

16/86

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

Aex

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT ELECTRONIQUE

المدرسة الوطنية للتكنولوجيا
BIBLIOTHEQUE - المكتبة
Ecole Nationale Polytechnique

PROJET DE FIN D'ETUDES

S U J E T

CALCUL DU GAIN
EN PREDICTION LINEAIRE
EN VUE DE LA SYNTHESE
DE LA PAROLE

Proposé par :
Guerti . M

Etudié par :
Ghobrini . M
Ghermoul . M

Dirigé par :
Guerti . M

PROMOTION :
Janvier 1986

T E S E

D'INGENIEUR D'ETAT
EN ELECTRONIQUE

-o-

Présentée à

L'ECOLE NATIONALE
POLYTECHNIQUE
D'El-Harrach



-o-

Thème du sujet

CALCUL DU GAIN
EN PREDICTION LINEAIRE
EN VUE DE LA SYNTHESE
DE LA PAROLE

-o-

Etudiée par

GHOBRINI Mourad
et
GHERMOUL Mohamed

-o-

PROMOTION
Janvier 1986

-o-o-o-o-o-o-o-o-o-o-

Védicaces

A mes parents

A mes frères et soeurs

A toute ma famille

A tous ceux qui me sont chers

-MOHAMED-



A mes parents

A mes frères et à ma soeur

A tous ceux que j'estime

A mes amis

-MOURAD-

.....

Remerciements

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

Nous tenons à exprimer nos plus vifs remerciements à notre promoteur Melle M.GUERTI, pour son aide et ses conseils tout au long de l'élaboration de ce projet de fin d'étude.

Nous tenons à remercier également Melle A.MOUSSAOUI, pour son aide bibliographique, qui nous a été très bénéfique.

Nous sommes particulièrement reconnaissant à Monsieur MERANI Laid, pour tous les efforts qu'il a fournis pour l'exécution de ce document.

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

Nous ne pouvons terminer, sans remercier tous ceux qui ont participé, de près ou de loin à la réalisation de ce travail.

.....

TABLE DES MATIERES



PAGES

INTRODUCTION GENERALE

CHAPITRE (I) : L'APPAREIL PHONATOIRE HUMAIN

I-1-Description	1
I-2-Fonctionnement acoustique	2
I-3-Caractéristiques acoustiques du signal de parole ...	4
I-4-Les différents types de sons	5
I-5-Classification des différents sons	6

CHAPITRE (II) : ELEMENTS DE TRAITEMENT DU SIGNAL

II-1-Transformation de Fourier	10
II-2-Echantillonnage et quantification	10
II-3-Convolution et corrélation	11
II-4-Filtres numériques	14
II-5-Fenêtrage	16

CHAPITRE (III) : ANALYSE ET SYNTHÈSE DE LA PAROLE

Introduction	19
III-1-Méthodes d'analyse	20
III-1-1-Analyse spectrale	20
III-1-2-Analyse temporelle	21
III-2-Méthodes et techniques de synthèse	28
III-2-1-Techniques de synthèse	29
III-2-2-Méthodes de synthèse	33
Conclusion	34

**CHAPITRE (IV) : PRINCIPALES METHODES DE PREDICTION LINEAIRE
ET LEURS SOLUTIONS**

Introduction35

IV-1-Méthode de Covariance36

IV-2-Méthode d'Autocorrélation37

IV-3-Erreur quadratique totale minimale de prédiction ..39

IV-4-Résolution du système de prédiction linéaire41

IV-4-1-Méthode de Cholesky41

IV-4-2-Algorithme de Durbin44

IV-5-Stabilité du filtre prédicteur46

Conclusion47

CHAPITRE (V) : DETERMINATION DU GAIN

Introduction49

V-1-Détermination du facteur du gain du modèle50

V-1-1-Calcul du gain pour les sons voisés51

V-1-2-Calcul du gain pour les sons non-voisés53

V-1-3-Remarque55

Conclusion55

V-2-Programmation56

CONCLUSION GENERALE59

ANNEXE61

BIBLIOGRAPHIE66

.....

INTRODUCTION GENERALE

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

Les hommes ont depuis toujours cherché à améliorer leurs moyens de communication et à s'affranchir des barrières du temps et de l'espace. Par contre ces dernières années ils se sont penchés de plus en plus sur la communication "homme-machine", surtout avec l'apparition des ordinateurs.

Quant au traitement de l'information il a été et demeure le privilège de l'être humain, toutefois depuis le milieu de ce siècle le développement prodigieux de l'informatique s'il ne remet pas en cause ce privilège confère à l'homme une puissance nouvelle et lui ouvre un champs immense de possibilités dont on ne mesure pas encore toute l'étendue.

Le signal de parole comme tous les autres signaux acoustiques peut-être transformé en un signal électrique et se trouve ainsi transmis, traité, conservé en mémoire sous forme analogique ou numérique. Le signal de parole peut enfin être un support d'informations émises ou reçues par une machine. Les transformations correspondantes sont respectivement, la synthèse et la reconnaissance de la parole.

L'intérêt donné à la synthèse de la parole n'est pas nouveau. Les premières recherches datent du "XVIIIème" siècle. Des travaux et manuscrits anciens font mentionner des machines parlantes. Aujourd'hui elle représente un secteur de recherche très important.

Les machines parlantes que nous connaissons reproduisent les sons enregistrés par l'intermédiaire de hauts-parleurs, mais depuis longtemps on a songé à construire des machines destinées non à reproduire mais à imiter la voix humaine. Leur histoire est déjà longue, les premières étaient mécaniques et acoustiques, elles sont désormais électroacoustiques et électroniques et leurs applications sont devenues très intéressantes dans de très nombreux domaines. On peut ainsi concevoir des machines à lire des textes et qui parlent, à l'intention des aveugles, de même qu'on peut songer à réaliser inversement des machines à écrire fonctionnant directement sous l'action de la parole. Les systèmes de production de la parole sont adaptés sur les ordinateurs, qui peuvent ainsi directement répondre à nos questions en langage parlé direct, sans avoir besoin d'aucune conversion.

Parmi les techniques de synthèse, nous parlerons en particulier de la "PREDICTION LINEAIRE" qui est l'une des plus utilisées, vu le grand intérêt qu'elle offre en raison de sa facilité de mise en oeuvre et du niveau de qualité qu'elle permet d'atteindre.

Bien que l'estimation des moindres carrés (ou prédiction) date de "Gauss" (1795), la première utilisation du terme "PREDICTION LINEAIRE" fut par "Wiener" (1949). Depuis ce temps elle a été appliquée dans plusieurs domaines et ce n'est qu'en 1966 que "Itakura" et "Saito" l'on

utilisé pour la première fois dans le domaine de la parole. Depuis ce temps elle est devenue l'une des techniques les plus puissantes dans l'analyse-synthèse de la parole. Elle est utilisée pour estimer les paramètres de base de la parole, tels que: les formants, le pitch, les coefficients prédicteurs et "LE GAIN" en particulier qui est l'objet même de notre étude.

Le travail que nous présentons consiste à déterminer le gain des sons voisés (ou des sons non voisés), qui intervient lors de la synthèse de la parole par la technique de prédiction linéaire.

-Le chapitre (I) traite l'anatomie et la physiologie de l'appareil vocal humain, ainsi que la classification des différents sons du langage.

-Le chapitre (II) est consacré à l'étude de quelques éléments de base du traitement du signal indispensable pour la suite.

-Au chapitre (III) nous verrons quelques techniques d'analyse et de synthèse du signal de parole en insistant bien sûr, sur la prédiction linéaire qui sera la technique utilisée par la suite.

-Le chapitre (IV) est consacré à la résolution du système de prédiction linéaire par les méthodes "D'AUTOCORRELATION" et de "COVARIANCE".

-Dans le chapitre (V) nous présenterons la méthode de détermination du gain, et le programme correspondant.

المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

.....

CHAPITRE I

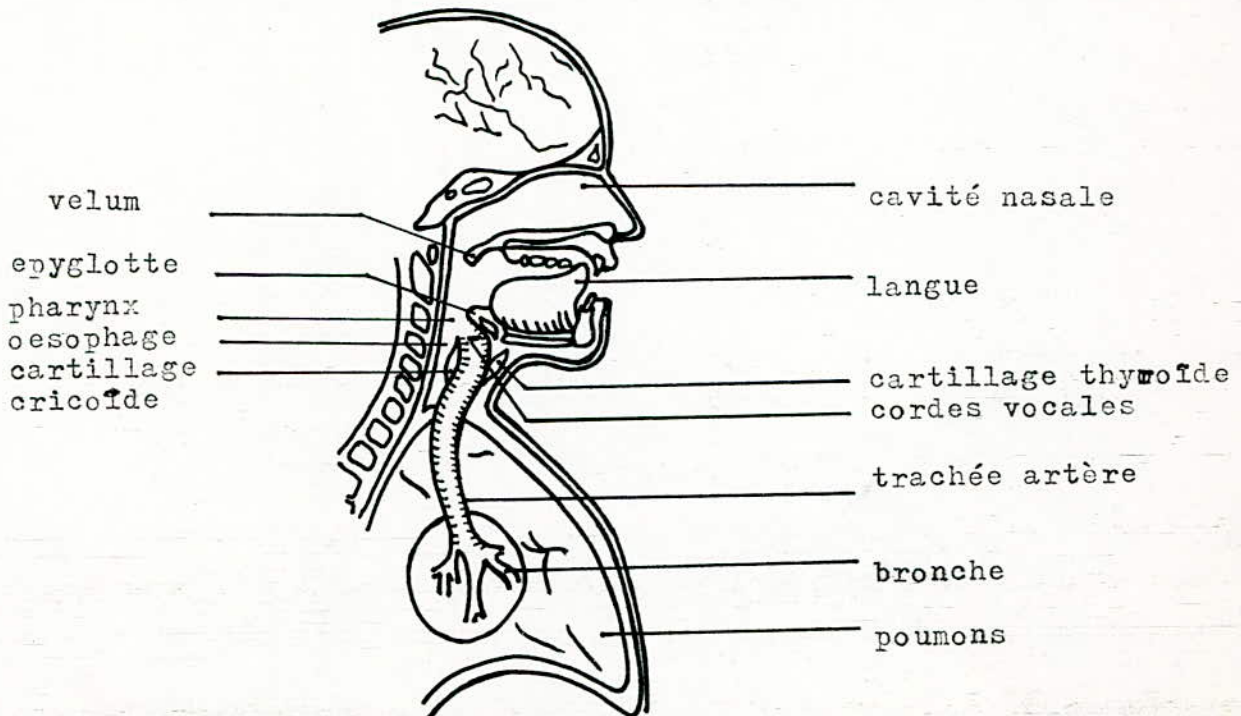
APPAREIL PHONATOIRE HUMAIN

Nous examinerons brièvement dans ce chapitre l'anatomie et la physiologie de l'appareil phonatoire humain ainsi que son fonctionnement acoustique.

Cette étude quelque peu simplifiée nous permettra de comprendre la formation des sons élémentaire, dans le but de l'analyse et la synthèse de la parole.

I-I-DESCRIPTION

La phonation n'est pas réalisée au moyen d'un appareil spécifique. Mis à part la larynx, les organes qu'elle met en jeu sont affectés aussi aux fonctions vitales: la respiration et la nutrition, fig(I.I)



Fig(I.I) : Appareil phonatoire humain

I-I-I-Les poumons:

Bien qu' étant les organes principaux de l'appareil respiratoire ils contribuent aussi lors de la production de la parole comme un générateur d'air.

I-I-2-La trachée artère:

La trachée artère est un conduit quasi cylindrique supportant à son extrémité supérieure un organe très important (jouant le triple rôle; de respiration, de nutrition, et de phonation) appelé: "LE LARYNX". Ce dernier supporte deux muscles: "LES CORDES VOCALES". L'espace circonscrit par celles-ci est "LA GLOTTE".

I-I-3-Le conduit vocal:

Le conduit vocal se compose de deux parties:

-La partie orale qui comprend les cavités pharyngale et buccale de formes et de volumes variables.

-La partie nasale composée de deux cavités fixes; les fosses nasales connectées en parallèle sur le conduit oral par l'intermédiaire du voile du palais (ou "VELUM").

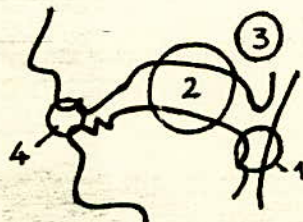
I-2-FONCTIONNEMENT ACOUSTIQUE

Un système acoustique peut-être généralement décomposé en trois parties fonctionnelles que l'on appelle; "GENERATEUR", "EXCITATEUR", et "RESONATEUR".

-La source d'énergie est représentée par les poumons.

-La source d'excitation est caractérisée par les cordes vocales.

-Le résonateur ou système résonnant est composé de quatre cavités; pharyngale, buccale nasale et labiale, fig(I.2).

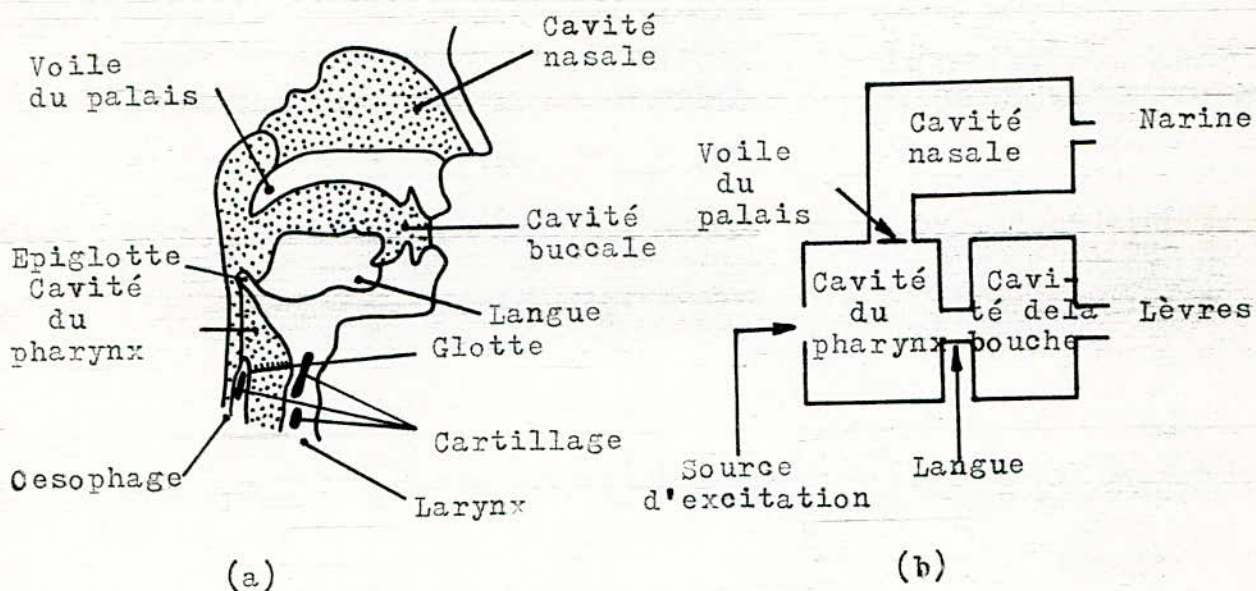


- 1 : cavité pharyngale
- 2 : cavité buccale
- 3 : cavité nasale
- 4 : cavité labiale

Fig(I.I) : Les quatre résonateurs de l'appareil phonatoire

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

Le fonctionnement acoustique du signal de parole est schématisé comme suit, fig(1.3).



Fig(1.3) : (a), le conduit vocal se compose essentiellement de la cavité du pharynx et de la cavité de la bouche. Le phénomène physique "PAROLE" résulte de l'excitation de ce système par l'ensemble poumons-cordes vocales pour les voyelles, et par un courant d'air agissant sur les parois du conduit pour certaines consonnes. La cavité nasale intervient également lors de la production des consonnes du type "m" ou "n" et des voyelles nasales du type "on".

Fig(1.3) : (b), on a schématisé ici les trois cavités couplées qui constituent l'appareil vocal humain. La cavité nasale, indéformable est couplée au pharynx par le voile du palais; la position de la langue règle le volume et le couplage des deux autres cavités. Chaque cavité, prise isolément, résonne à des fréquences déterminées par sa forme et son volume: le couplage de deux cavités fait apparaître des fréquences du système déplacées par rapport aux fréquences propres de chacune d'elles.

I-3-CARACTERISTIQUES ACOUSTIQUES DU SIGNAL DE PAROLE

-Le signal glottique a un spectre de raies périodique, sa fréquence est appelée "FREQUENCE FONDAMENTALE ou PITCH". Elle varie de 70 à 150KHZ pour les hommes, et de 150 à 300KHZ pour les femmes et peut dépasser 400KHZ pour les enfants.

-Le spectre transmis dans la parole s'étend de 50KHZ à 3KHZ pour les sons voisés et jusqu'à 6KHZ pour les sons non voisés.

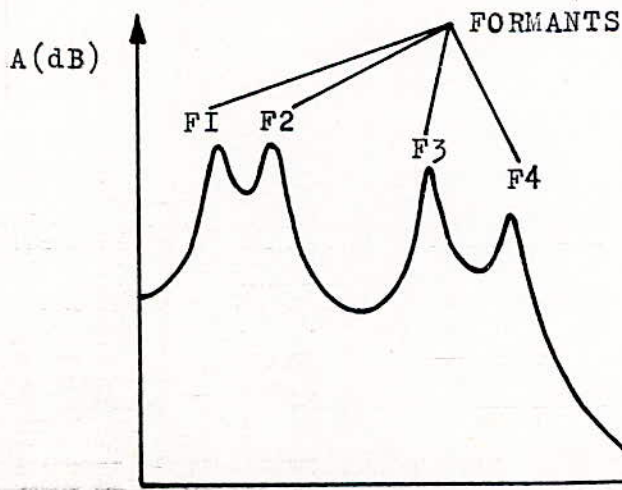
-La dynamique d'un signal peut-être définie comme étant l'écart en dB entre le niveau maximum et le niveau minimum de ce signal.

$$D = (N_{max} - N_{min}) \text{ dB}$$

La dynamique de la parole est inférieure à 50 dB.

-La fonction de transfert du conduit vocal présente un certain nombre de pôles appelés "FORMANTS", fig(I.4).

-La fonction de transfert de l'ensemble conduit vocal / conduit nasal présente également des zéros appelés "ANTIFORMANTS".



Fig(I.4) : Spectre vocalique présentant quatre résonances formantiques.

I-4-LES DIFFERENTS TYPES DE SONS

I-4-1-Définition d'un son:

Un son peut-être défini comme étant une variation de pression dans l'air, l'eau ou autre milieu quelconque pouvant être perçu par l'oreille humaine.

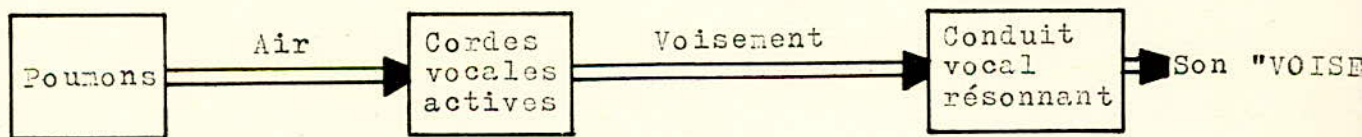
Dans le domaine de la phonétique un son est une vibration provoquée par l'appareil vocal dans un milieu matériel dans lequel elle peut se propager pour exciter notre ouïe.

I-4-2-Caractères sonore-sourd:

L'appareil vocal présente deux catégories de son:

-Sonore ou voisé:

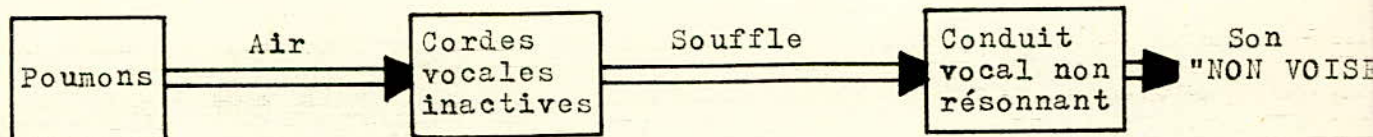
La production de ce type de son fait intervenir une source d'impulsions périodiques constituée par l'ensemble pouxons, cordes vocales, exemple: les voyelles, fig(I.5).



Fig(I.5) : Production d'un son voisé

-Sourd ou non voisé:

Ce type de son est engendré par une source de bruit. Ce dernier est produit par l'action du courant d'air issu des pouxons à travers une constriction située en un point quelconque du conduit vocal,fig(I.6).



Fig(I.6) : Production d'un son non voisé

I-4-3-Caractères oral-nasal:

On a vu précédemment que les cavités orale et nasale étaient connectées par l'intermédiaire du velum. En se relevant à l'horizontale le velum interdit toute communication entre les cavités orale et nasale, on a dans ce cas un "son oral". Par contre si le velum est abaissé on obtient un "son nasal".

I-5-CLASSIFICATION DES DIFFERENTS SONS

Cette classification est basée sur les critères suivants:

- Caractères sonore-sourd.
- Caractères oral-nasal.
- Le mode d'articulation.
- Le lieu d'articulation.

De ce fait les sons du langage ont été divisés en deux classes; les voyelles et les consonnes.

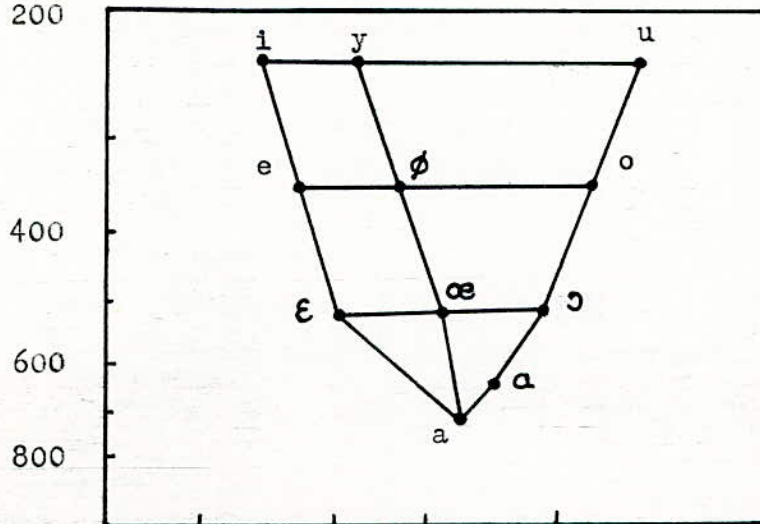
I-5-I-Les voyelles:

Les voyelles sont caractérisées par le passage libre de l'air. La source d'excitation étant la vibration des cordes vocales.

On distingue:

- Les voyelles orales, ex: /a/, /u/, /i/
- Les voyelles nasales, ex: /ã/, /õ/

F1(HZ) 200



Fig(I.7) : Fréquences des deux premiers formants (D'après Delattre)

I-5-2-Les consonnes:

Les consonnes sont caractérisées par une constriction ou fermeture du passage de l'air.

On distingue trois sortes de consonnes:

-Les fricatives ou constrictives:

Elles sont produites par le passage continu de l'air à travers un rétrécissement du conduit vocal. Celles-ci sont dites:

.voisées quand elles sont produites par l'association du signal bruité et de la vibration des cordes vocales, ex: /z/, /v/.

.non voisées quand elles sont dues uniquement à la source bruitée, ex: /f/, /s/.

-Les plosives ou occlusives:

Pour produire une plosive il faut procéder à une fermeture plus tenue en un point particulier du conduit vocal puis à un relâchement brusque.

.elles peuvent être aussi voisées, ex: /b/, /d/, /g/.

.ou non voisées, ex: /p/, /t/, /k/.

-Les nasales:

Elles sont dues à un abaissement du voile du palais, ex: /n/, /m/;

-Les semi-voyelles et les liquides:

Les semi-voyelles /w/, /y/, /j/ sont des transitions rapides du conduit vocal ce qui les apparente aux consonnes, mais celles-ci continuent à fonctionner en mode résonnant ce qui les apparente aux voyelles d'où leur nom.

Les liquides, /l/, /r/, font intervenir un phénomène de même nature que les semi-voyelles, mais ces deux consonnes sollicitent d'autres modes acoustiques, fig(1.8).

	OCCLUSIVES		CONSTRUCTIVES		LIQUI- DES	NASA- LES	SEMI- VOYEL- LES
	sourdes	sonores	sourdes	sonores	sonores	sonores	sonores
BI-LA- BIALES	p	b				m	w/y
LABIO- DENTALES			f	v			
DENTALES	t	d	s	z	l	n	
PALATO- ALVEO- LAIRES			ʃ	ʒ			
PALATALES							j
VELAIRES	k	g					
UVULAIRES					R		

Fig (I.8) : Tableau des consonnes françaises,
(D'après MALMBERG)

CHAPITRE II

ELEMENTS DE TRAITEMENT DU SIGNAL

INTRODUCTION

Les domaines d'application du traitement du signal sont de plus en plus nombreux. On peut citer le domaine médical pour l'analyse des signaux biologiques, comme l'électrocardiogramme et l'encéphalogramme ou pour des reconstitutions analytiques comme en systèmes radars et sonars, etc...

Cependant c'est dans le domaine des télécommunications que se trouve ses applications les plus importantes.

Notre but est que le lecteur trouve dans ce chapitre les notions qui nous sont apparues comme fondamentales, pour tout ceux qui ont à utiliser, de manière pratique, les méthodes du traitement du signal de la parole. Il reste bien entendu que la pratique n'exclue en aucun cas les connaissances théoriques indispensables.

Pour ces raisons nous commencerons par donner un très bref aperçu des éléments du traitement (numérique en particulier) afin d'aborder au chapitre suivant, deux points très important dans le traitement de la parole : "L'ANALYSE" et "LA SYNTHÈSE".

.....

II-1-TRANSFORMATION DE FOURIER (TF)

Un signal quelconque peut-être décrit par deux représentations possibles :

-Une représentation "TEMPS" de la forme : $y = f(t)$

-Une représentation "FREQUENCE" de la forme : $Y = F(f)$

Ces deux représentations sont reliées entre elles par la TF et il est indispensable de les manipuler correctement.

Par exemple l'autocorrélation d'un signal est définie dans le domaine temporel alors que son spectre de puissance est décrit dans le domaine des fréquences.

La transformation de Fourier directe fournit le spectre d'un signal défini dans le domaine temporel.

La transformation de Fourier inverse permet de revenir du spectre au signal initial.

-TF DES SIGNAUX CONTINUS:

$$X(f) = \int_{-\infty}^{+\infty} x(t) \cdot \exp(-2\pi jft) dt$$

$$x(t) = \int_{-\infty}^{+\infty} X(f) \cdot \exp(+2\pi jft) df$$

-TF DES SIGNAUX DISCRETS:

$$X(f) = \sum_{k=-\infty}^{+\infty} x(k) \cdot \exp(-2\pi jfk)$$

$$x(k) = \int_{-1/2}^{1/2} X(f) \cdot \exp(+2\pi jfk) df$$

Ces relations définissent la transformation de Fourier discrète (TFD), qui est à la base de la transformation de Fourier rapide (TFR).

II-2-ECHANTILLONNAGE ET QUANTIFICATION

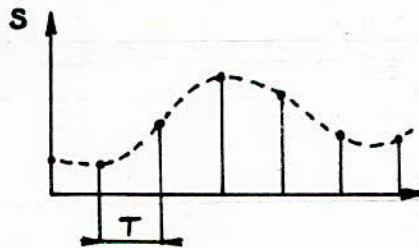
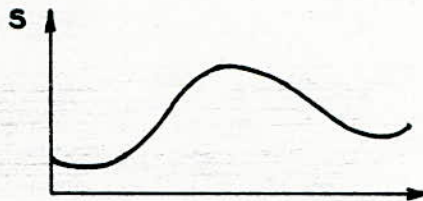
En traitement du signal les techniques numériques apportent des possibilités prodigieuses et présentent d'énormes avantages, par rapport aux techniques analogiques, de ce fait l'opération d'échantillonnage revêt une importance capitale.

II-2-I-Echantillonnage:

L'échantillonnage consiste à représenter une fonction du temps $s(t)$ par ses valeurs $S(nT)$ à des instants multiples d'une durée T , appelée période d'échantillonnage.

Echantillonner une fonction $s(t)$ revient à multiplier cette fonction par une suite d'impulsions de Dirac (Peigne de Dirac) de fréquence F_e , fig(2.1).

$$S_e(t) = S(t) \cdot \sum_{-\infty}^{+\infty} \delta(t - k/F_e)$$



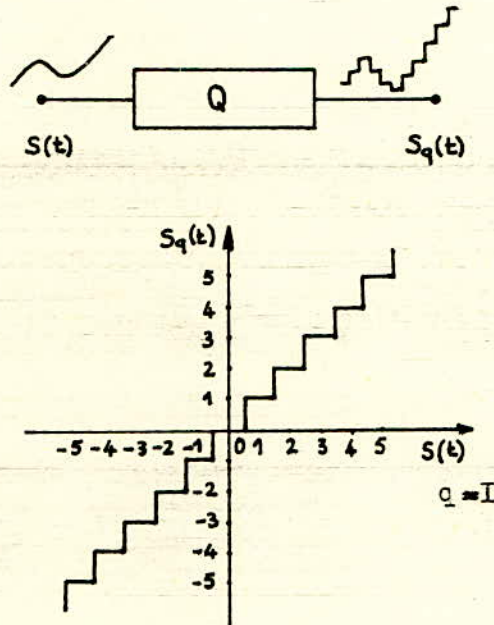
Fig(2.1) : l'opération d'échantillonnage

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

II-2-Quantification:

L'opération de quantification consiste à approcher par un multiple entier d'une quantité élémentaire " q ", chaque valeur $S(nT)$ dans l'espace des amplitudes.

" q " est appelé échelon de quantification, fig(2.2).



Fig(2.2) : l'opération de quantification

II-3-CONVOLUTION ET CORRELATION

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

II-3-I-Convolution:

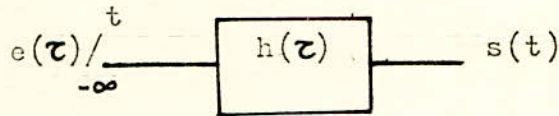
La convolution est définie à partir de la réponse impulsionnelle par une intégrale dans le cas des signaux analogiques. Ainsi si "e(t)" désigne le signal à filtrer, "h" la réponse impulsionnelle du filtre, le signal filtré "s(t)" est donné par :

$$s(t) = \int e(\tau) \cdot h(t-\tau) d\tau$$

$$\text{ou } s(t) = \int e(\tau) \cdot h(t-\tau) d\tau$$

Elle s'écrit symboliquement :

$$s(t) = e(t) * h(t)$$



Fig(2.3) : l'opération de convolution

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

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

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

C'est à dire qu'on peut faire le passage :

- Filtrage temporel = convolution fréquentielle
- Filtrage fréquentiel = convolution temporelle

II-3-2-Corrélation:

La définition même de la notion de corrélation montre son grand intérêt en traitement du signal.

La corrélation ou intercorrélation est une opération qui permet d'évaluer la ressemblance d'un signal avec un autre. pour cela on décale leurs échelles de temps d'un intervalle " " et on multiplie l'un par l'autre.

L'autocorrélation procède des mêmes opérations, mais exprime la ressemblance d'un signal avec lui-même.

Considérons deux signaux $x(t)$ et $y(t)$, leur fonction d'intercorrélation est définie par :

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

En particulier pour un même signal $x(t)$ sa fonction d'autocorrélation est définie comme suit :

$$C_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int x(t) \cdot x(t-\tau) dt$$

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

-La densité spectrale est la transformée de Fourier de la fonction de corrélation :

$$S_{xx}(f) = TF [C_{xx}(\tau)] ; S_{xy}(f) = TF [C_{xy}(\tau)]$$

II-4-FILTRES NUMERIQUES

II-4-I-Définition:

Les filtres numériques se rapportent à un système dans lequel un signal échantillonné est numérisé sous forme d'une suite de nombres, "X(nT)" ou plus simplement "Xn", (T est la période d'échantillonnage et "n" le rang de l'échantillon), est transformé en une suite de nombres "Yn", qui représente alors le signal de sortie.

Dans le cas d'un filtre linéaire la relation la plus générale entre les suites Xn et Yn est de la forme :

$$Y_n = \sum_{k=1}^M a_k \cdot y_{(n-k)} + \sum_{k=0}^M b_k \cdot x_{(n-k)}$$

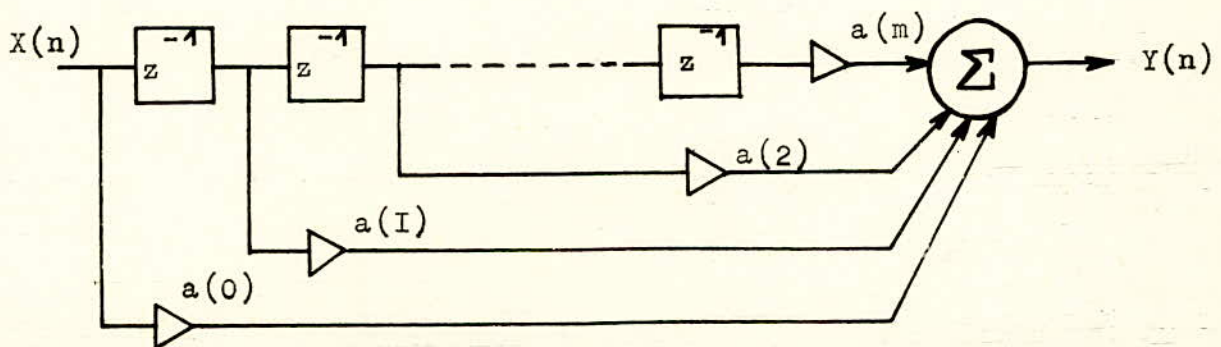
On dit qu'on a un filtre récursif quand l'un au moins des coefficients "ak" est non nul.

II-4-2-Les filtres non récursifs:

Un filtre non récursif est un filtre dont chaque échantillon de de sortie est une combinaison linéaire des échantillons présents et passés d'entrée.

Les suites d'entrée x(n) et de sortie y(n) sont reliées par une équation du type suivant, qui constitue la relation de définition :

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

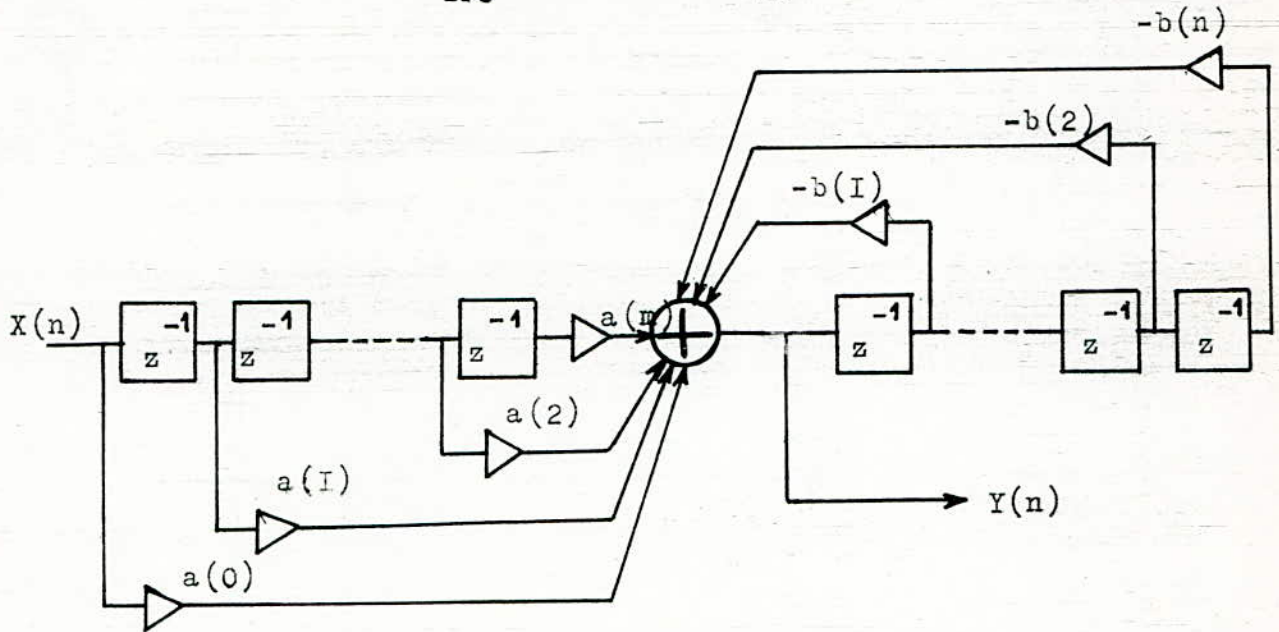


Fig(2.4) : Réalisation des filtres non récursifs

II-4-3-Les filtres récursifs:

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

$$Y_n = \sum_{l=0}^L a_l \cdot x_{(n-l)} - \sum_{k=1}^K b_k \cdot y_{(n-k)}$$



Fig(2.5) : Réalisation des filtres récursifs forme directe.

II-4-4-Fonction de transfert:

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

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

-Pour un filtre non récursif:

$$H(Z) = Y(Z)/X(Z) = \sum_{i=0}^{N-1} a_i \cdot z^{-i}$$

-Pour un filtre récursif:

$$H(Z) = Y(Z)/X(Z) = \sum_{l=0}^L a_l \cdot z^{-l} / (1 + \sum_{k=1}^K b_k \cdot z^{-k})$$

La fonction de transfert étant donnée sous la forme; $N(Z)/D(Z)$.
On appelle zéros de la fonction de transfert les racines de l'équation $N(Z) = 0$, et pôles les racines de l'équation $D(Z) = 0$.

Les filtres non récursifs n'ont que des zéros ce qui assure leur stabilité. Par contre les filtres récursifs ont des pôles et des zéros.

II-5-FENETRAGE

Si l'on veut étudier un signal sur un intervalle de durée finie soit " θ ", il suffit de multiplier ce signal par une fenêtre définie sur cette même durée.

Autrement dit, le rôle principale de la fenêtre est de limiter la durée d'un signal.

L'exemple le plus simple est la fenêtre rectangulaire définie telle que:

$$\phi(t) = \begin{cases} 1 & \text{to } -\theta/2 \text{ t to } +\theta/2 \\ 0 & \text{ailleurs} \end{cases}$$

Cette fenêtre n'est pas très utilisée vu qu'elle possède dans le domaine fréquentiel des lobes latéraux relativement amples de période " $1/T$ " qui amènent des distortions gênantes, ceci est connu sous le nom de phénomène de "Gibbs". Cet effet peut-être atténué avec l'utilisation d'autre type de fenêtres.

II-5-1-Le modèle de Hanning:

Ce modèle est défini dans le domaine temporel par la relation suivante:

$$\varphi(t) = 0.5 (1 + \cos 2\pi t / \theta) \quad 0 \leq t \leq \theta$$

Sa transformée de Fourier est donné par:

$$\phi(f) = 0.5 \text{ sinc}(2\pi f \theta) - 0.25 \text{ sinc}(2\pi(f + 1/\theta)\theta) + 0.25 \text{ sinc}(2\pi(f - 1/\theta)\theta)$$

II-5-2-Le modèle de Hamming:

Ce modèle, présenté fig(2.6), est l'un des plus utilisés. Il est défini en toute généralité par la relation paramétrique suivante:

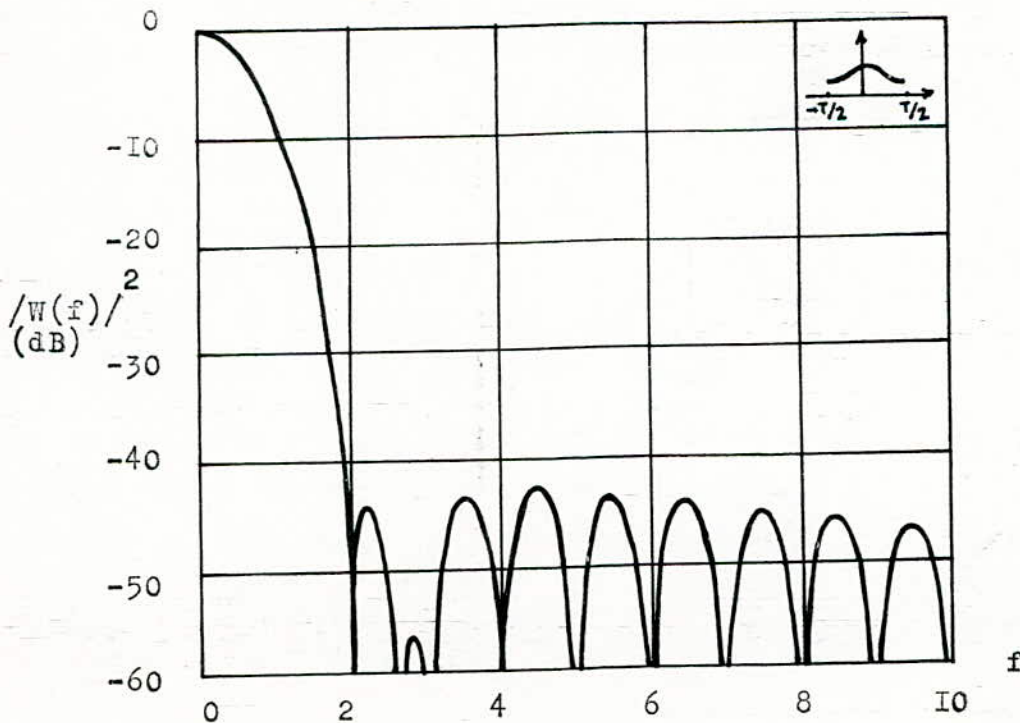
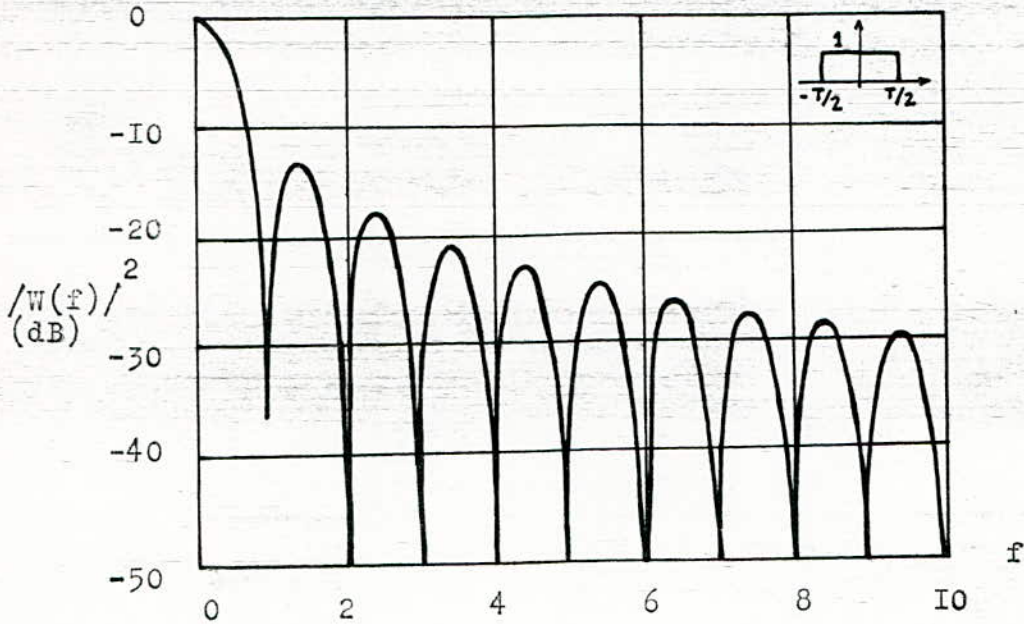
$$\varphi(t) = \alpha + (1 - \alpha) \cdot \cos(2\pi t / \theta) \quad t \in [-\theta/2, \theta/2]$$

-Pour $\alpha = 0.5$ on a la fenêtre de Hanning

-Pour $\alpha = 0.54$ on a la fenêtre de Hamming

La transformée de Fourier du modèle général de Hamming est :

$$\phi(f) = \alpha \text{sinc}(\pi f \theta) + (1-\alpha)/2 \cdot \text{sinc}(\pi(f-1/\theta)\theta) - (1-\alpha)/2 \cdot \text{sinc}(\pi(f-1/\theta)\theta)$$



Fig(2.6) : spectre de la fenêtre rectangulaire et spectre de la fenêtre de Hamming.

II-5-3-Caractéristiques et critères de choix:

Une fenêtre est caractérisée par deux paramètres principaux:

-La largeur de base du pic central

-Le rapport des amplitudes du premier lobe secondaire et du pic central. Ce rapport est exprimé en décibels;

$$W(f) = 20 \log_{10} \left[\frac{\phi(f)}{\phi(0)} \right]$$

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

.

CHAPITRE III

ANALYSE ET SYNTHÈSE DE

LA PAROLE

INTRODUCTION

Toutes les applications dans lesquelles la parole est le support physique d'information, font appel à l'analyse du signal vocal. L'objet de cette "ANALYSE" est d'extraire du signal acoustique, qui est très redondant, un nombre réduit de paramètres pertinents décrivant certains aspects de ce signal. À partir de ces paramètres, il est possible de reconstituer le signal original, c'est ce qu'on appelle l'opération de "SYNTHÈSE".

Dans le dialogue homme / machine, la qualité de la parole de synthèse dépend essentiellement de la qualité de représentation du signal après analyse.

L'ensemble des opérations d'analyse et de synthèse, doit pouvoir se faire en temps réel.

Après avoir abordé au chapitre II, les points essentiels du traitement du signal, nous allons passer en revue les techniques d'analyse et de synthèse du signal de parole, en insistant sur le modèle de "PREDICTION LINÉAIRE", qui est la technique utilisée par la suite.

.

III-I-METHODES D'ANALYSE

III-I-I-Analyse spectrale:

Le but de l'analyse spectrale est de connaître la disposition spectrale (ou fréquentielle) du signal de parole.

L'analyse spectrale se ramène en principe aux opérations mathématiques : les transformations de Fourier.

Nous allons aborder en particulier l'analyse spectrale du type numérique . Nous savons que l'essor du traitement numérique date de la découverte d'algorithmes de calcul rapide de la TFD. En effet cette transformation est à la base de l'étude des systèmes discrets et elle constitue dans le domaine numérique, l'équivalent de la TF dans le domaine analogique.

III-I-I-I-Analyse du spectre à court-terme par TFR

-Spectre à court-terme du signal de parole:

Le filtrage du signal glottique $u(t)$ par les cavités du conduit vocal se traduit par :

$$s(t) = u(t) * h(t) = \int_{-\infty}^{+\infty} u(\tau) \cdot h(t-\tau) d\tau \quad (3.1)$$

D'après le théorème de Plancherel $S(f)$ est donné par:

$$S(f) = U(f) \cdot H(f) \quad (3.2)$$

C'est ici qu'apparaît une difficulté dans l'analyse de la parole: les transformées de Fourier sont calculées sur un temps théoriquement infini alors que les caractéristiques de la source et du conduit vocal changent continuellement au cours du temps. Cependant du fait de l'inertie de l'appareil phonatoire, la configuration du conduit vocal et de la source n'évolue guère pendant un temps de l'ordre de ($T=25ms$). Ceci revient à appliquer une fenêtre au signal et à la déplacer par bond de " T " secondes.

En prenant la TF de la suite de fonctions on obtient une série de de spectre que l'on appelle spectres à court-terme.

La plupart des systèmes d'analyse de la parole sont basés sur cette notion fondamentale de spectre à court-terme.

-Principe:

La transformation de Fourier discrète (TFD) d'un signal est définie par la relation suivante:

$$X(n) = \sum_k x(k) \cdot \exp(-2\pi jnk/N) \quad n=0,1,\dots,N-1 \quad (3.3)$$

$$x(k) = \frac{1}{N} \sum_n X(n) \cdot \exp(+2\pi jnk/N) \quad k=0,1,\dots,N-1 \quad (3.4)$$

La transformation de Fourier rapide (TFR) est un algorithme de calcul permettant la résolution rapide et efficace de la TFD.

En effet le calcul de "N" échantillons d'une période spectrale exige un nombre d'opérations de l'ordre de "N²". Or, avec les algorithmes de la TFR il est possible de ramener ce nombre de l'ordre de "N log N".
2

Parmi les principaux algorithmes de la TFR nous citerons l'algorithme de Cooley et celui de Sande.

III-I-2-Analyse temporelle

Ces méthodes d'analyse privilégient l'évolution temporelle de certains paramètres. Prenons par exemple, le cas de la fermeture brusque du conduit vocal lors de la production d'une plosive, elle est mieux caractérisée par l'évolution temporelle du signal que par son spectre.

Les principales techniques d'analyse temporelle sont les suivantes

- Nombre de passage par zéro du signal
- Méthode d'autocorrélation
- Prédiction linéaire

III-I-2-I-Nombre de passage par zéro du signal:

Cette méthode est facile à utiliser. En effet il suffit de vérifier quand un signal $s(t)$ prend la valeur zéro. Ou en d'autre terme, quand il change de signe.

Par exemple lors de la production des fricatives (/f/,/s/) le taux de passage par zéro du signal (Par seconde) est très élevé.

Cette méthode permet la localisation des fréquences des premiers formants.

Elle présente cependant un inconvénient majeur qui est la perte de l'information relative à l'amplitude, puisqu'on ne s'intéresse qu'à son signe.

III-I-2-2-Méthode d'autocorrélation:

La fonction d'autocorrélation est une des méthodes de décalage. Elle est utilisée pour la mesure de la fréquence fondamentale "Fo".

Pour mettre en oeuvre cette méthode, le signal est découpé en fenêtres temporelles d'une longueur variant de 10 à 25 ms.

Théoriquement la fenêtre doit être suffisamment courte pour que le paramètre à mesurer puisse être considéré comme suffisamment constant et suffisamment longue pour qu'il soit mesurable.

Le principe de cette méthode est très simple, en effet il suffit de rechercher le maximum de la fonction d'autocorrélation:

$$f(d) = \sum_n X(n) \cdot X(n+d) \quad (3.5)$$

Ce maximum a lieu quand le décalage "d" est égal à la durée "To" de la période de Fo. On en déduit ainsi la fréquence fondamentale.

La fonction d'autocorrélation peut être générée à partir du signal analogique s(t) par trois ensembles :

- Retard
- Modulateur (multiplicateur)
- Sommateur

Ou par des procédés numériques à partir d'un signal échantillonné en utilisant par exemple des autocorrélateurs.

III-I-2-3-Prédiction linéaire:

La prédiction linéaire est l'une des techniques les plus utilisées dans le domaine de la parole, que ce soit en analyse ou en synthèse. Sa supériorité provient du fait qu'elle est fondée sur un modèle simple (filtre numérique ne possédant que des pôles).

a)Principe:

La méthode de prédiction linéaire repose sur le fait qu'un échantillon de parole peut être approximativement prédit par une combinaison linéaire

des échantillons précédents pondérés par des coefficients.

Le signal prédit est noté par:

$$\hat{S}(n) = \sum_{k=1}^P a(k) \cdot S(n-k) \quad (3.6)$$

p: ordre du prédicteur

a(k): coefficients du prédicteur

La connaissance des "a(k)" permet de calculer le signal de manière approchée, par récurrence à partir de l'un de ses échantillons et ceci tant que le spectre reste stable. On considère des intervalles de temps suffisamment faibles (de l'ordre de 10 à 25ms) pour que cette hypothèse reste vérifiée.

On suppose que pendant ces intervalles de temps, on peut approcher l'ensemble; source, conduit vocal, et rayonnement par un filtre linéaire invariable dans le temps et ne possédant que des pôles (filtre récursif).

L'erreur de prédiction est donnée par:

$$e(n) = S(n) - \hat{S}(n) \quad (3.7)$$

$$e(n) = S(n) - \sum_{k=1}^P a(k) \cdot S(n-k) \quad (3.8)$$

La transformée en Z de l'équation (3.6) nous donne:

$$\hat{S}(Z) = F(Z) \cdot S(Z) \quad (3.9)$$

Avec
$$F(Z) = \sum_{k=1}^P a(k) \cdot z^{-k} \quad (3.10)$$

F(Z) est défini comme étant un filtre linéaire de prédiction
 $\hat{S}(Z)$ TZ de $\hat{S}(n)$
 $S(Z)$ TZ de $S(n)$

La transformée en Z de l'expression (3.8) nous donne:

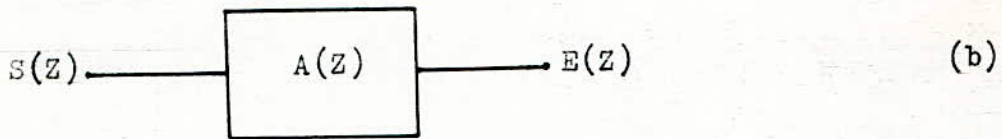
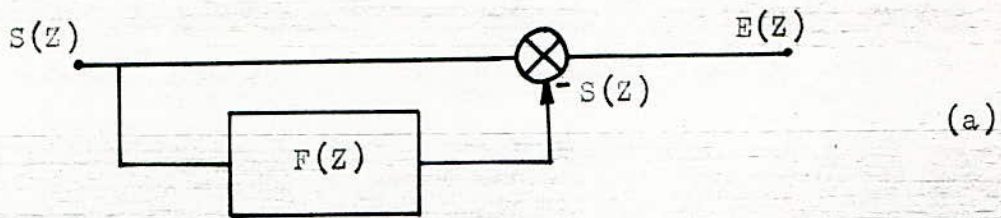
$$E(Z) = S(Z) \cdot [1 - F(Z)] \quad (3.11)$$

$$E(Z) = S(Z) \cdot A(Z) \quad (3.12)$$

Où
$$A(Z) = I - \sum_{k=1}^P a(k) \cdot z^{-k} \quad (3.13)$$

$$A(Z) = I - F(Z)$$

Nous donnons ci-dessus les schémas fonctionnels des modèles d'analyse, utilisant le filtre $F(Z)$ et le filtre inverse $A(Z)$, fig(3.I).



Fig(3.I) : Modèle de prédiction linéaire

(a) : Représentation du filtre prédicteur $F(Z)$

(b) : Représentation équivalente du filtre inverse $A(Z)$

b) Analyse-synthèse du signal de parole:

D'une manière générale on peut estimer tout signal temporel en terme d'un signal prédit et d'un signal d'erreur:

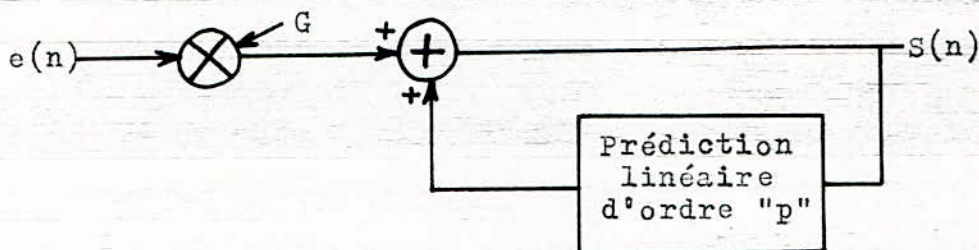
$$S(n) = \hat{S}(n) + G \cdot e(n)$$

En particulier le signal de parole peut être prédit par le modèle de prédiction linéaire:

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

D'où
$$S(n) = \sum_{k=1}^P a(k) \cdot S(n-k) + G \cdot e(n) \quad (3.14)$$

G : facteur de gain



Fig(3.2) : Modèle de production de la parole dans le domaine temporel.

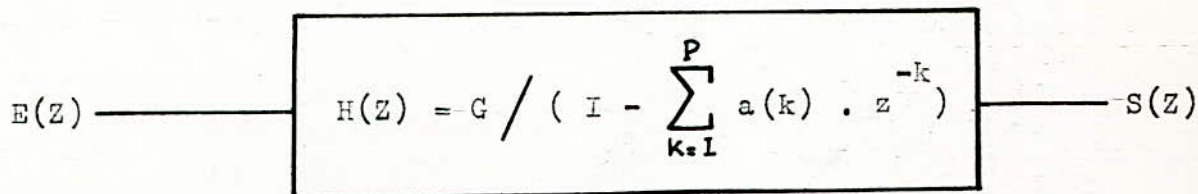
Prenons la TZ de l'expression (3.14):

$$S(Z) = S(Z) \cdot \left(\sum_{k=1}^P a(k) \cdot z^{-k} \right) + G \cdot E(Z)$$

Soit
$$S(Z) = G \cdot E(Z) / \left(1 - \sum_{k=1}^P a(k) \cdot z^{-k} \right) \quad (3.15)$$

$$H(Z) = S(Z)/E(Z) = G / \left(1 - \sum_{k=1}^P a(k) \cdot z^{-k} \right) \quad (3.16)$$

$$H(Z) = G/A(Z) \quad A(Z): \text{Filtre inverse}$$



Fig(3.3) : Modèle de production de la parole dans le domaine fréquentiel

On remarque que la fonction de transfert du modèle, ne comporte que des pôles. En réalité le système phonatoire comporte également des zéros, mais on peut montrer que chaque zéro peut être approché par un certain nombre de pôles supplémentaires (ces zéros étant situés à l'intérieur d'un cercle unité).

En effet pour s'en rendre compte il suffit de développer en série de Taylor la fonction complexe suivante:

$$I/N(Z) = I / (I - az^{-1})$$

Le zéro de N(Z) est $Z_0 = a$

Le développement en série de N(Z) converge si $|a| < 1$

$$I/(I - az^{-1}) = I + az^{-1} + a^2 z^{-2} + \dots + a^n z^{-n}$$

Notre système auditif étant beaucoup plus sensible à la localisation dans l'échelle des fréquences d'un maximum d'énergie qu'à celle d'un minimum : cela signifie que notre oreille est beaucoup plus sensible à un pôle qu'à un zéro.

En résumé le modèle de production de la parole utilisé dans la méthode de prédiction linéaire est fondé sur les trois hypothèses suivantes:

1-La source vocale et le conduit vocal peuvent être considérés comme constants pendant des intervalles de temps courts; de l'ordre de 10 à 25ms

2-Pendant ces intervalles de temps on suppose que l'on approche la fonction de transfert du conduit vocal, source vocale et rayonnement par la fonction de transfert d'un filtre récursif.

3-Le signal de parole de synthèse est constitué par le signal de sortie d'un filtre dont on renouvelle les coefficients $a(k)$, ($1 \leq k \leq p$) aux intervalles de temps déterminés ci-dessus.

c) Détermination du nombre de coefficients $a(k)$:

En admettant que le spectre de la parole puisse être décrit à l'aide de cinq résonances (D'après Makhoul) et du fait que chaque formant soit associé à deux coefficients, on aura besoin de dix coefficients pour représenter le conduit vocal. En ajoutant deux pôles supplémentaires pour représenter les influences de la source vocale et du rayonnement au niveau des lèvres, cela porte à douze (12) la valeur moyenne du nombre "p" du prédicteur.

d) Détermination des coefficients du prédicteur:

La quantité $e(n)$ est définie comme étant l'erreur entre l'échantillon $S(n)$ et le signal prédit $\hat{S}(n)$. De ce fait les coefficients $a(k)$ ($k=1, \dots, p$) sont calculés de façon à minimiser l'erreur $e(n)$. Dans ce cas on dit que les coefficients $a(k)$ sont optimaux.

Il existe plusieurs méthodes de minimisation de cette erreur mais celle qui donne de bons résultats dans l'analyse de la parole est la méthode des moindres carrés.

L'erreur quadratique totale minimale est définie par:

$$E = \sum_n e(n)^2 \quad (3.17)$$

$$E = \sum_n \left[S(n) - \sum_{k=1}^p a(k) \cdot S(n-k) \right]^2 \quad (3.18)$$

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

Pour minimiser "E", il faut annuler ses dérivées partielles par rapport à chaque coefficient $a(k)$ du prédicteur.

$$\partial E / \partial a(i) = 2 \sum_n \sum_{k=1}^p a(k) \cdot S(n-k) \cdot S(n-i) - 2 \sum_n S(n) \cdot S(n-i)$$

$$\text{Avec } \partial E / \partial a(k) = 0 \quad 1 \leq k \leq p \quad (3.19)$$

On obtient:

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

En interchangeant les signes de sommation on obtient finalement:

$$\sum_{k=1}^p a(k) \sum_n S(n-k) \cdot S(n-i) = \sum_n S(n) \cdot S(n-i) \quad (3.20)$$

Nous avons à résoudre un système de "p" équations à "p" inconnues dont la solution peut être aisément obtenue par les méthodes classiques telle que la triangularisation de Gauss, la méthode de Gauss-Jordan... Cependant il existe des techniques plus économiques en temps de calcul dont nous citerons les deux principales, (cf chapitre III).

CONCLUSION

Les techniques d'analyse de la parole sont très nombreuses. Nous avons passé en revue quelques unes d'entre elles.

Cependant toutes les méthodes étudiées souffrent de sérieuses limitations à cause des caractères non stationnaire et pseudo-périodique de la parole.

L'importance de la prédiction linéaire reside dans le fait qu'elle soit basée sur un modèle mathématique simple, qui constitue une bonne approximation du système phonatoire et des estimations précises des paramètres de prédiction.

Le traitement informatique est très rapide dans ce cas.

III-2-METHODES ET TECHNIQUES DE SYNTHÈSE

On entend par synthèse de la parole, la production de message parlés par assemblage d'éléments phonétiques préenregistrés (ex: phonèmes, mots, ...).

La synthèse représente l'opération inverse de l'analyse.

On distingue "4" principaux types de techniques de synthèse:

- Vocodeur à canaux
- Synthétiseur à formants
- Synthétiseurs à prédiction linéaire
- Synthèse par simulation du conduit vocal

Il existe "4" méthodes de synthèse:

- Synthèse par phrases
- Synthèse par mots
- Synthèse par règles
- Synthèse par diphtongues

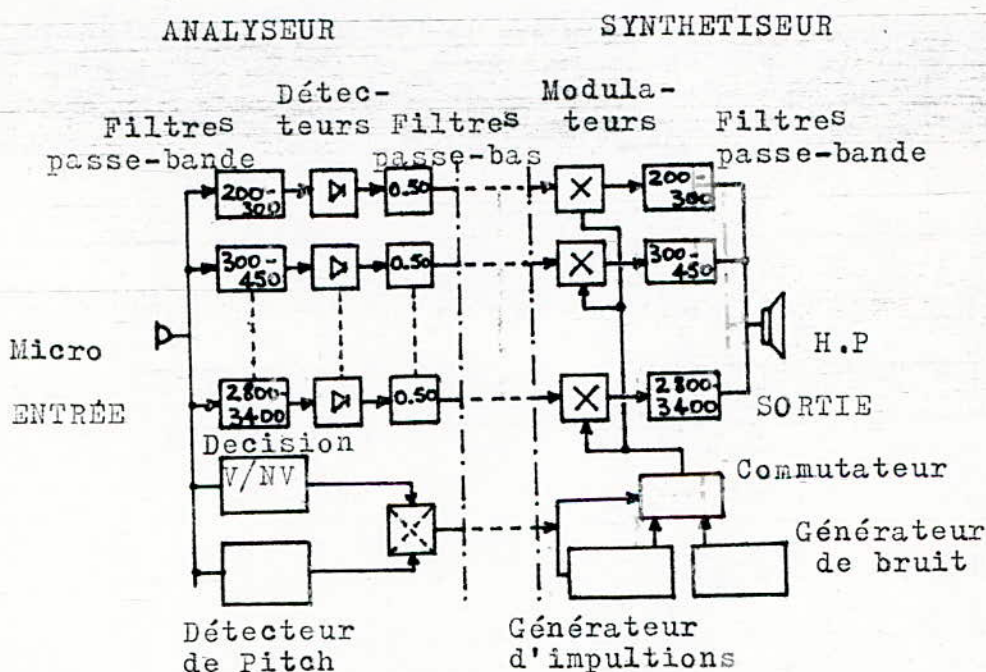
Nous allons effectuer un bref tour d'horizon des différentes techniques et méthodes citées de synthèse de la parole.

III-2-I-Les techniques de synthèse

Les procédés de synthèse actuellement en cours sont très nombreux nous ne citerons que les principaux.

III-2-I-I-Vocodeur à canaux:

Le vocodeur à canaux se compose essentiellement de deux grandes parties: l'analyseur et le synthétiseur, fig(3.4).



Fig(3.4) : Schéma d'un vocodeur à canaux

-L'analyse se fait à l'aide d'un banc de filtres passe-bande contigus couvrant l'étendue spectrale de la voix (bande téléphonique; 300 à 3300). Chaque filtre est suivi d'un étage de détection, lui-même suivi d'un étage de filtrage passe-bas qui a pour fonction de ne laisser que les variations lentes du signal.

-La synthèse est effectuée à l'aide d'un banc de filtres passe-bandes identiques à ceux de l'analyse.

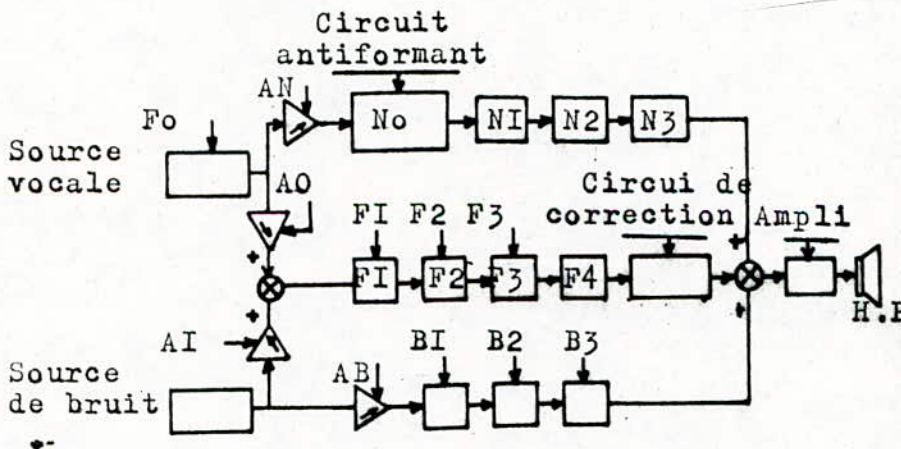
Pour chaque canal le signal basse fréquence issu du filtre d'analyse est multiplié par le signal d'excitation dans un modulateur. Selon que les sons soient voisés ou non, le signal d'excitation attaquant les modulateurs proviendra soit d'un générateur d'impulsions soit d'un générateur de bruit. Le signal de sortie est obtenu par addition des filtres de synthèse, on obtient ainsi la reconstitution du signal de parole.

III-2-I-2-Synthétiseurs à formants:

Comme dans les vocodeurs à canaux la synthèse est effectuée à l'aide de filtres passe-bandes (résonateurs) attaqués par un signal d'excitation. Les filtres sont en nombre réduit puisqu'il s'agit de modéliser les seuls formants, ils sont variables en ce sens que leurs fréquences centrales sont commandées par les paramètres issus de l'analyse des formants.

Les signaux issus d'une source d'impulsions périodiques et d'une source de bruit viennent attaquer les circuits résonnants.

On distingue deux types de synthétiseurs à formant, fig(3.5).



Fig(3.5) : Schéma d'un synthétiseur à formants " serie"
On trouve de bas en haut, la branche des formants de bruit, celles des formants vocaux, enfin la branche de nasalité.
Le synthétiseur est commandé à l'aide de onze "II" paramètres.

Dans la structure parallèle les filtres sont tous alimentés par le même signal, et leurs sorties sont additionnées.

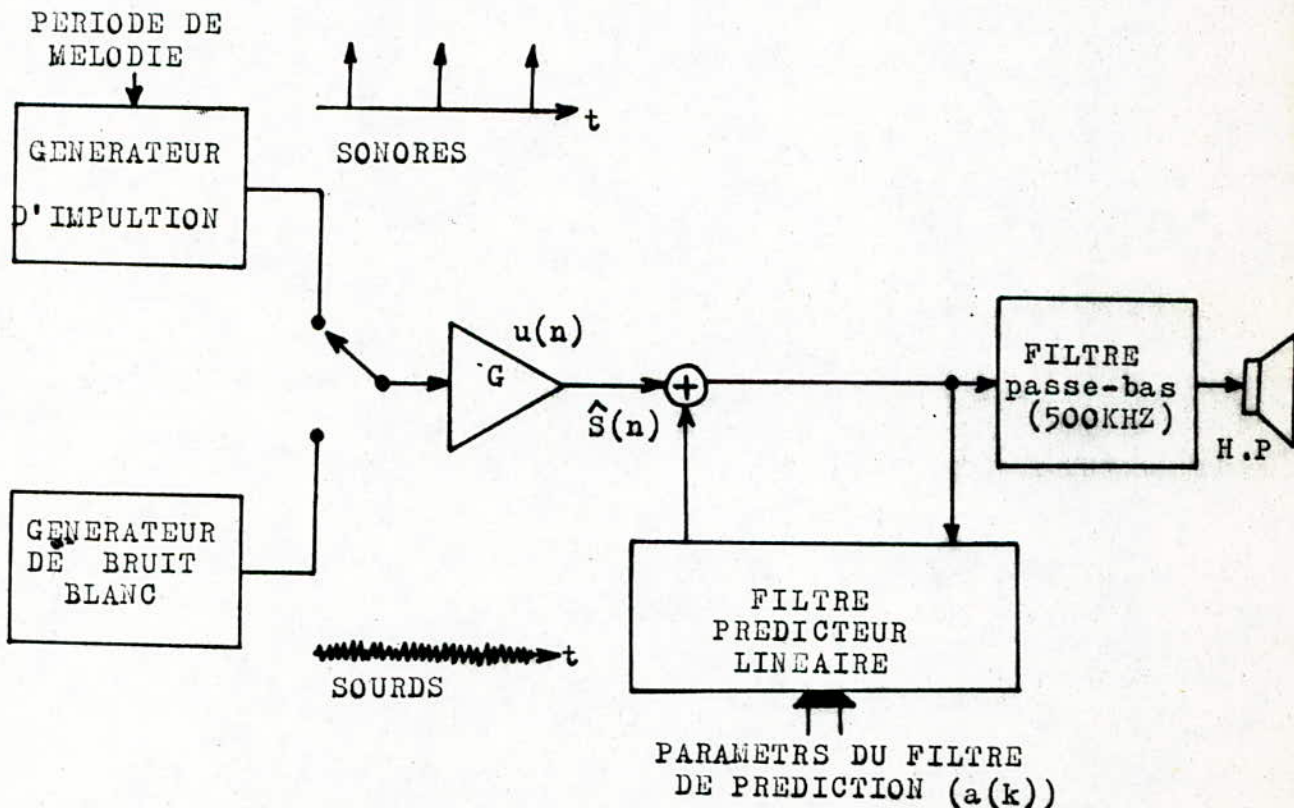
Dans la structure série les filtres sont placés en cascades.

Les synthétiseurs séries permettent une meilleure approximation du conduit vocal pour les sons non nasalisés. Ils présentent en outre l'avantage de ne pas nécessiter des contrôles individuels pour les amplitudes des formants. Par contre la structure parallèle permet de contrôler les amplitudes de chaque formants, ce qui est très important pour la synthèse des consonnes.

III-2-I-3-Synthétiseur à prédiction linéaire:

La prédiction linéaire est une technique de modélisation. L'équation de prédiction se traduit par un filtre récursif.

La fig(3.6) montre le principe de la réalisation d'un synthétiseur à prédiction linéaire.



Fig(3.6) : Synthèse prédictive. Le synthétiseur est contrôlé à l'aide des paramètres $a(k)$ de prédiction, d'un signal binaire fixant le type d'excitation (voisé ou non voisé), du gain G , et de la période T_0 relatif au fondamental dans le cas d'un son voisé.

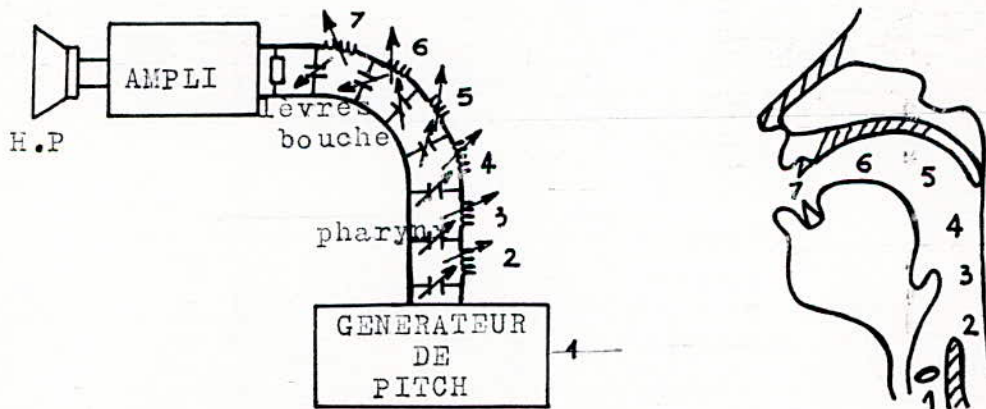
Les paramètres de contrôle fournis au synthétiseur sont:

- Les coefficients de prédiction
- La décision voisé / non voisé
- Le gain G
- La période T_0 du fondamental

Ces paramètres sont renouvelés toutes les 10 à 25ms. Comme il a été mentionné plus haut la synthèse se fait au moyen d'un filtre récur-

III-2-I-4-Synthèse par simulation du conduit vocal:

Tous les types de synthétiseurs simulent plus ou moins le conduit vocal. Mais dans le cas de la synthèse articuloire, on entend par cette simulation la caractérisation de dispositifs reproduisant les propriétés acoustiques d'un conduit de sections variables. C'est à dire que l'on peut simuler l'appareil phonatoire par une succession de tuyaux acoustiques ou de circuits électriques élémentaires. Chacun de ces éléments représentant une partie du conduit vocal. Un exemple en est donné par la fig(3.7).



Fig(3.6) : Simulation du conduit vocal à l'aide d'une ligne électrique. A chaque section du conduit est associé une cellule dont les éléments sont accordables; on simule ainsi la variation de la fonction d'aire du système phonatoire lorsqu'on se déplace de la glotte aux lèvres. Le conduit nasal, dont l'entrée est ici condamnée par le velum relevé, pourrait être simulé par un circuit en parallèle sur la ligne

III-2-2-Les méthodes de synthèse

Nous venons de décrire les principales techniques de synthèse de la parole. Ces techniques utilisent des méthodes diverses pour l'assemblage des éléments phonétiques préenregistrés.

III-2-2-I-Synthèse par phrases:

Il ne s'agit pas tout à fait d'une méthode de synthèse, c'est en fait un enregistrement de phrases sur magnétophone. La parole reproduite est de bonne qualité, mais le nombre de phrases enregistrées est très limité.

III-2-2-2-Synthèse par mots:

Actuellement cette méthode est la plus utilisée vu la simplicité de sa mise en oeuvre.

L'inconvénient majeur de cette méthode est qu'elle suppose la mise en mémoire préalable de tous les mots du vocabulaire, ce qui fait qu'elle soit très peu économique.

III-2-2-3-Synthèse par règles:

Cette méthode consiste à assembler des segments de parole (phonèmes) grâce à un ensemble de règles capables de reproduire les zones de transition d'un segment à l'autre.

Cette méthode donne une parole assez naturelle mais la réalisation des règles est très complexe.

III-2-2-4-Synthèse par diphonèmes ou diphones:

L'avantage de cette méthode tient en fait qu'elle rend inutile l'utilisation des règles. L'assemblage des unités de parole est réalisé par simple juxtaposition de ces unités. Ces dernières sont appelées diphonèmes.

En partant d'une trentaine de phonèmes (ce qui est le cas pour la langue française), en les associant deux par deux on obtient environ un millier de combinaisons.

CONCLUSION

Si nous comparons tout d'abord les vocodeurs à canaux par rapport aux vocodeurs à formants, on peut conclure que les premiers sont meilleurs au niveau de l'analyse grâce à leur simplicité. Par contre au niveau de la synthèse les seconds simulent plus fidèlement le fonctionnement de l'appareil vocal.

La prédiction linéaire a l'avantage d'être très simple au niveau de la réalisation (filtre récursif), son inconvénient majeur est le manque d'intelligibilité dans le cas de certains sons; (nasals, sonores, liquides).

Nous avons mentionné plus haut la supériorité des synthétiseurs à formants sur les vocodeurs à canaux au niveau de reconstitution du signal de parole. La synthèse par simulation du conduit vocal permet de reproduire plus fidèlement encore le fonctionnement de l'appareil phonatoire humain.

.

CHAPITRE IV

PRINCIPALES METHODES
DE PREDICTION LINEAIRE
ET LEURS SOLUTIONS

INTRODUCTION

On a vu au chapitre précédent que le problème en prédiction linéaire consiste à trouver les "p" coefficients "a(k)" afin que l'estimation du signal "S(n)" soit optimale. On a aboutit finalement à un système d'équations linéaires de p équations à p inconnues:

$$\sum_{k=1}^p a(k) \sum_n S(n-k) \cdot S(n-i) = \sum_n S(n) \cdot S(n-i) \quad (4.1)$$

Posons:

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

En remplaçant cette expression dans l'équation (4.1), on obtient:

$$\sum_{k=1}^p a(k) \cdot C(i,k) = C(i,0) \quad \begin{matrix} i = 1, 2, \dots, p \\ k = 1, 2, \dots, p \end{matrix} \quad (4.2)$$

Les p coefficients a(k) du prédicteur sont obtenus en résolvant le système linéaire de p équations, en utilisant les paramètres connus C(i,k); k=0,1,...,p, i=1,2,...,p.

Jusqu'ici nous n'avons pas tenu compte des bornes de sommation, pour les définir nous allons considérer à présent les deux principales méthodes de prédiction linéaire.

.

IV-1-METHODE DE COVARIANCE

En supposant que nous ayons une séquence de "N" échantillons de parole $S(n) = \{S(0), S(1), \dots, S(N-1)\}$, la méthode de covariance définit les bornes de sommation de $C(i,k)$ de "p" à N-1.

Le signal est non stationnaire à l'intérieur de l'intervalle d'analyse.

La résolution du système par la méthode de covariance se résume comme suit:

$$\sum_{k=1}^P a(k) \cdot C(i,k) = C(i,0) \quad \begin{matrix} i = 1, 2, \dots, p \\ k = 0, 1, \dots, p \end{matrix} \quad (4.3)$$

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

L'expression (4.3) se traduit par les relations matricielles suivantes :

$$\begin{bmatrix} C(1,1) & C(1,2) & \dots & C(1,P) \\ C(2,1) & C(2,2) & \dots & C(2,P) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ C(P,1) & C(P,2) & \dots & C(P,P) \end{bmatrix} \times \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ \vdots \\ a(P) \end{bmatrix} = \begin{bmatrix} C(1,0) \\ C(2,0) \\ \vdots \\ \vdots \\ C(P,0) \end{bmatrix} \quad (4.4)$$

qui devient sous forme condensée:

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

- C_p : matrice de covariance
- A_p : vecteur colonne de composantes $a(k)$
- C_a : vecteur colonne de composantes $C(i,0)$

La structure de C_p est symétrique et définie positive.

IV-2-METHODE D'AUTOCORRELATION

La méthode d'autocorrélation est fondée sur les hypothèses suivantes :

-L'erreur quadratique définie précédemment est minimisée sur une durée infinie.

-Le signal est défini sur une durée infinie, mais il est nul en dehors d'une fenêtre de "N" échantillons.

Ce qui conduit à la simplification suivante de C(i,k):

$$C(i,k) = \sum_{n=-\infty}^{+\infty} S(n-i) \cdot S(n-k) \quad (4.4)$$

En posant $n' = n-i$ dans l'expression (4.4) nous obtenons:

$$C(i,k) = \sum_{n=-\infty}^{+\infty} S(n) \cdot S(n+i-k) \quad (4.5)$$

Etant donné que le signal est limité par une fenêtre ;

$S(n) = 0$ pour $n < 0$ et $n > N-1$ (utilisant la fenêtre de Hamming)

$$C(i,k) = \sum_{n=0}^{N-1-|i-k|} S(n) \cdot S(n+i-k) \quad (4.6)$$

On définit les coefficients de corrélation par:

$$R(i-k) = \sum_{n=0}^{N-1-|i-k|} S(n) \cdot S(n+i-k) \quad (4.7)$$

Faisons un changement de variable; $i-k=l$

$$R(l) = \sum_{n=0}^{N-1-|l|} S(n) \cdot S(n+l) \quad (4.8)$$

La résolution du système par la méthode d'autocorrélation se résume comme suit:

$$\sum_{k=1}^P a(k) \cdot R / i-k/ = R(i) \quad \begin{matrix} i= 1,2,\dots,p \\ k= 1,2,\dots,p \end{matrix} \quad (4.9)$$

Avec $R(i) = \sum_{n=0}^{N-i-1} S(n) \cdot S(n+1) \quad i \geq 0$

L'expression (4.9) se traduit par la relation matricielle suivante:

$$\begin{bmatrix} R(0) & R(1) & \dots & R(P-1) \\ R(1) & R(0) & \dots & R(P-2) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ R(P-1) & R(P-2) & \dots & R(0) \end{bmatrix} \times \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ \vdots \\ a(p) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ \vdots \\ R(p) \end{bmatrix} \quad (4.10)$$

qui devient sous forme condensée:

$$[R_p] \cdot [A_p] = [r_p]$$

R_p : matrice d'autocorrélation

A_p : vecteur colonne de composante $a(k)$

r_p : vecteur colonne de composante $R(i)$

Nous remarquons que la structure de R_p est symétrique, définie positive et les éléments situés sur une même diagonale parallèle à la diagonale principale sont égaux (matrice de Toeplitz).

IV-3-ERREUR QUADRATIQUE TOTALE MINIMALE DE PREDICTION

On a montré précédemment l'égalité suivante:

$$E(p) = \sum_n \left(S(n) - \sum_{k=1}^P a(k) \cdot S(n-k) \right)^2$$

$$E(p) = \sum_n \left(S(n)^2 - 2S(n) \cdot \sum_{k=1}^P a(k) \cdot S(n-k) + \sum_{k=1}^P \sum_{i=1}^P a(k) \cdot a(i) \cdot \frac{S(n-k) \cdot S(n-i)}{S(n-i)} \right)$$

$$E(p) = \sum_n S(n)^2 - 2 \sum_{k=1}^P a(k) \sum_n S(n-k) \cdot S(n) + \sum_{k=1}^P \sum_{i=1}^P a(k) \sum_n a(i) \sum_n S(n-k) \cdot S(n-i) \quad (4.II)$$

Or, d'après la relation (4.I), on peut écrire:

$$\sum_{i=1}^P a(i) \sum_n S(n-k) \cdot S(n-i) = \sum_n S(n) \cdot S(n-k) \quad (4.I2)$$

En remplaçant cette expression dans l'équation (4.II) on obtient:

$$E(p) = \sum_n S(n)^2 - 2 \sum_{k=1}^P a(k) \sum_n S(n) \cdot S(n-k) + \sum_{k=1}^P a(k) \sum_n S(n) \cdot S(n-k)$$

D'où finalement:

$$E(p) = \sum_n S(n)^2 - \sum_{k=1}^P a(k) \sum_n S(n) \cdot S(n-k) \quad (4.I3)$$

a) Pour la méthode de covariance:

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

$$D'où: \sum_{n=0}^{N-1} S(n)^2 = C(0, 0) \quad (4.I4)$$

$$Et \sum_{n=0}^{N-1} S(n) \cdot S(n-k) = C(0, k) \quad (4.I5)$$

Soit en remplaçant ces deux dernières expressions dans l'équation (4.14), on aboutit à:

$$E(p) = C(0,0) - \sum_{k=1}^P a(k) \cdot C(0,k) \quad (4.16)$$

b) Pour la méthode d'autocorrélation:

$$R(k) = \sum_{n=0}^{N-1-k} S(n) \cdot S(n+k) \quad \text{soit} \quad R(0) = \sum_{n=0}^{N-1} S(n)^2 \quad (4.17)$$

D'où l'expression de "E" devient:

$$E(p) = R(0) - \sum_{k=1}^P a(k) \cdot R(k) \quad (4.18)$$

On définit "l'erreur normalisée" $V(p)$, comme étant le rapport de l'énergie représentée par les échantillons de l'erreur de prédiction et de l'énergie représentée par les échantillons du signal de parole pour la séquence considérée.

$$V(p) = E(p) / R(0) \quad (4.19)$$

$$V(p) = 1/R(0) \left[R(0) - \sum_{k=1}^P a(k) \cdot R(k) \right]$$

$$V(p) = 1 - (1/R(0)) \cdot \sum_{k=1}^P a(k) \cdot R(k)$$

En posant, $r(k) = R(k) / R(0)$ (coefficients normalisés de la fonction d'autocorrélation)

$V(p)$ devient:

$$V(p) = 1 - \sum_{k=1}^P a(k) \cdot r(k) \quad (4.20)$$

IV-4-RESOLUTION DU SYSTEME DE PREDICTION LINEAIRE

On a vu précédemment que pour les deux méthodes étudiées, on aboutit à une équation du type $AX = B$, pour la détermination des coefficients $a(k)$. La résolution de ce système est possible par des méthodes directes telle que la triangularisation de Gauss ou la méthode de Gauss-Jordan. Seulement le nombre d'opérations est de l'ordre de " $p^3/3$ " ce qui est considérable. Pour remédier à cela on utilise d'autres méthodes exploitant les formes particulières des matrices mises en jeu et qui peuvent réduire considérablement le nombre d'opérations.

Plusieurs méthodes découlent de la matrice de covariance et plusieurs autres de la matrice d'autocorrélation. Pour notre part on ne s'intéresse qu'à la décomposition de "CHOLESKY" (pour la méthode de covariance) et à l'algorithme de "DURBIN" (pour la méthode d'autocorrélation). Ces méthodes sont considérées parmi les plus intéressantes, vu leur rapidité d'exécution et leur nombre réduit d'opérations.

IV-4-I-Méthode de Cholesky:

On a vu que système à résoudre, par la méthode de covariance s'écrit sous forme condensée :

$$[C] \cdot [A] = [C_0] \quad (4.21)$$

C matrice de covariance est définie positive, alors elle peut être décomposée sous la forme suivante:

$$C = V \cdot D \cdot V^T \quad (4.22)$$

Où "V" est une matrice triangulaire inférieure dont les éléments de la diagonale principale sont égaux à un (1).

"D" est matrice diagonale.

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ v(2,1) & 1 & 0 & & 0 \\ v(3,1) & v(3,2) & 1 & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v(p,1) & v(p,2) & \vdots & & 1 \end{bmatrix} = [V]$$

Soit:

$$C = \begin{bmatrix} I & 0 & \dots & 0 \\ V(2,I) & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ V(P,I) & V(P,2) & \dots & I \end{bmatrix} \times \begin{bmatrix} d(I) & 0 & \dots & 0 \\ 0 & d(2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d(P) \end{bmatrix} \times \begin{bmatrix} I & V(2,I) & \dots & V(P,I) \\ 0 & I & \dots & V(P,2) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I \end{bmatrix}$$

A partir de ce système on tire l'équation réccurente:

$$C(i,j) = \sum_{k=I}^j V(i,k) \cdot d(k) \cdot V(j,k) \quad I \leq j \leq i-I \quad (4.23)$$

Qui peut s'écrire:

$$C(i,j) = \sum_{k=I}^{j-I} V(i,k) \cdot d(k) \cdot V(j,k) + V(i,j) \cdot d(j) \cdot V(j,j) \quad (4.24)$$

Ayant $V(j,j) = I$ par hypothèse, on peut écrire d'après (4.24);

$$V(i,j) \cdot d(j) = C(i,j) - \sum_{k=I}^{j-I} V(i,k) \cdot d(k) \cdot V(j,k) \quad I \leq j \leq i-I$$

Pour $k=I$ nous obtenons les éléments de la diagonale de C , soit:

$$\begin{aligned} C(i,i) &= \sum_{k=I}^i V(i,k) \cdot d(k) \cdot V(i,k) \\ &= \sum_{k=I}^i V(i,k)^2 \cdot d(k) \\ &= \sum_{k=I}^{i-I} V(i,k)^2 \cdot d(k) + V(i,i)^2 \cdot d(i) \end{aligned}$$

Or, $V(i,i) = I$, ce qui conduit à écrire:

$$d(i) = C(i,i) - \sum_{k=I}^{i-I} V(i,k)^2 \cdot d(k) \quad i \geq 2$$

Avec comme condition initiale:

$$d(I) = C(I, I)$$

C'est à dire en récapitulant, on a:

$$C = V \cdot D \cdot V^T$$

V et D peuvent être connues à l'aide des formules suivantes:

$$\left\{ \begin{array}{l} V(i, j) \cdot D(j) = C(i, j) - \sum_{k=I}^{j-I} V(i, k) \cdot d(k) \cdot V(j, k) \quad j = I, \dots, i-I \quad (4.25) \\ D(i) = C(i, i) - \sum_{k=I}^{i-I} V(i, k) \cdot D(k) \quad i \geq 2 \quad (4.26) \\ \text{Avec } D(I) = C(I, I) \end{array} \right.$$

D'après (4.21) et (4.22) le système à résoudre s'écrit:

$$V \cdot D \cdot V^T \cdot A = C_0$$

Qu'on décompose comme suit:

$$V \cdot Y = C_0 \quad (4.27)$$

$$\text{Avec } D \cdot V^T \cdot A = Y \quad (4.28)$$

En utilisant la forme particulière de V, le système $V \cdot Y = C$ peut être résolu par itération:

$$Y(i) = C_0(i) - \sum_{j=I}^{i-I} V(i, j) \cdot Y(j)$$

$Y(I) = C_0(I)$ étant la condition initiale

On obtient finalement la matrice A des coefficients inconnus par la relation:

$$a(i) = Y(i)/D(i) - \sum_{j=i+I}^P V(j, i) \cdot a(j) \quad i=I, \dots, P \quad (4.29)$$

L'erreur quadratique de la méthode de covariance étant:

$$E(p) = C(0, 0) - \sum_{k=I}^P a(k) \cdot C(0, k) \quad (4.30)$$

Elle devient sous forme condensée:

$$E(p) = c(0,0) - A^T \cdot C_0 \quad (4.31)$$

Exprimons "E" en fonction de Y et D;

De l'équation (4.28), on peut tirer:

$$(V \cdot A)^T = (D^{-I} \cdot Y)^T$$

En utilisant les propriétés de V et D nous pouvons écrire:

$$A^T \cdot V = Y^T \cdot (D^{-I})^T = Y^T \cdot D^{-I}$$

D'où $A^T = Y^T \cdot D^{-I} \cdot V^{-I}$

En substituant cette expression de A^T dans l'équation (4.31) nous obtenons:

$$E(p) = c(0,0) - Y^T \cdot D^{-I} \cdot V^{-I} \cdot C_0 \quad (4.32)$$

Remplaçons C_0 par son expression (réf équation (4.27)):

$$E(p) = c(0,0) - Y^T \cdot D^{-I} \cdot Y \quad (4.33)$$

De cette expression nous pouvons déduire la relation finale suivante:

$$E(p) = c(0,0) - \sum_{k=1}^P Y^2(k) / d(k) \quad (4.34)$$

IV-4-2-Algorithmme de Durbin:

Pour faciliter les calculs par la suite, on prendra pour coefficient de prédiction inconnus; $\alpha(k) = -a(k)$. Tenant compte de cette hypothèse, le système à résoudre par la méthode d'autocorrélation s'écrit comme suit

$$\begin{bmatrix} R(0) & R(1) & \dots & R(P-1) \\ R(1) & R(0) & \dots & R(P-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(P-1) & \dots & \dots & R(0) \end{bmatrix} \times \begin{bmatrix} \alpha(1) \\ \alpha(2) \\ \vdots \\ \alpha(P) \end{bmatrix} = \begin{bmatrix} -R(1) \\ -R(2) \\ \vdots \\ -R(P) \end{bmatrix} \quad (4.35)$$

$$\text{Avec } E(p) = R(0) + \sum_{k=1}^p \alpha(k) \cdot R(k) \quad (4.36)$$

En prenant $\alpha(0) = 1$, avec le système:

$$R(0) + \sum_{k=1}^p \alpha(k) \cdot R(k) = E(p)$$

$$\begin{bmatrix} R(1) & R(0) & R(1) & \dots & R(p-1) \\ R(2) & R(1) & R(0) & \dots & R(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p) & R(p-1) & \vdots & \dots & R(0) \end{bmatrix} \times \begin{bmatrix} \alpha(1) \\ \alpha(2) \\ \vdots \\ \alpha(p) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.37)$$

Le système à résoudre devient finalement:

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p) \\ R(1) & R(0) & \dots & R(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(p) & R(p-1) & \dots & R(0) \end{bmatrix} \times \begin{bmatrix} \alpha(1) \\ \vdots \\ \alpha(p) \end{bmatrix} = \begin{bmatrix} E \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.38)$$

On utilise l'algorithme de "Durbin" pour résoudre ce système.

L'énoncé de cet algorithme est le suivant:

$$E(0) = R(0) \quad (4.39)$$

$$K(i+1) = - \left(\sum_{j=0}^i \alpha(j) \cdot R(i-j+1) \right) / E(i) \quad 1 \leq i \leq p \quad (4.40)$$

$$\alpha(i+1)^{(i+1)} = K(i+1) \quad (4.41)$$

$$\alpha(j)^{(i+1)} = \alpha(j)^{(i)} + K(i+1) \cdot \alpha(i-j+1)^{(i)} \quad j=1, \dots, i \quad (4.42)$$

$$E(i) = \left(1 - K(i+1) \right) \cdot E(i-1) \quad (4.43)$$

p : nombre de coefficients du filtre.

$\alpha(k)$: coefficients du filtre.

$E(i)$: erreur de prédiction à l'ordre "i".

$K(i)$: coefficients de corrélation partielle; ou coefficients de réflexion.

$R(i)$: coefficients d'autocorrélation, $0 \leq i \leq P$.

Utilisant la relation (4.43), l'erreur quadratique totale s'exprime dans ce cas par:

$$E(p) = R(0) \prod_{i=1}^p (1 - K^2(i)) \quad (4.44)$$

IV-5-STABILITE DU FILTRE PREDICTEUR

La stabilité d'un filtre de la forme " $I/A(Z)$ ou $P(Z)/A(Z)$ " est déterminée entièrement par les racines du dénominateur $A(Z)$. Si ces racines se trouvent à l'intérieur d'un cercle unité, le filtre est dit "stable". Dans le cas contraire il est dit instable.

La condition nécessaire et suffisante pour la stabilité de " $I/A(Z)$ " est que les coefficients de réflexion vérifient la condition suivante:

$$|K(i)| < 1 \quad i = 1, \dots, P \quad (4.45)$$

-Cette condition est théoriquement satisfaite pour la méthode d'autocorrélation.

En effet, le test de stabilité: $|K(i)| < 1 \quad i = 1, \dots, p$ se traduit par le test de la positivité de l'erreur $E(i)$.

D'après la relation (4.43), $E(i)$ est donné par:

$$E(i) = (1 - K^2(i+1)) \cdot E(i-1)$$

Or, $E(i)$ est une erreur quadratique, c'est-à-dire que c'est une quantité positive quelque soit "i".

De ce fait; $(1 - K^2(i+1)) \geq 0$

Soit $K^2(i+1) = |K(i+1)|^2 \leq 1$

D'où $\forall i \quad / K(i) / \leq I$

Ce qui implique que la stabilité du filtre "I/A(Z)" est théoriquement assurée dans le cas de la méthode d'autocorrélation.

-Par contre pour la méthode de covariance, la stabilité n'est généralement pas assurée. Donc il faut tester à chaque fois la relation de stabilité (4.45). Pour cela on calculera les coefficients de réflexion K(i) à partir des coefficients de prédiction a(i) par les relations récurentes suivantes:

$$K(i) = a^{(i)}(i)$$

$$a^{(i-1)}(j) = \left(a^{(i)}(j) - a^{(i)}(i) \cdot a^{(i)}(i-j) \right) / \left(I - K^2(i) \right) \\ j=1, \dots, i-1 \quad i=1, \dots, p$$

Les conditions initiales sont:

$$a^{(p)}(j) = a(j) \quad j=1, \dots, p$$

CONCLUSION

Nous avons développé dans ce chapitre les méthodes principales de résolution du système de prédiction linéaire:

-La méthode d'autocorrélation suppose le signal stationnaire à l'intérieur d'un intervalle. L'avantage principal de cette méthode est qu'elle assure la stabilité du filtre I/A(Z), mais en revanche elle pose un problème de précision des résultats, du fait de la troncature du signal originale.

-La méthode de covariance suppose le signal non stationnaire à l'intérieur de l'intervalle d'analyse. La stabilité du filtre I/A(Z) n'est pas assurée dans ce cas. Ce dernier tend à être plus stable quand le nombre d'échantillons "N" augmente.

La production de la parole étant un processus non stationnaire, par conséquent les hypothèses formulées dans le cas de la méthode de covariance sont plus réalistes que celles de la méthode d'autocorrélation. Néanmoins, dans un intervalle de temps relativement court, le signal peut être considéré quasi-stationnaire. De ce fait, l'hypothèse de stationnarité n'est pas fautive.

Du point de vue traitement informatique, il est intéressant de comparer les algorithmes utilisés, ainsi que la méthode de triangulation de Gauss, suivant ces deux critères;

-Taille mémoire

-Nombre d'opérations

	TAILLE MEMOIRE	NOMBRE D'OPERATIONS
TRIANGULARISATION DE GAUSS	P^2	$P(2P^2 + 6P - 2) / 6$
DECOMPOSITION DE CHOLESKY	$P(P+1) / 2$	$P(P^2 + 6P/11) / 6$
ALGORITHME DE DURBIN	$2P$	$P(P+1)$

Comparaison du nombre d'opérations nécessaires dans chaque méthodes (D'après Makhoul, 1972)

.....

CHAPITRE V

DETERMINATION

DU

GAIN

INTRODUCTION

On a vu que lors de la synthèse par la technique de prédiction linéaire, il fallait introduire un certain nombre de paramètres (toutes les IO à 25ms), pour pouvoir reconstituer de la parole synthétique.

Parmi ces paramètres on distingue un facteur très important, qui est le gain "G". En effet, le contrôle de ce gain permet de déterminer les différentes amplitudes du signal d'excitation, lors de la synthèse du signal de parole.

Dans une première étape, nous allons voir comment déterminer l'expression du gain pour les sons voisés et les sons non voisés.

La seconde étape est consacrée à la programmation. Le programme est exécuté en langage "BASIC", le calculateur utilisé étant le "RAINBOW 100", disponible au centre de calcul.

.

VI-I-DETERMINATION DU FACTEUR DE GAIN DU MODELE

Il est possible de trouver une relation entre le signal d'excitation $G.U(n)$ et l'erreur de prédiction $e(n)$.

En effet, en se référant aux expressions (3.8) et (3.14), le signal d'excitation peut-être exprimé comme suit:

$$G.U(n) \hat{=} S(n) - \sum_{k=1}^P a(k).S(n-k) \quad (5.1)$$

Or:

$$e(n) = S(n) - \sum_{k=1}^P a(k).S(n-k) \quad (5.2)$$

Dans ce cas: $e(n) = G.U(n)$

C'est-à-dire le signal d'entrée est proportionnel au signal d'erreur avec comme constante de proportionnalité le gain "G".

Le calcul de "G" est basé sur l'hypothèse selon laquelle l'énergie du signal de synthèse doit être égale à l'énergie totale de la séquence correspondante du signal analysé. En d'autres termes le signal de sortie du filtre $I/A(Z)$ doit avoir la même énergie que le signal à analyser, $S(n)$.

L'énergie totale contenue dans la réponse impulsionnelle de " $H(Z)$ " est donnée par sa fonction d'autocorrélation $\hat{R}(0)$.

A ce stade, pour la détermination du gain "G" il faut faire des hypothèses sur le signal d'entrés " $U(n)$ ".

Deux sortes d'entrée seront considérées: une déterministe et une autre aléatoire.

-La première consistera en une impulsion de Dirac: soit $U(n) = \delta(n)$

-La seconde est un bruit blanc stationnaire.

V-I-I-Calcul de "G" pour les sons voisins:

Nous supposons dans ce cas que le signal est analysé sur une période du fondamental, c'est-à-dire qu'à la synthèse il n'y aura qu'une seule impulsion d'excitation du filtre (pendant le temps correspondant).

L'entrée du filtre est alors une impulsion de Dirac:

$$U(n) = \delta(n) = \begin{cases} 1 & n=0 \\ 0 & \text{ailleurs} \end{cases}$$

On posera :

$$\hat{S}(n) = h(n)$$

On a alors en sortie du filtre "I/A(Z)":

$$h(n) = - \sum_{k=1}^P \alpha(k) \cdot h(n-k) + G \cdot \delta(n) \quad (5.3)$$

En multipliant les deux membres de l'expression (5.3) par $S(n-i)$, on obtient:

$$h(n) \cdot S(n-i) = - \sum_{k=1}^P \alpha(k) \cdot h(n-k) \cdot S(n-i) + G \cdot \delta(n) \cdot S(n-i) \quad (5.4)$$

En effectuant la sommation sur "n" de l'expression (5.4), nous aurons:

$$\sum_n h(n) \cdot S(n-i) = - \sum_n \sum_{k=1}^P \alpha(k) \cdot h(n-k) \cdot S(n-i) + G \sum_n \delta(n) \cdot S(n-i)$$

Or, $\sum_n h(n-i) \cdot h(n-k)$ est un coefficient d'autocorrélation de la réponse impulsionnelle. Notons le " $\hat{R}(i-k)$ ".

En distinguant les cas; $n = 0$ et $n \geq 1$, on aura les expressions suivantes:

$$\hat{R}(i) = - \sum_{k=1}^P \alpha(k) \cdot \hat{R}(i-k) \quad i \neq 0 \quad (5.5)$$

$$\hat{R}(0) = - \sum_{k=1}^P \alpha(k) \cdot \hat{R}(k) + G^2 \quad i = 0 \quad (5.6)$$

En comparant les expressions (5.5) et (4.9), il est clair que:

$$\hat{R}(i) = C \cdot R(i) \quad 0 \leq i \leq p \quad (5.7)$$

"C" étant un coefficient de proportionnalité .

L'utilisation de l'hypothèse selon laquelle l'énergie du signal de sortie "h(n)" est égale à celle du signal "S(n)", soit: $\hat{R}(0) = R(0)$, conduit à:

$$\hat{R}(i) = R(i) \quad 0 \leq i \leq p \quad (5.8)$$

Les $(p+1)$ coefficients d'autocorrélation de la réponse impulsionnelle sont égaux aux coefficients d'autocorrélation déjà calculés , (cf. éq. 4.9).

L'expression du gain sera dans ce cas:

$$G^2 = R(0) + \sum_{k=1}^P \alpha(k) \cdot R(k) \quad (5.9)$$

Or, d'après l'expression (4.18), on en déduit:

$$G^2 = E(p) \quad (5.10)$$

V-I-2-Calcul de "G" pour les sons non voisés:

Dans ce cas l'entrée $U(n)$ est un bruit blanc formé d'une suite d'échantillons non corrélés; de moyenne nulle et de variance unité:

$$E[U(n)] = 0 \quad \text{et} \quad E[U(n) \cdot U(n-m)] = \delta(m)$$

Ala sortie du filtre "I/A(Z)", on a le signal:

$$\hat{S}(n) = g(n) = - \sum_{k=1}^P \alpha(k) \cdot g(n-k) + G \cdot U(n) \quad (5.II)$$

En multipliant les deux membres de l'équation (5.II) par " $g(n-m)$ " et en prenant l'espérance mathématique des deux membres, nous obtenons:

$$E[g(n) \cdot g(n-m)] = - \sum_{k=1}^P \alpha(k) \cdot E[g(n-k) \cdot g(n-m)] + G \cdot E[U(n) \cdot g(n-m)] \quad (5.I2)$$

Comme les signaux " $g(n)$ " et " $U(n)$ " sont décorrélés (corrélacion nulle), c'est-à-dire:

$$E[g(n) \cdot g(n-m)] = 0 \quad m \neq 0$$

Nous remarquons que:

$$E[g(n) \cdot g(n-m)] = \hat{R}(m) \quad \text{et} \quad E[g(n-m) \cdot g(n-k)] = \hat{R}(m-k)$$

En distinguant les cas; $m=0$ et $m \neq 0$, l'expression (5.I2) devient:

$$\hat{R}(m) = - \sum_{k=1}^P \alpha(k) \cdot \hat{R}(m-k) \quad m \neq 0 \quad (5.I3)$$

$$\hat{R}(0) = - \sum_{k=1}^P \alpha(k) \cdot \hat{R}(k) + G \cdot E[U(n) \cdot g(n)] \quad m=0 \quad (5.I4)$$

Calculons la valeur de l'expression; $E[U(n).g(n)]$;

$$E[U(n).g(n)] = E[U(n).(-\sum_{k=1}^P \alpha(k).g(n-k) + G.U(n))]$$
$$E[g(n).U(n)] = -\sum_{k=1}^P \alpha(k).E[U(n).g(n)] + G.E[U^2(n)]$$

En utilisant les hypothèses suivantes:

-U(n),g(n) décorrélés, soit: $E[U(n).g(n-k)] = 0$

-Variance de U(n) égale à l'unité: $E[U^2(n)] = 1$

L'expression (5.I4) devient:

$$\hat{R}(0) = -\sum_{k=1}^P \alpha(k).\hat{R}(k) + G^2 \quad (5.I5)$$

En appliquant l'hypothèse d'égalité des énergies, de la réponse impulsionnelle et du signal S(n), on aboutit à:

$$G^2 = R(0) + \sum_{k=1}^P \alpha(k).R(k) = E(p) \quad (5.I6)$$

Nous pouvons exprimer ce gain en fonction de l'erreur normalisée, "v(p)".

En effet, ayant:

$$v(p) = E(p) / R(0)$$

On trouve:

$$G^2 = R(0).v(p) \quad (5.I7)$$

V-I-3-Remarque:

Pour déterminer le gain pour les sons voisés, on a supposé que le signal était analysé sur une période de mélodie (analyse synchrone). Par conséquent à la synthèse il n'y aurait qu'une seule impulsion d'excitation du filtre.

Mais dans le cas où le signal est analysé sur une période différente de celle de la mélodie (analyse asynchrone), le nombre d'impulsions d'excitation du filtre va être supérieur à un (dépendant du rapport entre la valeur de la période de mélodie T_m et la durée de la fenêtre T_f). Dans l'hypothèse d'une mélodie constante.

$$C \hat{=} T_m/T_f \quad \text{avec } T_m < T_f$$

La relation d'égalité des énergies s'écrit:

$$\hat{R}(0) = C.R(0) \quad C < 1 \quad (5.18)$$

Dans ce cas la constante "C" intervenant dans l'équation (5.7) n'est plus égale à un, mais possède la valeur définie ci-dessus.

Tenant compte des expressions (5.6), (5.7), (5.8), l'équation du gain devient:

$$G^2 = C.R(0) - C \sum_{k=1}^P a(k).R(k)$$
$$G^2 = C \left[R(0) - \sum_{k=1}^P a(k).R(k) \right] \quad (5.19)$$

CONCLUSION

Nous remarquons que l'expression du gain est identique selon que les sons sont voisés ou non, ceci facilitera considérablement la partie programmation.

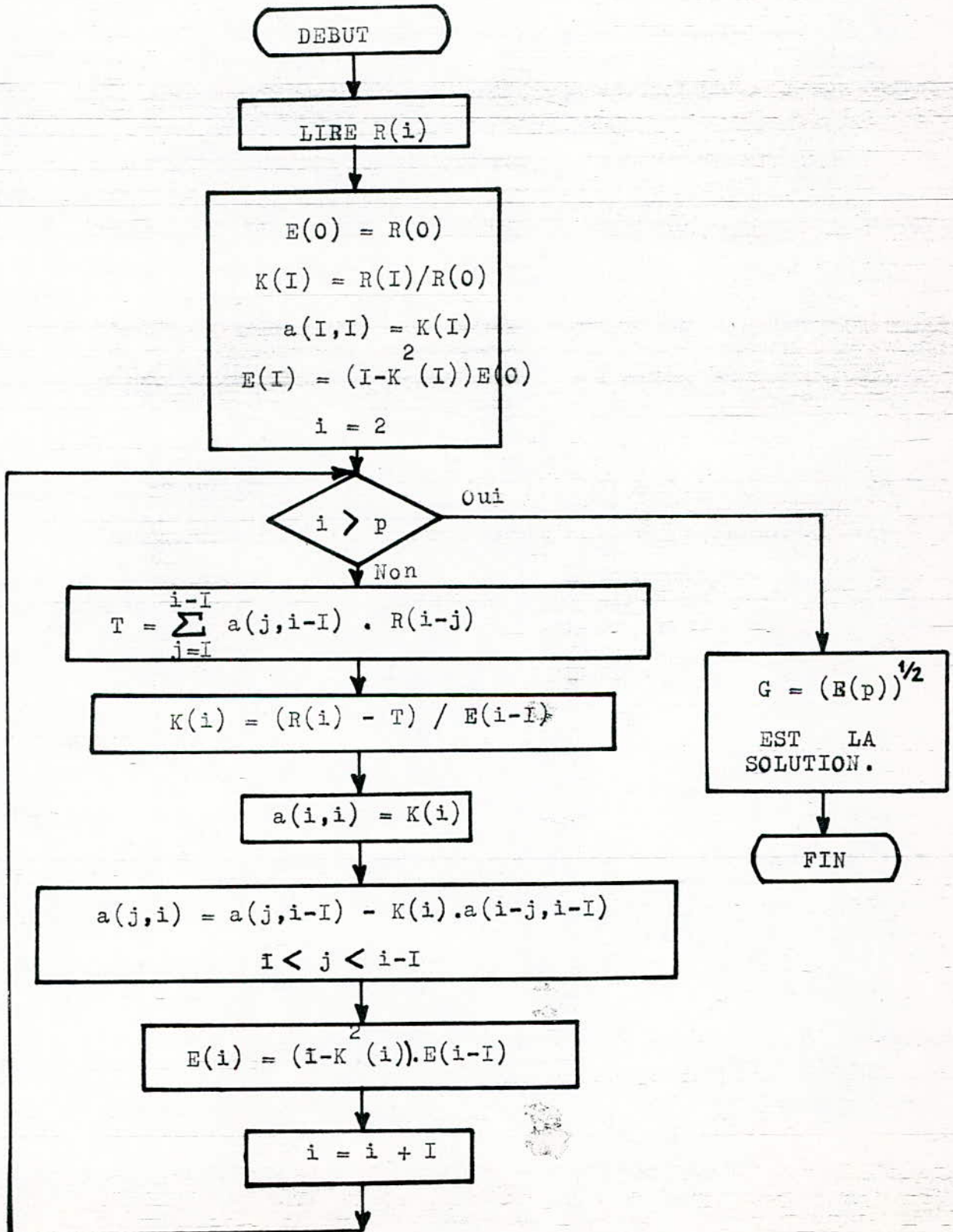
Dans le cas d'une analyse asynchrone, le gain "G" garde la même expression. Avec néanmoins une différence au niveau de la constante "C" qui n'est plus égale à un, mais au rapport entre la période " T_m " du fondamental et la durée de la fenêtre " T_f ".

V-2-

PROGRAMMATION

Nous présentons, dans cette partie l'organigramme et le programme pour le calcul du gain des différents sons. Le calculateur utilisé étant le "RAINBOW 100".

ORGANIGRAMME DE CALCUL DE GAIN PAR LA METHODE D'AUTOCORRELATION



```
100 REM*****
110 REM*      CALCUL DU GAIN PAR LA METHODE D'AUTOCORRELATION      *
120 REM*****
130 INPUT "ENTREZ LA VALEUR DE N=" ,N
140 INPUT "ENTREZ L'ORDRE DU PREDICTEUR K=" ,K
150 DIM S(N)
160 FOR I=1 TO N
170 S(I)=2*.9899999^(I-K)-.9899999^(2*(I-K))
180 NEXT I
190 DIM R(K)
200 R(0)=0
210 FOR J=1 TO N
220 P=S(J)^2
230 R(0)=R(0)+P
240 NEXT J
250 P=0
260 FOR I=1 TO K
270 R1=0
280 FOR J=1 TO N-I
290 P=S(J)*S(J+I)
300 R1=R1+P
310 NEXT J
320 R(I)=R1
330 PRINT "R(" , I , ")=" ;R(I)
340 NEXT I
350 REM RECHERCHE DES COEFFICIENTS DE RELEXION L(K)
360 DIM E(K),L(K),A(K,K)
370 E=0
380 A=0
390 L=0
400 E(0)=R(0)
410 L(1)=R(1)/R(0)
420 A(1,1)=L(1)
430 PRINT "L(1)=" ;L(1)
440 PRINT "A(1,1)=" ;A(1,1)
450 PRINT "E(0)=" ;E(0)
460 E(1)=(1-L(1)^2)*E(0)
470 FOR I= 2 TO K
480 T=0
490 FOR J=1 TO I-1
500 T=T+A(J,I-1)*R(I-J)
510 NEXT J
520 L(I)=(R(I)-T)/E(I-1)
530 PRINT "L(" , I , ")=" ,L(I)
540 REM RECHERCHE DES COEFFICIENTS DE PREDICTIONS A(J,I)
550 A(I,I)=L(I)
560 FOR J=1 TO I-1
570 A(J,I)=A(J,I-1)-L(I)*A(I-J,I-1)
580 PRINT "A(J,I)=" ,A(J,I)
590 NEXT J
600 REM RECHERCHE DE L'ERREUR QUADRATIQUE E(K) ET DU GAIN G
610 E(I)=(1-L(I)^2)*E(I-1)
620 PRINT "E(" , I , ")=" ;E(I)
630 NEXT I
640 G=SQR(E(K))
650 PRINT "GAIN=" ,G
660 END
```


CONCLUSION GENERALE

Dans ce projet, nous avons étudié quelques aspects du traitement du signal intervenant dans le cadre d'une application du dialogue "HOMME/MACHINE". Nous avons pris en considération l'analyse et la synthèse de la parole, en mettant l'accent sur le problème d'extraction de paramètres pertinents, tels que; les coefficients de prédiction et le gain "G" en particulier.

Cette étude nous a permis d'élargir nos connaissances dans le domaine de la parole, qui est un foyer nouveau de la recherche.

Nous avons montré que lors de la production de la parole, on pouvait adopter un filtre modèle ne possédant que des pôles (filtre récursif). C'est la théorie même de la technique de "prédiction linéaire".

Deux méthodes de cette dernière ont été étudiées en détail. Les hypothèses et les conditions d'application propres à chacune d'elles ont été énoncées.

Nous avons opté pour la méthode d'autocorrélation afin de déterminer le gain. Les raisons qui ont motivé notre choix sont les suivantes:

-Si sur le plan théorique, les hypothèses de la méthode de covariance sont plus réalistes, nous avons mentionné que l'hypothèse sur laquelle est fondée la méthode d'autocorrélation, à savoir la stationnarité du signal de parole dans un laps de temps relativement court n'était pas fausse.

-Une propriété importante de cette dernière est la stabilité du filtre prédicteur. Par conséquent les paramètres calculés peuvent être utilisés directement en synthèse. Ce qui n'est pas le cas pour l'autre méthode; il faut ajouter un algorithme de détection de stabilité, ce qui alourdit considérablement le programme.

Après avoir déterminé les paramètres du filtre prédicteur, nous avons proposé un algorithme de calcul de gain, en utilisant la relation entre ce dernier et l'erreur quadratique de prédiction. Nous avons remarqué que cette expression est la même, selon que l'on parte d'hypothèse de sons voisés ou de sons non voisés.

Nous avons vu en dernier lieu, comment était affecté ce gain dans le cas de l'analyse "asynchrone". Le résultat final est similaire au précédent. La seule différence réside dans la constante "C", qui n'est plus égale à "1" dans ce cas, mais au rapport entre la période du fondamental "Tm" et la durée de la fenêtre "Tf"; $C = T_m/T_f$.

Il aurait été souhaitable d'avoir des échantillons de parole, on aurait pu faire dans ce cas un classement des différents sons du langage.

Dans le cadre de notre travail, nous nous sommes intéressés à la synthèse par la technique de prédiction linéaire et en particulier à l'extraction du gain "G". Cette technique a bien d'autres applications dans le domaine de la parole, qui peuvent faire l'objet de travaux de projets de fin d'études d'ingénieur. Nous citerons en particulier les systèmes d'identification du locuteur.

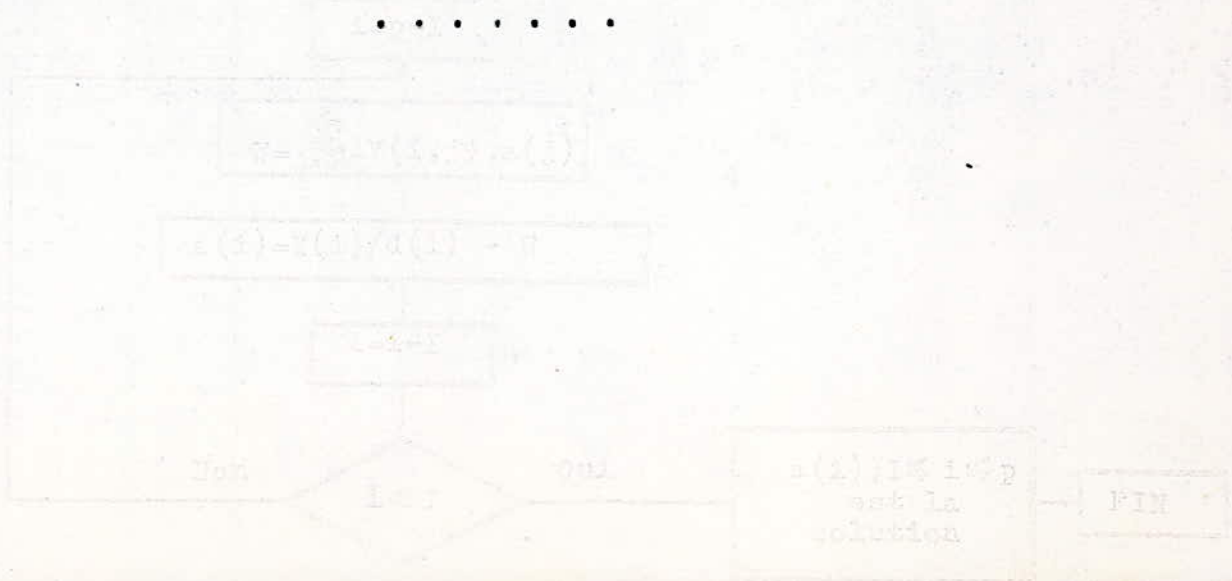
.

... nous présentons, dans cette partie, l'organigramme et le programme pour le calcul de gain des différents sons. Le calculateur utilisé est le... 1967.

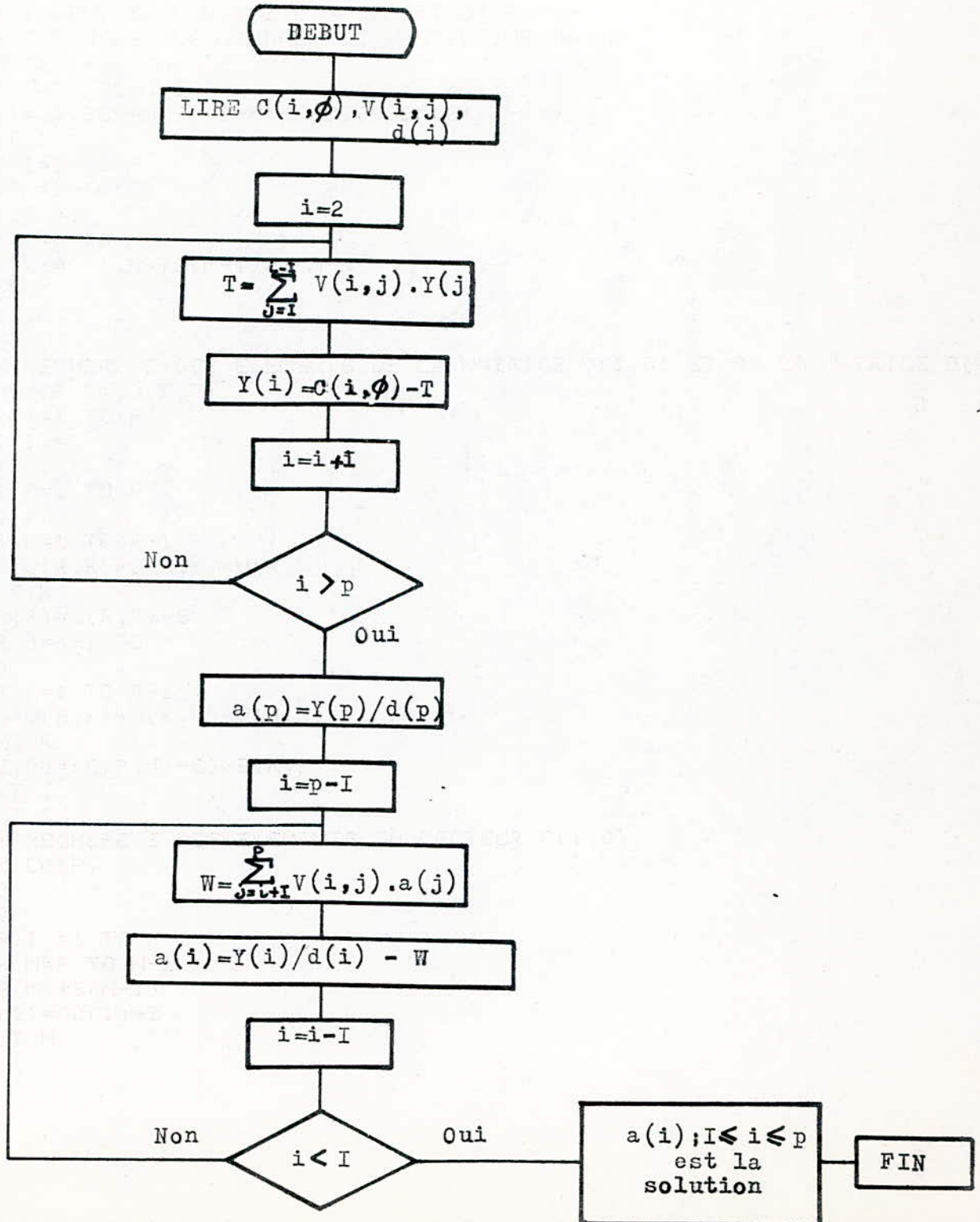
ANNEXE

Notre choix s'étant porté sur la méthode d'autocorrélation pour le calcul du gain, on a été amené à utiliser l'algorithme de "DURBIN" correspondant au calcul des coefficients prédictifs.

Vu qu'on a étudié aussi la résolution du système de prédiction linéaire par la méthode de covariance, nous présentons dans cet annexe l'organigramme et le programme correspondant à la décomposition de "CHOLESKY".



ORGANIGRAMME DE CALCUL DES COEFFICIENTS DU FILTRE PREDICTEUR PAR LA METHODE DE COVARIANCE




```
100 REM *****
110 REM * RECHERCHE DES COEFFICIENTS PAR LA METHODE DE COVARIANCE *
120 REM *****
130 REM RECHERCHE DES ELEMENTS DE LA MATRICE C(I,J)
140 INPUT "ENTREZ L'ORDRE DU PREDICTEUR P=",P
150 INPUT "ENTREZ LE NOMBRE D'ECHANTILLONS N=",N
160 DIM C(P,P),S(N)
170 FOR I=P TO N-1-P
180 S(I)=2*.9899999^(I-P)-.9899999^(2*(I-P))
190 NEXT I
200 FOR I=1 TO P
210 FOR K=1 TO P
220 C(I,K)=0
230 FOR M=P TO N-1
240 C(I,K)=C(I,K)+S(M-I)*S(M-K)
250 NEXT M
260 NEXT K
270 NEXT I
280 REM RECHERCHE DES ELEMENTS DE LA MATRICE V(I,J) ET DE LA MATRICE D(J)
290 DIM V(P,P),D(P,P)
300 FOR I=1 TO P
310 V(I,I)=1
320 NEXT I
330 FOR R=1 TO P
340 S=0
350 FOR K=1 TO R-1
360 S=S+V(R,K)*D(K,K)*V(R,K)
370 NEXT K
380 D(R,R)=C(R,R)-S
390 FOR J=R+1 TO P
400 S=0
410 FOR K=1 TO R-1
420 S=S+V(R,K)*D(K,K)*V(J,K)
430 NEXT K
440 V(J,R)=(C(R,J)-S)/D(R,R)
450 NEXT J
460 NEXT R
470 REM RECHERCHE DES ELEMENTS DU VECTEUR C(I,0)
480 DIM C0(P)
490 C=0
500 Z=0
510 FOR I =1 TO P
520 FOR M=P TO N-1
530 Z=S(M)*S(M-I)
540 C0(I)=C0(I)+Z
550 NEXT M
```

```
560 NEXT I
570 REM RECHERCHE DES ELEMENTS DU VECTEUR Y
580 DIM Y(P)
590 Y(1)=C0(1)
600 FOR I=2 TO P
610 G=0
620 FOR J=1 TO I-1
630 G=G+V(I,J)*Y(J)
640 NEXT J
650 Y(I)=C0(I)-G
660 NEXT I
670 REM RECHERCHE DES COEFFICIENTS DE PREDICTION A(I)
680 A(K)=Y(K)/D(K,K)
690 I=P-1
700 H=0
710 FOR J=I+1 TO P
720 H=H+V(J,I)*A(J)
730 NEXT J
740 A(I)=Y(I)/D(I,I)-H
750 I=I-1
760 IF I<1 THEN 780
770 GOTO 700
780 PRINT "COEFFICIENTS DE PREDICTION A(I)"
790 FOR I=1 TO P
800 PRINT "A(";I;")=";A(I)
810 NEXT I
820 REM RECHERCHE DE L'ERREUR QUADRATIQUE E
830 C1=0
840 FOR M=P TO N-1
850 C1=C1+S(M)^2
860 NEXT M
870 T=0
880 FOR K=1 TO P
890 T=T+Y(K)^2/D(K,K)
900 NEXT K
910 E=C1-T
920 PRINT "L'ERREUR DE PREDICTION EST";E
930 END
```


BIBLIOGRAPHIE

- 1-BELLANGER M. (1981)
"Traitement Numérique Du Signal." MASSON.
- 2-DESCOUT R.(1982)
"Les Techniques de Synthèse de la Parole."
Documentation Bibliographique Regroupée. CNET-LANNION.
- 3-EL MALAWANYI (1975)
"Contribution aux Recherches sur la Communication Parlée:
Etude de Vocoders à Prédiction Linéaire.Détermination de
l'Intervalle de Fermeture de la Glotte.Détection de la mé-
lodie.Extraction de la Fonction d'Aire du Conduit Vocal.
Thèse de Docteur-Ingenieur. Université de GRENOBLE.
- 4-FONDANECHÉ P.,GILBERTAS P.(1981)
"Filtres Numériques,Principes et Réalisations." MASSON.
- 5-GUERTI M (1983)
"Contribution à la Synthèse de la Parole en Arabe Standard."
Thèse de Magister.Université d'ALGER.
- 6-GUIBERT J. (1979)
"La Parole.Comprehension et Synthèse par les Ordinateurs."
- 7-KUNT M. (1981)
"Traitement Numérique des Signaux." DUNOD.
- 8-LE GUYADER A.(1978)
"Etude d'un Vocodeur à Excitation Vocale et à base de
Prédiction Linéaire".
Thèse de Doctorat de 3 cycle. Université de RENNES.
- 9-LIENARD J.S (1977)
"Les Processus de la Communication Parlée.Introduction
à l'Analyse et à la Synthèse de la Parole." MASSON.
- 10-LIFERMANN J (1980)
"Les Méthodes Rapides de Transformation du Signal:Fourier,
- 11- MARKEL J.D -GRAY,A.H (1976)
"Linear Prediction of Speech" SPRINGER-VERLAG.
- 12-MAX J (1981)
"Méthodes et Techniques de Traitement du Signal et Appli-
cations aux Mesures Physiques".tome 1. MASSON.

- 13-RABINER L.R ,SCHAFER R.W (1978)
"Digital Processing of Speech Signals." PRINTICE HALLS.
- 14-SERIGNAT J.F (1974)
"Application de la Prédiction Linéaire à l'Analyse de la Parole."
Bulletin de l'Institut de Phonétique de GRENOBLE.Vol 3
pp.23-52.
- 15-SONG J.M (1983)
"Relation entre le Gain et le Pitch en L.P.C."
Rapport de Stage.Université de PARIS 4.
- 16-STELLA M. (1983)
"Techniques de Prédiction Lineaire." CNET-LANNION.
- 17-THOMAS P.,J.LEBLANC J. (1981)
"Principes de Prédiction et Synthèse." MASSON.
- 18-THOMAS P. (1985)
"Contribution à la Synthèse de la Parole en Mode Standard."
Thèse de Magister, Université de ALBES.
- 19-THOMAS P. (1989)
"La Parole, Codement, Cod et Synthèse par les Ordinateurs."
Thèse de Magister, Université de ALBES.
- 20-THOMAS P. (1990)
"Principes de Synthèse de la Parole." MASSON.
- 21-LE SYRARD M. (1976)
"Etude de la Vitesse de l'Extraction Vocale et à Base de Prédiction Linéaire."
Thèse de Doctorat de 3 cycle, Université de ALBES.
- 22-THOMAS P. (1977)
"Les Procédés de la Communication Parlée, Introduction à l'Analyse et à la Synthèse de la Parole." MASSON.
- 23-LIFTON J. (1980)
"Les Méthodes Rapides de Transformation du Signal:Fourier,"
- 24-MARTEL J.G -URAY A.M (1976)
"Linear Prediction of Speech" SPRINGER-VERLAG.
- 25-THOMAS P. (1981)
"Méthodes et Techniques de Traitement du Signal et Applications aux Mémoires Électroniques", tome I, MASSON.