

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT d'Electronique



**PROJET DE FIN D'ETUDES**

**S U J E T**

**ETUDE D'UN VOCODER  
A PREDICTION LINEAIRE**

Proposé par :  
M.Guerti

Etudié par :  
A.Hamza  
D.Sedira

Dirigé par :  
M.Guerti

PROMOTION : Juin 86

A mon grand père BOUHALLEL Mokhtar.

A mes parents.

A Hafid, Sofiane, Adel.

A mes soeurs.

A mes tantes, Hafida et Sophia.

A tous ceux qui me sont chers.

ALILA

A mon père.

A ma mère.

A mes frères, Ghassan, Hamoud, Boualem, Smail.

A mes soeurs, Bahia, Samia, Saida, Malika.

A tous ceux que j'estime.

BDELKRIM

## S O M M A I R E

المدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

|   | Pages |
|---|-------|
| - <u>INTRODUCTION</u>   |       |
| - <u>CHAPITRE I : ETUDE DE LA PAROLE</u>  |       |
| Introduction  | 1     |
| I-1 : Anatomo-physiologie de l'appareil vocal   | 1     |
| I-2 : Principe de fonctionnement de l'appareil phonatoire   | 1     |
| I-3 : L'appareil phonatoire en tant que système acoustique  | 2     |
| I-4 : Les différents types de sons  | 2     |
| I-5 : Classification des différents sons  | 3     |
| I-6 : Caractéristiques physiques de la parole   | 4     |
| I-7 : Conclusion  | 4     |
| - <u>CHAPITRE II : APPLICATION DE LA PREDICTION LINEAIRE A L'ANALYSE-<br/>    SYNTHESE DE LA PAROLE</u> |       |
| Introduction  | 8     |
| II-1 : Principe de la prédiction linéaire   | 8     |
| II-2 : Analyse et synthèse de la parole par prédiction linéaire   | 8     |
| II-3 : Modèle de la production de la parole   | 10    |
| II-4 : Le modèle de prédiction linéaire   | 10    |
| II-5 : Optimisation des coefficients $a(k)$   | 12    |
| II-6 : Erreur quadratique minimale de prédiction  | 13    |
| II-7 : Calcul du gain   | 16    |
| II-8 : Stabilité du filtre prédicteur   | 17    |
| II-9 : Conclusion   | 18    |
| - <u>CHAPITRE III : PRINCIPAUX ALGORITHMES DE PREDICTION LINEAIRE</u>                                   |       |
| III-1 : Algorithme de la méthode d'autocorrélation  | 20    |
| III-2 : Algorithme de prédiction d'ITAKURA-SAITO  | 21    |
| III-3 : Comparaison des principaux algorithmes  | 23    |
| III-4 : Mode d'analyse  | 23    |
| III-5 : Nombre optimal de coefficients  | 24    |

|  |           |
|--|-----------|
| <b>- <u>CHAPITRE IV : SYNTHESE DE LA PAROLE</u></b>        |           |
| IV-1 : Les différentes méthodes de synthèse                | 27        |
| IV-2 : Les diverses techniques de synthèse                 | 28        |
| IV-2-1 : Vocoder à canaux                                  | 28        |
| IV-2-2 : Vocoder à formants                                | 31        |
| IV-3 : Conclusion  | 35        |
| <br>   |           |
| <b>- <u>CHAPITRE V : VOCODER A PREDICTION LINEAIRE</u></b> |           |
| Introduction   | 37        |
| V-1 : Choix de la méthode de résolution                    | 37        |
| V-2 : Elements constitutifs d'un vocoder                   | 38        |
| V-3 : L'analyseur du vocoder                               | 38        |
| V-3-1 : Le prétraitement                                   | 40        |
| V-3-2 : Le calcul de la fonction de transfert              | 41        |
| V-3-3 : La détection de voisement                          | 41        |
| V-4 : Le codage  | 46        |
| V-5 : La transmission                                      | 47        |
| V-6 : Le décodage  | 47        |
| V-7 : Le synthétiseur du vocoder                           | 47        |
| V-8 : Conclusion   | 52        |
| - PROGRAMMATION  | 53        |
| <br>   |           |
| <b>- <u>CONCLUSION</u></b>                                 | <b>56</b> |

— I N T R O D U C T I O N —

La parole étant le premier outil de communication entre les personnes et le moyen le plus facile pour l'échange des informations, l'homme a de tout temps essayé de trouver des procédés qui lui permettent de produire de la parole, fournissant ainsi des possibilités de communication à longue distance.

Les machines parlantes de VON KEMPLEN (1871) et FABER (1835) ont été conçues de manière à ressembler le plus possible au conduit vocal ; elles reproduisaient une parole compréhensible mais de caractère artificiel, ce qui s'expliquait par le fait que dans le message parlé, les sons sont enchaînés les uns aux autres et la parole est le résultat d'un mouvement continu de l'appareil vocal et non d'une suite de positions comme était le cas de la parole fournie par ces machines.

D'autres tentatives ont été menées au fil des années pour aboutir en 1939 au vocoder (voice coder) de DUDELY, qui marque le début des recherches modernes.

Avec l'apparition de l'ordinateur, ces recherches se sont tournées vers les techniques numériques qui présentent par rapport aux techniques analogiques divers avantages.

Parmi ces techniques numériques, et surtout dans le domaine de l'analyse-synthèse de la parole, nous aborderons dans ce projet l'étude de la prédiction linéaire, car elle permet l'estimation de tous les paramètres qui caractérisent le mieux possible le signal de parole.

Le travail que nous présentons, fait l'étude d'un vocoder (voice coder) utilisant la prédiction linéaire pour l'analyse et la synthèse de la parole.

Le premier chapitre de ce projet traite de la physiologie de l'appareil phonatoire humain.

Le deuxième aborde l'étude de la prédiction linéaire en tant que méthode d'analyse de la parole.

Le chapitre 3, est consacré à la résolution des équations de prédiction linéaire.

Le chapitre 4, regroupe les différentes méthodes et techniques de synthèse.

Le chapitre 5, présente l'étude d'un vocoder à prédiction linéaire.

Une programmation concernant la détermination des paramètres entrant dans le vocoder est incluse à la fin de ce chapitre.

← CHAPITRE I →

- INTRODUCTION :

- I-1 : ANATOMO-PHYSIOLOGIE DE L'APPAREIL VOCAL.
- I-2 : PRINCIPE DE FONCTIONNEMENT DE L'APPAREIL PHONATOIRE.
- I-3 : L'APPAREIL PHONATOIRE EN TANT QUE SYSTEME ACOUSTIQUE.
- I-4 : LES DIFFERENTS TYPES DE SONS.
- I-5 : CLASSIFICATION DES DIFFERENTS SONS.
- I-6 : CARACTERISTIQUES PHYSIQUES DE LA PAROLE.
- I-7 : CONCLUSION.

- INTRODUCTION :

Nous étudierons dans ce chapitre l'anatomie et la physiologie de l'appareil phonatoire humain en expliquant son fonctionnement acoustique.

Cette étude nous permettra de comprendre la formation des différents sons et leur enchainement dans la parole courante.

I-1 : ANATOMO-PHYSIOLOGIE DE L'APPAREIL VOCAL :

L'ensemble du système vocal se compose des poumons, du conduit trachéo-bronchique, du larynx et du conduit vocal. (Fig.1.1).

- 1) Les poumons et le conduit trachéo-bronchique :

La trachée artère est un conduit cylindrique dont l'extrémité supérieure supporte le larynx, tandis que son extrémité inférieure est liée aux bronches des poumons.

- 2) Le larynx :

C'est l'ensemble des cartilages articulés, ligaments, muscles et muqueuses qui surmontent la trachée artère et pénètre dans le pharynx. Il supporte deux muscles appelés "cordes vocales".

- 3) Le conduit vocal :

Il est formé du pharynx, de la cavité buccale de forme et de volume variables et de la cavité nasale.

Les cavités buccales et nasales peuvent être connectées en parallèle par le voile du palais.

I-2 : PRINCIPE DE FONCTIONNEMENT DE L'APPAREIL PHONATOIRE :

L'air est expulsé par la contraction de la cage thoracique vers les cordes vocales qui sont appliquées l'une sur l'autre avant son arrivée.

La pression sous glottique les fait vibrer. Le signal issu des cordes vocales chemine ensuite dans le conduit vocal. Ce dernier est constitué par les cavités pharyngales, buccales, nasales et labiales, où il sera filtré donnant naissance aux divers sons du langage.

I-3 : L'APPAREIL PHONATOIRE EN TANT QUE SYSTEME ACOUSTIQUE :

On peut décomposer le système acoustique en trois parties :

- Source d'énergie : constituée par les poumons.
- Source d'excitation : constituée par les cordes vocales.
- Système résonant : constitué par les cavités du conduit vocal.

L'excitateur délivre un signal de source modifié par la fonction de transfert du résonateur. (Fig. 1.2).

I-4 : LES DIFFERENTS TYPES DE SONS :

1 - Sons voisés :

Lorsqu'on a une vibration périodique des cordes vocales, les sons émis sont dits voisés ou sonores, (Fig. 1.3).

2 - Sons non voisés :

Ils sont dûs à un passage turbulent de l'air à travers une constriction très étroite du conduit vocal, (Fig. 1.4).

3 - Opposition orale-nasale :

Les cavités nasales et buccales peuvent être isolées par le voile du palais, cet état est celui qu'on observe lors de la production des sons "oraux". Lors du couplage des deux cavités, on obtient un signal dit "nasal"

I-5 : CLASSIFICATION DES DIFFERENTS SONS :

Les phonèmes sont définis comme étant les sons simples dont l'enchaînement forme le langage.

La classification des phonèmes peut s'effectuer d'après les critères suivants :

- critère de phonation : opposition orale-nasale, voisée non voisée.
- mode d'articulation : permet de montrer la façon dont l'air s'écoule à travers le conduit respiratoire.

La classification générale utilisée en linguistique permet de distinguer les deux sous ensembles suivants :

1- les voyelles : elles sont dûes à une vibration laryngienne.

- voyelles orales : l'influence de la cavité nasale est nulle. (|a|, |e|, |u|).

- voyelles nasalisées : le rayonnement se fait par les narines et la bouche. (|o|, |oe|).

2- les consonnes : elles sont dûes à une constriction du passage de l'air.

- consonnes fricatives : c'est le bruit de l'écoulement de l'air dans des constriction importantes du conduit vocal. (|s|, |f|).

Certaines fricatives sont dites voisées. (|j|, |v|).

- consonnes plosives : elles proviennent de l'occlusion du conduit vocal suivie d'une brusque ouverture. Elles sont voisées comme (|t|, |d|). Elles sont non voisées comme (|p|, |t|, |k|).

- consonnes nasales : la transmission et le rayonnement se fait principalement à travers le conduit nasal (|m|).

3- Les semi-voyelles et liquides : leur comportement est assez voisin des voyelles orales ( l , r ).

I-6 : CARACTERISTIQUES PHYSIQUES DE LA PAROLE :

1- les formants : se définissent comme étant les maximum de la courbe de réponse en fréquences du conduit vocal (LIENARD, 1977) (Fig. 1.5).

2- les antiformants : les minimas du spectre du signal acoustique provenant des couplages entre cavités nasales et buccales sont appelés antiformants.

3- hauteur : la fréquence la plus basse dans le signal de parole est la fréquence fondamentale de vibrations des cordes vocales, elle varie approximativement de 100 à 150 Hz pour les hommes et de 200 à 250 Hz chez les femmes et jusqu'à 400 Hz pour les enfants (CATIER, 1983).

I-7 : CONCLUSION :

Nous avons fais une étude brève de l'ensemble du système vocal en nous limitant aux points susceptibles de jouer un rôle dans la phonation.

Par ailleurs, nous avons examiné la production des sons élémentaires de la parole (voyelles, consonnes, semi-voyelles) considérés isolement. Certaines caractéristiques de la parole ont été également abordées afin d'avoir une vue complète sur ses différents aspects.

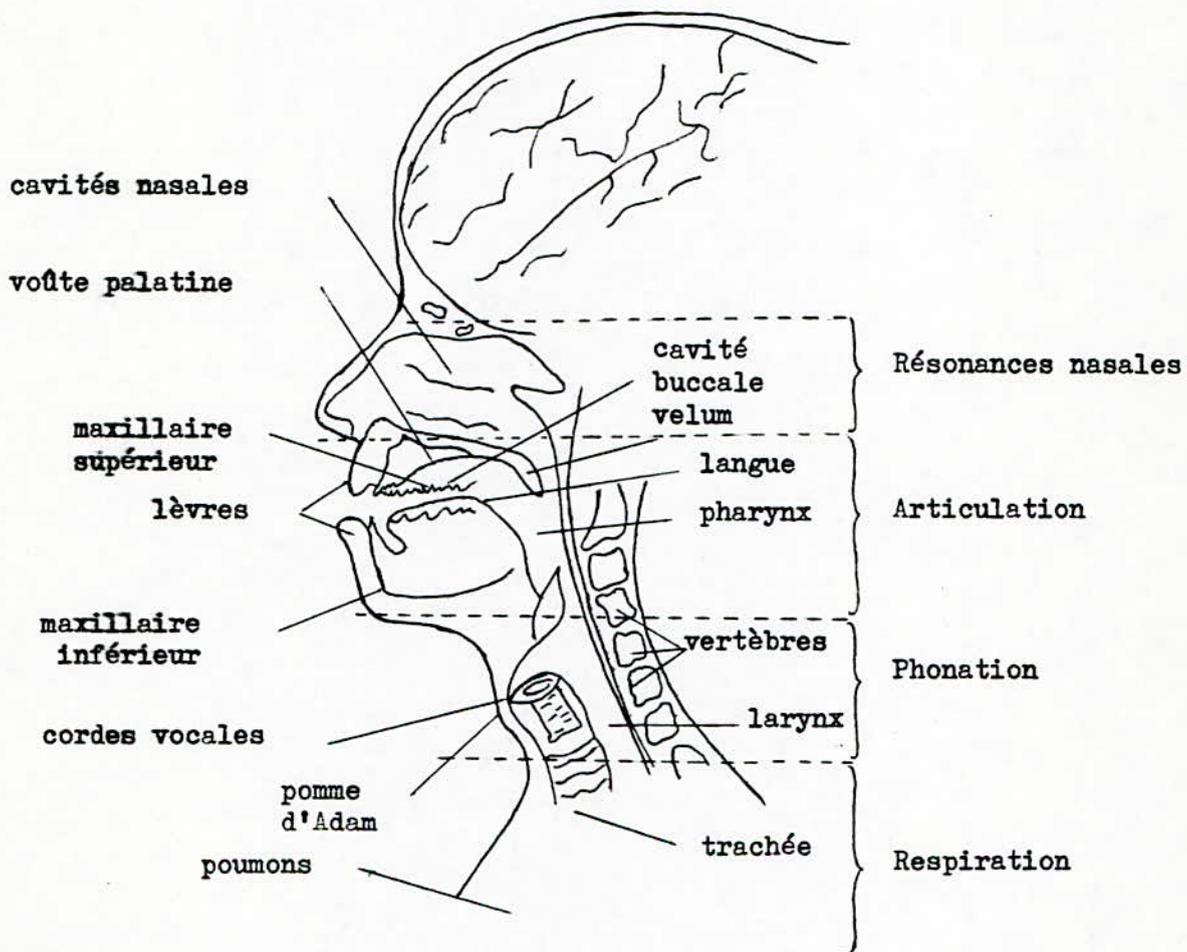


Fig. 1.1 : Organes de la parole.

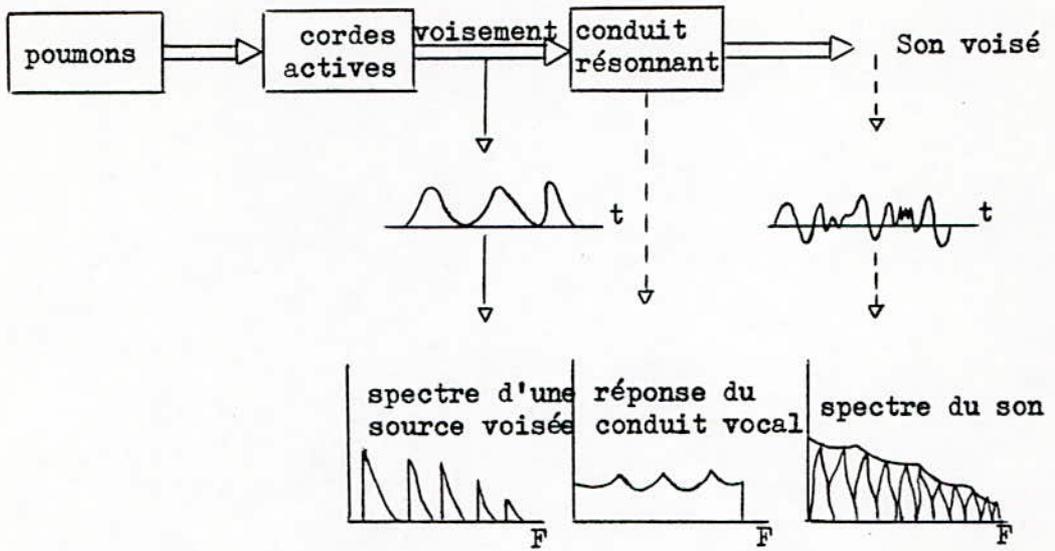


Fig. 1.4 : Production d'un son voisé.

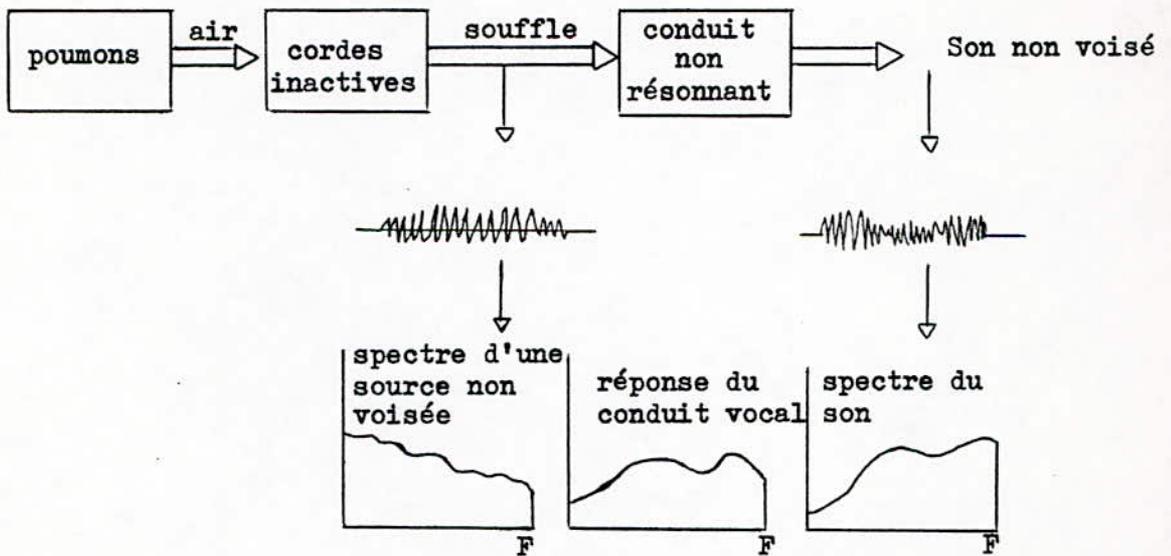


Fig. 1.5 : Production d'un son non voisé.

## CHAPITRE II

### APPLICATION DE LA PREDICTION LINEAIRE

#### A L'ANALYSE SYNTHÈSE DE LA PAROLE

#### - INTRODUCTION :

- II - 1 Principe de la prédiction linéaire.
- II - 2 Analyse et synthèse de la parole par prédiction linéaire.
- II - 3 Modèle de la production de la parole.
- II - 4 Modèle de prédiction linéaire.
- II - 5 Optimisation des coefficients  $a(k)$ .
- II - 6 Erreur quadratique totale minimale de prédiction.
- II - 7 Calcul du gain.
- II - 8 Stabilité du filtre prédicteur.

## INTRODUCTION :

Les méthodes classiques d'analyse de la parole (transformée de Fourier, corrélation, convolution, ...) souffrent de limitations à cause du caractère non stationnaire et pseudo-périodique de la parole.

La prédiction linéaire qui peut être considérée comme une méthode d'analyse temporelle ou spectrale souffre des mêmes limitations; sa supériorité provient du fait qu'elle est fondée sur un modèle simple (filtre numérique n'ayant que des pôles).

### II - 1 : PRINCIPE DE LA PREDICTION LINEAIRE :

La méthode est basée sur le fait qu'un échantillon de paroles  $S(nT)$  peut être prédit par une combinaison linéaire d'un certain nombre d'échantillons prélevés à des instants précédents.

La valeur prédite de  $S(n)$  s'obtient par une somme pondérée linéairement de  $p$  échantillons :

$$\hat{S}(n) = \sum_{k=1}^p a(k) \cdot S(n-k) \dots \dots \dots \text{(II-1-1)}$$

### II - 2 : ANALYSE ET SYNTHESE DE LA PAROLE PAR PREDICTION LINEAIRE :

L'idée essentielle de l'analyse synthèse de la parole est de considérer les signaux de parole comme la réponse d'un filtre linéaire.

#### II - 2 - 1 : L'analyse de la parole :

On considère le signal de parole  $S(t)$  ;  $t \in [0, T]$

Le signal  $S(t)$  pris dans son ensemble n'est pas stationnaire, pour cela on découpe le temps en intervalles de 10 à 25 ms, appelés fenêtres temporelles, ainsi la restriction du signal à chacun de ces  $n$  intervalles sera considérée comme stationnaire.

L'échantillonnage du signal sur chaque fenêtre donnera une suite d'échantillons  $S(n)$ ,  $n \in [0, N]$  .  $N$  étant le nombre d'échantillons.

La prédiction linéaire permet d'écrire :

$$S(n) = - \sum_{k=1}^P a(k) \cdot S(n-k) + G \cdot U(n) \dots \dots \dots (II - 2 - 1)$$

où :

- a(k) représente les coefficients du filtre prédicteur.
- G représente un facteur de gain.

L'équation (II - 2 - 1) donne :

$$S(n) + \sum_{k=1}^P a(k) S(n-k) = G \cdot U(n) \dots \dots \dots (II - 2 - 2)$$

En prenant la transformée en Z de (II - 2 - 2) :

$$S(Z) + \sum_{k=1}^P a(k) \cdot Z^{-k} \cdot S(Z) = G \cdot U(Z)$$

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

d'où :

$$\frac{S(Z)}{U(Z)} = \frac{G}{1 + \sum_{k=1}^P a(k) \cdot Z^{-k}} = \frac{G}{\sum_{k=0}^P a(k) \cdot Z^{-k}}$$

$$\frac{S(Z)}{U(Z)} = \frac{G}{\sum_{k=0}^P a(k) \cdot Z^{-k}} \dots \dots \dots (II - 2 - 3)$$

avec a(0) = 1

définissons le filtre inverse par :

$$A(Z) = \sum_{k=0}^P a(k) \cdot Z^{-k}$$

Le signal S(n) est donc la sortie d'un filtre linéaire de fonction de transfert G/A(Z), excité par U(n).

II - 2 - 2 : Synthèse de la parole :

L'excitation du filtre linéaire G/A(Z) par un signal d'entrée U(n) permet d'obtenir en sortie des échantillons de parole synthétiques  $\hat{S}(n)$ , la nature de U(n) est liée aux caractéristiques des différents sons produits (voisés - non voisés). (Fig 1)

## II - 3 : MODELE DE LA PRODUCTION DE LA PAROLE :

La production de la parole est caractérisée par deux types d'excitation :

- Le premier type d'excitation est produit par la vibration des cordes vocales. L'onde produite a approximativement une forme triangulaire qui peut être assimilée à un train d'impulsions de période  $p$  appelée "pitch" (figure 2).

- Le deuxième type d'excitation est produit par le passage d'air turbulent dans une constriction du conduit vocal, il peut être assimilé à du bruit blanc (spectre plat).

## II - 4 : LE MODELE DE PREDICTION LINEAIRE :

Le modèle de production de la parole utilisé dans les méthodes de prédiction linéaire repose sur les hypothèses suivantes :

1 - Pendant des intervalles de temps de 10 à 25 ms, la source vocale et le conduit vocal peuvent être considérés comme constants.

2 - Pendant chaque intervalle de stationnarité, la fonction de transfert de l'ensemble de l'appareil vocal peut être représentée par un filtre numérique récursif à coefficients constants, ne possédant que des pôles.

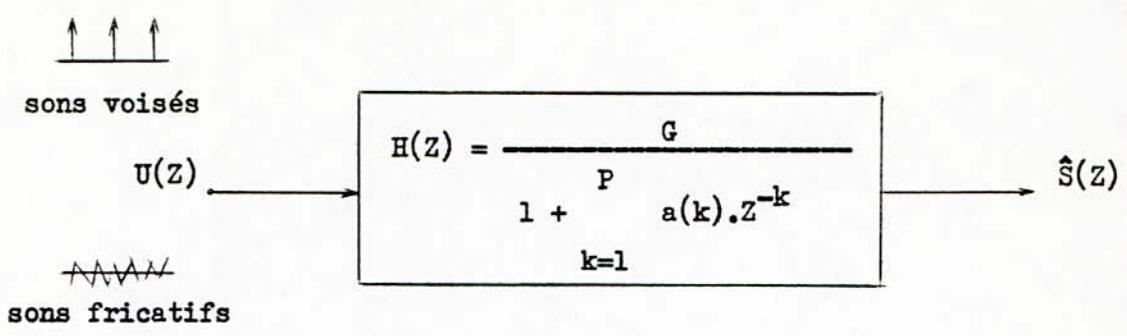
3 - L'entrée du filtre est un signal impulsionnel de période égale au fondamental pour les sons voisés, et une séquence de bruit blanc pour les sons sourds.

4 - Les coefficients du filtre numérique ou les paramètres déduits de ce dernier sont renouvelés toutes les 5 à 20 ms de façon à ne pas obtenir des variations brusques des paramètres.

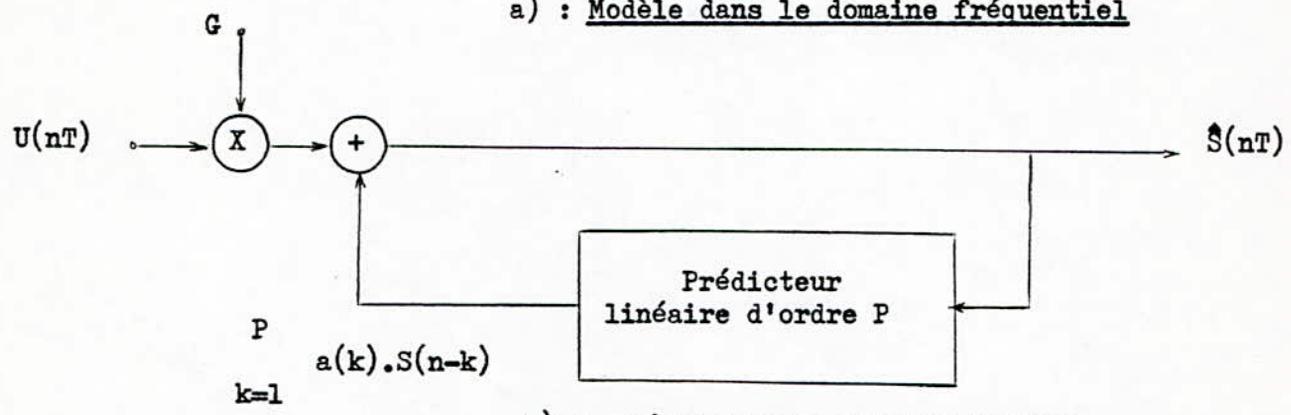
### Remarque :

Pour certains sons (nasalisés et sourds) la fonction de transfert du système phonatoire comporte des zéros; ceci n'est pas gênant car : - l'oreille est plus sensible à un maximum d'énergie qu'à un minimum.

- du point de vue mathématique, un zero peut être approché par un nombre limité de pôles.



a) : Modèle dans le domaine fréquentiel



b) : Modèle dans le domaine temporel

Fig. 3 : Modèle de production de la parole

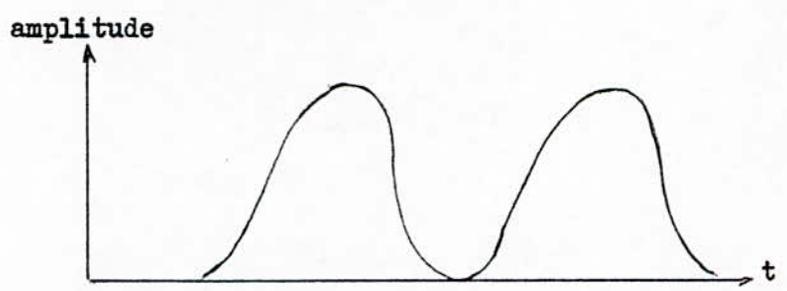


Fig. 2 : Forme de l'onde glottique

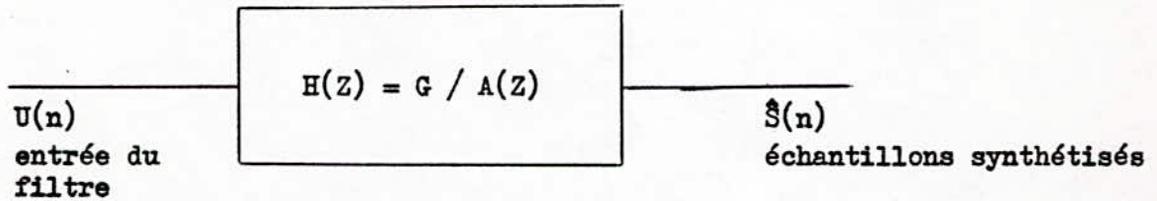


Fig. 1 : Production du signal de parole

II - 5 : OPTIMISATION DES COEFFICIENTS DU PREDICTEUR :

L'optimisation des coefficients  $a(k)$  consiste à trouver les  $a(k)$  qui donnent en sortie du modèle un signal, le plus proche possible du signal à analyser.

L'erreur entre le signal  $S(n)$  et le signal prédit  $\hat{S}(n)$  est :

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

D'après (II - 1 - 1) :

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

$$e(n) = \sum_{k=0}^P a(k) \cdot S(n-k) \quad a(0) = 1$$

L'optimisation des coefficients  $a(k)$  revient à minimiser l'erreur quadratique moyenne  $E$ .

$$E = \sum_{n=n_0}^{n_1} e(n)^2$$

$$E = \sum_{n=n_0}^{n_1} \left( S(n) - \sum_{k=1}^P a(k) \cdot S(n-k) \right)^2 \dots \dots \dots \text{(II-5-1)}$$

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

$$\partial E / \partial a(j) = 0 \quad \text{pour } j = 1, 2, \dots \dots \dots P$$

$$\frac{\partial}{\partial a(j)} \left[ \sum_{n=n_0}^{n_1} \left( S(n) - \sum_{k=1}^P a(k) \cdot S(n-k) \right)^2 \right] = 0 \dots \text{(II-5-2)}$$

En dérivant (II-5-2) on obtient :

$$\begin{aligned} 2 \sum_{n=n_0}^{n_1} \left[ S(n) - \sum_{k=1}^P a(k) \cdot S(n-k) \right] \cdot S(n-j) &= 0 \\ \sum_{n=n_0}^{n_1} S(n) \cdot S(n-j) - \sum_{n=n_0}^{n_1} \sum_{k=1}^P a(k) \cdot S(n-k) \cdot S(n-j) &= 0 \\ \sum_{n=n_0}^{n_1} \sum_{k=1}^P a(k) \cdot S(n-k) \cdot S(n-j) &= \sum_{n=n_0}^{n_1} S(n) \cdot S(n-j) \\ \sum_{k=1}^P a(k) \sum_{n=n_0}^{n_1} S(n-k) \cdot S(n-j) &= \sum_{n=n_0}^{n_1} S(n) \cdot S(n-j) \dots \text{(II-5-3)} \end{aligned}$$

posons :

$$C(j,k) = \sum_{n=n_0}^{n_1} S(n-k) \cdot S(n-j) \quad j = 1, 2, \dots, p$$

$$C(j,0) = \sum_{n=n_0} S(n) \cdot S(n-j)$$

L'équation (II-5-3) donne :

$$\sum_{k=1}^P a(k) \cdot C(j,k) = - C(j,0)$$

Les équations (II-5-3) constituent un ensemble de "p" équation à "p" inconnues, que l'on peut résoudre pour obtenir les coefficients prédicteurs a(k).

## II - 6 : L'ERREUR QUADRATIQUE TOTALE MINIMALE DE PREDICTION E(P) :

C'est la valeur minimale de E déterminée une fois les coefficients a(k) calculés.

A partir de (II-5-1) et (II-5-3) on démontre que :

$$E = \sum_{n=n_0}^{n_1} \left[ S(n) - \sum_{k=1}^P a(k) \cdot S(n-k) \right]^2$$

$$E = \sum_{n=n_0}^{n_1} \left[ S(n)^2 - 2 \sum_{k=1}^P a(k) \sum_{j=1}^P a(j) \sum_{n=n_0}^{n_1} S(n-k) \cdot S(n-j) \right] \dots (II-6-1)$$

d'après (II-5-3) :

$$\sum_{j=1}^P a(j) \sum_{n=n_0}^{n_1} S(n-j) \cdot S(n-k) = \sum_{n=n_0}^{n_1} S(n) \cdot S(n-j)$$

En remplaçant dans (II-6-1) on a :

$$E = \sum_{n=n_0}^{n_1} S(n)^2 - 2 \sum_{k=1}^P a(k) \sum_{n=n_0}^{n_1} S(n) \cdot S(n-k) +$$

$$\sum_{k=1}^P a(k) \sum_{n=n_0}^{n_1} a(k) \sum_{n=n_0}^{n_1} S(n) \cdot S(n-k)$$

$$E = \sum_{n=n_0}^{n_1} S(n)^2 - \sum_{k=1}^P a(k) \sum_{n=n_0}^{n_1} S(n) \cdot S(n-k)$$

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

$n_0, n_1$  sont les limites de sommation dont le choix conduit aux méthodes suivantes :

II - 6-1 : METHODE D'AUTOCORRELATION :

Les hypothèses de base de cette méthode sont les suivantes :

- le signal est défini pour toutes les valeurs du temps, il est nul en dehors d'une fenêtre de N échantillon.
- l'erreur quadratique minimale est minimisée sur toute l'échelle des temps.

La relation (II - 5 - 3) s'écrira donc :

$$\sum_{k=1}^p a(k) \sum_{n_0=-\infty}^{\infty} S(n-k) \cdot S(n-j) = \sum_{n=-\infty}^{\infty} S(n) \cdot S(n-j) \dots \quad (\text{II-6-2})$$

En posant  $m = n-j$ , l'équation (II-6-2) devient :

$$\sum_{k=1}^p a(k) \sum_{m=-\infty}^{\infty} S(m + |j-k|) \cdot S(m) = \sum_{m=-\infty}^{\infty} S(m+j) \cdot S(m)$$

L'autocorrelation de  $S(n)$  est définie par :

$$R(j) = \sum_{n=-\infty}^{\infty} S(n) \cdot S(n+j) \quad \text{avec } R(j) = R(-j)$$

$$R(j) = \sum_{k=1}^p a(k) \cdot R(j-k)$$

donc  $R(j) = \sum_{k=1}^p a(k) \cdot R(|j-k|) \quad 1 \leq j \leq p \dots \dots \dots (\text{II-6-3})$

$$R(j) = \sum_{n=0}^{N-1-j} S(n) \cdot S(n+j)$$

car le signal est limité par une fenêtre et  $S(n) = 0$  si  $n \notin [0, N-1]$ .

Le système s'écrit sous forme matricielle :

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \cdot \\ \cdot \\ \cdot \\ a(p) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \cdot \\ \cdot \\ \cdot \\ R(p) \end{bmatrix}$$

Remarque : La matrice carrée  $R(p)$  est symétrique, de plus les éléments situés sur les diagonales parallèles à la diagonale principale sont égaux.

II - 6 - 2 : METHODE DE COVARIANCE :

Les hypothèses de base sont :

- le signal est déterminé uniquement sur une séquence de N échantillons.

- l'erreur quadratique est minimisée sur les (N-M) derniers échantillons.

donc :  $n_0 = 0$  ,  $n_1 = N-1$

d'où : 
$$C(j,k) = \sum_{n=p}^{N-1} S(n-j).S(n-k) \quad j,k = 0, \dots, p$$

Le système s'écrit sous la représentation matricielle suivante :

$$\begin{bmatrix} C(1,1) & C(1,2) & \dots & C(1,p) \\ C(2,1) & C(2,2) & \dots & C(2,p) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ C(p,1) & C(p,2) & \dots & C(p,p) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \cdot \\ \cdot \\ \cdot \\ a(p) \end{bmatrix} = \begin{bmatrix} C(1,0) \\ C(2,0) \\ \cdot \\ \cdot \\ \cdot \\ C(p,0) \end{bmatrix}$$

II - 6 - 3 : COMPARAISON DES DEUX METHODES :

La méthode de covariance ne considère pas le signal comme stationnaire, elle n'assure pas la stabilité du modèle.

La méthode d'autocorrélation suppose le signal stationnaire à l'intérieur de la fenêtre, elle garantit la stabilité du filtre.

| Propriétés | METHODE         |            |
|------------|-----------------|------------|
|            | Autocorrélation | Covariance |
| Fenêtre    | Nécessaire      | non        |
| Stabilité  | Garantie        | non        |

II - 7 : CALCUL DU GAIN :

Pour synthétiser la parole, le modèle autorégressif est excité soit par une suite de bruit blanc, soit par des impulsions..

On peut calculer le gain en faisant appel au critère suivant :

"L'énergie totale contenue dans la réponse impulsionnelle doit être égale à l'énergie totale du signal analysé. (EL MALAWANY, 1975)".

Le signal reconstitué est donné par la formule :

$$S(n) = - \sum_{k=1}^P a(k) \cdot S(n-k) + G \cdot \delta(n) \dots \dots \dots \text{(II-7-1)}$$

$\delta(n)$  est une impulsion de Dirac.

et

$$\begin{cases} \hat{S}(n) = 0 & \text{pour } n < 0 \\ \hat{S}(0) = G & \text{pour } n = 0 \\ \hat{S}(n) = \sum_{k=1}^P a(k) \cdot \hat{S}(n-k) & \text{pour } n > 0 \end{cases}$$

L'autocorrélation du signal prédit est :

$$\begin{aligned} \hat{R}(j) &= \sum_{n=-\infty}^{\infty} \hat{S}(n) \cdot \hat{S}(n+j) \\ \hat{R}(0) &= \sum_{n=0}^{\infty} S(n)^2 \quad \text{pour } j = 0 \end{aligned}$$

En remplaçant  $S(n)$  par sa valeur de (II-7-1) et en distinguant les cas  $n = 0$  et  $n \geq 1$ , l'expression de  $R(0)$  devient :

$$\hat{R}(0) = \hat{S}(0)^2 + \sum_{n=1}^{\infty} \hat{S}(n) \sum_{k=1}^P a(k) \cdot \hat{S}(n-k)$$

En posant  $m = n-k$

$$S(m) = 0 \quad \text{pour } m < 0$$

On a ::

$$\begin{aligned} \hat{R}(0) &= G^2 + \sum_{k=1}^P a(k) \cdot \left( \sum_{m=1-k}^{\infty} \hat{S}(m) \cdot \hat{S}(m+k) \right) \\ \hat{R}(0) &= G^2 + \sum_{k=1}^P a(k) \cdot \hat{R}(k) \end{aligned}$$

D'où :

$$\begin{aligned} G^2 &= \hat{R}(0) - \sum_{k=1}^P a(k) \cdot \hat{R}(k) \\ \hat{R}(j) &= \sum_{n=0}^{\infty} \hat{S}(n) \cdot \hat{S}(n+j) \quad \text{pour } j > 1 \\ \hat{R}(j) &= \sum_{k=1}^P a(k) \cdot \hat{R}(j-k) \end{aligned}$$

Nous pouvons conclure que l'autocorrélation du modèle est proportionnelle à celle du signal dans l'intervalle  $0 \leq j \leq p$ .

$$\hat{R}(j) = c R(j)$$

En égalisant les énergies de  $S(n)$  et  $\hat{S}(n)$  on a :

$$\hat{R}(0) = R(0)$$

Et par conséquent :  $c = 1$   $\hat{R}(j) = R(j)$   $0 \leq j \leq p$

$$G^2 = R(0) - \sum_{k=1}^p a(k) \cdot R(k)$$

## II - 8 : STABILITE DU FILTRE A(Z) :

Il est important de savoir si le filtre est stable ou non, notamment dans l'hypothèse où ce filtre est utilisé à des fins de synthèse.

$H(Z)$  est stable si tous les pôles sont situés à l'intérieur du cercle unité, dans le plan  $Z$ .

L'autocorrélation garantie la stabilité de  $H(Z)$ , cette propriété résulte de la théorie des polynômes orthogonaux (EL MALAWANI, 1975).

Pour la méthode de covariance, la stabilité n'est pas garantie, cependant, on peut réduire les possibilités d'instabilité du prédicteur en augmentant le nombre d'échantillons contenus dans la séquence du signal à analyser. Pour cela, on augmente la valeur  $N$ .

## II - 9 : CONCLUSION :

Dans ce chapitre, ont été étudié les différents aspects de la prédiction linéaire. Dans cette méthode, le modèle de production de la parole est constitué par un filtre numérique récursif ne possédant que des pôles.

L'analyse-synthèse de la parole par cette méthode consiste à représenter le signal de parole par un nombre de paramètres se rapportant à la fonction de transfert du conduit vocal et aux caractéristiques du signal source.

L'avantage de la prédiction linéaire sur les autres méthodes tient au fait qu'elle permet de reconstituer en partie, à la synthèse, l'information perdue sur l'enveloppe au cours de l'analyse; en ce sens que le modèle utilisé effectue une interpolation entre les points fournis par l'analyse.

## CHAPITRE III

### PRINCIPAUX ALGORITHMES DE PREDICTION LINEAIRE

III - 1 : Algorithme de la méthode d'autocorrélation.

III - 2 : Algorithme de prédiction d'ITAKURA - SAITO.

III - 3 : Comparaison des principaux algorithmes.

III - 4 : Mode d'analyse.

III - 5 : Nombre optimal de coefficients.

PRINCIPAUX ALGORITHMES DE PREDICTION LINEAIRE :

Les différentes méthodes de prédiction linéaire font appel à une résolution de "p" équation à "p" inconnues.

Plusieurs algorithmes permettent d'effectuer une telle résolution tels que les méthodes de GAUSS - SEIDEL, JACOBI, ...

III - 1 : ALGORITHME DE LA METHODE D'AUTOCORRELATION :

Le système à résoudre est le suivant :

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \cdot \\ \cdot \\ \cdot \\ a(p) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \cdot \\ \cdot \\ \cdot \\ R(p) \end{bmatrix}$$

$$\text{Avec } E(p) = R(0) - \sum_{k=1}^p a(k) \cdot R(k)$$

L'algorithme de DURBIN permettant la résolution du système est le suivant :

$$\begin{aligned} & E(0) = R(0) \\ & \text{pour } I \text{ allant de } 0 \text{ à } p \\ ! & K(I) = a_i^i \\ ! & K(I) = - (R(i) - \sum_k a_k^{i-1} \cdot R(i-k)) / E(i-1) \\ ! & a_k^i = a_k^{i-1} + K(I) \cdot a_{i-k}^{i-1} \quad 1 \leq k \leq i-1 \\ ! & E(I) = (1 - K(I)^2) \cdot E(I-1) \end{aligned}$$

L'algorithme de DURBIN fait apparaître de nouveaux coefficients, les coefficients de réflexion dont le module d'amplitude est inférieur à 1 pour les filtres stables.

III - 2 : ALGORITHME DE PREDICTION D'ITAKURA-SAITO :

Définissons  $S_m^+(n)$  comme l'erreur de prédiction de  $S(n)$  à partir des "m" échantillons précédants.

$$S_m^+(n) = \sum_{k=0}^m a(m,k) \cdot S(n-k)$$

et  $S_m^-(n)$  comme l'erreur de prédiction de  $S(n-m-1)$  à partir des mêmes échantillons.

$$S_m^-(n) = \sum_{k=1}^{m+1} a(m, m+1-k) \cdot S(n-k)$$

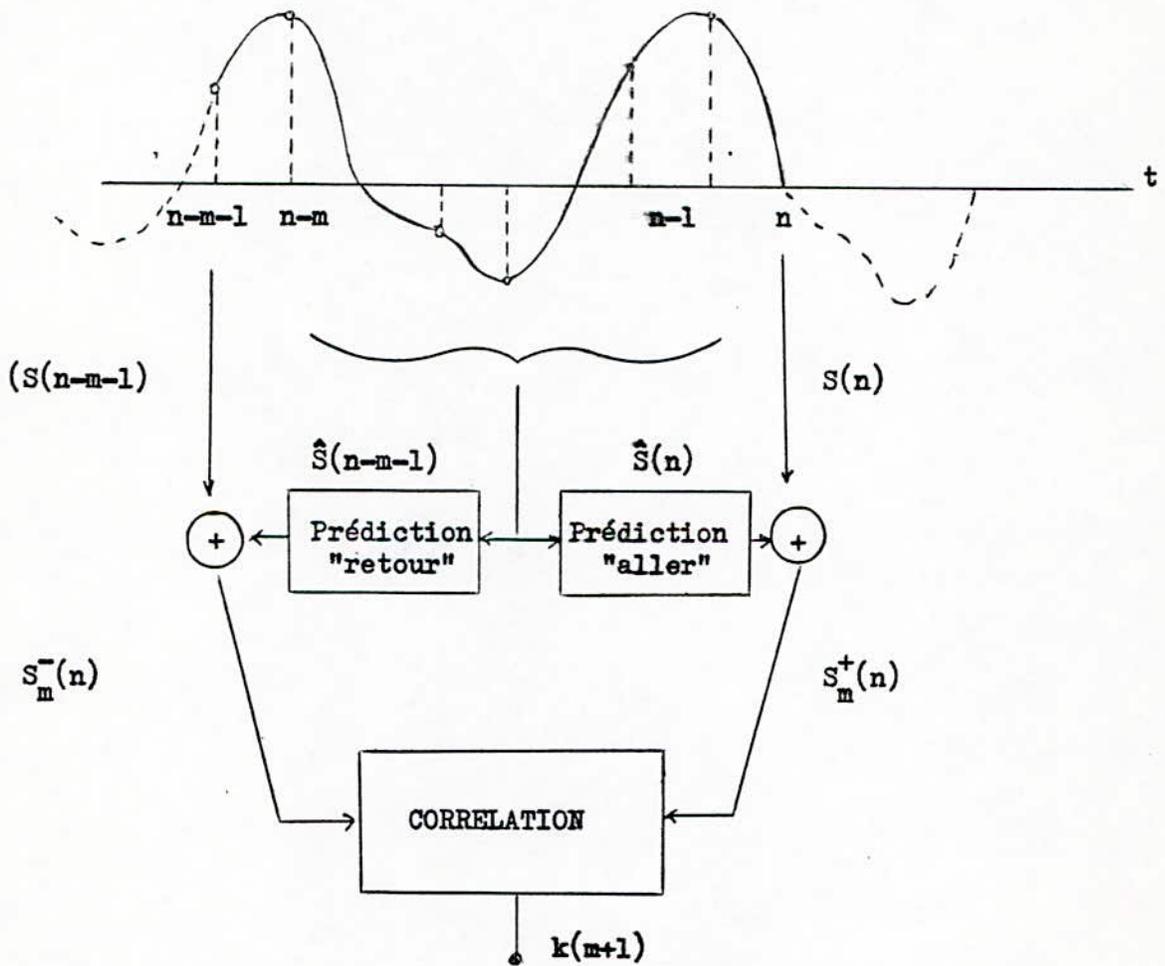


fig 3.1

Fig. 3.1 : Définition des coefficients  $k(m)$

ITAKURA et SAITO définissent les coefficients de corrélation partielles (parcor) par la corrélation entre les signaux  $S_m^+(n)$  et  $S_m^-(n)$ . (Fig. 3.1).

$$k_{(m+1)} = - \frac{\sum_{n=0}^{N+m-1} S_m^+(n) \cdot S_m^-(n)}{\sum_{n=0} S_m^+(n)^2 \cdot \sum_{n=0} S_m^-(n)^2}^{\frac{1}{2}}$$

Si le conduit vocal est modélisé par un tube acoustique ayant M sections successives, le passage de la cellule d'analyse m à la cellule m+1 se traduit par les relations suivantes :

$$S_{m+1}^+(n) = S_m^+(n) + k_{(m+1)} \cdot S_m^-(n) \dots \dots \dots (II-9-1)$$

$$S_{m+1}^-(n) = S_m^-(n) + k_{(m+1)} \cdot S_m^+(n) \dots \dots \dots (II-9-2)$$

Les relations (II-9-1) et (II-9-2) permettent de réaliser un filtre en trellis, modèle du conduit vocal. (Fig. 3.2).

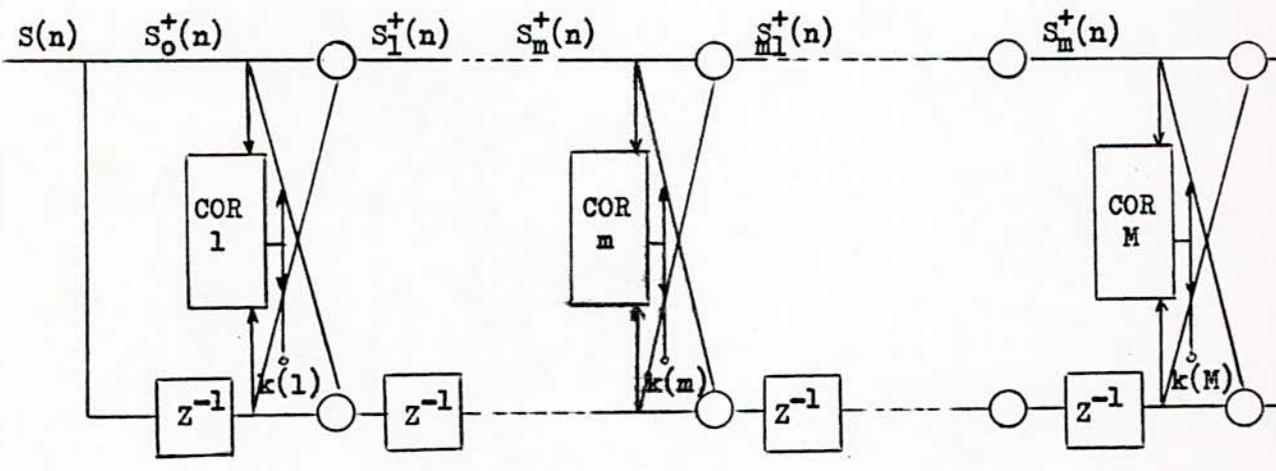


Fig. 3.2 : Filtre numérique pour le calcul des paramètres  $k(m)$

### III - 3 : COMPARAISON DES PRINCIPAUX ALGORITHMES DE PREDICTION LINEAIRE :

Si P est l'ordre du prédicteur et N le nombre d'échantillons, la méthode d'autocorrélation nécessite P.N multiplications pour le calcul des coefficients de corrélation.

Dans la méthode d'ITAKURA-SAITO, le nombre de multiplications pour obtenir les "P" coefficients  $k(m)$  est de l'ordre de 5P.N

Ces deux méthodes garantissent théoriquement la stabilité du filtre de synthèse.

Elles nécessitent toute les deux un fenêtrage sauf, que pour celle d'ITAKURA, il n'y a aucune hypothèse sur la nature du signal en dehors de la fenêtre.

### III - 4 : LES MODES D'ANALYSE :

La méthode d'autocorrélation permet d'obtenir des pôles avec une grande précision.

Les résultats obtenus par cette méthode dépendent essentiellement du type du signal à analyser et du choix de la fenêtre.

Pour les fricatives, la largeur et la position de la fenêtre ne sont pas des paramètres critiques.

Avec une fenêtre de 10 à 25 ms, on obtient de bons résultats; par contre pour les sons voisés, il est nécessaire d'examiner l'influence de la largeur et de la position de la fenêtre d'analyse.

#### III-4-1 : Analyse asynchrone :

Dans l'analyse asynchrone, la longueur de la séquence du signal analysé et sa position sont indépendantes des impulsions d'excitation du conduit vocal.

La fenêtre la plus utilisée dans le domaine de la parole est celle de HAMMING, dont l'expression est :

$$F_h(t) = \begin{cases} 0,54 + 0,46 \cos \frac{\pi t}{T_0} & t \quad T_0 \\ 0 & t \quad T_0 \end{cases}$$

III-4-2 : Analyse synchrone :

Le mode d'analyse synchrone implique le recours à un moyen de détection des instants correspondant aux débuts de fermeture de la glotte.

Si on utilise l'autocorrélation à une séquence du signal de durée  $T_m$ , il faut alors que la fenêtre d'analyse soit rectangulaire et qu'elle coïncide en largeur et en position avec une période d'excitation du conduit vocal

III - 5 : NOMBRE OPTIMAL DE COEFFICIENTS DU PREDICTEUR :

Il est possible de déterminer le nombre optimal de coefficients du prédicteur en se référant à la courbe d'évolution de l'erreur normalisée en fonction du nombre de pôles.

L'erreur normalisée  $V(p)$  de prédiction est définie par :

$$V(p) = E(p)/R(o) = 1 - \sum_{k=1}^p a(k).R(k)/R(o)$$

De même cette erreur de prédiction  $V(p)$  est liée au spectre de puissance  $P(w)$  par la relation :

$$V(p) = \frac{\exp T/2 \cdot \int_{-\pi/T}^{\pi/T} \log \hat{P}(w) dw}{T/2 \cdot \int_{-\pi/T}^{\pi/T} \hat{P}(w) dw} \dots\dots\dots (III.5.1)$$

où  $\hat{P}(w)$  est le spectre de puissance approché du signal de parole.

À partir de là, des considérations sur l'approximation spectrale relèvent que si :

- p est petit alors  $\hat{P}(w)$  n'est pas proche de  $P(w)$ .
- p est grand  $\hat{P}(w)$  approche  $P(w)$ .
- p tend vers l'infini alors  $\hat{P}(w)$  est identique à  $P(w)$

(Fig. 3.3)

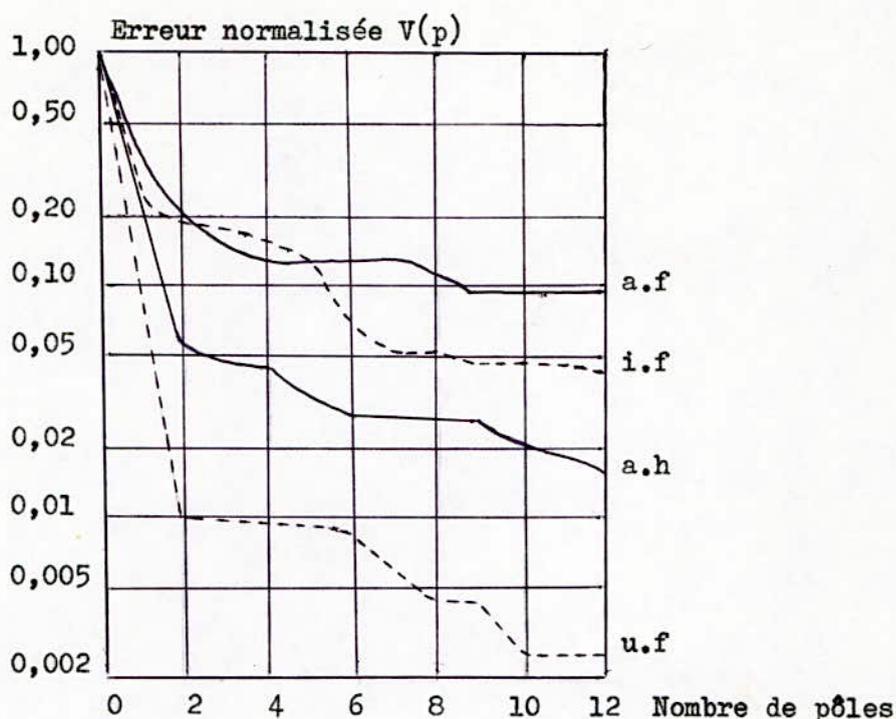


Fig. 3.3. : Erreur normalisée pour 4 voyelles (f désigne voix de femme et h voix d'homme)

L'erreur de prédiction donne donc une indication pour le choix de l'ordre du prédicteur, cependant et en admettant que pour chaque formant il faut associer deux coefficients et en admettant que le spectre de parole peut être décrit à l'aide de 5 résonances, 10 coefficients suffisent donc, pour représenter le conduit vocal. On adjoint en général 2 coefficients pour représenter le spectre de la source d'excitation et du rayonnement au niveau des lèvres.

## CHAPITRE IV

### SYNTHESE DE LA PAROLE

#### IV - 1 : LES DIFFERENTES METHODES DE SYNTHESE.

IV-1-1 : La synthèse par phrases.

IV-1-2 : La synthèse par mots.

IV-1-3 : La synthèse par règles.

IV-1-4 : La synthèse par diphtongues.

#### IV - 2 : LES DIVERSES TECHNIQUES DE SYNTHESE.

IV-2-1 : Vocodeur à canaux.

IV-2-2 : Vocodeur à formants.

#### IV - 3 : CONCLUSION.

#### IV - 1 : LES DIFFERENTES METHODES DE SYNTHESE :

Nous entendons par synthèse, la reproduction artificielle de la parole à partir d'éléments phonétiques pré-enregistrés et analysés.

Il existe quatre types de synthèses :

- La synthèse par phrases.
- La synthèse par mots.
- La synthèse par règles.
- La synthèse par diphtongues.

##### IV-1-1 : La synthèse par phrases :

La synthèse par phrases consiste à mémoriser un ensemble de phrases, puis à les reproduire à la demande suivant les besoins.

Le temps d'accès à l'information doit être très court, ce qui implique un vocabulaire limité et des moyens modernes de mémorisation.

##### IV-1-2 : La synthèse par mots :

On utilise des mots pré-enregistrés et mémorisés, la réalisation de n'importe quel message se fait par juxtaposition des mots désirés.

##### IV-1-3 : La synthèse par règles :

Des études ont permis de déterminer un ensemble de règles de transitions d'un son élémentaire au suivant (LIENARD, 1977).

##### IV-1-4 : La synthèse par diphtongues :

Le diphtongue est défini comme étant le segment qui s'étend de la zone stable d'une réalisation phonétique à la zone stable suivante et qui protège en son centre toute la zone de transition. Tout message est susceptible d'être découpé en diphtongues. Il suffit pour reconstituer un mot ou une phrase d'assembler les diphtongues nécessaires en tenant compte des zones de stabilité et des zones de transition entre phonèmes.

#### IV - 2 : LES DIVERSES TECHNIQUES DE SYNTHÈSE :

Le stockage et la restitution de la parole sous forme d'enregistrement analogique est caractérisé par un temps d'accès très long et par un nombre limité de messages. Ces contraintes ont permis l'apparition et le développement d'autres techniques de synthèse de la parole moins lourdes et assurant une bonne qualité à la parole de synthèse. Parmi ces techniques, et plus particulièrement dans le domaine spectral, on peut citer les vocodeurs à canaux, à formants, et à prédiction linéaire.

##### IV-2-1 : Vocodeur à canaux :

###### - Structure du vocodeur à canaux :

Le vocodeur à canaux est constitué d'un organe d'analyse et d'un organe de synthèse.

###### 1 - L'analyseur du vocodeur :

La fonction d'analyse de l'enveloppe spectrale est effectuée à l'aide d'un banc de filtres passe-bandes couvrant l'étendue de la bande téléphonique (300 - 3400 Hz).

Ainsi, le message à transmettre est décomposé en bandes de largeur d'environ 300 Hz, chaque canal traite une bande de fréquences déterminée.

Après chaque filtre, il y a un étage de détection, lui même suivi d'un étage de filtrage passe-bas limité à 25 Hz environ. Cette limitation des variations d'énergie à 25 Hz a pour effet de modifier ou de supprimer les transitions rapides du niveau d'énergie.

L'analyseur comporte aussi un détecteur de voisement, il permet de différencier les sons voisés des non voisés et de donner dans le premier cas la valeur du pitch. Les signaux issus de la sortie de l'analyseur sont codés avant d'être transmis au synthétiseur.

## 2 - Codage des signaux :

Le codage désigne le passage de la représentation analogique à la représentation numérique caractérisée par son débit exprimé en bit/seconde.

Le codage permet le stockage et la transmission numérique de l'ensemble des informations relatives au spectre d'amplitude et à la structure fine de ce dernier.

Cet ensemble d'informations numériques est appelé "échantillon vocodeur". Fig.42. (EMERARD, 1977).

Les informations numériques nécessaires pour le codage d'un échantillon vocodeur sont réparties comme suit :

- concernant le spectre :

(nombre de canaux) X (4 bits par canal)

- concernant le pitch (période codée): on utilise 8 bits.

Pour un nombre de canaux égal à 16, et pour une fréquence d'échantillonnage de 50 Hz le débit est de :

$$((16 \times 4) + 8) \times 50 = 3600 \text{ bits/sec. (ZURSHER, 1983)}$$

Une fois le codage effectué, les informations numériques sont transmises et stockées dans la mémoire du ordinateur; si l'on désire une reconstitution différée du signal de parole.

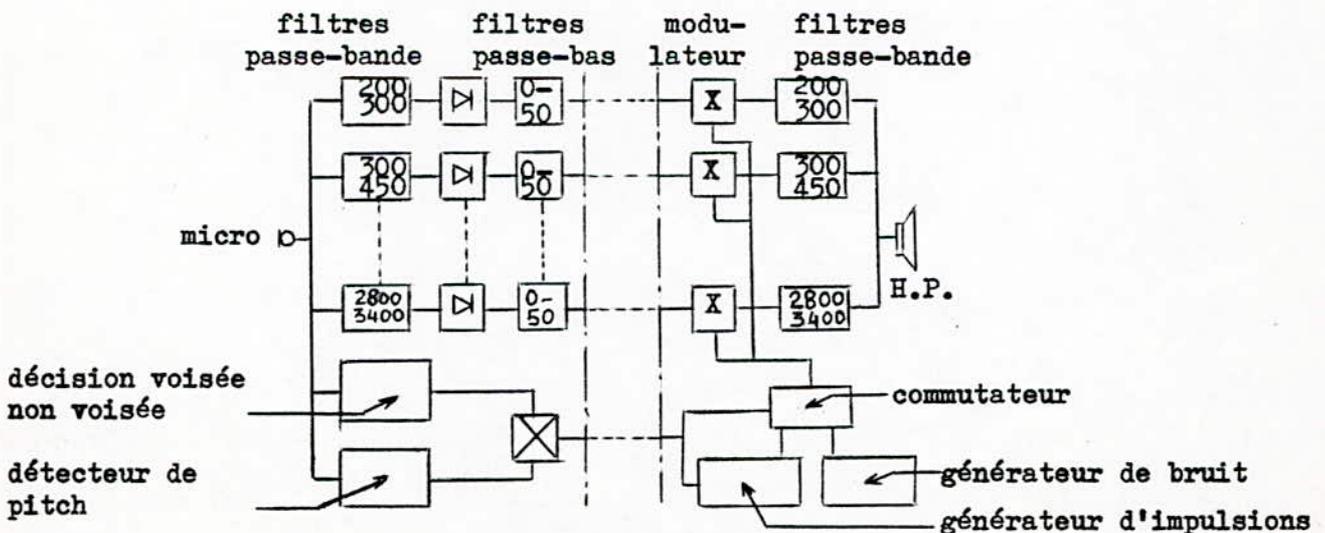


Fig. 4.1 : Vocodeur à canaux.



### 3 - Le synthétiseur du vocodeur :

La synthèse est effectuée à l'aide d'un banc de filtres passe-bande identique à ceux de l'analyseur. Ainsi, le signal issu des filtres d'analyse est multiplié par le signal d'excitation dans un modulateur. Le signal d'excitation est délivré soit par un générateur d'impulsions (sons voisés), soit par un générateur de bruit (sons non voisés).

Le signal de sortie est obtenu par addition des signaux des filtres de synthèse.

### IV-2-2 : Vocodeur à formants :

Dans le vocodeur à formants on transmet des paramètres porteurs d'informations précises concernant l'emplacement et l'amplitude des formants.

#### 1 - L'analyseur :

L'analyse consistera donc, à déterminer les formants par diverses techniques dont on citera quelques unes.

La première consiste à rechercher les 3 premiers maximum du spectre instantané à l'aide d'un banc de filtres passe-bande contigus (entre 30 et 50 ). Cette technique est sujette à des erreurs dues aux maximum parasites qui peuvent apparaître dans le spectre (ZURSHER, 1983). De plus, le premier formant peut être confondu avec la fréquence du fondamental  $F_0$ .

La deuxième technique consiste à distinguer par filtrage passe-bande, les zones d'évolution des 3 formants.

L'inconvénient de la deuxième technique réside dans le fait que les zones d'évolution des formants sont largement variables d'une personne à une autre et plusieurs formants peuvent apparaître dans une même zone.

Une troisième technique consiste en l'extraction des formants par prédiction linéaire ; en effet, les formants correspondent aux résonances du conduit vocal, c'est-à-dire à des pôles de la fonction de transfert du conduit vocal. Chaque formant correspond à une paire de racines complexes conjuguées ayant la forme suivante :

$$Z_k, Z_k^* = Z_{kr} + j.Z_{ki}$$

La fréquence centrale  $F_k$  et la bande passante  $F_k$  du formant sont données par :

$$F_k = w(k) / 2 = (1 / 2T) \cdot \arctg (Z_{ki} / Z_{kr})$$

$$F_k = - (k) / = - (1 / 2T) \cdot \ln(Z_{kr}^2 + Z_{ki}^2)$$

## 2 - Synthétiseur à formants :

Il est constitué de plusieurs résonateurs dont on commande la fréquence et l'amplitude.

Une branche parallèle de fréquence fixe simule les résonances nasales.

Le signal d'excitation est soit un signal périodique, soit du bruit blanc.

Il existe deux façons de connecter les filtres principaux, au nombre de trois en général :

### - Synthétiseur parallèle :

Les filtres sont tous alimentés par le même signal qu'ils renforcent dans des bandes différentes.

L'amplitude d'un filtre n'influe pas sur celle d'un autre.

Le signal de sortie est la parole synthétique. (Fig. ).

- Synthétiseur série :

Dans ce type de synthétiseur, les filtres sont placés en cascades. Le filtre F2 renforce une partie du signal délivré par le filtre F1, de même pour le filtre F3 qui renforce le signal délivré par F1 et F2. (Pg 44)

Comme dans le vocodeur à canaux, on retrouve deux sources d'excitation à savoir, un générateur d'impulsions périodiques pour les sons voisés et un générateur de bruit pour les fricatives.

L'enveloppe du spectre est déterminée par 3 branches montées en parallèle:

- Canal des formants vocaux :

Il correspond à la branche médiane de la figure 44. On y trouve 3 résonances variables qui donneront naissance aux 3 premiers formants F1, F2, F3.

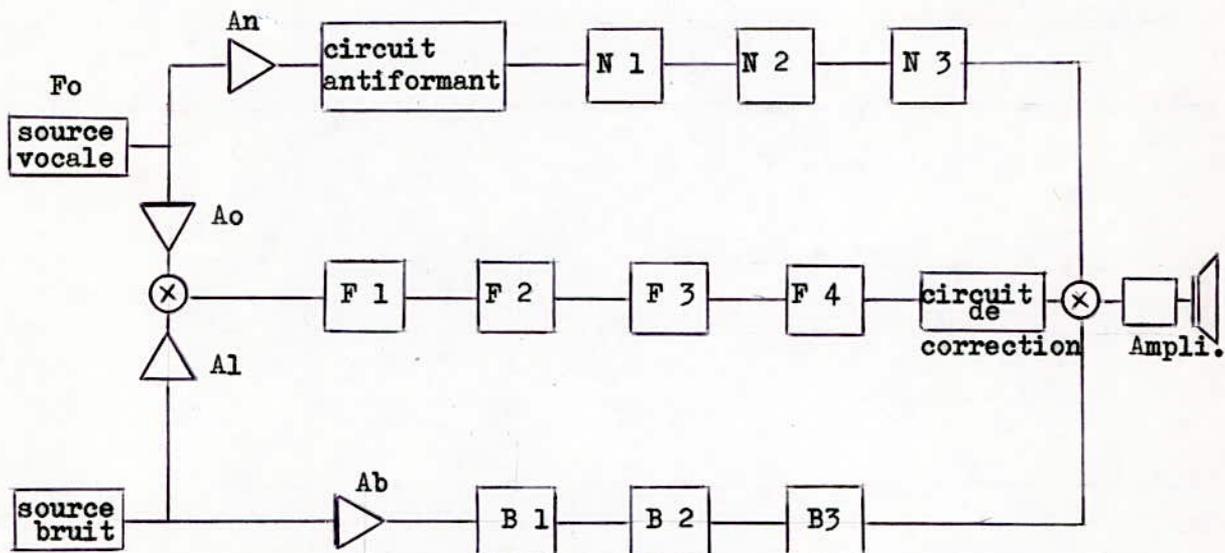
La fréquence du quatrième formant est généralement fixe. Le circuit de correction est relatif aux résonances d'ordre plus élevé.

- Canal de nasalité :

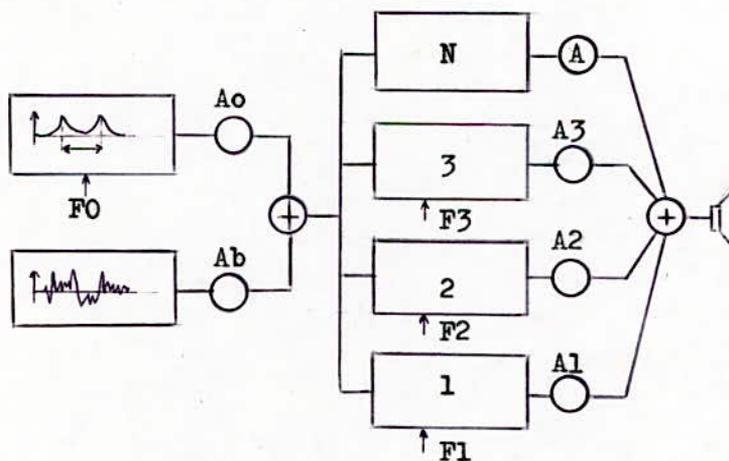
Il occupe la branche supérieur de la figure 44, il contient les circuits de formants de nasalité. La mise en fonction du conduit nasal fait apparaître des anti-résonances, qui se traduisent sur l'enveloppe du spectre par des minimum plus ou moins marqués.

- Canal de bruit :

Il est destiné à la production des sons bruités (fricatives); il comporte 3 circuits de formants de bruit B1, B2, B3. et est attaqué par un générateur de bruit.



**Fig. 4.4 : Synthétiseur à formants en série.**



**Fig. 4.3 : Synthétiseur à formants parallèle.**  
(d'après LIENARD)

#### IV - 3 : CONCLUSION :

Il a été question dans ce chapitre de l'étude des différentes méthodes et techniques de synthèse.

Parmi ces techniques, nous avons cité les vocodeurs à canaux et les vocodeurs à formants.

Les vocodeurs à formants permettent d'avoir une parole plus naturelle que celle produite par les vocodeurs à canaux, car ils simulent plus fidèlement le fonctionnement de l'appareil vocal.

## CHAPITRE V

### VOCODER A PREDICTION LINEAIRE

#### - INTRODUCTION

- V - 1 : CHOIX DE LA METHODE DE RESOLUTION.
- V - 2 : ELEMENTS CONSTITUTIFS D'UN VOCODER.
- V - 3 : L'ANALYSEUR DU VOCODER.
  - V-3-1 : LE PRETRAITEMENT.
  - V-3-2 : LE CALCUL DE LA FONCTION DE TRANSFERT.
  - V-3-3 : LA DETECTION DE VOISEMENT.
- V - 4 : LE CODAGE.
- V - 5 : LA TRANSMISSION.
- V - 6 : LE DECODAGE.
- V - 7 : LE SYNTHETISEUR DU VOCODER.
- V - 8 : CONCLUSION.

- INTRODUCTION :

L'étude des vocoders mettant en oeuvre la prédiction linéaire consiste, en premier lieu, à sélectionner la méthode qui est de réalisation simple et à définir les conditions d'application les plus favorables pour l'analyse-synthèse afin, de parvenir à une parole synthétique de qualité satisfaisante aux faibles débits.

Les vocoders à faible débit sont destinés principalement à la synthèse de la parole.

- V - 1 : CHOIX DE LA METHODE DE RESOLUTION :

La méthode de prédiction linéaire la plus apte à satisfaire les objectifs cités est la méthode d'autocorrélation. Ce choix est motivé par les raisons suivantes :

- La méthode d'autocorrélation garantit la stabilité du prédicteur linéaire, ce qui permet l'utilisation directe des paramètres calculés, en synthèse.

- La stationnarité du signal dans un laps de temps court (10 à 25 ms) n'est pas injustifiée.

- Les autres méthodes nécessitent un algorithme de détection de stabilité du filtre, et dans le cas d'instabilité d'un algorithme de correction ce qui alourdit considérablement le vocoder.

- La détermination des coefficients du filtre est plus rapide, lors de la résolution du système d'équations linéaires par la méthode d'autocorrélation.

- L'autocorrélation permet la détermination des coefficients de réflexion  $k(i)$  à partir des  $a(k)$  et ceci dans le cours normal des calculs. Les  $a(k)$  étant les coefficients d'un filtre numérique récursif représentant la fonction de transfert d'un conduit équivalent à l'appareil vocal.

Par contre les  $k(i)$  sont les paramètres d'un filtre en trellis moins sensible à la quantification des coefficients que les filtres classiques.

- V - 2 : ELEMENTS CONSTITUTIFS D'UN VOCODER :

Les composantes principales d'un système de compression de la parole sont représentées par la figure (5.1).

L'analyseur extrait du signal de parole  $s(t)$  un vecteur de paramètres  $x(t)$ . Ces paramètres sont ensuite quantifiés et codés ce qui produit le vecteur  $y(t)$  qui est transmis à l'aide d'un canal de transmission.

A la réception le vecteur  $y'(t)$  est décodé ce qui permet d'avoir l'estimation  $x'(t)$  des paramètres d'analyse. Ces derniers commandent le synthétiseur pour produire le signal de parole  $s'(t)$  qui est une approximation du signal réel  $s(t)$ .

Le synthétiseur est constitué de la source d'excitation et de la fonction de transfert de l'appareil vocal.

- V - 3 : L'ANALYSEUR DU VOCODER :

Il comporte les trois blocs de traitement suivants :

- Prétraitement : consiste à choisir l'intervalle d'analyse, la pondération par une fenêtre et à effectuer la préaccentuation.

- Calcul des paramètres de la fonction de transfert : ces paramètres sont le gain  $G$ , les coefficients  $a(k)$  ou les coefficients  $k(i)$  .

- Décision voisée / non voisée : consiste à déterminer si le signal est sourd ou sonore et à donner dans le dernier cas la période de mélodie.

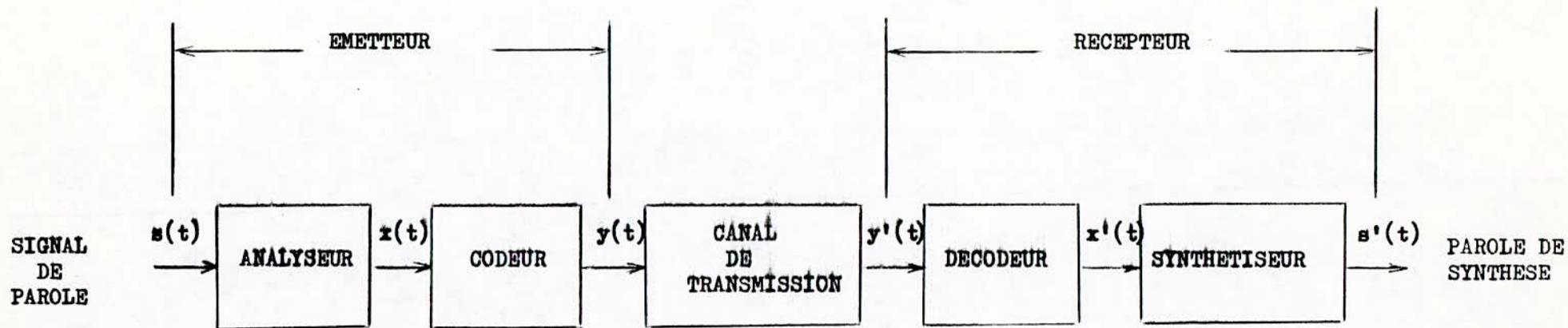


Fig. 5.1 : Eléments constitutifs d'un vocoder (d'après EL-MALAWANI)

## V - 3 - 1 : LE PRETRAITEMENT :

### 1 - Mode et intervalle d'analyse :

L'intervalle de temps d'analyse est basée sur l'estimation de la durée pendant laquelle les caractéristiques de l'articulation peuvent être considérées comme constantes, c'est-à-dire 10 à 25 ms.

Le mode d'analyse asynchrone en plus de sa simplicité, permet d'avoir une parole synthétique de bonne qualité.

La fenêtre d'analyse utilisée est celle de HANNING (Cf. III.4.1), car l'énergie est essentiellement concentrée dans le lobe principal.

### 2 - La préaccentuation :

Afin de gagner de la précision sur l'enveloppe spectrale dans les hautes fréquences, le signal de parole est envoyé avant l'analyse dans un filtre de préaccentuation pour compenser les influences de la source d'excitation (-12 dB/octave) et du rayonnement (+ 6 dB/octave).

Le filtre de préaccentuation est de la forme :

$$Q(Z) = 1 - mZ^{-1}$$

Pour les sons non voisés "m" est proche de 0, par contre, il est proche de 1 pour les sons voisés. (MARKEL, 1976).

La valeur de m sera transmise au synthétiseur pour réaliser la désaccentuation correspondante.

Il existe deux types de préaccentuation :

#### 2 - 1 : Préaccentuation conditionnelle :

Elle consiste à envisager plusieurs valeurs de m. Dans la pratique, deux valeurs de m sont généralement considérées suivant que les sons sont voisés ou non.

## 2 - 2 : Préaccentuation fixe :

Dans ce mode  $m$  a une seule valeur. L'application de cette préaccentuation n'introduit pas de dégradation notable de la qualité des fricatives de la parole synthétique (EL-MALAWANI, 1976).

## V - 3 - 2 : LE CALCUL DE LA FONCTION DE TRANSFERT :

Ceci consiste à déterminer les coefficients du filtre prédicteur et le facteur de gain.

Les coefficients de réflexion du filtre sont donnés par l'algorithme de DURBIN (Cf. III.1)

$$k(i) = (R(i+1) - \sum_{k=1}^{i-1} a(k).R(i+1-k)) / E(i-1)$$

## V - 3 - 3 : LA DETECTION DE VOISEMENT :

Dans le vocoder la détection du voisement et de la mélodie est effectuée en parallèle pour éviter toute influence de l'algorithme retenu sur le reste du système (Fig. 5.2).

### - Décision voisée/non voisée :

Il existe plusieurs méthodes pour la décision du voisement, mais elles souffrent toutes de limitations, car ni la période, ni la forme de l'onde d'excitation n'est constante.

1- Méthode utilisant l'erreur normalisée  $V(P)$  :

L'erreur normalisée  $V(P)$  est reliée au spectre de puissance approché du signal par la relation (III-5-1).

En général, les sons non voisés possèdent une erreur normalisée plus importante que les sons voisés car leur énergie est répartie sur tout le spectre à l'inverse des sons voisés où elle est concentrée dans les basses fréquences.

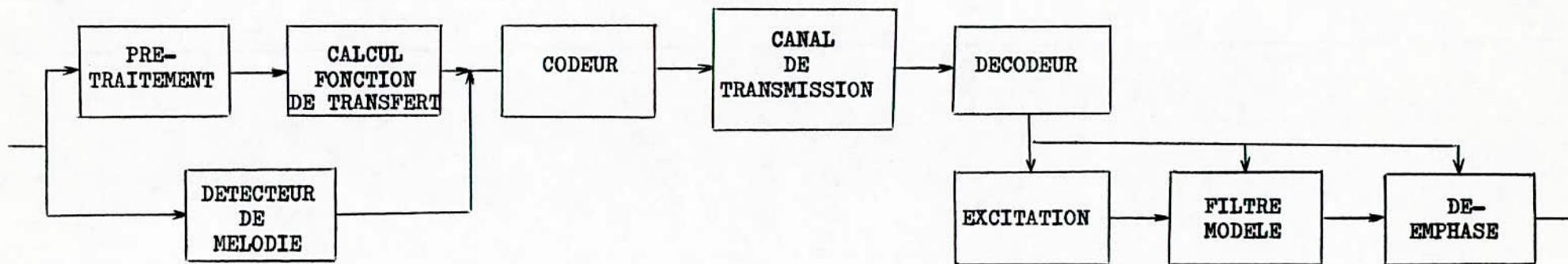


Fig. 5.2. : Synoptique d'un vocoder à prédiction linéaire.

A partir de cette remarque, l'erreur normalisée  $V(P)$  peut servir de détecteur de voisement à condition que le signal ne soit pas préaccentué. En effet, la préaccentuation tend à rendre le spectre des sons voisés plus plat ce qui augmente la valeur de  $V(P)$ .

## 2- Méthode utilisant le filtrage inverse :

La fonction d'autocorrélation du signal de parole, peut avoir des oscillations d'amplitudes élevées dues au premier formant. Ces oscillations peuvent être confondues avec la composante représentant la période de mélodie.

De ce fait, on élimine la réponse du conduit vocal avant le calcul de la fonction d'autocorrélation en appliquant au signal de parole un filtrage inverse. En procédant ainsi, on obtient le signal d'erreur  $e(n)$ , dans lequel les résonances du conduit vocal sont nettement affaiblies.

La fonction d'autocorrélation est alors calculée sur  $e(n)$  pour des retards allant de 0 à 25 ms (représentant les valeurs extrêmes que peut prendre la mélodie). Le signal d'erreur peut être considéré comme un bruit blanc gaussien de valeur moyenne nulle. De cette façon, la fonction d'autocorrélation  $R(n)$  tend vers 0 pour  $n$  différent de 0. (EL-MALAWANI, 1975).

Partant de là, on peut déterminer un seuil  $\alpha$ , tel que pour une certaine probabilité aucun échantillon de la fonction d'autocorrélation ne dépasse la valeur de ce seuil, excepté celui de l'origine. (Fig. 5.3).

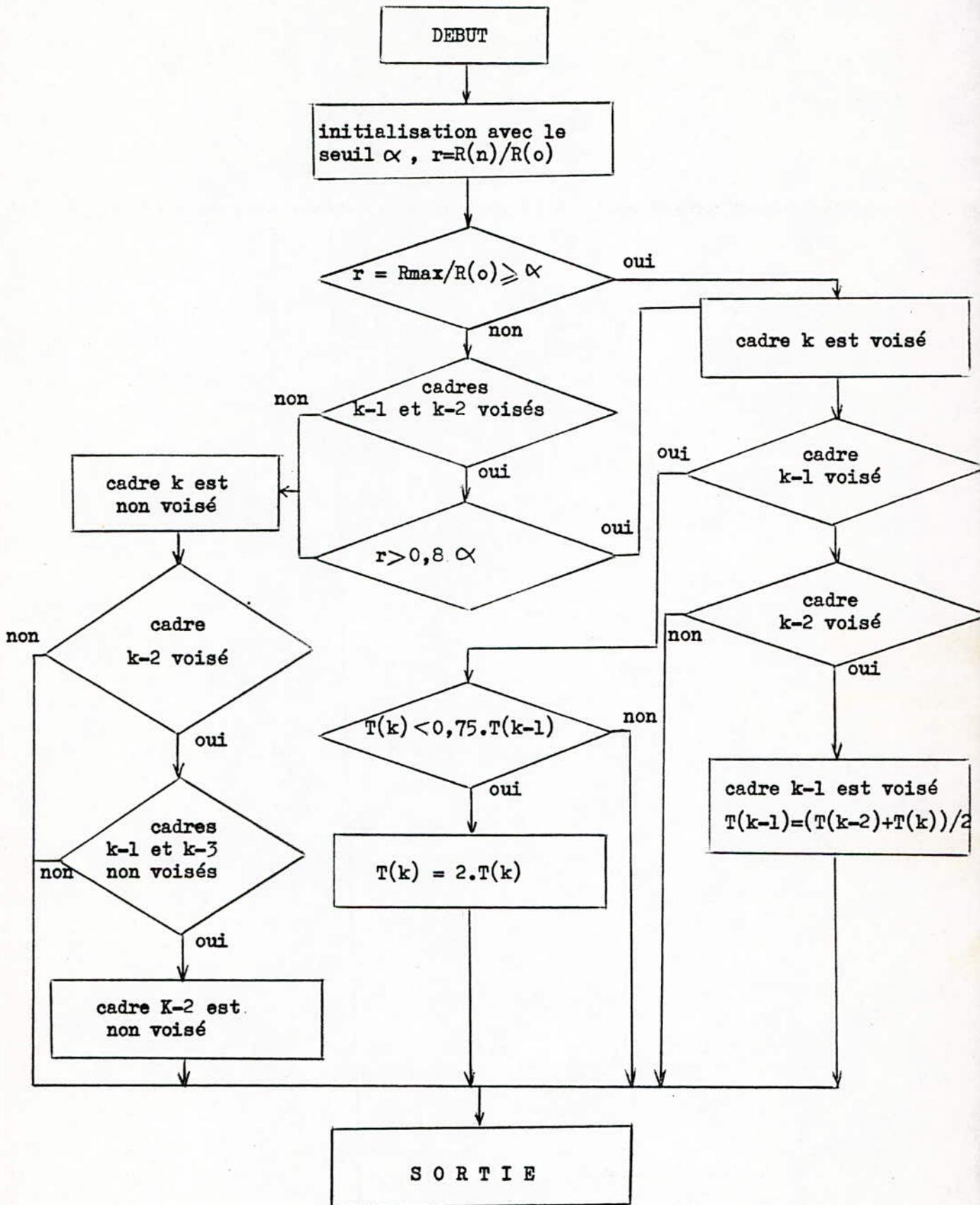


Fig. 5.4 : Organigramme de l'algorithme de décision voisé/non voisé

$R(n)$  représente le maximum de la fonction d'autocorrélation.  
 $r = R(n) / R(0)$

La période fondamentale est déterminée par l'intervalle de temps séparant le maximum de la fonction d'autocorrélation à l'origine du maximum le plus important de cette fonction.

L'organigramme de l'algorithme est donné par la figure 5.4.

- Explication de l'algorithme :

- Le seuil alpha est multiplié par un facteur de 0,8 si les deux cadres précédents sont voisés. Ceci est lié au fait que dans la fin d'un segment de période voisée, les périodes du signal sont faiblement corrélées ce qui réduit la valeur du maximum de la fonction d'autocorrélation.

- Si  $r \geq \alpha$ , la séquence  $k$  est voisée et  $T(k) = n \cdot T_e$   
 $T_e$  étant la période d'échantillonnage.  $n \in [0, N - 1]$ .  
Cette décision permet la correction de deux erreurs

possibles :

1) si  $k - 2$  et  $k$  sont voisés alors,  $k - 1$  est déclaré voisé avec une période qui est la moyenne de celle des cadres  $k - 2$  et  $k$ .

2) si la période de mélodie  $T(k)$  correspond par erreur à un formant, le seul moyen de s'en rendre compte est de comparer  $T(k)$  à  $T(k-1)$  dans la mesure où il est voisé. Compte tenu que la variation de la mélodie dépasse très rarement 20 % entre cadres consécutifs, le test  $T(k) < T(k-1) \cdot 0,75$  permet de corriger l'erreur.

#### V - 4 : LE CODAGE :

Le codage est effectué sur les paramètres à transmettre au synthétiseur, à savoir le gain, la mélodie, les paramètres de la fonction de transfert, l'indication de voisement est incluse dans l'information de mélodie.

##### 1 - Codage du facteur de gain G :

Afin que l'erreur relative de quantification ne dépasse pas un seuil constant quelle que soit la valeur du gain, il est préférable d'adopter un codage linéaire du logarithme du gain (EL-MALAWANI, 1975).

##### 2 - Codage de la mélodie :

La période de mélodie  $T(m)$  s'inscrit dans l'intervalle 2,5 - 20 ms correspondant à une fréquence  $f$  allant de 50 à 400 Hz.

L'erreur de quantification doit être minimisée sur la totalité de l'intervalle. Une loi de codage logarithmique satisfait à ce critère.

Le paramètre est mis à 0 en cas de non voisement.

##### 3 - Codage des paramètres de la fonction de transfert :

Les paramètres de la fonction de transfert doivent assurer la stabilité du filtre après leur quantification.

Les logarithmes des rapports d'aires définis par :

$$\log B(i) = \log (A(i) / A(i+1))$$

$$B(i) = \frac{1+k(i)}{1-k(i)}$$

représentent un ensemble de coefficients optimal pour la quantification (MARKEL, 1976).

## V - 5 : LA TRANSMISSION :

Les paramètres transmis sont :

- Le facteur de gain codé avec N bits.
- La mélodie et le voisement codés avec Q bits.
- Les coefficients du filtre codés avec M bits.
- La préaccentuation codée avec 1 ou 0 bits, suivant qu'elle soit fixe ou non.

Les valeurs de N, Q, M déterminent le débit du vocoder.

## V - 6 : LE DECODAGE :

Il consiste à effectuer la transformation inverse du codage des paramètres transmis.

## V - 7 : LE SYNTHETISEUR :

Il est constitué de trois éléments à savoir, la source d'excitation, le filtre modèle et la desaccentuation.

Afin de minimiser le bruit dû aux transitions brusques des paramètres entre deux séquences d'analyse consécutives, on effectue un lissage des paramètres.

### 1 - L'excitation :

Suivant que la séquence est voisée ou non, un commutateur sélectionnera le type d'excitation correspondant. En cas de voisement, le commutateur branchera un générateur d'impulsions de période  $T_m$ . Dans le cas de non voisement, un générateur de bruit blanc est branché.

### 2 - Le filtre modèle :

La parole synthétique provient de l'excitation d'un filtre récursif de fonction de transfert  $1 / A(Z)$ , soit par des impulsions, soit par du bruit blanc.

Plusieurs structures de filtres de synthèse peuvent être utilisées. La plus simple à réaliser étant celle de la figure 5.5 qui réalise la fonction suivante :

$$S(n) = e(n) + \sum_{k=1}^p a(k).S(n-k)$$

$e(n)$  : échantillons d'entrée.

$S(n)$  : échantillons de sortie.

$a(k)$  : coefficients du filtre.

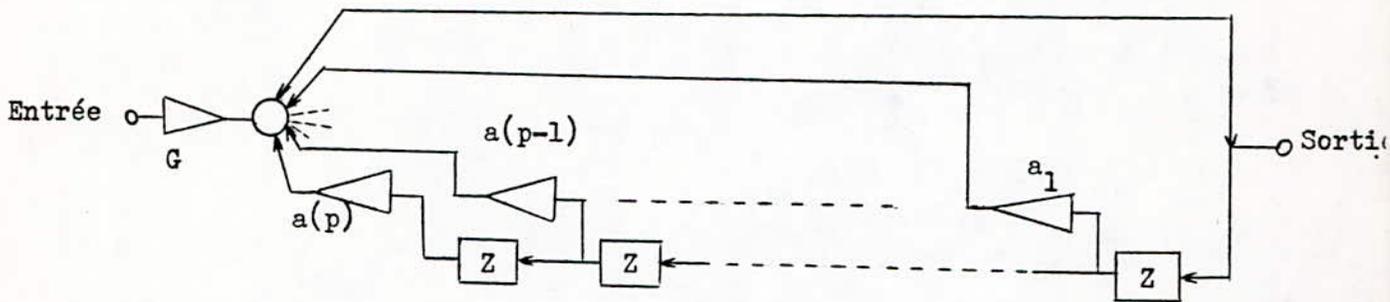


Fig. 5.5 : Filtre récursif classique.

D'autres structures peuvent être utilisées, notamment les structures en treillis.

- Filtres en treillis à 2 multiplieurs :

Les relations (II.9.1) et (II.9.2) permettent de réaliser le filtre  $1/A(Z)$  sous forme de filtre en treillis, modèle du conduit vocal non pas à partir des coefficients de prédiction mais à partir des coefficients de réflexion.

La structure d'un filtre en treillis est de la forme (Fig. 5.6.) :

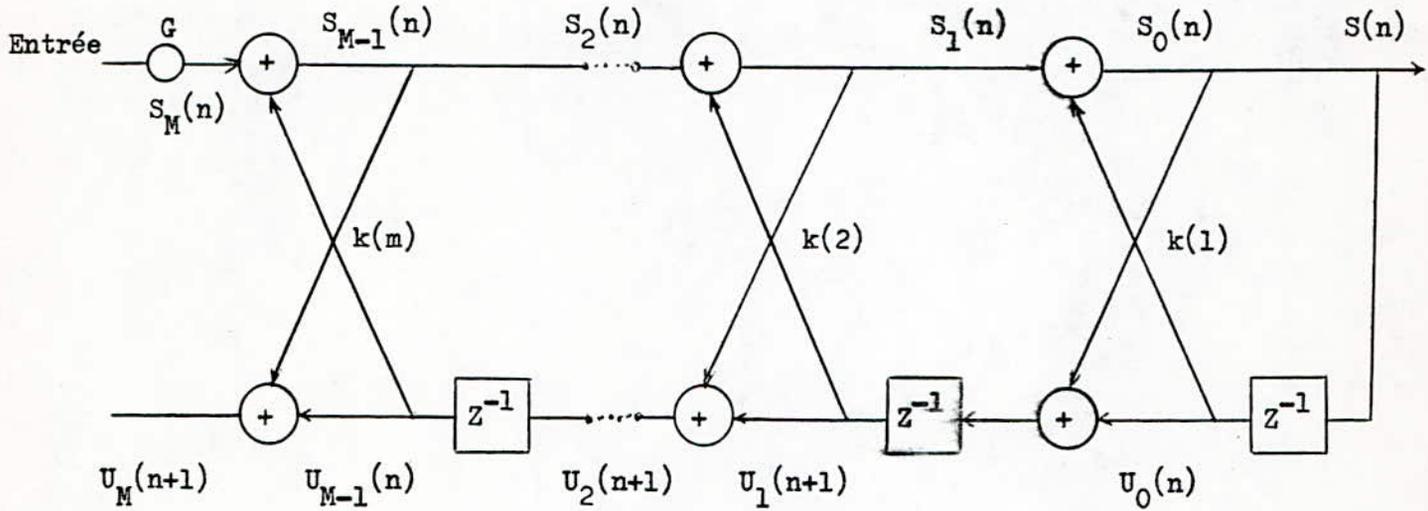


Fig. 5.6. : Filtre en treillis à 2 multiplieurs.

les suites  $S_1(n)$  et  $S_1(n+1)$  sont reliées à la suite  $S(n)$  par les relations :

$$S(n) = S_1(n) - k(1) \cdot S(n-1) \dots \dots \dots (V.7.1)$$

$$U_1(n+1) = k(1) \cdot S(n) + S(n-1) \dots \dots \dots (V.7.2)$$

de même les suites  $S_2(n)$  et  $U_2(n+1)$ ,  $S_1(n)$  et  $U_1(n+1)$

vérifient :

$$S_1(n) = S_2(n) - k(2) \cdot U_1(n)$$

$$U_2(n+1) = k(2) \cdot S_1(n) + U_1(n)$$

$$S_2(n) = S_1(n) + k(2) \cdot U_1(n)$$

$$\text{or } \begin{cases} S_1(n) = S(n) + k(1).S(n-1) \\ U_1(n) = k(1).S(n-1) + S(n-2) \end{cases}$$

de là :

$$S_2(n) = S(n) + k(1).S(n-1) + k(2).(k(1).S(n-1) + S(n-2))$$

$$S_2(n) = S(n) + k(1).(1+k(2)).S(n-1) + k(2).S(n-2)$$

en passant à la transformée en Z, on obtient :

$$S_2(Z) = S(Z) + k(1).(1+k(2)).Z^{-1}.S(Z) + k(2).Z^{-2}.S(Z)$$

Il en résulte entre l'entrée  $S_2(n)$  et la sortie  $S(n)$  la fonction de transfert  $H_2(Z)$  telle que :

$$H_2(Z) = S(Z)/S_2(Z) = \frac{1}{1+k(1).(1+k(2)).Z^{-1}+k(2).Z^{-2}} \dots(V.7.3)$$

de même entre  $U_2(n+1)$  et  $S(n)$

$$U_2(n+1) = k(2).(S(n) + k(1).S(n-1)) + k(1).S(n-1) + S(n-2)$$

il apparait donc la fonction de transfert  $U_2(Z)$  telle que :

$$U_2(Z) = k(2) + k(1).(1+k(2)).Z^{-1} + Z^{-2} \dots(V.7.4)$$

par itération, il apparait que les suites  $S_M(n)$  et  $S(n)$  d'une part,  $U_M(n+1)$  et  $S(n)$  d'autre part sont liées par les relations :

$$S(n) = S_M(n) - \sum_{i=1}^M a(i).S(n-i)$$

$$U_M(n+1) = \sum_{i=0}^{M-1} a(M-i).S(n-i) + S(n-M)$$

Il en résulte les fonctions de transfert  $H_M(Z)$  et  $U_M(Z)$  telles que :

$$H_M(Z) = \frac{1}{1 + \sum_{i=1}^M a(i) \cdot Z^{-i}} = \frac{1}{D_M(Z)} \dots\dots (V.7.5)$$

$$U_M(Z) = \sum_{i=0}^{M-1} a(M-i) \cdot Z^{-i} + Z^{-M} \dots\dots\dots (V.7.6)$$

En désignant par  $H_m(Z)$  et  $U_m(Z)$  les fonctions de transfert correspondant aux sorties de la cellule de rang  $m$ , on peut écrire les relations matricielles suivantes :

$$\begin{bmatrix} D_m(Z) \\ U_m(Z) \end{bmatrix} = \begin{bmatrix} 1 & k(m) \cdot Z^{-1} \\ k(m) & Z^{-1} \end{bmatrix} \cdot \begin{bmatrix} D_{m-1}(Z) \\ U_{m-1}(Z) \end{bmatrix}$$

La stabilité du filtre en treillis est facile à réaliser, il suffit que :

$$|k(i)| < 1 \quad 1 \leq i \leq M$$

3 - La désaccentuation :

C'est l'opération inverse de la préaccentuation. Le filtre de désaccentuation a une fonction de transfert du type :

$$F(Z) = 1 / Q(Z) = 1 / 1 - m \cdot Z^{-1}$$

où  $m$  est le zéro de la préaccentuation.

## V - 8 : CONCLUSION :

Dans ce chapitre ont été étudiés les différents aspects d'un vocoder à prédiction linéaire faisant appel à la méthode d'auto-corrélation.

Cette méthode se prête bien à la réalisation d'un système d'analyse-synthèse fonctionnant en temps réel et permet, en même temps une bonne représentation du signal de parole à de faibles débits.

Les conditions les plus favorables à l'analyse, au codage et à la synthèse de la parole ont été définies.

L'utilisation des coefficients  $a(k)$  du filtre modèle conduit à un synthétiseur d'une extrême simplicité et permet d'obtenir une parole de synthèse de bonne qualité à des débits moyens (6400 bits/s). (EL-MALAWANI, 1975).

La nécessité d'atteindre des débits plus faibles sans risques d'instabilité du filtre conduit à l'utilisation des coefficients  $k(i)$  dont les logarithmes des rapports d'aires représentent des paramètres optimaux pour la quantification.

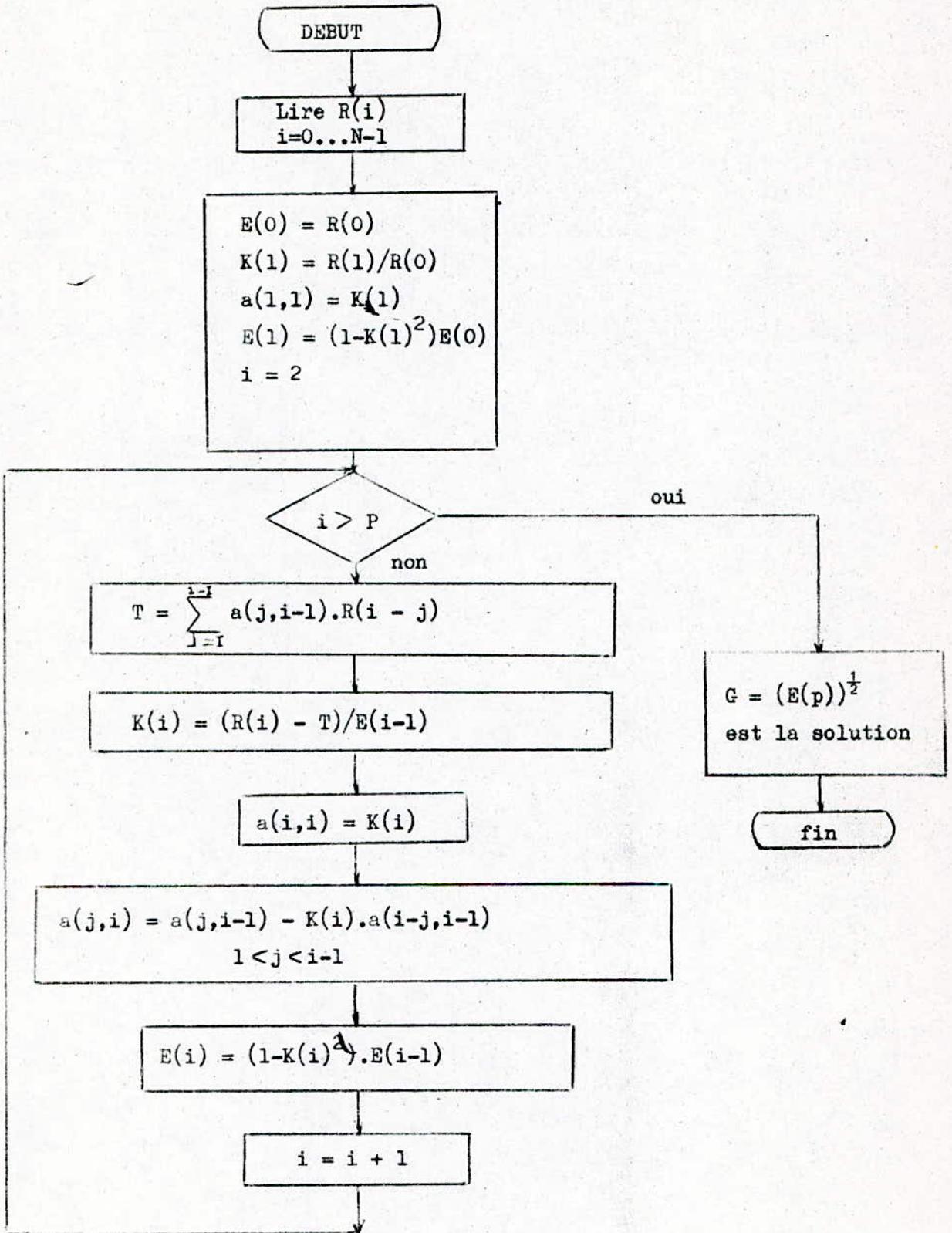
PROGRAMMATION

---

Nous présentons dans cette partie deux programmes.

Le premier concerne la détermination des coefficients d'autocorrélation et le calcul du gain. Le deuxième traite de la décision de voisement et de la détermination de la mélodie.

La programmation a été faite sur le micro OLIVETTI M24.



```

1 CLS
5 REM *****
10 REM * CALCUL DES COEFFICIENTS D'AUTOCORRELATION *
11 REM *****
20 INPUT "DONNER LA VALEUR DE N";N
30 INPUT "DONNER LA VALEUR DE P:";P
40 DIM S(N+P)
45 FOR I=0 TO N-1
48 PRINT "S(";I;")=":INPUT S(I)
50 NEXT I
60 DIM R(P)
65 FOR K=0 TO N-1
68 R(K)=0
70 FOR I=0 TO N-1
75 R(K)=R(K)+S(I)*S(I+K)
80 NEXT I
85 NEXT K
100 REM" METHODE AUTOCOR"
120 DIM A(P,P),K(P),E(P)
170 E(0)=R(0)
180 K(1)=R(1)/R(0)
190 A(1,1)=K(1)
200 E(1)=(1-(K(1)*K(1)))*E(0)
210 FOR I=2 TO P
220 T=0
230 FOR J=1 TO I-1
240 T=T+A(J,I-1)*R(I-J)
250 NEXT J
260 K(I)=(R(I)-T)/E(I-1)
270 A(I,I)=K(2)
280 FOR J=1 TO I-1
290 A(J,I)=A(J,I-1)-K(I)*A(I-J,I-1)
300 NEXT J
310 E(I)=(1-(K(I)*K(I)))*E(I-1)
320 NEXT I
330 FOR I=1 TO P-1
340 FOR J=1 TO P-1
350 PRINT "E(";I;")=";E(I),"K(";I;")=";K(I)
360 PRINT"A(";J;",";I;")=";A(J,I)
370 NEXT J
380 NEXT I
385 REM *****
390 REM * CALCUL DU GAIN *
400 REM *****
410 G=SQR(E(I))
420 PRINT "GAIN=";G
430 END

```

```

90 CLS
100 REM *****DECISION VOISEE-NON VOISEE*****
110 REM *****MESURE DE LA FREQUENCE*****
115 REM *****FONDAMENTALE*****
120 PRINT "DONNER LA VALEUR DU SEUIL ALPHA";
130 INPUT ALPHA
140 PRINT "DONNER LE NOMBRE DE SEQUENCES K";
150 INPUT M
160 PRINT "DONNER LA PERIODE D'ECHANTILLONAGE T";
170 INPUT T
180 PRINT "ENTRER LA VALEUR DU PREDICTEUR P";
190 INPUT P
200 DIM R(P)
204 K=1
210 FOR I=0 TO P
220 PRINT "DONNER LES VALEURS DE R(";I;")=";
230 INPUT R(I)
240 NEXT I
250 GOSUB 590
260 R=(RMAX/R(0))
270 IF R>=ALPHA THEN 420
280 IF K<=2 THEN 720
290 REM IF V(K)=1 SEQUENCE VOISEE
300 REM IF V(K)=0 SEQUENCE NON VOISEE
310 IF V(K-1)=1 AND V(K-2)=1 THEN 330
320 GOTO 340
330 IF R>=.8*ALPHA THEN 420
340 PRINT "LA SEQUENCE, K, EST NON VOISEE"
350 IF V(K-2)=1 GOTO 380
360 GOTO 680
370 IF K=3 GOTO 690
380 IF V(K-1)=0 AND V(K-3)=0 GOTO 400
390 GOTO 680
400 PRINT "LA SEQUENCE K-2 EST NON VOISEE"
410 GOTO 680
420 PRINT " LA SEQUENCE K EST VOISEE"
430 TO(K)=J*T
440 PRINT "LE FONDAMENTAL EST TO(K) ="; "TO(";K;")="; TO(K)
450 V(K)=1
460 IF K<=2 GOTO 690
470 IF V(K-1)=1 GOTO 500
480 IF V(K-2)=1 GOTO 540
490 GOTO 680
500 IF TO(K)<.75*TO(K-1) GOTO 540
510 GOTO 680
520 TO(K)=2*TO(K)
530 GOTO 680
540 PRINT "LA SEQUENCE K-1 EST VOISEE"
550 TO(K-1)=(TO(K-2)+TO(K))/2
560 PRINT "TO(K-1)"; TO(K-1)
570 GOTO 680
580 REM ***RECHERCHE DU PREMIER MAXIMUM A PARTIR DE R(0)***
590 J=1
600 RMAX=R(J)
610 FOR I=J+1 TO P
620 IF R(J)<R(I) GOTO 650
630 NEXT I
640 RETURN
650 RMAX=R(I)
660 J=I
670 GOTO 610
680 IF K=M THEN 710
690 K=K+1
700 GOTO 210
710 END
720 PRINT "LA SEQUENCE K EST NON VOISEE"
730 190

```

## CONCLUSION

Dans ce projet, nous avons fait l'étude d'un vocoder à prédiction linéaire.

Dans ce cadre, nous avons été amenés à explorer les deux grandes étapes de base qui sont, l'analyse et la synthèse de la parole.

L'étape d'analyse avait pour but de représenter le signal de parole à l'aide d'un nombre pertinent de paramètres, elle consistait à rechercher un modèle qui reconstitue une sortie la plus proche possible du signal à analyser.

Parmi les différentes méthodes d'analyse répondant à ce critère, la prédiction linéaire est apparue comme la plus appropriée et la plus performante du fait de la simplicité de son modèle et des bons résultats qui en découlent.

L'étape de synthèse, consiste à déterminer le modèle récursif à partir des paramètres issus de l'analyse.

Ces deux étapes nous ont permis d'aborder ensuite l'étude des différents aspects d'un vocoder à base de prédiction linéaire.

Dans ce but, nous avons opté pour la méthode d'autocorrélation qui s'est révélée la plus pratique dans ce cadre d'application, car elle permettait d'utiliser directement, les paramètres calculés, à la synthèse.

Par ailleurs, nous avons donné les conditions les plus favorables pour le calcul, le codage des différents paramètres à savoir le gain, le fondamental et les coefficients de prédiction en vue de leur transmission au synthétiseur.

Le synthétiseur du vocoder étant choisi de telle manière à simuler le mieux possible la parole humaine.

L'utilisation de vocoder ou de machine parlante offre des intérêts considérables dans divers domaines, notamment dans les applications médicales où la synthèse de la parole apporte une aide précieuse aux handicapés facilitant ainsi leur intégration dans la société.

D'autres études sur le vocoder peuvent être envisagées, visant, notamment l'amélioration de la qualité de la parole comme le vocoder à excitation vocale.

1- BELLANGER M. (1981).

"Traitement numérique du signal". MASSON.

2- CATIER E. (1984).

"La parole : analyse et synthèse". Revue Toute l'Electronique n° 490.

3- EL-MALAWANI I. (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  
mélodie. Extraction de la fonction d'aire du conduit vocal!"  
Thèse de Docteur - Ingénieur. Université de Grenoble.

4- EMERARD F. (1977).

"Synthèse par diphone et traitement de la prosodie".  
Thèse de Doctorat 3ème cycle. Université de Grenoble.

5- FONDANECHÉ P. GILBERTAS P. (1981).

"Filtre numériques, principes et réalisations". MASSON.

6- GUERTI M. (1983).

"Contribution à la synthèse de la parole en arabe standard".  
Thèse de magister. Université d'Alger.

7- GUIBERT J. (1979).

"La parole, compréhension et synthèse par les ordinateurs".

8- HAVET P. (1981).

"Contribution à l'étude de la source vocale : détermination  
des paramètres de commande d'un modèle de source vocale et  
étude de l'onde glottique par filtrage inverse".  
Thèse de Doctorat de 3ème cycle. Université de Rennes.

9- KUNT M. (1981).

"Traitement numérique des signaux". DUNOD.

- 10- LE GUYADER A. (1978).  
"Etude d'un vocoder à excitation vocale et à base de prédiction linéaire".  
Thèse de Doctorat de 3ème cycle. Université de Rennes.
- 11- LIENARD J.S. (1977).  
"Les processus de la communication parlée. Introduction à l'analyse et à la synthèse de la parole". MASSON.
- 12- LIFERMANN J. (1980).  
"Les méthodes rapides de transformation du signal".
- 13- MARKEL J.D. - GRAY A.H. (1976).  
"Linear prediction of speech". SPRINGER - VERLAG.
- 14- MAX J. (1981).  
"Méthodes et techniques de traitement du signal et application aux mesures physiques". MASSON.
- 15- RABINER L.R., CHAFFER R.W. (1978).  
"Digital processing of speech signals". PRINTICE HALLS.
- 16- 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.
- 17- STELLA M., ZURCHER F. (1983).  
"Codage et analyse-synthèse de la parole".  
"Document remis aux élèves de l'E.N.S.T" C.N.E.T. - LANNION.