

# DEUXIEME PARTIE

# PROGRAMMATION.

Dans cette partie nous donnons les organigrammes des deux méthodes de calcul des coéfficients de prédiction (autocorrélation;covariance)et leurs programmes respec--tifs ainsi que l'organigramme de la décision Voisé/nonVoisé.

**Organigramme pour la méthode d'autocorrélation**

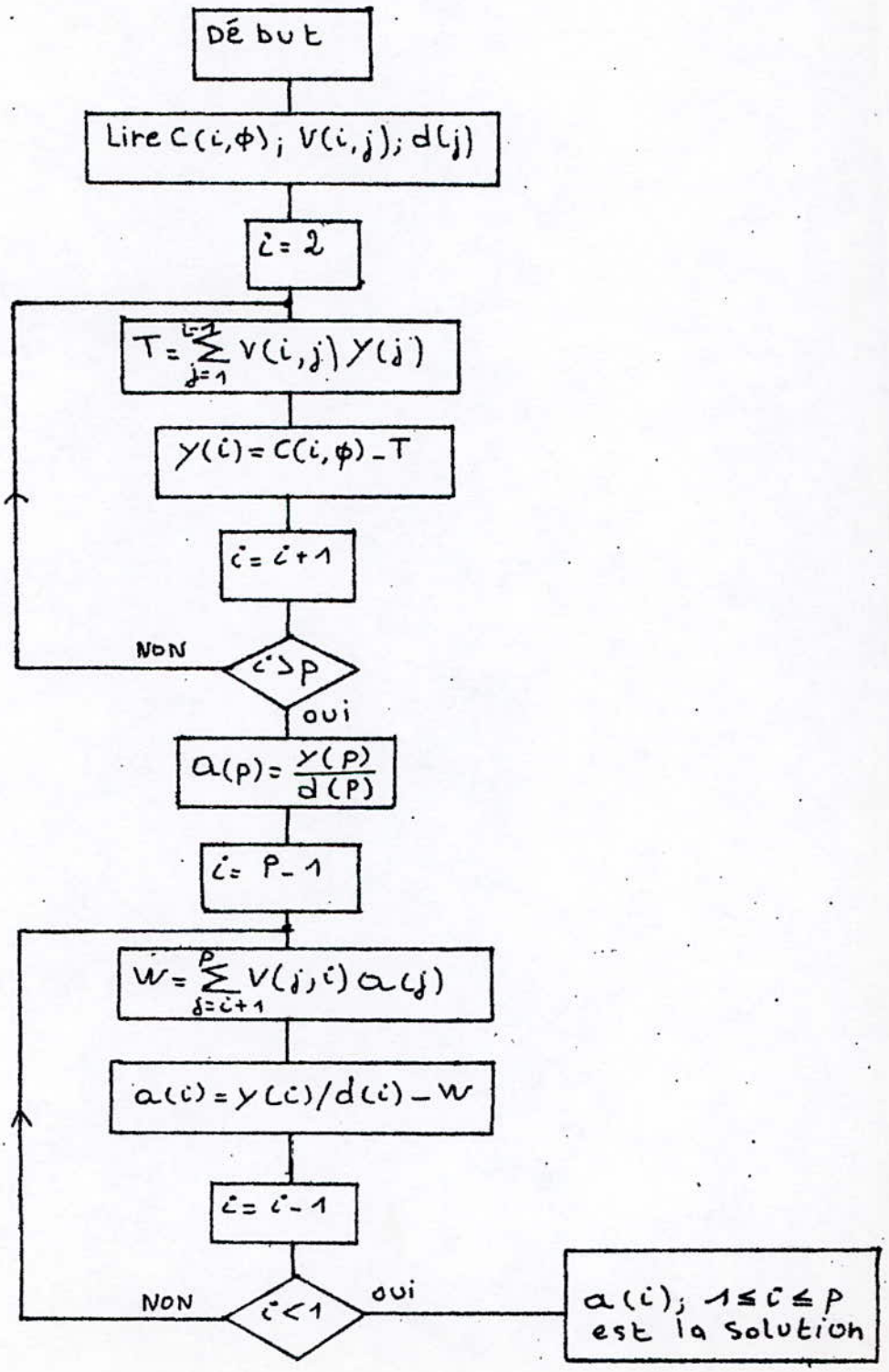


```
LIST

100     INIT
110     PAGE
120     REM ***CALCUL DES COEFFICIENTS DE PREDICTION A(I) PAR LA
                METHODE ***
130     REM *** D'AUTOCORRELATION ***
140     DELETE S, R, R1
150     PRINT "CALCUL DES COEFFICIENTS PAR LA METHODE D'AUTO";
160     PRINT    "CORRELATION"
170     "ENTREZ LA VALEUR DE N =        ";
180     INPUT   N
190     PRINT   "ENTREZ L' RDRE DU PREDICTEUR  K=  ";
200     INPUT   K
210     DIM   S(N)
220     FOR  I = 1 TO N
230     S(I) =  2 *0,99 ↑(I-K) - 0, 99 ↑(2 *(I-K))
240     NEXT I
250     DIM   R(K)
260     N1 =N-K-1
270     PRINT " LA VALEUR DE  N1=  "" ; N1
280     I= ø
290     R ø = ø
300     FOR J=1 TO  N
310     P= S(J)↑2
330     R ø =  R ø  + P
335     NEXT   J
340     P= ø
350     FOR  I=1 TO  K
360     R1 = ø
370     FOR  J=1 TO N-I
```

```
380     P= S (J) * S ( J + I )
390     R1 =  R 1 + P
400     NEXT  J
410     R ( I ) = R1
420     PRINT  "R" , I ,  " ) =  " ; R (I)
430     NEST  I
440     REM RECHERCHE DES COEFFICIENTS DE REFLEXION L (K)
450     DIM  E (K) , L (K),  A (K., K)
460     E= ϕ
470     A= ϕ
480     L= ϕ
490     E ϕ'= R ϕ
500     L (1) = R (1) / R ϕ
510     A (1,1) = L (1)
520     E (1) = (1-L (1)↑ 2)  E ϕ
530     FOR   I=2 TO  K
540     T=0
550     FOR   J= 1  TO I-1
560     T =T+ A (J , I-1) *R (I-J)
570     NEXT  J
580     L (I) = (R (I)- T) / E (I-1)
590     REM RECHERCHE DES COEFFICIENTS DE  PREDICTION A  (J,I)
600     A II,I) =L(I)
610     FOR   J=1 TO  I-1
620     A (J,I) =A (J,I-1)-L(I) *A(I-J,I-1)
630     NEXT   J
640     REM RECHERCHE DE L'ERREUR QUADRATIQUE E(K)
650     E (I) = (1- L (I)↑ 2) *E(I- 1)
660     NEXT  I
670     END
```

**Organigramme pour la méthode de covariance.**



```
LIST
100       INIT
110       PAGE
120       PRINT " RECHERCHE DES COEFFICIENTS DE PREDICTION PAR LA METHODE";
130       PRINT" DE COVARIANCE  3
140       REM RECHERCHE DES ELEMENTS DE LA MATRICE C (I,J)
150       PRINT3 ENTREZ L'ORDRE DU  PREDICTEUR K=  ",
160       INPUT  K
170       PRINT" ENTREZ LE NOMBRE D'ECHANTILLONS N=  ",
180       INPUT  N
190       DELETE  C,V,D
200       DIM C (K,K), S N)
210       FOR  I=1 TO  N
220       S(I)=2*0,99↑(I-K)-0,99↑(2*(I-K))
230       NEXT  I
240       C=0
250       FOR  I=1 TO  K
80        FOR  J =1 TO K
270       C (I,J) =  ϕ
280       FORM = K+1 TO N
290       C(I,J) = C(I,J) +S(M-I)*S(M-J)
300       NEXT  M
310       NEXT  J
320       NEXT  I
330       REM RECHERCHE DES ELEMENTS DE LA MATRICE V(K,K)
340       REM RECHERCHE DE LA MATRICE  D(K)
350       DIM V(K,K),D(K),Y(K), A(K), B (K,K)
360       V= ϕ
370       B= ϕ
380       D(1) = C(1,1)
390       J= 1
400       V(J,J) =1
410       FOR I=J+ 1 TO  K
420       V(I,I) = 1
430       V(I,J) = C(I,J) / D(J)
```

```
440         NEXT I
450         D(J+1) =C(J+1 ,J+1)-V(J+1,J)↑ 2 *D(J)
460         J= J+1
470         FOR  I= J+1  TO K
480         T=0
490         FOR  P=1  TO J-1
500         T=T + V(I,P) * D(P) * V(J,P)
510         NEXT  P
520         V(I,J) = (C (I,J)-T) / D(J)
530         NEXT  I
540         F= ø
550         FOR  P=1 TO  J
560         F=F+V (J+1,P)↑ 2 * D(P)
570         NEXT P
580         D(J+1) = C(J+1, J+1)-F
590         IF J < K-1 THEN 460
600         FOR I=  1 TO K
610         B(I,I) = D(I)
620         NEXT I
630         REM RECHERCHE DES ELEMENTS DE LA MATRICE CØ(K)
640         DIM CØ(K)
650         CØ = Ø
660         Z = Ø
670         FOR I = 1 TO K
680         FOR M = K + 1 TO N
690         Z = S(M) * S(M-I)
700         CØ(I) = CØ(I) + Z
710         NEXT M
720         NEXT I
730         REM RECHERCHE DES ELEMENTS DE LA MATRICE Y(K)
740         Y(1) = CØ(1)
750         FOR I = 2 TO K
760         G = Ø
770         FOR J = 1 TO I-1
780         G = G + V(I,J) * Y(J)
```

```
790       NEXT J
800       Y(I) = CØ(I) - G
810       NEXT I
820       REM RECHERCHE DES COEFFICIENTS DE PREDICTION A(I)
830       A(K) = Y(K)/D(K)
840       I = K-1
850       H = Ø
860       FOR J = I + 1 TO K
870       H = H + V(J, I)*A(J)
880       NEXT J
890       A(I) = Y(I)/D(I) - H
900       I = I-1
910       IF I < 1 THEN 925
920       GO TO 850
925       PRINT "COEFFICIENTS DE PREDICTION A"
926       FOR I = 1 TO K
927       PRINT "A("; I; ") =      "; A(I)
928       NEXT I
930       REM RECHERCHE DE L'ERREUR QUADRATIQUE  E
940       C1 = Ø
950       FOR M= K+1 TO N
960       C1 = C1 + S(M)↑2
970       NEXT M
980       W = 0
990       FOR P = 1 TO K
1000      W = W + Y(P)↑2/D(P)
1010      NEXT P
1020      E = C1 -W
1030      END.
```

Commentaire des organigrammes de la methode
d'autocorrélation et de la methode de covariance.

1/Methode d'autocorrélation.

L'organigramme est initialisé avec les valeurs R(i) de la fonction d'autocorrélation. La methode d'autocorrélation permet de calculer les coéfficients prédictifs de l'ordre p du prédicteur ainsi que tous ceux des ordres inferieurs à p.De même que nous pouvons avoir la valeur de l'erreur quadratique et celle de tous les coefficients de reflexion de chaque prédicteur.

Exemple : pour le signal ;

$$S(n) = 2 \times 0{,}99^{(n-4)} - 0{,}99^{2(n-4)} \qquad \text{avec} \quad 1 \leq n \leq N$$

et pour un nombre d'échantillons N =24 et l'ordre de prédicteur p=4 les résultats sont les suivants :

| i | $a_i$ | $E_i$ | $K_i$ |
|---|---|---|---|
| 1 | $a_1$= 0,958879931568 | 1,89323864532 | 0,958879931568 |
| 2 | $a_1$= 0,98074697128<br>$a_2$= -0,0228047735615 | 1,89225405199 | -0,0228047735615 |
| 3 | $a_1$=0,980218833825<br>$a_2$= -9,158115466E -5<br>$a_3$= - 0,0231590747379 | 1,89123915526 | - 0,0231590747379 |
| 4 | $a_1$= 0,97967383941<br>$a_2$= -9,373630198 E- 5<br>$a_3$= -9,19273821 E- 5<br>$a_4$= - 0,0235326506284 | 1,89019181417 | - 0,0235326506284 |

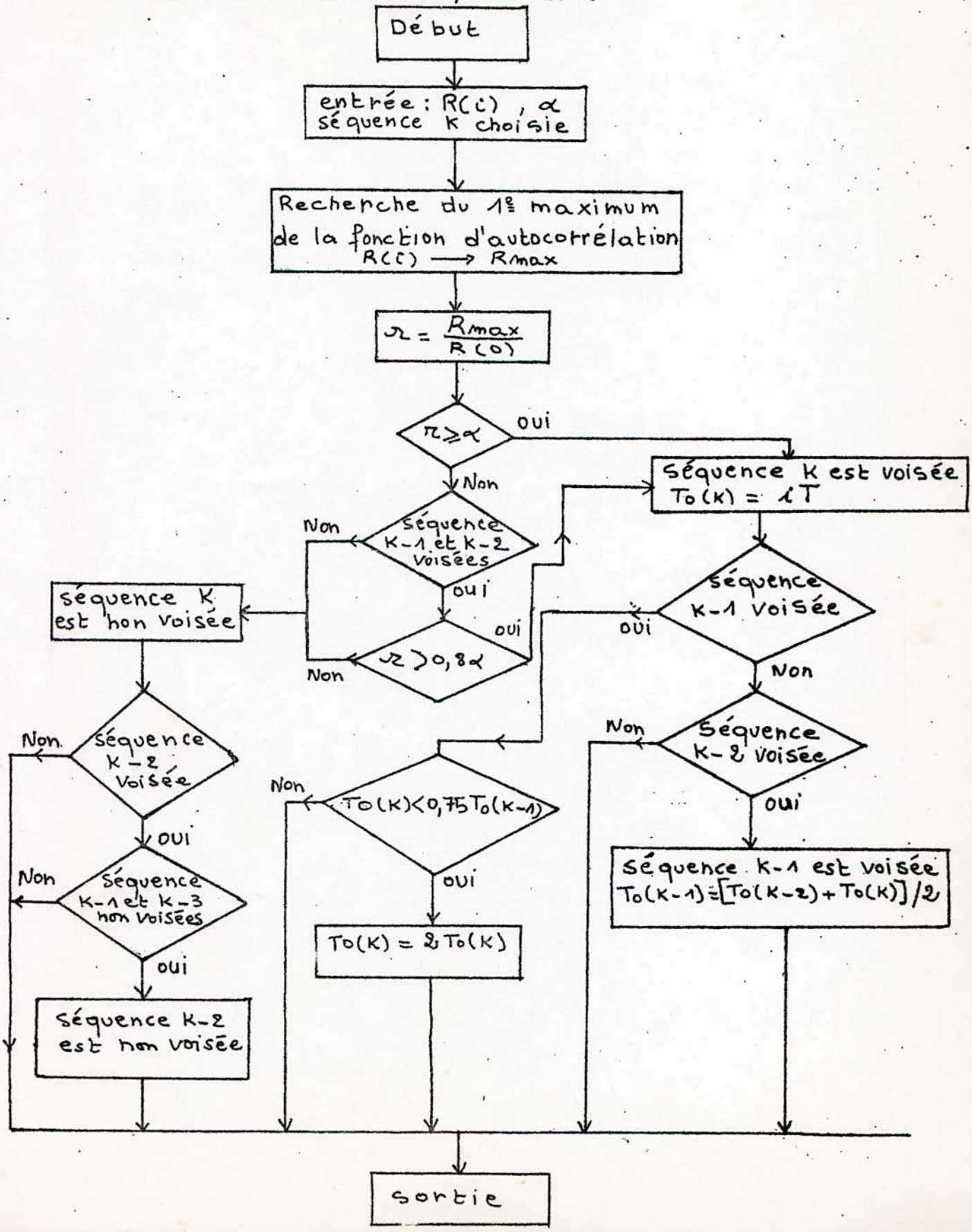
Nous remarquons que les coéfficients de reflexion sont tels que $-1 < K(i) < 1$ donc le filtre numérique construit avec les coéfficients ai est stable .Ceci vérifie bien la propriété de la stabilité de la méthode d'autocor- -rélation.

2/Méthode de covariance

Dans ce cas l'organigramme est initialisé avec les valeurs des matrices C(i,O);V(i,j); d(j) calculees à partir de la matrice C(i,j) de covariance.Nous avons donc d'abord calculé les éléments de la matrice C(i,j) en connaissant les valeurs du signal S(n) .

Contrairement à la méthode d'autocorrélation, la covariance ne permet de calculer que les coefficients et l'erreur quadratique de l'ordre du prédicteur choisi.

Organigramme de la décision voisé / non voisé .



Commentaire de l'organigramme de la décision voisé/non voisé

L'algorithme est initialisé avec les coefficients "$R_i$" et le seuil $\alpha$ . Après la recherche du premier maximum de la fonction d'autocorrélation, nous établissons le rapport $r = \frac{Rmax}{R(0)}$ . Si r est supérieur ou élage à $\alpha$ , la séquence k est dite voisée de période $T_0(k) = iT_e$ avec $T_e$ : période d'échantillonnage et i variant de 0 à N-1. Cette décision nous permet de corriger deux erreurs possibles :

- Si la séquence k-1 est déclarée non-voisée alors que les séquences k et k-2 sont voisées nous avons une anomalie et nous la corrigeons en déclarant k-1 voisée avec une période $T_0$ (k-1) égale à la moyenne de celles des deux séquences k et k-2.

- La période $T_0(k)$ peut correspondre au premier formant et non au fondamental, il est possible de détecter cette erreur en comparant $T_0(k)$ à la valeur $T_0(k-1)$. Puisque la différence entre les périodes de deux séquences consécutives ne dépasse pas 20 %, le test se fait comme suit : $T_0(K) < 0{,}75\ T_0(K-1)$ car $T_0(K) - 0{,}75\ T_0(K) = 25\ \%\ T_0(K)$. La correction sera faite en prenant le double de $T_0(K)$.

Si les séquences précédant "k" sont voisées, le seuil $\alpha$ est pondéré par un facteur 0,8 et si r est supérieur à $0{,}8\,\alpha$ la séquence k est considérée voisée.

Si les deux séquences k-1 et k-2 sont non voisées ou bien $r < 0{,}8\,\alpha$ , la séquence k est non voisée. Cependant une vérification doit se faire sur k-2, car si elle est déclarée voisée il faut que k-1 et k-3 ainsi que k soient voisées.

# CONCLUSION

Le but de notre travail était la détection de la fréquence fondamentale du signal de la parole par la technique prédictive.Cette étude nous a permis d'élargir nos connaissances au domaine de l'étude de la parole qui est un nouveau foyer de rec-herches.

Les deux principales méthodes utilisées pour le calcul des coefficients predictifs sont l'autocorrelation et la covariance.Dans le cas de l'algorithme du S.I.F.T pour l'estimation du "pitch", le choix s'est porté sur la méthode d'autocorrélation car elle assure la stabilité du filtre. Ceci diminue l'encombrement de l'algorithme et du programme car ils ne nécessitent pas la détection de la stabilité.

Le problème rencontré dans l'estimation du pitch est que la fréquence $F_o$ peut correspondre au premier formant et non au fondamental.Ce problème se pose surtout pour les voix de femmes pour lesquelles le premier formant est trés proche du fondamental.Dans l'algorithme utilisé un test de comparaison entre périodes de deux séquences consécutives est utilisé ($T_o(K) < T_o(K-1) \times 0,75$) pour detecter cette erreur et la corriger.Néanmoins elle peut subsister surtout dans le cas des sons voisés riches en harmoniques et pour lesquels le filtre inverse ne permet pas d'affaiblir suffisamment les resonances du conduit vocal.Ainsi on peut envisager une amélioration de cet algorithme par une methode plus efficace de détection et correction de cette erreur possible.

Il est aussi trés interéssant d'envisager une réalisation pratique ex-ploitant l'algorithme du S.I.F.T pour la détection du fondamental de certaines voyelles par exemple.Il aurait été souhaitable que nous approfondissions notre étude dans ce sens mais la durée consacrée à ce travail n'aurait pas été suffisante pour mener à bien une telle étude.

Nous avons donc étudié le principe du S.I.F.T en donnant l'organigramme de la décision voisé/non voisé.Une étude détaillée de chaque bloc du schéma fonc-tionnel du détecteur du fondamental permettrait d'écrire un programme à l'aide duquel on pourraitcalculer la valeur de la période du fondamental.Ceci peut être l'objet d'un autre sujet de projet de fin d'études.

# BIBLIOGRAPHIE


BELLANGER,M. (1981) "Traitement Numérique Du Signal".EDITION MASSON.

CINARE,F. & FERRETI,M. (1983) "Synthése,Reconnaissance de la Parole." EDITION TESTS

EL MALAWANY,I. (1975) "Contribution aux Recherches sur la Communication Parlée;Etude deVocoders à Prédiction Linéaire.Determination de l'Intervalle de Fermeture de la Glotte;.Detection de la Mélodie.Extraction de la Fonction D'aire du Conduit Vocal."Thése de Docteur-Ingénieur,Université Scientifique et Médi--cale,GRENOBLE.

EMERIT,E. (1977) "Cours de Phonétique Acoustique."S.N.E.D ALGER.

FONDANECHE,P. & GILBERTAS,P. (1981) "Filtres Numériques,Principes et Réalisation." MASSON.

GUERTI,M. (1983) "Contribution à la Synthèse de la Parole en Arabe Standard."Thése de Magister Univ.d'ALGER.

GUIBERT,J. (1979) "La Parole.Compréhension et Synthèse Par les Ordinateurs."P.U.F.

KUNT,M. (1981) Traitement Numérique Des Signaux."3ème édition DUNOD.

LEGUYADER,A. (1980) "Algorithmes et Thechniques de Codage du Signal de Parole."Note Technique NT/LAA/TSS/18 CNET-LANNION Page 6.

LENARD,J.S. (1977) "Les Processus de la Communication Parlée.Introduction à L'analyse et à laSynthèse de la Parole."MASSON.

LONCHAMP,F. (1978) "Recherche sur les Indices Perceptifs des Voyelles Orales et Nas--ales "Thèse de Doctorat de 3ème Cycle.Univ.NANCY II Chap. III pp-167-178.

RRABINER,L.R. & SCHAFER,R.W. (1978) "Digital Processing of SSpeech Signal;."Prentice Halls.

SERIGNAT,J.F (1974) "Application de la Prédiction Linéaire à L'analyse de la Parole." Bulletin de l'Institut de Phonétique de GRENOBLE Vol III pp23-52

SONG,J.M. (1983) "Relation Entre le Gain et le Pitch ParL.P.C."Rapport de StageUniv. de PARIS. XI .

# A N N E X E

# A N N E X E I

## 1/ Fenêtrage

### 1.1/ Introduction

Le rôle principal de la fenêtre est de limiter la durée d'un signal infini. La position de celle-ci est choisie de manière à conserver les échantillons importants du signal à traiter. Par exemple, en ce qui concerne un signal du type exponentiel, $x(k) = a^{/k/}$ avec $/a/ < 1$, les échantillons de grande amplitude se trouvent au voisinage de l'origine.

La fenêtre sera donc placée autour de l'origine.

Quant au choix de la forme, il depend principalement de la largeur B du pic central de W( f) et l'amplitude des lobes secondaires. Nous allons étudier quelques fenêtres qui sont le plus souvent utilisées à savoir la fenêtre rectangulaire, la fenêtre triangulaire et enfin la fenêtre de Hamming. Celle-ci, par ses caractéristiques, a été choisie dans l'analyse de notre signal de parole.

### 1.2/ Etude de quelques fenêtres, choix de la fenêtre

#### 1.2.1 Fenêtre rectangulaire

$$X_N(k) = X(k) \cdot \text{rect}_N\left(k + \frac{N}{2}\right) \quad (1)$$

ou

$$X_N(k) = \begin{cases} X(k) & \text{pour } |k| \leq N/2 \\ 0 & \text{partout ailleurs} \end{cases} \quad (2)$$

Effet de la limitation de durée

$$X_N(f) = \int_{g_0}^{g_0+1} X(g) \cdot W_R(f-g)\, dg \quad (3)$$

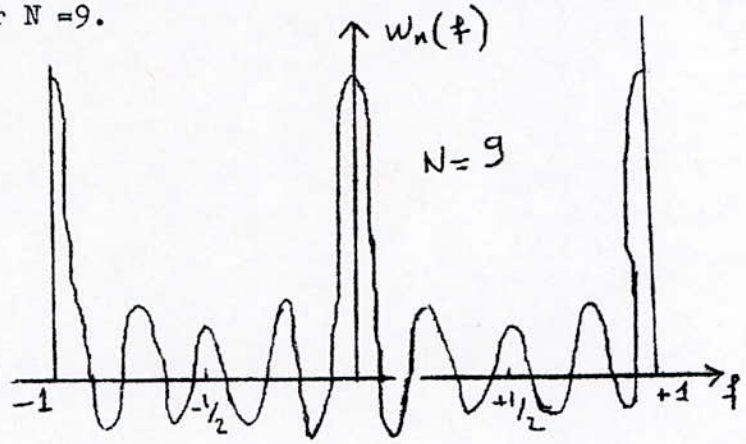
où $X_N(f)$; $X(f)$; et $W_R(f)$ sont respectivement les transformées de Fourier des signaux $X_N(k)$: $x(k)$ et $W_R(k) = \text{rect}_N(k+N/2)$

$W_R(f)$ est une fonction qui correspond à la transformée de Fourier d'une fenêtre temporelle appelée fenêtre spectrale.

$$W_R(f) = \sum_{k=-N/2}^{N/2} e^{-j2\pi f k} = \frac{\sin \pi f n}{\sin \pi f}$$

Traçons la courbe repésentant cette fenêtre spectrale :

pour N =9.




Caracterisation des fenêtres spectrales

Nous remarquons que cette fonction possède un pic central et des lobes secondaires. Deux parametres principaux permettent de les caracteriser. Le premier est la largeur de base du pic central et le second est le rapport de l'amplitude ( du premier lobe secondaire et celle du pic central). Ce dernier est exprimé en décibels de la manière suivante:

$$\lambda_i = 20 \log_{10} \left| \frac{W_i(f_s)}{W_i(0)} \right|$$

Où $f_s$ est la frequence au milieu du lobe secondaire de la fonction $W_i(f)$. Par exemple pour la fenêtre rectangulaire ayant pour fenêtrespectrale $W_R(f)$, pour N = 9 , on a :

$$\begin{cases} D_R = 20 \log_{10} \frac{1}{4,5} \simeq 13\,dB \\ \left| \frac{W_R(0)}{W_R(1,5/N)} \right| = 4,5 \end{cases}$$

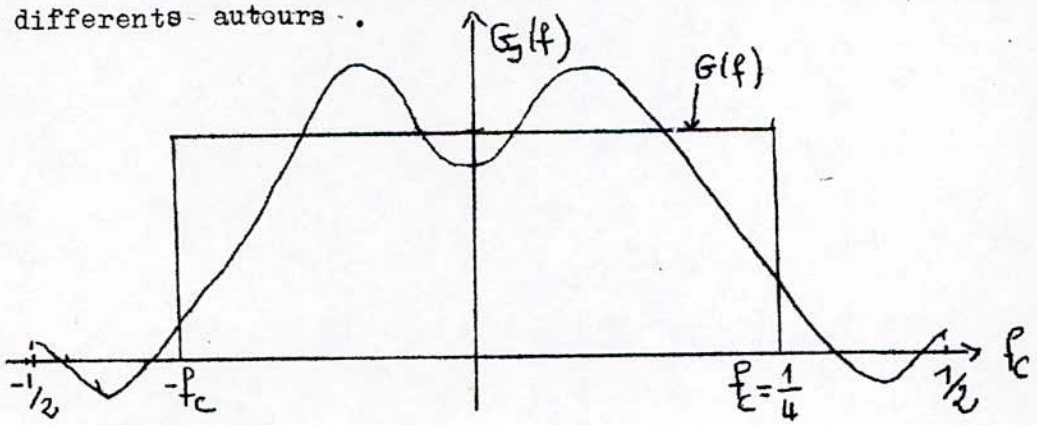
Première remarque: 1/ L'approximation d'une transformée de Fourier X(f) d'un signal à durée illimitée par la fonction $X_N(f)$, obtenue aprés la limitation de la durée, fait apparaitre les oscillations surtout autour des discontinuités de X(f) . Ceci est connu sous le nom de phénomène de Gibbs.

Deuxième remarque: Le rapport des amplitudes du pic central et du premier lobe secondaire de $W_R$ (f) varie trés peu en fonction de N.

Pour N = 50 ce rapport vaut 4,705 et pour N= 100 il vaut 4,711 et pour N =trés grand, un seuil de 4,712 est atteint. Alors que l'augmentation de N a pour effet d'augmenter la frequence des oscillations.

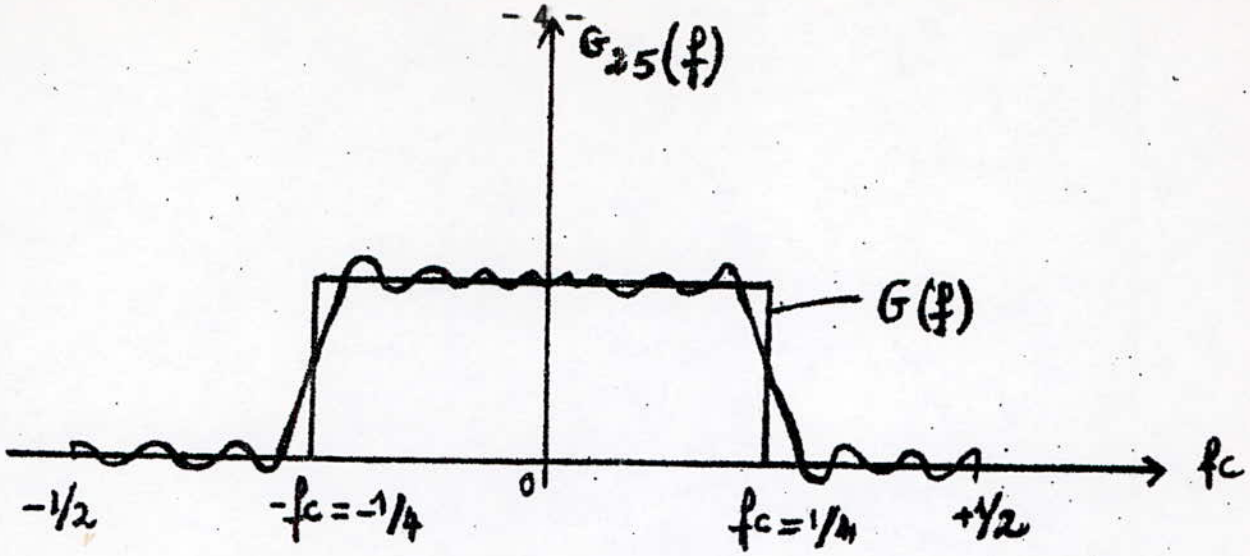
D'aprés cette brève étude, nous constatons que la fenêtre rectangulaire, malgré sa simplicité, n'est pas adaptée à l'analyse de la parole .Aussi, allons nous pouvoir étudier d'autres fenêtres spectrales ayant des caracteristiques susceptibles d'être utilisées en analyse de la parole.

Néanmoins, nous donnerons un aperçu sur les fenêtres utilisées par differents autours .

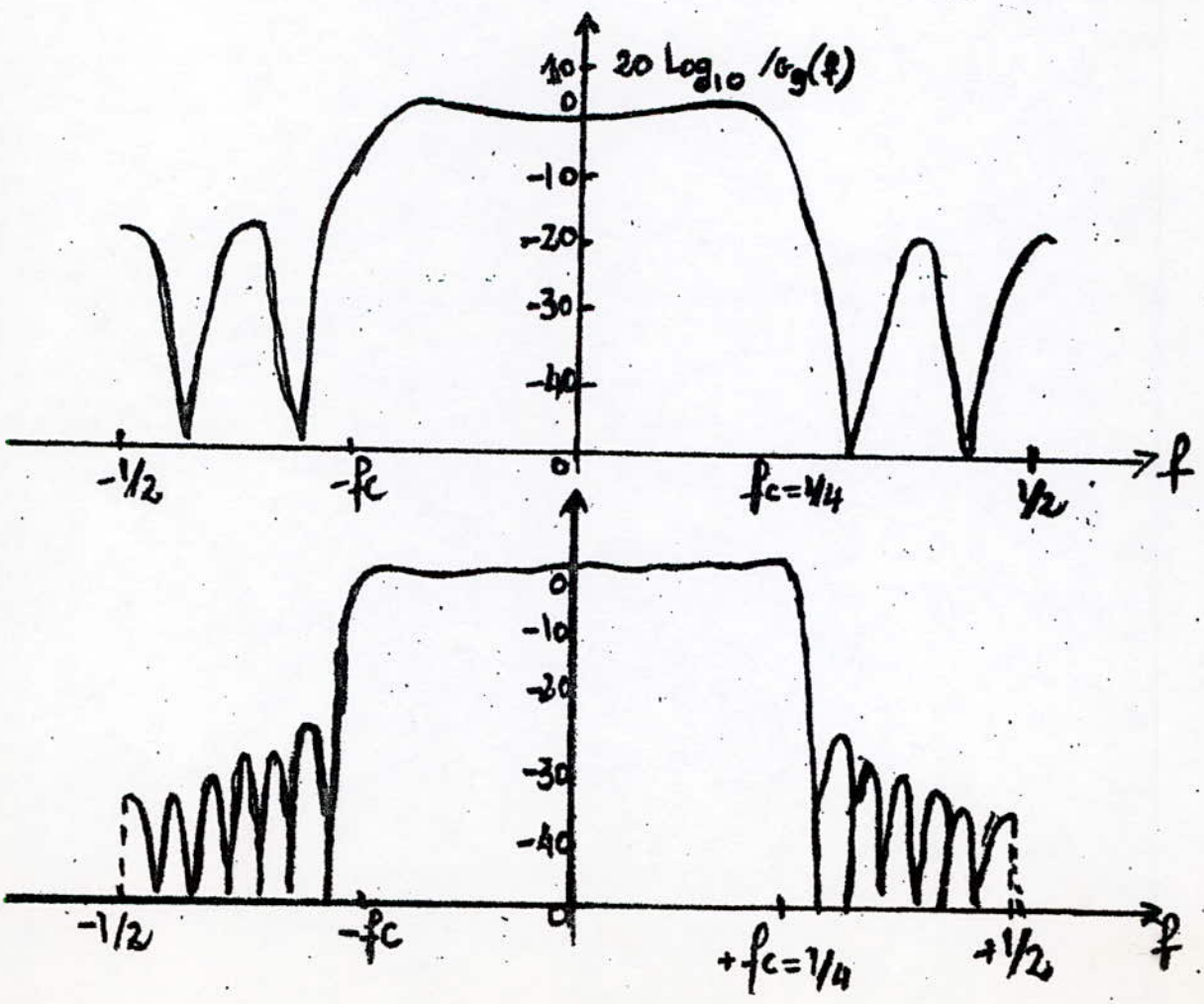



1.2.2/Fenêtre triangulaire

$$W_T(k) = \begin{cases} 1 - 2\frac{|K|}{N} & |K| \leq \frac{N}{2} \\ 0 & \text{ailleurs.} \end{cases}$$




Les erreurs d'approximation $G_N(f)$ sont mieux appréciées par une représentation logarithmique des amplitudes.



Produit de convolution

$$y(k) = \sum_{\ell=-\infty}^{+\infty} x(\ell) \cdot g(k-\ell) \cdot$$

$$y(k) = X(k) * g(k)$$

$$N_T(k) = \frac{2}{N} \text{ rect } \frac{N}{2}(k + N/4) * \text{rect } \frac{N}{2}(k + N/4)$$

produit de convolution d'un signal rectangulaire par lui même. Pour calculer sa transformée de Fourier; il suffit d'élever au carré la transformée de Fourier de la fenêtre rectangulaire

$$W_T = \frac{2}{N} \left( \frac{\sin \pi f N/2}{\sin \pi f} \right)^2$$

Le facteur 2/N provient du fait que $\int_{-\ell/2}^{+\ell/2} W(f)\, df = 1$ sur une période.

La comparaison des figures de $W_R(f)$ et $W_T(f)$ montre clairement l'atténuation des lobes secondaires.

1.2.9/ Fenêtre de Hamming

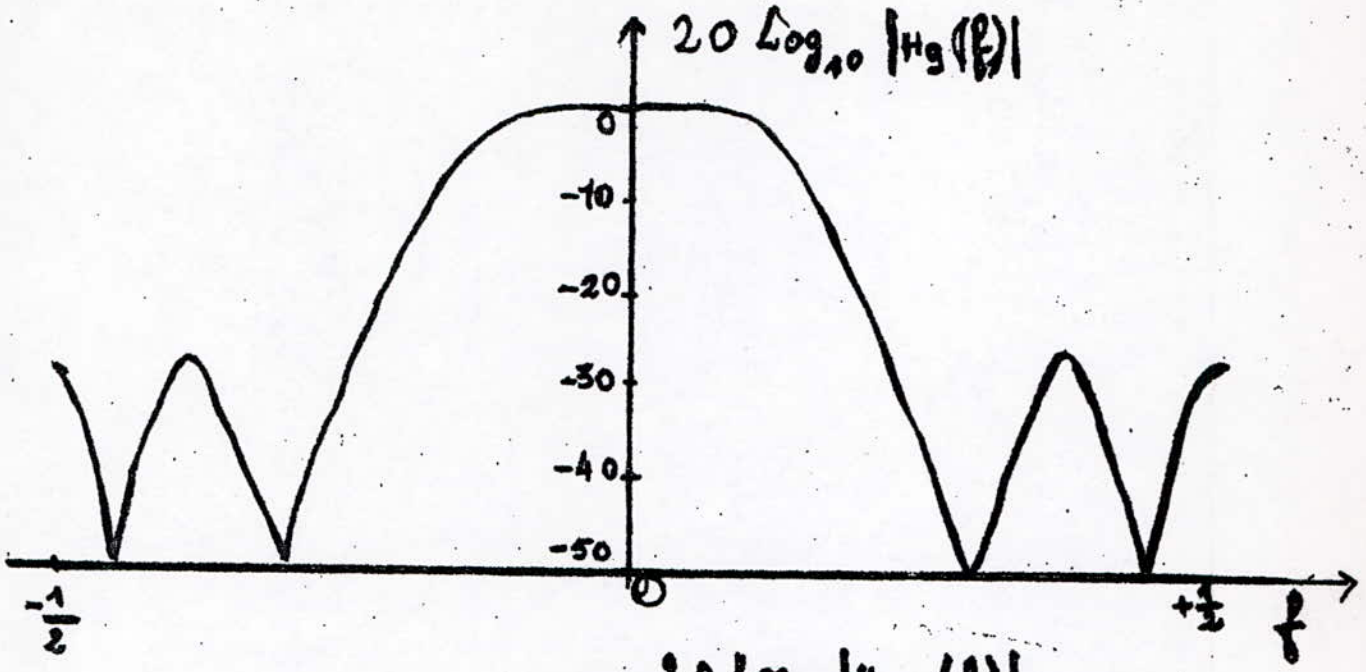
L'atténuation des lobes secondaires sera plus importante en choisissant une fenêtre.

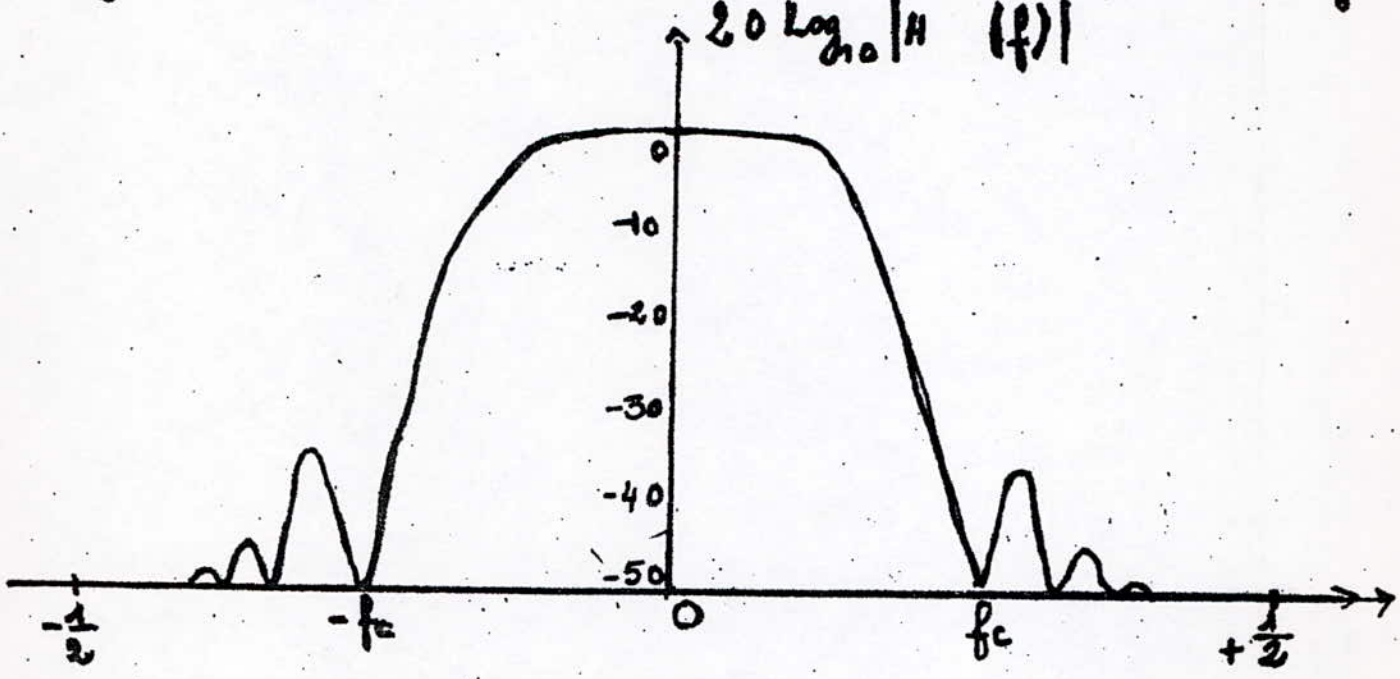
$$W_h(k) = -\frac{1}{4} \exp\left[ \left( j \frac{2\pi k}{N} \right) + 2 + e^{j \frac{2\pi k}{N}} \right]$$

$$= \frac{1}{2} \left( 1 + \cos \frac{2\pi k}{N} \right) \quad \text{pour } |k| \leq -\frac{N}{2} \cdot$$

On peut generaliser la fenêtre de " Hamming " de la manière suivante:

$$W_H(k) = \begin{cases} \alpha + (1-\alpha) \cos\left(\frac{2\pi k}{N}\right) & \text{pour } k \leq -\frac{N}{2} \cdot \\ 0 & \text{ailleurs} \cdot \end{cases}$$


20 Log10 |Hg(f)|
0
-10
-20
-30
-40
-50
-1/2
0
+1/2
f


20 Log10 |H (f)|
0
-10
-20
-30
-40
-50
-1/2
-fc
0
fc
+1/2

Pour $\alpha = 1/2$ on réobtient la fenêtre de Hamming.

La forme générale ci-dessus dépendant du parametre $\alpha$ est appelé fonction de fenêtre de Hamming généralisée.

Si $\alpha = 0,54$, on obtient la fenêtre de Hamming.

Remarque: pour $\alpha = 1$, on obtient la fenêtre rectanguleire.

Conclusion

L'étude comparative des différentes fenêtres d'analyse, montre clairement sur quoi doit se porter notre choix.

Les élémants determinants sont:

- La position
- La forme
- La taille (nombre d'échantillons)

Nous remarquons que la position est choisie de manière à conserver les échantillons importants du signal à traiter. Quant à la forme , son choix dépend de la largeur du pic central et de l'amplitude des lobes secondaires. En ce qui concerne la taille des échantillons, il suffit de réduire les oscillations en limitant N (nombre d'échantillons de la fenêtre).

A cause des propriétés favorables de la fenêtre de Hamming, nous l'utilisons pour l'analyse de la parole.

# A N N E X E  II
## T R A N S F O R M E E  E N  Z.

La transformée en Z est l'outil mathématique permettant l'étude théorique des séquences de nombre $\{ x(k) \}$, elle joue dans le domaine numérique le même rôle fondamental que la rtransformée de Laplace dans le domaine analogique.

x(k) étant un signal échantillonné défini aux instants k = 0 , 1, .....

Par définition, la transformée en Z de ce signal notée X( Z ) est :

$$X(Z) = x(0)\, Z^{0} + x(1)\, Z^{-1} + x(2)\, Z^{-2} + \ldots\ldots\ldots + = \sum_{k=0}^{\infty} x(k)\, Z^{-k}$$

Z est un nombre complexe quelconque. Pour le calcul de la fonction de transfert d'un filtre il est utile de lui donner la valeur $e^{St}$ ( S= jw).

Considérons un signal digitalisé x(k) qui est nul pour $k < 0$ et un autre signal défini par y(k) qui est égal à x ( k- 1). On a donc;

$$y(0) = x(-1)$$
$$y(1) = x(0)$$
$$y(2) = x(1)$$
$$\vdots$$

Le signal y(k) est identique à x(k) à l'exception du fait ou'il est décalé d'un échantillon sur l'axe des temps. En utilisant la définition de la transformée en Z, on peut écrire.

$$Y(Z) = y(0)\, Z^{0} + Y(1)\, Z^{-1} + y(2)\, Z^{-2} + y(3)\, Z^{-3} + \ldots\ldots$$
$$= x(-1)\, Z^{0} + x(0)\, Z^{-1} + x(1)\, Z^{-2} + x(2)\, Z^{-3} + \ldots\ldots$$

or x(k) = 0 pour $k < 0$

$$Y(Z) = X(0)\, Z^{-1} + x(1)\, Z^{-2} + x(2)\, Z^{-3}$$

$$= Z^{-1} \left[ X(0) + x(1)\, Z^{-1} + x(2^{\circ})\, Z^{-2} + \ldots \right]$$

Donc le signal $y(k) = x(k-1)$ a pour transformée en Z:

$$Z^{-1}\, X(Z)$$

on peut généraliser ce résultat à $x(K-m)$ on a:

$$Z^{-m}\, X(Z).$$

# ANNEXE III

# L'ANALYSE DE FOURIER

L'analyse de Fourier est un moyen de décomposer un signal en une somme de signaux élémentaires particuliers qui ont la propriété d'être faciles à mettre en oeuvre et à observer. L'intérêt de cette décomposition repose sur le fait que la réponse au signal d'un système obéissant au principe de supperposition peut être déduite de la réponse aux signaux élémentaires. Ces derniers sont périodiques et complexes afin de permettre une étude en amplitude et en phase des systèmes.

## 3.1/ Transformée de Fourier des fonctions périodiques

Soit x(t) une fonction du temps périodique, de période T. Le développement en série de Fourier de cette fonction est le suivant :

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left( an \cos \frac{2\pi}{T} nt + bn \sin \frac{2\pi}{T} nt \right). \qquad (3.1)$$

avec

$$an = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos \frac{2\pi}{T} nt \; dt \qquad (3.2)$$

$$bn = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin \frac{2\pi}{T} nt \; dt \qquad (3.3)$$

On pose :

$$X(nf_o) = \frac{1}{2} (an - jbn) \; ; \; f_o = \frac{1}{T} \qquad (3.4)$$

On a :

$$X(nf_o) = \frac{1}{T} \int_{-T/2}^{T/2} x(t).e^{-2\pi jnfot} \; dt \qquad (3.5)$$

$X(nf_o)$ est le spectre de fréquence qui peut se décomposer en :

- spectre d'amplitude :

$$/X(nf_o)/ = \tfrac{1}{2}\sqrt{an^2 + bn^2} \qquad (3.6)$$

- spectre de phase :

$$\varphi(nf_o) = \mathrm{Arctg}(-\frac{bn}{an}) \qquad (3.7)$$

Réciproquement on a :

$$X(f) = \sum_{n=-\infty}^{+\infty} /X(nf)/ e^{-j\varphi(nf_o)} \qquad (3.8)$$

$$x(t) = e^{2\pi jnf_o t} \frac{1}{T} \int_{-T/2}^{+T/2} x(\sigma) e^{-2\pi jfn\sigma} d\sigma \qquad (3.9)$$

Le spectre d'une fonction périodique de période T est composé de raies dont l'écart est sur l'axe des fréquences $f = \frac{1}{T}$.

3.2/Transformation de Fourier des fonctions non périodiques.

Soit x(t) une fonction non périodique de la variable t.

On a :
$$x(t) = \int_{-\infty}^{+\infty} X(f) e^{j2\pi ft} df \qquad (3.10)$$

avec :
$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft} dt \qquad (3.11)$$

La fonction X(f) est la transformée de Fourier de x(t). X(f) est appelé spectre du signal x(t).

3.3/Produit de convolution.

Soient deux fonctions x(t) et h(t) dont les transformées de Fourier sont respectivement X(f) et H(f).Le produit de convolution y(t) est défini par :

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(t - \tau) \cdot h(\tau)\, d\tau \qquad (3.12)$$

La transformée de Fourier de ce produit s'écrit :

$$Y(f) = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} x(t - \tau) \cdot h(\tau)\, d\tau \right) e^{-j2\pi ft} dt \qquad (3.13)$$

$$Y(f) = \int_{-\infty}^{+\infty} h(\tau) e^{-j2\pi ft} d\tau \; x(t - \tau) e^{-j2\pi ft} dt$$

$$Y(f) = \int_{-\infty}^{+\infty} h(\tau) e^{-j2\pi ft} d\tau \int_{-\infty}^{+\infty} x(u) e^{-j2\pi fu} du \qquad (3.14)$$

avec : $u = t - \tau$

$$Y(f) = H(f) \cdot X(f) \qquad (3.15)$$

Conclusion :la transformée de Fourier d'un produit de convolution est un produit simple .Réciproquement on montre que la transformée de Fourier d'un produit simple est un produit de convolution.

3.4/Fonction d'autocorrélation.

La fonction d'autocorrélation permet de comparer un signal à un instant donné à ce même signal à un autre instant.

Soit x(t) une fonction de la variable t. Sa fonction d'autocorrélation est :

$$g(\tau) = \int_{-\infty}^{+\infty} x(t).\, x(t-\tau)\, dt. \qquad (3.16)$$

ou

$$g(\tau) = \int_{-\infty}^{+\infty} P(f) e^{j2\pi f\tau}\, df \qquad (3.17)$$

P(f) étant la densité spectrale de puissance de x(t)

$$P(f) = \int_{-\infty}^{+\infty} g(\tau) e^{-2\pi j f\tau}\, d\tau = / X(f) /^{2} \qquad (3.18)$$

Conclusion : La densité spectrale de puissance de x(t) est égale à la transformée de Fourier de sa fonction d'autocorrélation.

3.5/ La Transformation de Fourier Discrète

La Transformation de Fourier Discrète s'introduit quand il s'agit de calculer la transformée de Fourier d'une fonction à l'aide d'un calculateur numérique. Il s'ensuit que la transformée de Fourier :

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft}\, dt$$

doit être adaptée, d'une part en remplaçant le signal x(t) par des nombres x(nT),

et d'autre part en limitant l'ensemble des nombres sur lesquels portent les calculs à une valeur finie N. Le calcul fournit alors des nombres $X^*(f)$ définis par :

$$X^*(f) = \sum_{n=0}^{N-1} x(nT)e^{-j2\pi fnT} \qquad (3.19)$$

Comme le calculateur est limité dans sa puissance de calcul, il ne peut fournir ces résultats que pour un nombre limité de valeurs de la fréquence f, qui est choisi multiple d'un certain pas de fréquence $\Delta f$. Alors :

$$X^*(K\ f) = \sum_{n=0}^{N-1} x(nT)e^{-j2\pi K\Delta fnT} \qquad (3.20)$$

Pour plus de simplification, un choix consiste à prendre :

$$\Delta f = \frac{1}{NT} \qquad (3.21)$$

Dans ce cas il existe seulement N valeurs différentes dans la suite des $X^*(\frac{K}{NT})$ qui est une suite périodique de période N. Cette suite est obtenue par transformation de Fourier de la suite des x(nT) qui est une suite périodique de période nT.

### 3.5.1/ Définition de la TFD et propriétés

Soient deux suites de nombres complexes x(n) et X(K) périodiques et de période N. La transformée de Fourier discrète et la transformée inverse établissent entre ces deux suites les relations suivantes respectivement :

$$X(K) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)e^{-j2\pi\frac{nK}{N}} \qquad (3.22)$$

$$x(n) = \sum_{n=0}^{N-1} X(k)e^{j2\pi\frac{Kn}{N}} \qquad (3.23)$$

Cette transformation possède les propriétés suivantes :

- Linéarité :

si $x_1(n) \longleftrightarrow X_1(K)$

$x_2(n) \longleftrightarrow X_2(K)$

alors :

$$a\,x_1(n) + b\,x_2(n) \longleftrightarrow aX_1(K) + b\,X_2(K)\ ,\ \forall\ a \text{ et } b \qquad (3.24)$$

- Symétrie : si la suite x(n) est réelle les nombres X(K) et X(N-K) sont complexes et conjugués.

$$\overline{X}(N-K) = X(K) \qquad (3.25)$$

- Décalage temporel :

Une translation des x(n) entraine une rotation de phase des X(K)

$$x(n-n_0) \longleftrightarrow X(K)e^{-j2\pi n_0 K/N} \qquad (3.26)$$

- Décalage fréquentiel :

$$x(n)e^{j2\pi K_0 n/N} \longleftrightarrow X(K-K_0) \qquad (3.27)$$

- Fonction réelle paire :

$$\text{si } x(N-n) = x(n) \Rightarrow X(N-K) = X(K) \qquad (3.28)$$

- Fonction réelle impaire :

$$\text{si } x(N-n) = -x(n) \Rightarrow X(K) = -X(N-K) \qquad (3.29)$$

- Egalité de Parseval : La puissance du signal est égale à la somme des puissances de ses harmoniques

$$\frac{1}{N} \sum_{n=0}^{N-1} /x(n)/^2 = \sum_{K=0}^{N-1} /X(K)/^2 \qquad (3.30)$$

### 3.5.2/ La transformation de Fourier Rapide (F.F.T)

On appelle transformée de Fourier rapide un algorithme permettant de calculer d'une façon particulièrement performante la transformée de Fourier discrète.

La transformée de Fourier discrète peut s'écrire sous une forme matricielle en posant :

$$W = e^{-j2\pi/N} \tag{3.31}$$

Les nombres $W^n$ sont appelés coefficients de la T.F.D. Leurs affixes se trouvent sur le cercle unité. Ce sont les racines de l'équation $Z^N - 1 = 0$.

L'équation matricielle est la suivante :

$$\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{N-1} \end{bmatrix} = 1/N \begin{bmatrix} 1 & 1 & 1 \ldots\ldots\ldots & 1 \\ 1 & W & W^2 \ldots\ldots\ldots & W^{N-1} \\ 1 & W^2 & W^4 \ldots\ldots\ldots & W^{2(N-1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & W^{N-1} & W^{2(N-1)} \ldots\ldots & W^{(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{(N-1)} \end{bmatrix} \tag{3.40}$$

Pour la transformée inverse, il suffit de retirer le scalaire $1/N$ et de changer $W^n$ en $W^{-n}$ .

La matrice carrée d'ordre N désignée par $T_N$ présente des particularités évidentes . Les lignes et les colonnes de même indice ont les mêmes éléments et ces éléments sont des puissances d'un nombre de base W tel que $W^N = 1$. Des simplifications importantes peuvent être envisagées dans ces conditions, conduisant à des algorithmes de calcul rapide. Quand la TFD est calculée à l'aide de tels algorithmes on dit que l'on effectue une transformation de Fourier Rapide (T.F.R).