

ECOLE NATIONALE POLYTECHNIQUE



DEPARTEMENT D'ELECTRONIQUE

## PROJET DE FIN D'ETUDES

S U J E T

# ETUDE & REALISATION D'UN SYNTHETISEUR

Proposé par :

B. BOUSSEKSOU

Etudié par :

N. HASSAINE

M. OTMANI

Dirigé par :

B. BOUSSEKSOU

PROMOTION : JANVIER 1986

# ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT



## PROJET DE FIN D'ETUDES

S U J E T

# ETUDE & REALISATION D'UN SYNTHETISEUR

Proposé par :

**B. BOUSSEKSOU**

Etudié par :

**N. HASSAINE**

**M. OTMANI**

Dirigé par :

**B. BOUSSEKSOU**

PROMOTION : JANVIER 1986

DEDICACES

-----

*Je dédie ce modeste travail :*

- \* *A la mémoire de ma grand-mère Hadja Atika.*
- \* *A mon père et ma mère : en signe de reconnaissance pour tous les sacrifices consentis à mon égard, pour leur soutien moral , matériel et leurs encouragements tout le long de mes études.*
- \* *A mes frères et soeurs.*
- \* *A toute ma famille.*
- \* *A tous mes amis.*
- \* *Aux familles : LARBES, DJEBLOUN, OTMANI, ESCID et HAMZA.*

Monsieur HASSAINE Nacer



D E D I C A C E S

-----

*Je dédie ce modeste travail :*

- \* *A la mémoire de mon père.*
- \* *A ma mère : en signe de reconnaissance pour tous les sacrifices consentis à mon égard.*
- \* *Ames oncles : Saïd, Ahmed et Meziane, pour leur soutien moral, matériel et leurs encouragements tout le long de mes études.*
- \* *A chaque membre de ma famille.*
- \* *A tous mes amis.*
- \* *Aux familles : NAIT-CHERIF, SIHADI et HASSAINE, (particulièrement Hadj Mokhtar).*

*Monsieur OTMANI Mohamed*



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

=====

*Nous tenons à exprimer nos plus vifs remerciements à notre promoteur, Monsieur Boualem BOUSSEKSOU, pour son aide dévouée, efficace, et ses précieux conseils tout au long de l'élaboration de ce projet de fin d'études.*

*Que Monsieur A. BOULARAS, trouve ici le témoignage de notre reconnaissance pour l'aide qu'il nous a apportée en programmation.*

*Nous adressons aussi nos remerciements à Monsieur A. MENACER (Docteur-Ingénieur), pour son aide et ses précieux conseils.*

*Nous remercions également Monsieur N. BELBERKANI et Madame N. TIGZIRI, qui nous ont aidé à l'élaboration du 3<sup>eme</sup> programme.*

*La frappe de ce polycopié à été soigneusement faite par Mademoiselle N. ABDENNOURI. Qu'elle trouve ici l'expression de notre réelle et sincère reconnaissance.*

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

S O M M A I R E



PREMIERE PARTIE

Pages

* Introduction .....	1
* <u>Chapitre I</u> : l'appareil phonatoire humain .....	2
1.1) Introduction .....	2
1.2) La production de la parole : Aspects articulatoires..	2
1.3) Fonctionnement général - Aspect acoustique .....	3
1.4) Caractéristiques de la parole .....	3
* <u>Chapitre II</u> : Modèle de production de la parole .....	7
1) Introduction .....	7
2) Principe d'analyse - synthèse .....	7
a) Analyse de la parole	
b) La synthèse	
3) Détermination du modèle .....	9
a) Introduction	
b) Détermination du modèle	
4) Principales méthodes de prédiction linéaire .....	12
a) Exacte	
b) Covariance	
c) Autocorrelation	
5) Résolution des équations de prédiction linéaire par la méthode d'autocorrelation .....	13
a) Etablissement des équations .....	13
b) Erreur quadratique totale minimale de prédiction..	15
c) Calcul du facteur de gain A .....	16
d) Résolution du système d'équation selon la méthode de Durbin .....	18
- Algorithme de Durbin .....	19
* <u>Chapitre III</u> :	21
1) Introduction .....	21
2) Influence de l'excitation en prédiction linéaire sur la qualité de la parole synthétisée .....	21

a) Différents formes d'excitation	
- Bi-impulsionnelle	
- Triangulaire	
- Bi-triangulaire	
- Excitation visant une approche de l'onde glottique (polynomiale)	
- Excitation mixte	
3) Générateur de bruit .....	27
a) Principe	
b) Réalisation	
4) Modélisation du conduit vocal par un filtre en treillis.	31
5) Détection du voisement et mesure de la période du fondamental dans le cas d'une séquence sonore .....	34
* <u>Chapitre IV</u> : Les synthétiseurs à formants .....	38
1) Introduction .....	38
2) Définition des formants .....	38
3) Les résonateurs .....	39
a) Définition	
b) Propriétés dans résonateurs acoustiques	
c) Application en filtre sonore	
4) Les synthétiseurs à formants .....	41
a) Structure serie	
b) Structure parallèle	
5) Comparaison entre les deux structures .....	45
6) Détermination de trois premiers <sup>formants</sup> et la fréquence fondamentale .....	46
 <u>DEUXIEME PARTIE</u>	
Programmation .....	47
Conclusion.	



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

Dés la fin du XVIII<sup>eme</sup> siècle, avec le Baron VON KEMPELEN et sa machine parlante, l'homme a commencé à étudier de manière scientifique la production de la parole. Avant d'aborder la rédaction d'un mémoire de thèse, il nous semble important d'évoquer les raisons pour lesquelles l'homme s'intéresse depuis aussi longtemps à ce sujet.

L'homme est curieux : il essaie de faire progresser la connaissance pour elle-même, et pour le plaisir de savoir. Par ailleurs il éprouve le besoin de dominer son environnement : démontrer et comprendre les mécanismes de la nature lui font espérer le moyen de s'en rendre maître et de les détourner de leurs fonctions primitives. Voilà, nous pensons, les deux raisons fondamentales qui poussent l'homme vers cette étude approfondie de la nature.

La recherche sur la communication parlée n'échappe pas à ce double principe : l'étude du signal de parole, l'étude des processus de production et de perception de celui-ci servent à faire les connaissances de l'homme sur lui-même, et lui permettent de contrôler et de simuler les mécanismes de communication. En témoignent de nombreuses applications telles que le dialogue homme-machine, les transmissions à bas débit, ou l'aide aux handicapés, pour n'en citer que quelques-unes.

Notre travail est composé de deux parties.

La première est consacrée à l'étude théorique du sujet et la deuxième à l'élaboration des programmes.

CHAPITRE I : Description de l'appareil phonatoire humain et quelques caractéristiques de la parole.

CHAPITRE II : Il est consacré à la détermination de notre modèle de production de la parole par la méthode de prédiction linéaire.

Dans le CHAPITRE III nous présentons :

- Les différentes formes d'excitations et leur influence sur la qualité de la parole (pour les sons voisés).
- Générateur de bruit (séquence pseudo-aléatoire).
- Modélisation du conduit vocal par les filtres en treillis.
- Décision du voisement et la mesure du fondamental.

Le CHAPITRE IV : est consacré à l'étude de la synthèse formantique



## Chapitre I : L' APPAREIL PHONATOIRE HUMAIN



### I.1. Introduction :

Le rôle de ce chapitre est d'exposer les éléments essentiels sur lesquels s'appuie la suite de ce mémoire, on va donner une description de l'appareil phonatoire humain, et voir son fonctionnement général, et à la fin on va donner quelques caractéristiques de la parole.

### I.2. La production de la parole : Aspects articulatoires

Seuls sont rappelés ici les points essentiels : Aspect anatomo-physiologique du système vocal humain et fonctionnement général sous l'angle acoustique.

#### I.2.1. Aspect anatomo-physiologique :

La figure (1.1) schématise l'ensemble du système vocal humain, qui peut être décomposé en trois parties :

##### I.2.1.1. Les poumons et le conduit trachéo-bronchique :

La trachée-artère est un conduit cylindrique qui réalise la liaison entre le larynx et les bronches qui se ramifient à l'intérieur des poumons. Au cours de la phonation, la diminution du volume thoracique tend à chasser l'air des poumons ; le larynx est alors alimenté par de l'air à une pression atmosphérique, la pression subglottique, qui met en vibration des cordes vocales.

##### I.2.1.2. Le larynx - les cordes vocales :

Le larynx est une boîte formée de cartilages articulés, ligaments, muscles et muqueuses qui surmontent la trachée-artère. A la base du larynx sont attachées les cordes vocales, sortes de lèvres placées symétriquement à gauche et à droite de la ligne médiane du larynx (voir figure 1.2). Lors de l'émission de sons de parole voisés, les cordes vocales vibrent principalement dans un plan horizontal, obéissant à un phénomène d'oscillation de relaxation.

##### I.2.1.3. Le conduit vocal :

Le conduit vocal se compose de deux cavités :

###### - Le conduit nasal :

Est formé des forces nasales qui sont deux cavités de forme fixe dont la communication avec la cavité orale est commandée par le voile du palais, prolongement mobile du palais dur.

###### - Le conduit buccal :

Est formé du pharynx, cavité qui se situe au dessus du larynx, et de la cavité buccale. Le conduit oral possède un volume

et une géométrie extrêmement variable grâce à la grande mobilité de la langue essentiellement, et du maxillaire inférieur.

### I.3. Fonctionnement général - Aspect acoustique :

Le comportement acoustique du système vocal humain peut être schématisé de la manière suivante :

Une "excitation" délivre un signal de source qui est modifié par la fonction de transfert du "résonateur". On compte deux modes fondamentaux du fonctionnement du système vocal humain, selon la nature du signal excitateur.

#### I.3.1. Excitation pseudo-périodique :

Ce premier type d'excitation est produit par la vibration des cordes vocales sous l'action de la pression de l'air en provenance des poumons. L'onde produite par cette vibration a une forme approximativement triangulaire que l'on peut assimiler à un train d'impulsions, appelés fréquence fondamentale ou fréquence de mélodie (pitch en anglais) varie en moyenne :

- de 100 à 150 HZ pour les hommes.
- de 200 à 250 HZ pour les femmes.

Les sons produits par ce type d'excitation sont dits voisés, ce sont par exemple les voyelles.

#### I.3.2. Excitation du conduit vocal par un bruit continu :

Le deuxième type d'excitation est provoqué par la génération d'un écoulement d'air turbulent dans le canal vocal, l'entrée de ce filtre peut être assimilée à un bruit acoustique approximativement blanc.

Les sons ainsi produits sont appelés non-voisés comme certaines consonnes "s" ; "ch" ; et "f".

Remarque : La production de certains sons nécessite l'intervention simultanée des deux phénomènes ; il en est ainsi par exemple de "z" ; "j" ; et "v".

### I.4. Caractéristiques de la parole :

Les cordes vocales constituent un générateur de son en hauteur. Les autres organes influent sur les variations timbre et les transitoires constituent l'articulation.

Les principales caractéristiques de la parole sont :

#### - Le niveau :

Compte tenu des instants de silence, l'énergie moyenne développée au cours d'une conversation est d'environ 10 pw, mais selon les cas elle peut varier de 1 à 500 pw.

- La hauteur :

La fréquence la plus basse dans le signal de la parole est la fréquence fondamentale des vibrations des cordes vocales (pitch) elle peut varier de 90 HZ pour une voix basse à 300 HZ pour une voix très aigüe et à 500 HZ pour les voix d'enfants.

- Le timbre :

C'est la caractéristique de deux sons de même fréquence. Il peut être considéré comme un paramètre du son. Il se caractérise par l'amplitude harmonique. Chaque harmonique a son amplitude propre. C'est pourquoi il est important que chaque harmonique soit respectée pour retrouver toute l'intégrité et la richesse du timbre.

- Bande passante :

Quand les fréquences entre 100 HZ et 10 KHZ sont correctement rendus, on peut admettre qu'une transmission de parole est fidèle.

- La dynamique :

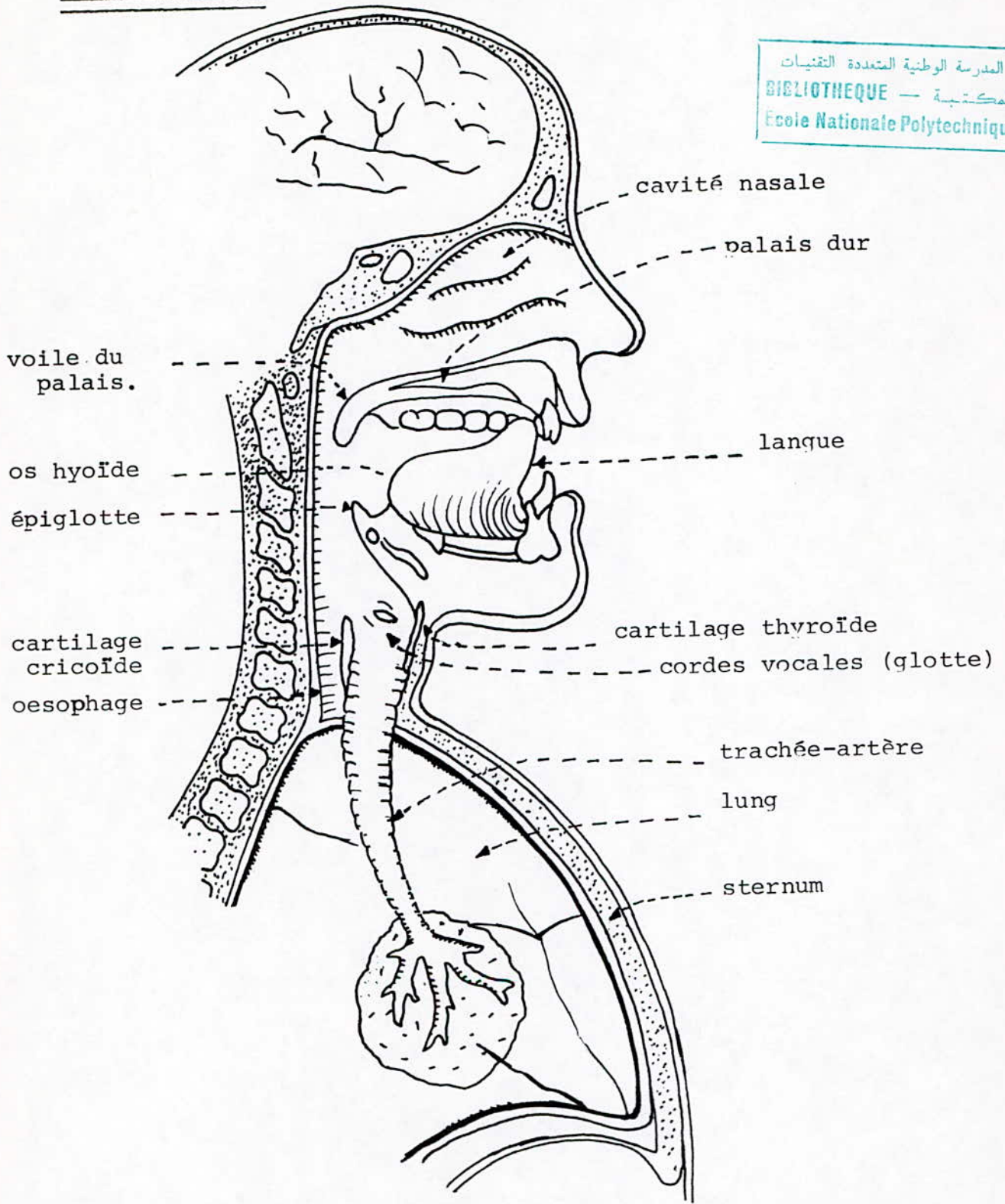
La dynamique d'un signal peut être définie comme étant l'écart en "dB" entre le niveau maximum et le niveau minimum de ce signal.

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

Globalement l'énergie de la parole varie au plus entre  $1/10^{\text{ème}}$  et 1000 pw, soit une dynamique de valeur extrême 40 dB.

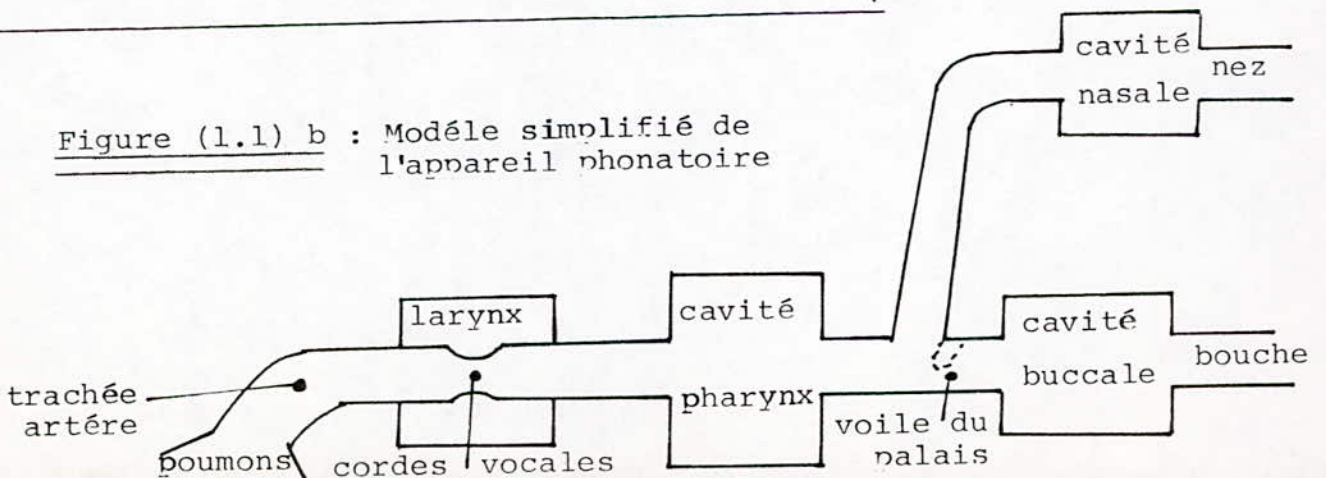
$1/10^{\text{ème}}$  et 1000 pw.

Figure (1.1) a : Anatomie-physiologie de l'appareil vocal.



المدرسة الوطنية المتعددة التقنيات  
 BIBLIOTHEQUE — المكتبة  
 Ecole Nationale Polytechnique

Figure (1.1) b : Modèle simplifié de l'appareil phonatoire



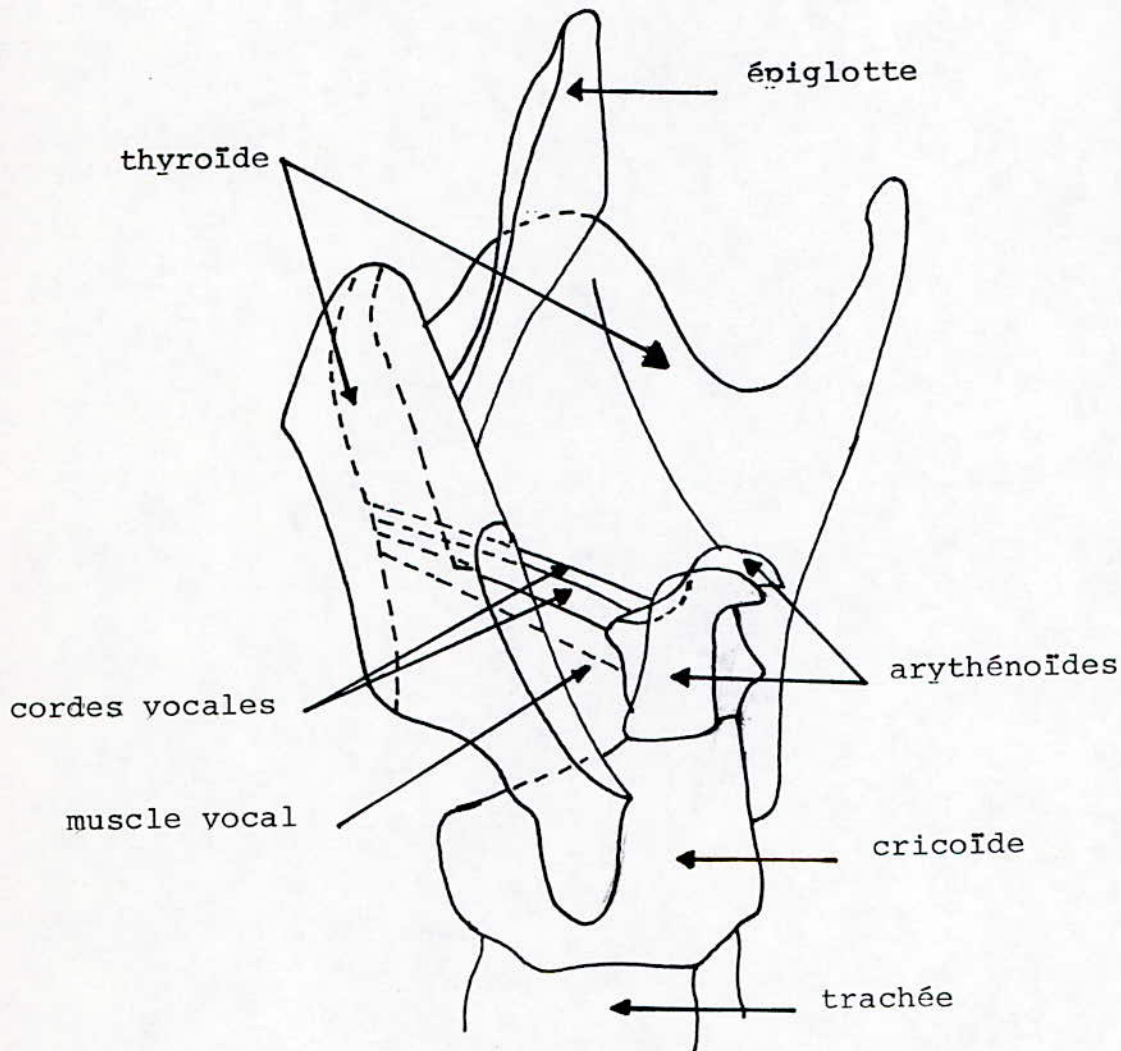


Figure (1.2) : Coupe horizontale du larynx au niveau des cordes vocales.

Chapitre II : MODELE DE PRODUCTION DE LA PAROLE .

II.1. Introduction :

L'objectif de notre étude est de procéder à une analyse de type production sur le signal de la parole, et d'essayer ensuite d'effectuer une synthèse de haute qualité.

Une analyse de type production est une analyse qui ne décrit pas simplement le signal par ses propriétés spectrales, par exemple, mais vise à une "déconvolution" du signal par rapport à un modèle de production.

II.2. Principe d'analyse - synthèse :

2.1. Analyse de la parole :

Soit un signal de la parole  $S(t)$  tel que  $t \in [0, T]$  . La première opération d'analyse consiste à tronçonner le temps en intervalles d'environ 10 à 20 ms. Ces intervalles de temps sont appelés "fenêtres temporelles". L'analyse du signal de parole repose sur l'hypothèse fondamentale suivante :

- Le signal à analyser est stationnaire pendant toute la durée de la fenêtre d'analyse, c'est à dire que ses propriétés, en particulier spectrales, ne varient pas sur la longueur de la fenêtre. L'analyse sera faite successivement sur chacune des fenêtres. Le signal est ensuite échantillonné et l'on obtient ainsi pour chaque fenêtre une suite d'échantillons  $S(n)$  ;  $n \in [0, N]$  et  $N$  est généralement de l'ordre de 200. La prédiction linéaire consiste à exprimer la valeur du  $n$ ème échantillon en fonction des  $P$  échantillons précédents et de l'entrée  $U(nT)$  du système.

Il s'agit d'obtenir une formule du type :

$$\hat{S}(nT) = \sum_{k=1}^P a(k) \hat{S}(nT - kT) + A U(nT) \dots\dots (2.1)$$

Où  $T$  : est la période d'échantillonnage.

Les coefficients  $a(k)$  et  $A$  reproduisent le mieux possible le comportement du canal vocal, compte tenu de l'entrée  $U(nT)$  appliquée.

Remarque : On obtient une formule de type (2.1) pour chaque fenêtre ce qui signifie que les coefficients  $a(k)$  et  $A$  doivent être recalculés sur chaque fenêtre.

Pour simplifier la formulation on va poser  $\hat{S}(nT) = \hat{S}(n)$  alors la relation (2.1) devient :

$$\hat{S}(n) = \sum_{k=1}^P a(k) \hat{S}(n - k) + A U(n) \dots\dots\dots (2.2)$$

En terme de transformée en Z la relation (2.2) donne :

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

$$\hat{S}(Z) \left[ 1 - \sum_{k=1}^P a(k) Z^{-k} \right] = A U(Z)$$

On pose :  $H(Z) = 1 - \sum_{k=1}^P a(k) Z^{-k}$

d'où  $\hat{S}(Z) = \frac{A}{H(Z)} U(Z)$

Le signal  $\hat{S}(Z)$  apparait donc comme la réponse du filtre  $A/H(Z)$  ne possédant que des poles à l'entrée  $U$ .

Une fois  $A/H(Z)$  est connu, la phase d'analyse est terminée.

## 2.2. La synthèse :

On dispose pour chaque fenêtre temporelle des paramètres définissant le filtre  $A/H(Z)$ . Il reste maintenant à exciter ce filtre, afin d'obtenir en sortie des échantillons de parole synthétique  $\hat{S}(n)$ .

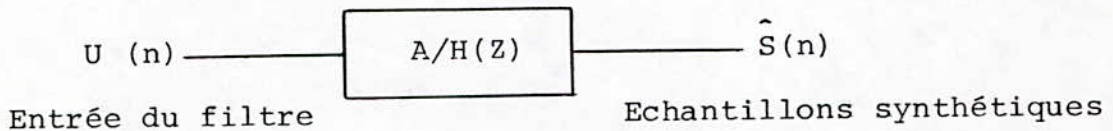


Figure (2.1) : Production du signal synthétique

Ayant déterminées les deux types d'excitations  $U(n)$  du filtre, qui rappelons-les, sont un train périodique d'impulsions et un bruit acoustique blanc.

Nous disposons maintenant de tous les paramètres nécessaires à l'obtention du signal synthétique.

II.3. Determination du modèle :

3.1. Introduction :

L'analyse par prédiction linéaire est une méthode de modélisation de type essentiellement spectral, même si cette méthode revient à considérer au niveau temporel, qu'un échantillon peut être "prédit" comme fonction linéaire des P échantillons précédents (d'où le nom de prédiction linéaire).

La méthode de prédiction linéaire repose principalement sur deux hypothèses :

a) Les variations du conduit vocal peuvent être approchées par une succession de configuration stationnaires dont la durée (dite période "trame") est de l'ordre de 10 à 25 ms.

b) Pendant chaque trame, la source d'excitation est constante (série d'impulsions ou bruit blanc) et la fonction de transfert de l'appareil vocal est représentée par un filtre récursif ne contenant que des poles.

Soient :

T : La période d'échantillonnage

S(n) = S(nT) : L'échantillon du rang n du signal à analyser.

S(Z) : La transformée en Z de ce signal.

$\hat{S}(n) = \hat{S}(nT)$  : Le signal modélisé

$\hat{S}(Z)$  : Sa transformée en Z.

3.2. Determination du modèle de production de la parole :

On désigne par S'(n) la valeur prédite de S(n) qui s'écrit :

$$S'(n) = \sum_{k=1}^P a(k) S(n-k) \dots\dots\dots (2.3)$$

où a(k) :  $1 \leq k \leq P$  est un ensemble de coefficients réels, que l'on appelle généralement coefficients de prédiction P est l'ordre du prédicteur.

L'erreur de prédiction entre le signal original et le signal prédit est représentée par :



$$e(n) = S(n) - S'(n) \dots\dots\dots (2.4)$$

$$e(n) = S(n) - \sum_{k=1}^P a(k) S(n-k) \dots\dots (2.5)$$

Nous allons maintenant appliquer la transformée en Z à l'équation (2.5) :

$$E(Z) = S(Z) - \sum_{k=1}^P a(k) Z^{-k} S(Z) \dots\dots (2.6)$$

$$E(Z) = S(Z) \left[ 1 - \sum_{k=1}^P a(k) Z^{-k} \right] \dots\dots (2.7)$$

$$E(Z) = S(Z) \cdot H(Z) \dots\dots\dots (2.8)$$

Où E(Z) et S(Z) sont les transformées en Z de e(n) et S(n) respectivement.

Nous avons  $H(Z) = 1 - \sum_{k=1}^P a(k) Z^{-k}$  avec  $1 \ll P \ll k$

H(Z) : désigne le filtre inverse

De l'équation (2.8) nous déduisons  $S(Z) = \frac{E(Z)}{H(Z)}$

On désigne par G(Z) la fonction de transfert approximée :

$$G(Z) = \frac{A}{H(Z)} = \frac{A}{1 - \sum_{k=1}^P a(k) Z^{-k}}$$

Cette fonction de transfert est celle d'un filtre ne possédant que des pôles (modèle spectral "tous pôles")

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

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

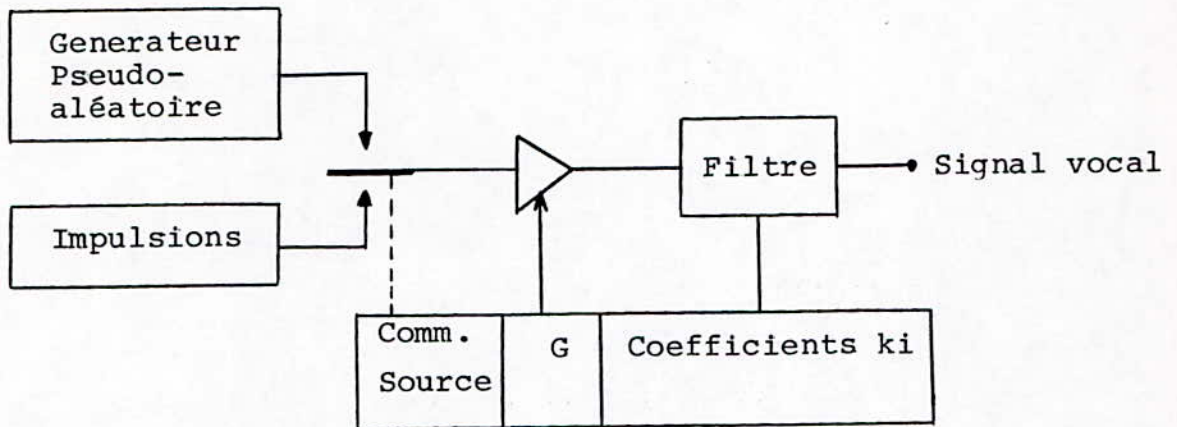


Figure (2.2) : Représentation simplifiée du mécanisme de production de la parole.

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

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

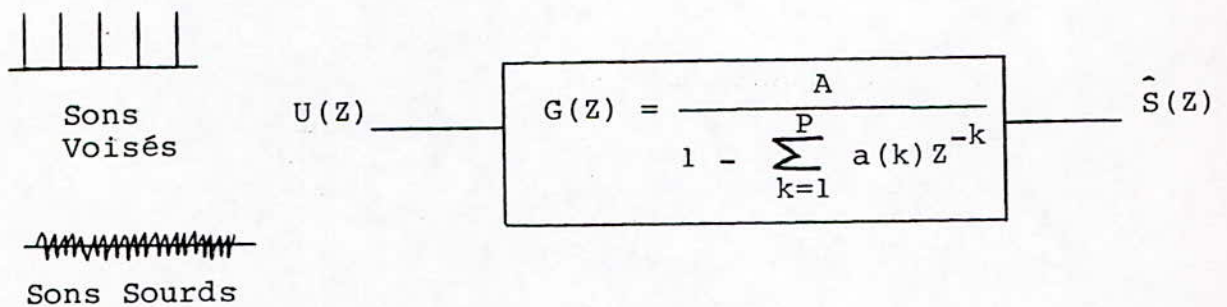


Fig.(2.3) a : Modèle dans le domaine fréquentiel.

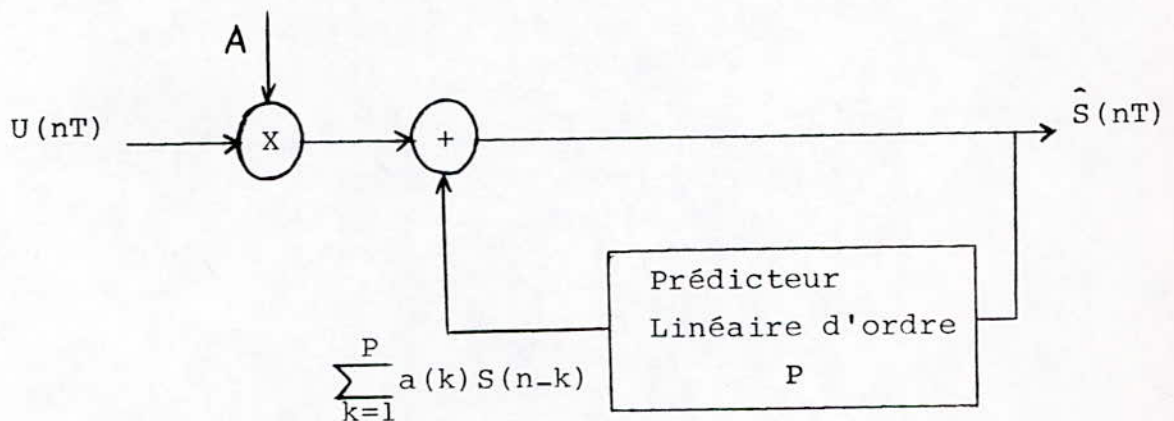


Fig.(2.3) b : Modèle dans le domaine temporel.

Figure (2.3) : Modèle de production de la parole utilisé dans les méthodes de prédiction linéaire.

A partir de la figure (2.3) a on a :

$$\hat{S}(Z) = U(Z) \cdot G(Z) = \frac{A U(Z)}{1 - \sum_{k=1}^P a(k) Z^{-k}} \dots\dots\dots (2.9)$$

Où U(Z) et S(Z) sont respectivement les transformées en Z de U(nT) et S(nT).

De l'équation (3.7) on a :

$$\hat{S}(Z) = \hat{S}(Z) \sum_{k=1}^P a(k) Z^{-k} + A U(Z) \dots\dots\dots (2.10)$$

On prend la transformée inverse de l'équation (2.10) ce qui donne :

$$\hat{S}(n) = \sum_{k=1}^P a(k) \cdot \hat{S}(n-k) + A U(n) \dots\dots\dots (2.12)$$

Ayant déterminé le modèle de production de la parole utilisé dans les méthodes de prédiction linéaire, reste à déterminer les paramètres du prédicteurs.

II.4. Principales méthodes de prédiction linéaire :

On a vu que le signal prédit étant

$$S'(n) = \sum_{k=1}^P a(k) S(n-k) \dots\dots\dots (2.13)$$

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

La relation (2.13) s'adapte bien au signal de la parole car comme on l'a vu il est possible de l'appliquer à un modèle de production de la parole.

Ils existent plusieurs possibilités de déterminer ces paramètres a(k).

4.1. La méthode exacte :

Cette méthode est fondée sur les hypothèses suivantes :

- Le signal est défini exactement pour "2p" échantillons consécutifs.
- Un échantillon de signal de parole peut être prédit exactement à partir des "p" échantillons précédents.
- La deuxième hypothèse est valable pour les "p" échantillons qui suivent.

4.2. La méthode de covariance :

Cette méthode suppose que le signal est non stationnaire à l'intérieur d'une trame, c'est à dire qu'elle tient compte

de la variation spectrale due à un décalage du signal à l'intérieur de cette trame.

Nous citons les hypothèses fondamentales sur lesquelles est fondée la méthode :

- Le signal est défini pour "P+N" échantillons consécutifs avec N entiers.

- Un échantillon du signal de parole peut être prédit approximativement à partir des "P échantillons précédents.

- La deuxième hypothèse est valable pour les N échantillons consécutifs qui suivent.

- L'écart quadratique total, entre le signal original et sa valeur prédite, est minimisée pour l'ensemble des "N" échantillons consécutifs.

#### 4.3. La méthode d'autocorrelation :

Cette méthode considère le signal stationnaire dans un intervalle de temps. Elle suppose que le spectre à court terme du signal est invariant dans la trame considérée. Ceci est réalisé par un fenêtrage temporel préalable du signal.

Les suppositions de la méthode d'autocorrelation sont les suivantes :

- Le signal est défini pour toutes les valeurs du temps, il est identiquement nul en dehors d'une séquence de N échantillons où N est un entier, ceci équivaut à multiplier le signal de parole par une fenêtre finie correspondant à N échantillons.

- La deuxième hypothèse est valable pour toutes les valeurs du temps (pour tous les échantillons du signal).

- L'écart quadratique total entre le signal original et le signal prédit est minimisé sur toute l'échelle du temps.

Remarque : La méthode exacte suppose que l'erreur  $e(n)$  est identiquement nulle à chaque instant. Cela entraîne que dans le modèle de production de la parole les échantillons du signal de source  $U(nT)$  sont nuls ce qui implique qu'il n'y a pas d'impulsions de mélodie pendant la période d'échantillonnage.

La méthode de covariance nécessite pour l'analyse moins d'échantillons que l'autocorrelation mais avec elle de sérieuses instabilités sont susceptibles de survenir.

## II.5. Résolution des équations de prédiction linéaire par la méthode d'autocorrelation.

### 5.1. Etablissement des équations d'autocorrelation :

Nous avons vu précédemment, d'après l'équation (2.5) que l'erreur  $e(n)$  entre le signal original de parole et le signal prédit s'écrit :

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

L'approximation peut être appliquée à n'importe quel signal, et on peut utiliser n'importe quel critère d'optimisation à l'intérieur d'un intervalle, il s'agira toujours d'approcher le plus  $S'(n)$  de  $S(n)$ .

Le critère classique est la minimisation de l'énergie de l'erreur de prédiction dans un intervalle.

Ce qui nous conduit à choisir le critère de moindres carrés.

Désignons par  $E$  l'erreur quadratique totale :

$$E = \sum_n e^2(n) = \sum_n [S(n) - S'(n)]^2 \dots\dots\dots (2.14)$$

$$E = \sum \left[ S(n) - \sum_{k=1}^P a(k) S(n-k) \right]^2 \dots\dots\dots (2.15)$$

Il faut rendre  $E$  le plus petit possible donc minimaliser  $E$  ; il existe pour cela des méthodes mathématiques : On calcule les dérivées partielles de  $E$  par rapport aux coefficients de pondération  $a(k)$  et l'on annule chacune d'entre elles ; on écrit donc pour l'une quelconque des dérivées partielles (de l'erreur  $E$  par rapport au coefficient  $a(i)$  servant de témoin) :

$$\frac{\partial E}{\partial a(i)} = \sum_n 2 \left[ S(n) - \sum_{k=1}^P a(k) S(n-k) \right] \cdot S(n-i) = 0 \quad (2.16)$$

$$\sum_n \sum_{k=1}^P a(k) \cdot S(n-k) \cdot S(n-i) = \sum_n S(n) \cdot S(n-i) \quad (2.17)$$

$$\sum_{k=1}^P a(k) \sum_n S(n-k) \cdot S(n-i) = \sum_n S(n) \cdot S(n-i) \quad (2.18)$$

Les équations (2.18) constituent un ensemble de  $P$  équations à  $P$  inconnues. A partir des hypothèses faites sur cette méthode (i.e la méthode d'autocorrelation), le signal  $S(n)$  sera défini comme suit :

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

De cette manière l'équation (2.18) s'écrira :

$$\sum_{k=1}^P a(k) \sum_{n=-\infty}^{+\infty} S(n-k) \cdot S(n-i) = \sum_{n=-\infty}^{+\infty} S(n) \cdot S(n-i) \quad (2.19)$$

Posant  $m = n-i$

$$\sum_{k=1}^P a(k) \sum_{m=-\infty}^{+\infty} S(m) \cdot S(m+|i-k|) = \sum_m S(m+|i|) \cdot S(m) \quad (2.20)$$

La fonction d'autocorrelation R(i) est définie comme suit :

$$R(i) = \sum_{n=-\infty}^{+\infty} S(n) \cdot S(n-i) \dots (2.21) \text{ avec } R(i) = R(-i)$$

Or  $S(n) = 0$  pour  $n \notin [0, N-1]$

$$\text{Donc } R(i) = \sum_{n=0}^{N-1-|i|} S(n) \cdot S(n+|i|) \dots (2.22)$$

$$R(i) = \sum_{k=1}^P a(k) R(|i-k|) \dots (2.23) \quad 1 \leq i \leq P$$

$$\Rightarrow R(i) = \sum_{k=1}^P a(k) R(|i-k|) \dots (2.24)$$

Ecriture de l'équation (2.24) sous forme matricielle.

$$\begin{pmatrix} R(0) & R(1) & R(2) & \dots & R(P-1) \\ R(1) & R(0) & R(1) & \dots & R(P-2) \\ R(2) & R(1) & R(0) & \dots & R(P-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R(P-1) & R(P-2) & R(P-3) & \dots & R(0) \end{pmatrix} \begin{pmatrix} a(1) \\ a(2) \\ a(3) \\ \vdots \\ \vdots \\ a(P) \end{pmatrix} = \begin{pmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ \vdots \\ R(P) \end{pmatrix} \quad (2.25)$$

On remarque que cette matrice carrée, appelée matrice de TOEPLITZ, elle est symétrique et que sa diagonale est constituée par des éléments identiques.

### 5.2. Erreur quadratique totale minimale de prédiction:

On a

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

$$= \sum_n \left[ S(n)^2 - 2 S(n) \sum_{k=1}^P a(k) S(n-k) + \sum_{k=1}^P \sum_{i=1}^P a(k) a(i) S(n-k) S(n-i) \right] \quad (2.26)$$

$$= \sum_n S(n)^2 - 2 \sum_{k=1}^P a(k) \sum_n S(n) S(n-k) + \sum_{k=1}^P a(k) \sum_{i=1}^P a(i) \sum_n S(n-k) S(n-i) \quad (2.27)$$

Or nous savons que :

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

De même :

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

On remplace dans (2.27)

$$\Rightarrow E(P) = \sum_n S(n)^2 - \sum_{k=1}^P a(k) \sum_n S(n) S(n-k) \dots\dots\dots (2.28)$$

Sachant que pour la méthode d'autocorrelation le signal  $S(n) = 0$  pour  $n \notin [0, N-1]$ .

et

$$\left\{ \begin{aligned} R(k) &= \sum_{n=0}^{N-1-|k|} S(n) \cdot S(n+|k|) \\ R(0) &= \sum_{n=0}^{N-1} S(n)^2 \end{aligned} \right. \dots\dots\dots (2.29)$$

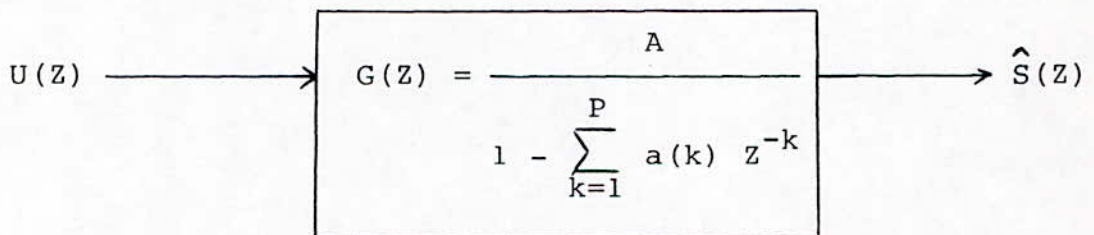
En remplaçant ces deux équations dans l'expression de  $E(P)$  nous obtenons :

$$E(P) = R(0) - \sum_{k=1}^P a(k) R(k) \dots\dots\dots (2.30)$$

### 5.3 Calcul du facteur de gain A :

Le calcul du gain est nécessaire pour obtenir l'information sur l'énergie du signal dans une trame.

On peut calculer la valeur de  $A$  en faisant appel au critère qui dit que : l'énergie totale contenue dans une trame du signal synthétique  $\hat{S}(n)$  doit être égale à l'énergie totale correspondante du signal original  $S(n)$ .



On sait que :  $\hat{S}(n) = \sum_{k=1}^P a(k) \hat{S}(n-k) + A U(n)$

On posera :  $U(n) = \delta(n)$

Donc :  $\hat{S}(n) = \sum_{k=1}^P a(k) S(n-k) + A \delta(n) \dots\dots\dots (2.31)$

Avec :  $\delta(n) = \begin{cases} 0 & n \neq 0 \\ 1 & n = 0 \end{cases} \dots\dots\dots (2.32)$

$$\Rightarrow \begin{cases} \hat{S}(n) = \sum_{k=1}^P a(k) S(n-k) & \text{si } n \geq 1 \\ \hat{S}(n) = 0 & \text{qd } n < 0 \\ \hat{S}(n) = A & \text{qd } n = 0 \end{cases} \dots\dots\dots (2.33)$$

Par définition : La fonction d'autocorrelation  $\hat{R}(i)$  est donnée par

$$\hat{R}(i) = \sum_{n=-\infty}^{+\infty} \hat{S}(n) \hat{S}(n+i) \quad \forall i \quad \dots\dots\dots (2.34)$$

Et comme :  $\hat{S}(n) = 0$  pour  $n < 0$

$$\hat{R}(i) = \sum_{n=0}^{\infty} \hat{S}(n+i) \hat{S}(n) \quad i \geq 0 \quad \dots\dots\dots (2.35)$$

On peut écrire :  $\hat{S}(n+i) = \sum_{k=1}^P a(k) \hat{S}(n+i-k)$

$$\text{Alors : } \hat{R}(i) = \sum_{n=0}^{+\infty} \hat{S}(n) \sum_{k=1}^P a(k) \hat{S}(n+i-k)$$

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

Et comme :  $\hat{R}(|i-k|) = \sum_{n=0}^{\infty} \hat{S}(n) \hat{S}(n+i-k)$

$$\text{d'où : } \hat{R}(i) = \sum_{k=1}^P a(k) \hat{R}(|i-k|) \quad \text{pour } i \geq 1 \quad \dots\dots\dots (2.36)$$

D'après l'équation (2.34) on a :

$$\hat{R}(0) = \sum_{n=0}^{\infty} \hat{S}^2(n) = \hat{S}(0)^2 + \sum_{n=1}^{\infty} \hat{S}^2(n)$$

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

On posera :  $n = m + k$  dans  $\hat{R}(0)$



Alors :  $\hat{R}(0) = A^2 + \sum_{k=1}^P a(k) \sum_{m=1}^{\infty} \hat{S}(m) \hat{S}(m+k)$

Et comme :  $\hat{S}(m) = 0$  pour  $m < 0$

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

Finalement :  $\hat{R}(0) = A^2 + \sum_{k=1}^P a(k) \hat{R}(k) \dots\dots\dots (2.37)$

Il est possible de déterminer les coefficients  $a(k)$  uniquement à partir des valeurs de  $\hat{R}(i)$   $0 \leq i \leq P$  compte tenu du fait que les équations de  $R(i)$  et  $\hat{R}(i)$  sont du même type. Néanmoins, il existe une différence qui tient au domaine de variation limitée de l'indice  $i$  pour  $R(i)$ . Cependant les fonctions d'auto-correlation  $\hat{R}(i)$  et  $R(i)$  sont représentées par la même équation matricielle (2.2\_).

De ce fait, les fonctions d'autocorrelation  $\hat{R}(i)$  et  $R(i)$  doivent satisfaire l'équation suivante :

$$\hat{R}(i) = C R(i) \quad 0 \leq i \leq P \quad \dots\dots\dots (2.38)$$

Où  $C$  est une constante à déterminer.

Afin de conserver l'égalité des énergies de la réponse impulsionnelle de  $\hat{S}(Z)$  et du signal, il faut que  $\hat{R}(0) = R(0)$ . Par conséquent  $C = 1$  et nous obtenons le résultat suivant :

$$\hat{R}(i) = R(i) \quad 0 \leq i \leq P \quad \dots\dots\dots (2.39)$$

Et la relation (2.37) devient :

$$R(0) = A^2 + \sum_{k=1}^P a(k) R(k) \quad \dots\dots\dots (2.40)$$

d'où  $A^2 = R(0) - \sum_{k=1}^P a(k) R(k) \quad \dots\dots\dots (2.41)$

En comparant (2.41) avec (2.30) on aura :

$$E(P) = A^2 = R(0) \quad V(P) = R(0) \left[ 1 - \sum_{k=1}^P a(k) r(k) \right] \quad (2.42)$$

Pour  $k = 0$  ;  $A^2 = E(0) = R(0)$

$A^2$  est donc égale à l'erreur quadratique totale minimale.

5.4. Résolution du système d'équation selon la méthode de "Durbin".

De nombreuses méthodes permettent de résoudre le système d'équations linéaires telles que :

- Méthode de GAUSS - SEIDEL.
- Méthode de JACOBI.
- Méthode de GAUSS - JORDAN.
- Méthode de DURBIN.

Mais en tenant compte de la rapidité d'exécution et de l'encombrement mémoire, nous avons choisis pour résoudre le système d'équation d'autocorrelation la méthode de DURBIN (récursive).

Elle utilise  $P(P+1)$  opérations et occupe une place mémoire réduite :

Algorithme de DURBIN :

$$E(0) = R(0)$$

$$K(i) = \left[ R(i) - \sum_{j=1}^{i-1} a(j) R(i-j) \right] / E(i-1) \quad j \leq i \leq P$$

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

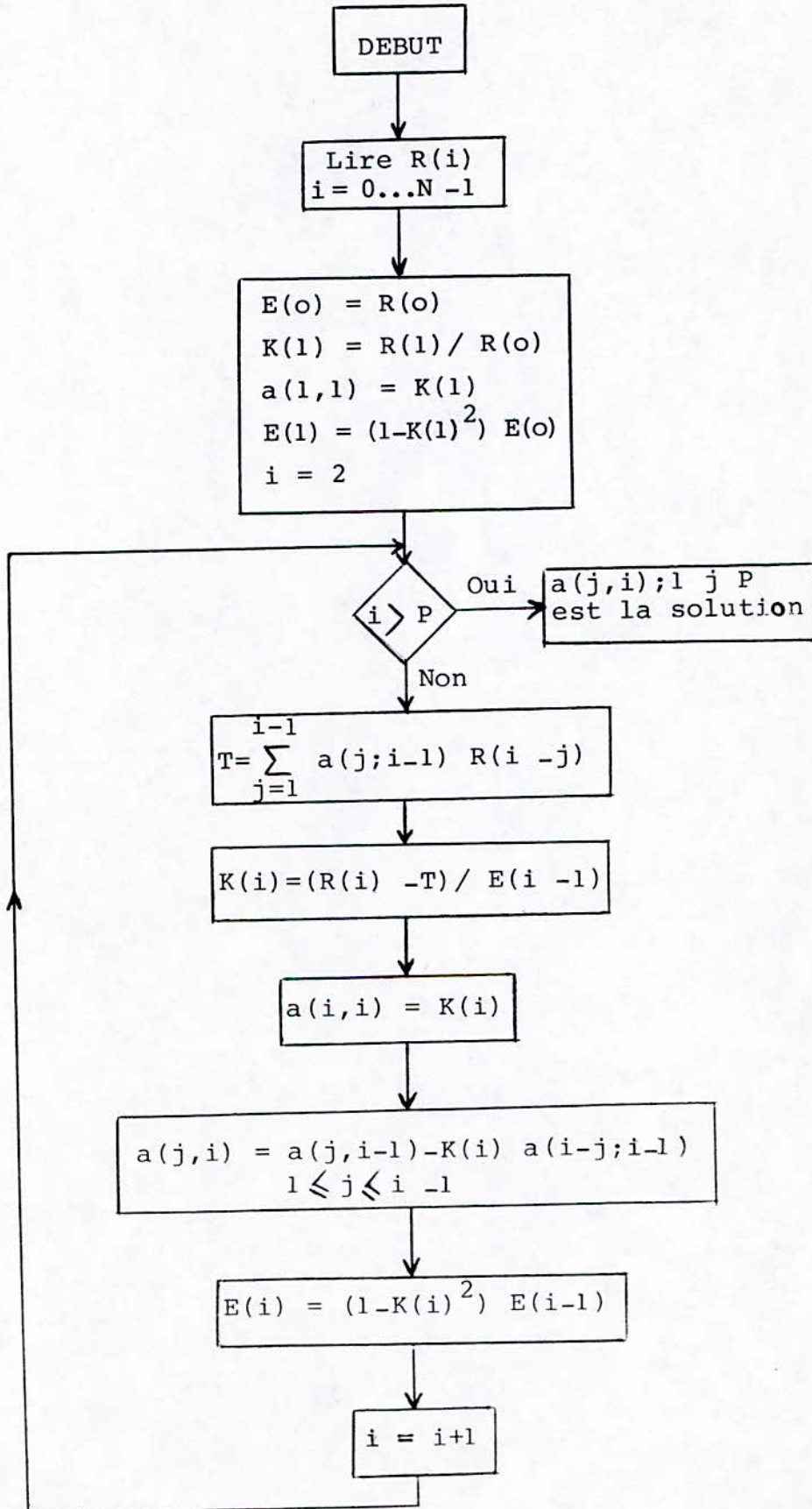
$$a_{(j)}^{(i)} = a_{(j)}^{(i-1)} - K(i) a_{(i-1)}^{(i-1)} \quad 1 \leq j \leq i-1$$

$$E(i) = \left[ 1 - K(i)^2 \right] E(i-1)$$

Les coefficients  $K(i)$  dits coefficients de réflexion de l'onde à la jonction entre deux sections du conduit vocal sont liés d'une manière récursive aux coefficients de prédiction par les relations précédentes.

Nous remarquons que pour obtenir les coefficients  $a(j)$  d'un prédicteur d'ordre  $P$  on doit calculer tous les coefficients des prédicteurs d'ordre inférieur à  $P$ .

Organigramme pour la méthode d'autocorrelation.



## Chapitre III :

### III.1. Introduction :

D'une façon générale, une machine à produire de la parole est constituée de sources d'excitations qui sont : Une source d'impulsions pour les sons voisés ; une source de bruit pour les non-voisés et un simulateur du conduit vocal.

On crée ainsi un "modèle qui est supposé stable sur des intervalles de temps de l'ordre de la vingtaine de milli-secondes et qui reçoit également des paramètres de commandes toutes les 20 msec. environ.

Ce modèle peut être réalisé par un nombre plus ou moins important de filtres électriques dont les caractéristiques reproduisent les propriétés résonnantes du conduit vocal (synthèse par formant) il est également réalisable numériquement par un unique filtre numérique (synthèse à prédiction linéaire). Voir figure (3.1).

Une autre méthode consiste à modéliser directement le conduit vocal lui-même et non plus ce qui est généré, on tient compte par exemple des inerties de la mâchoire et de la langue dans ce modèle dit de "synthèse articulatoire". On devrait obtenir une parole de très haute qualité.

### III.2. Influence de l'excitation en prédiction linéaire sur la qualité de la parole synthétisée :

L'avantage de la prédiction linéaire se traduit par une bonne qualité de la voix de synthèse ainsi que par la réduction du débit de transmission.

L'inconvénient est que dans le cadre d'application la parole synthétique est altérée par "un bourdonnement".

Dans ce paragraphe nous allons voir les différentes formes d'excitations proposées dans le but d'atténuer ce défaut principal.

#### III.2.1. Différentes formes d'excitation :

Ces différentes formes tendent de réduire le bourdonnement en annulant la moyenne du signal d'excitation ou en répartissant cette excitation sur plusieurs points. Ces excitations ne sont utilisées que sur des trames voisées, les trames non voisées étant excitées par le conventionnel bruit blanc.

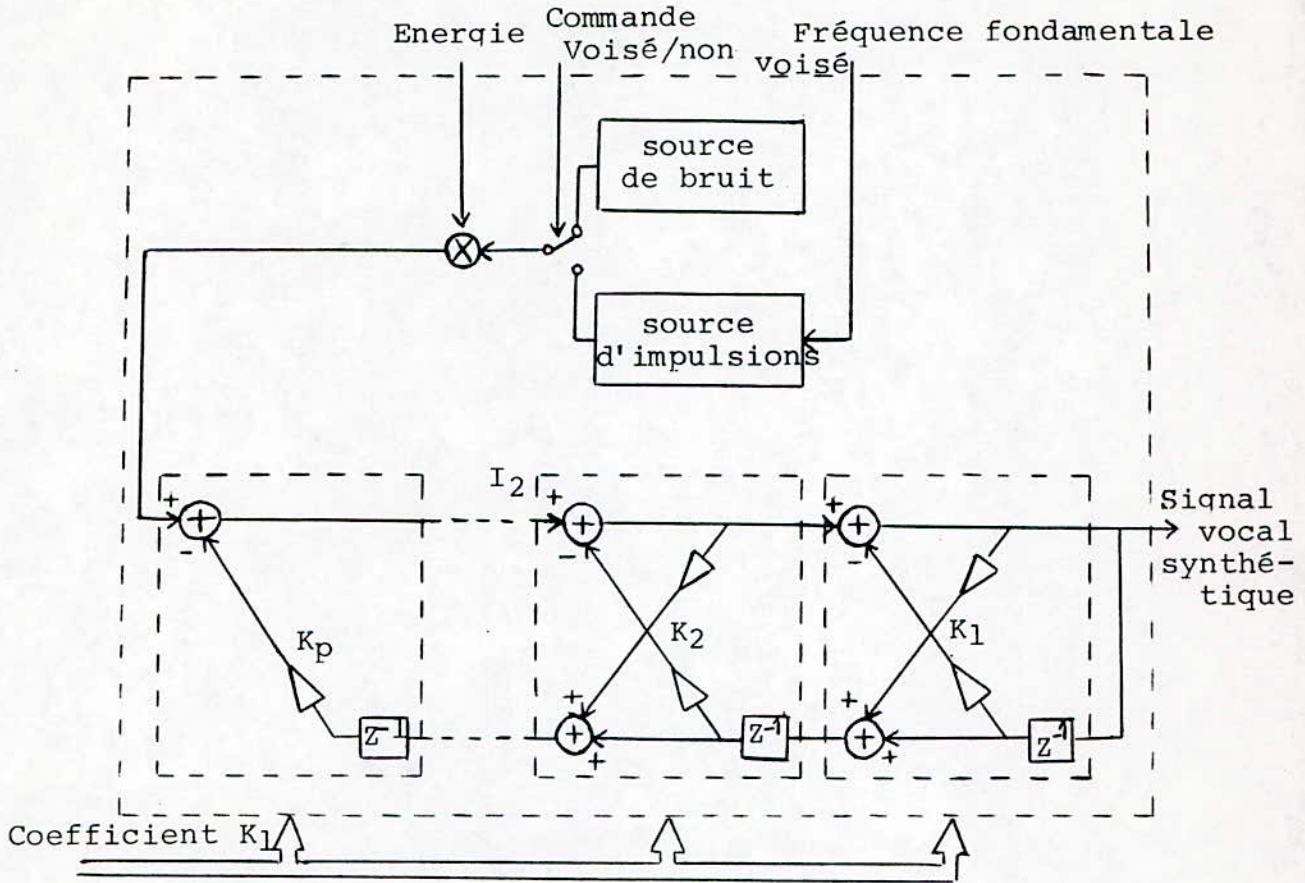


Figure (3.1) : Filtre en treillis à deux multiplieurs pour un synthétiseur à prédiction linéaire.

III.2.1.1. Excitation bi-impulsionnelle :

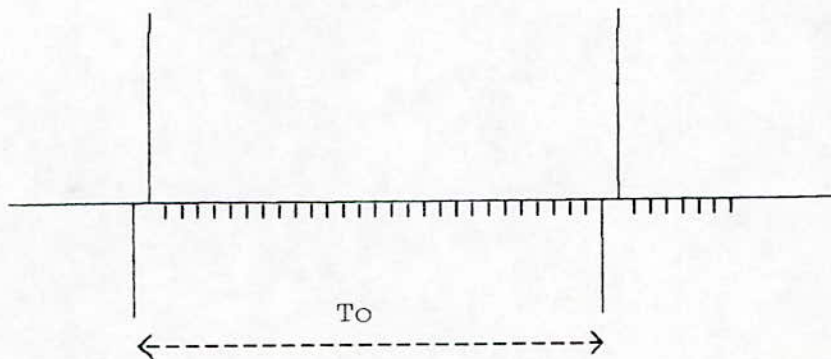


Figure (3.2) : Excitation bi-impulsionnelle.

Elle permet de réduire un peu le bourdonnement sur une période de  $[n = 0 ; n = N_T]$  elle est définie de la manière suivante :

$$\left\{ \begin{array}{ll} E = -\frac{G}{2} & n = 0 \\ E = G & n = 1 \dots\dots\dots (3.1) \\ E = \frac{-G}{2(N_T - 1)} & n \gg 1 \end{array} \right.$$

Où G est l'énergie d'excitation sur la trame.

III.2.1.2. Excitation triangulaire :

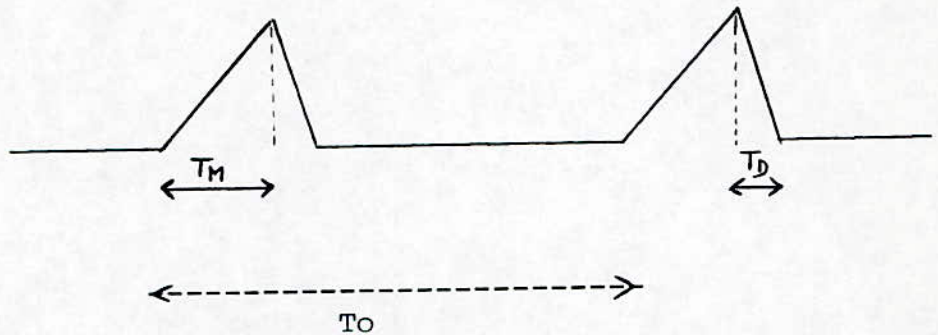


Figure (3.3) : Excitation triangulaire.

Sur une trame  $[n = 0 ; N]$  cette excitation est définie par un temps de montée,  $T_M$  et un temps de descente  $T_D$ . Voir figure (3.3).

Elle est définie comme suit :

$$\left\{ \begin{array}{ll} t \in [0 ; T_M] & ; E(t) = \frac{a}{T_M} t \\ t \in [T_M ; T_M + T_D] & ; E(t) = a(1 - \frac{t - T_M}{T_D}) \dots\dots (3.2) \\ t \in [T_M + T_D ; T_0] & ; E(t) = 0 \end{array} \right.$$

En répartissant l'énergie d'excitation sur plusieurs points on diminue efficacement le facteur de crête.

Par contre la parole synthétique est d'autant plus "étouffée" que  $T_M$  et  $T_D$  augmentent.

Il y a donc un compromis à trouver entre le bourdonnement et l'aspect étouffé du filtrage bas, l'optimum est pour  $T_M / T_0 = (3.6)\%$

et  $T_D / T_0 = 1\%$

### III.2.1.3. Excitation bi-triangulaire :

Elle est répartie sur deux triangles l'un positif l'autre négatif (voir figure 3.4).

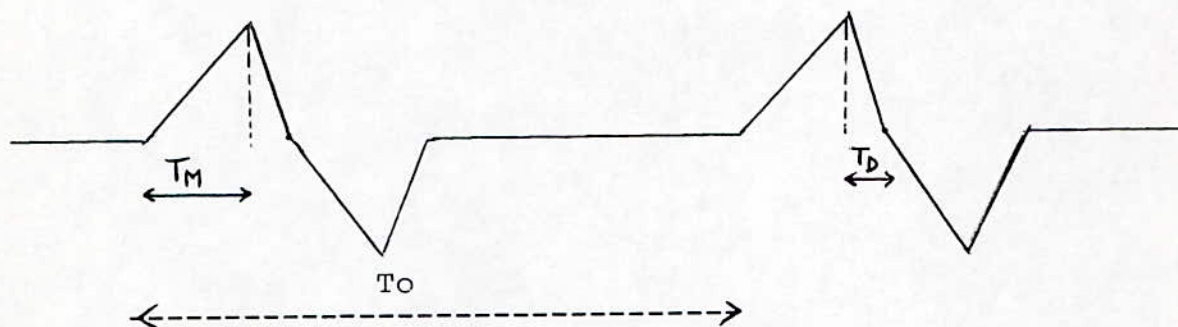


Figure (3.4) : Excitation bi-triangulaire

Cette forme d'excitation est donc définie de la manière suivante :

$$\begin{cases}
 t \in [0 ; T_M] & ; E(t) = \frac{a}{T_M} t = E_0(t) \\
 t \in [T_M ; T_M + T_D] & ; E(t) = a(1 - \frac{t - T_M}{T_D}) = E_1(t) \\
 t \in [T_M + T_D ; 2T_M + T_D] & ; E(t) = E_0(t - (T_M + T_0)) \\
 t \in [2T_M + T_D ; 2T_M + 2T_D] & ; E(t) = E_1(t - 2T_M - T_D)
 \end{cases} \quad (3.3)$$

C'est cette excitation qui réduit le plus le facteur de crête. Cependant, l'aspect spectral du signal synthétique diffère de celui du signal original.

#### III.2.1.4. Excitation visant une approche de l'onde glottique :

Pour modéliser l'onde glottique ROSENBERG a effectué sur des segments de parole une analyse par filtrage inverse et on a déduit une mesure de l'onde glottique. Remplaçant l'onde mesurée par diverses modelisations, il en a déduit que la meilleure représentation de l'onde glottique est une onde dont la dérivée présente une seule discontinuité à la fermeture de la glotte. Voir figure (3.5).

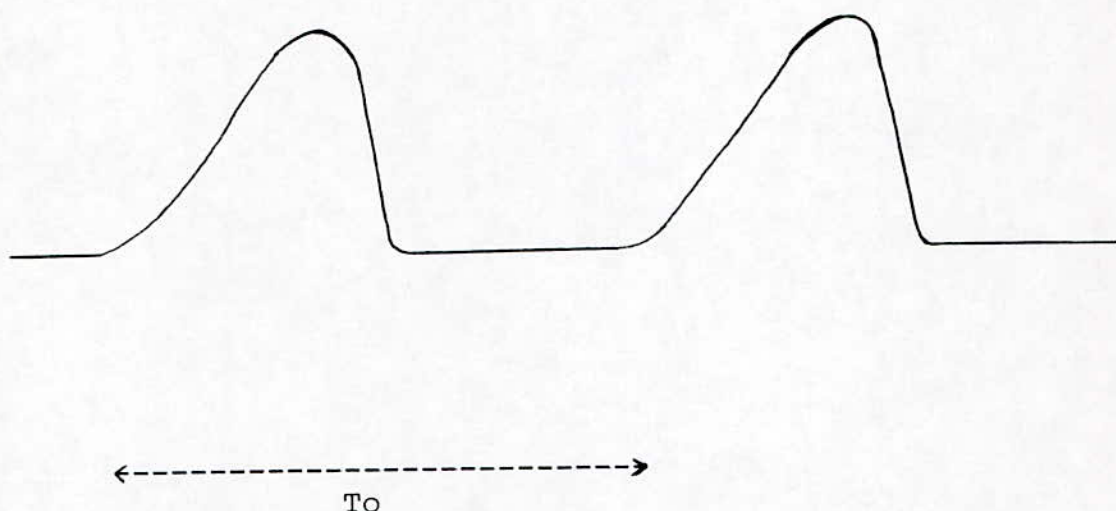


Figure (3.5) : Excitation polynomiale.

Il ne faut <sup>pas</sup> que  $T_M / T_0$  dépasse 35% si non l'indication formatique s'efface et le signal perd de son naturel.

Remarque : Un test subjectif montre que l'excitation la plus agréable d'écoute est l'excitation polynomiale ( $T_M / T_D = 20\%$  ;  $T_D / T_0 = 10\%$  ).

Les excitations bi-impulsionnelle et triangulaire de qualité inférieur sont sensiblement équivalentes entre elles.

#### III.2.1.5. Excitation mixte :

L'excitation idéale étant le signal résiduel. Ce dernier (selon FUJIMA) peut être considéré comme périodique jusqu'à une fréquence  $F_C$  et apériodique aux fréquences supérieures.  $F_C$  est nulle pour un signal non-voisé.

C'est un modèle que MAKOUL a adapté pour l'excitation des synthétiseurs à prédiction linéaire.



Il est constitué d'un train d'impulsions aux fréquences inférieures à la fréquence de coupure  $F_C$ , et d'un bruit blanc aux fréquences supérieures ( voir figure 3.6 ).

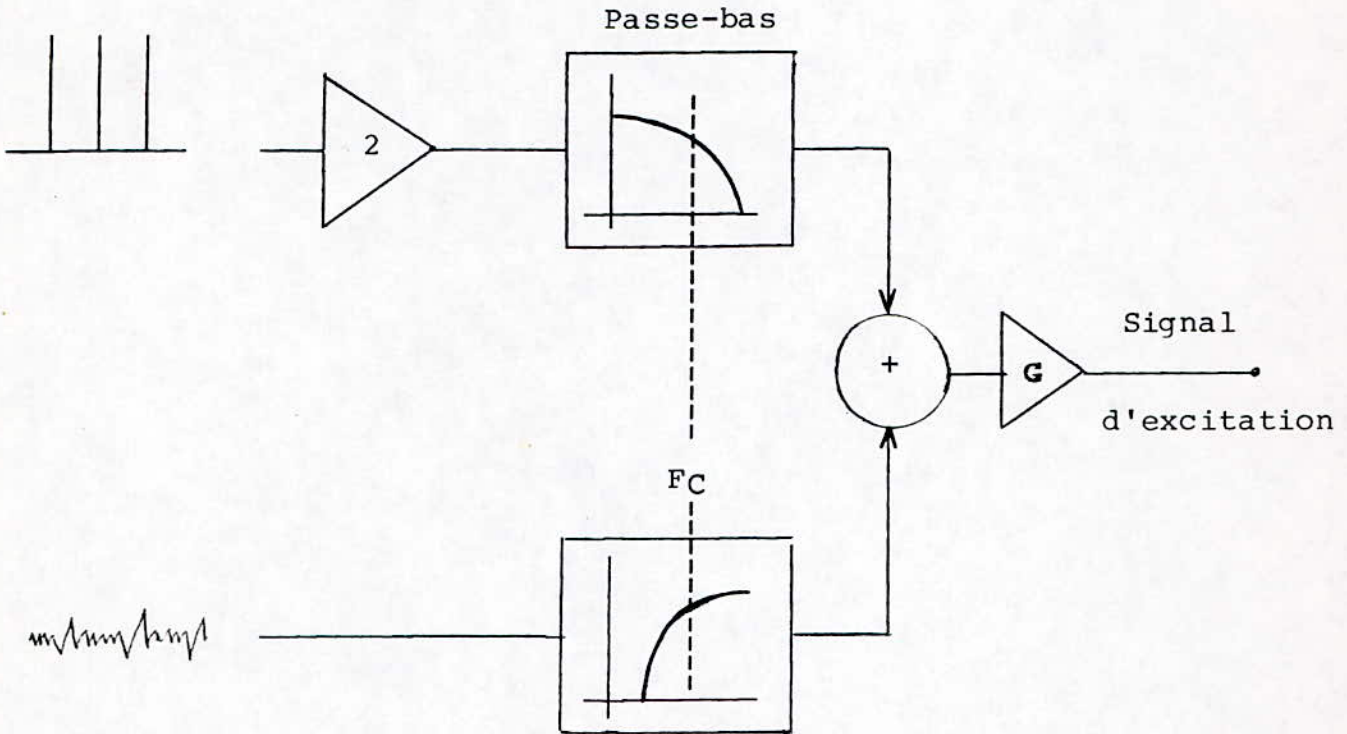


Figure (3.6) : Production d'une source mixte d'excitation

Cette forme d'excitation nécessite la transmission des paramètres usuels de prédiction linéaire et de la fréquence  $F_C$  (dont le calcul est très difficile).

L'algorithme proposé par MAKOUL examine l'écart fréquentiel entre deux maxima du spectre et  $F_C$  est la plus haute fréquence où écart se situe dans un intervalle de confiance donné. cet

Le résultat de cette forme d'excitation en est une diminution de bourdonnement et une augmentation du naturel de la voix synthétique.

Son inconvénient est que les calculs de  $F_C$  et des filtres sont complexes.

### III.3. Générateur de bruit :

On distingue deux types de générateurs de bruit qui sont :

- Générateur analogique.
- Générateur à séquence numérique.

Le deuxième type présente l'avantage d'être précis, commode et sans grand frais car il est réalisé avec des circuits intégrés numériques d'une grande fiabilité.

#### III.3.1. Principe :

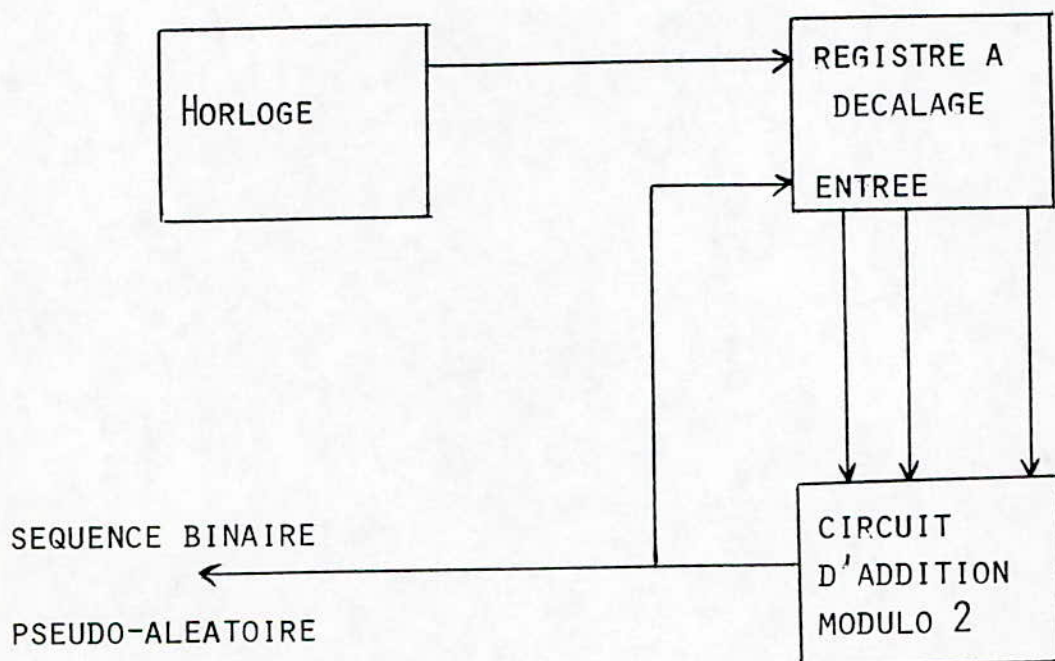


Figure (3.7) : Synoptique d'un générateur de S.P.A.

Le générateur de bruit pseudo-aléatoire ainsi réalisé comprendra :

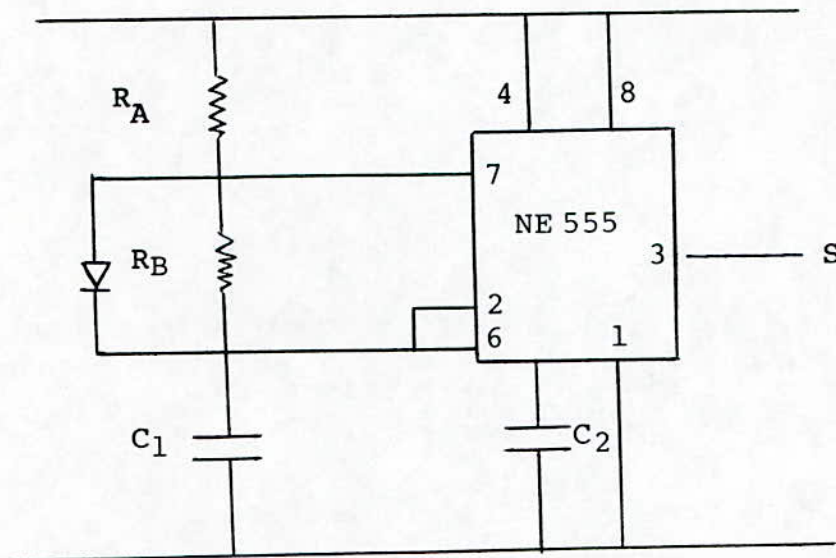
Une horloge, un registre à décalage et un additionneur modulo 2.

En attaquant l'entrée série du registre à décalage par le résultat de l'addition modulo 2 des sorties d'un certain nombre d'étages du registre, on obtient à la sortie une S.P.A.

a) Horloge :

L'horloge permet de générer des signaux carrés pour synchroniser le décalage du registre.

- Montage en multivibrateur :



Fonctionnement :

L'horloge est construite autour de CI NE 555 monté en multivibrateur, la charge de  $C_1$  se fait à travers  $R_A$  et  $R_B$ . Sa décharge à travers  $R_B$  seulement.

Remarque : (1) Pour une bonne symétrie on ajoute une diode.

(2) Le temps de charge de la capacité est donné par :  
$$t_1 = 0,7 (R_A + R_B) \cdot C_1$$

Le temps de décharge est donné par :  
$$t_2 = 0,7 \cdot R_B \cdot C_1$$

La période est  $T = t_1 + t_2 = 0,7 (R_A + 2R_B) \cdot C_1$

b) Régistre à décalage :

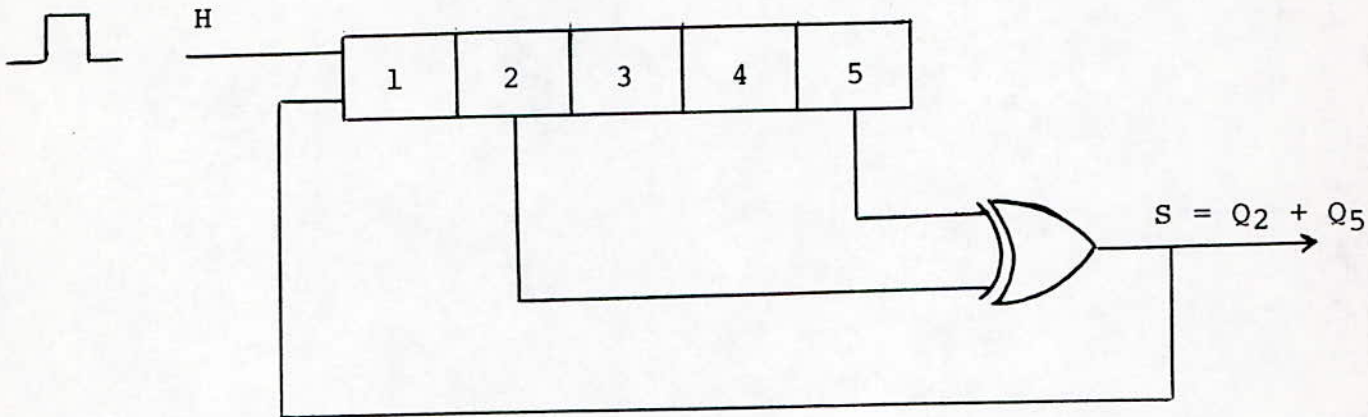
Entièrement intégré c'est le SN 7496 contenant 5 bascules R5.

Il effectue le décalage vers la droite sous la synchronisation des tops d'horloge de l'entrée série qui est la sortie de l'additionneur (Voir schéma synoptique).

c) L'additionneur modulo 2 :

C'est un "OU exclusif" pris dans le SN 7496 (quadripole "ou exclusif" à 2 entrées). Il effectue l'addition binaire des sorties Q<sub>2</sub> et Q<sub>5</sub> du registre à décalage, sortie donnant une séquence de longueur maximale (31 états).

$$S = Q_2 + Q_5$$



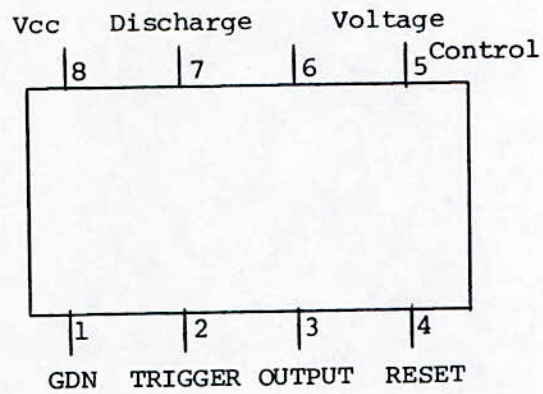
d) l'alimentation :

Elle est de type classique, redressement double alternance. On utilise deux régulateurs intégrés : l'un pour le 12 V dont la stabilité reste recéssaire pour le NE 555 et le CD 4030 l'autre pour l'alimentation en 5 V ducircuit TTL.

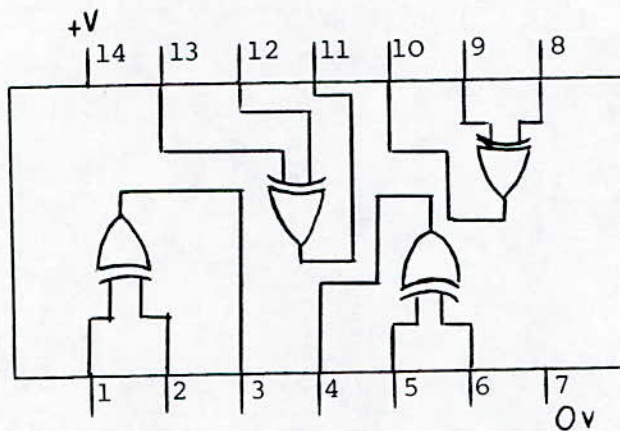
R<sub>1</sub>-C<sub>4</sub> reduisent considérablement le taux d'ondulation de la tension continue de sortie. (Voir schéma ci-dessous).

Brochages des différents circuits intégrés :

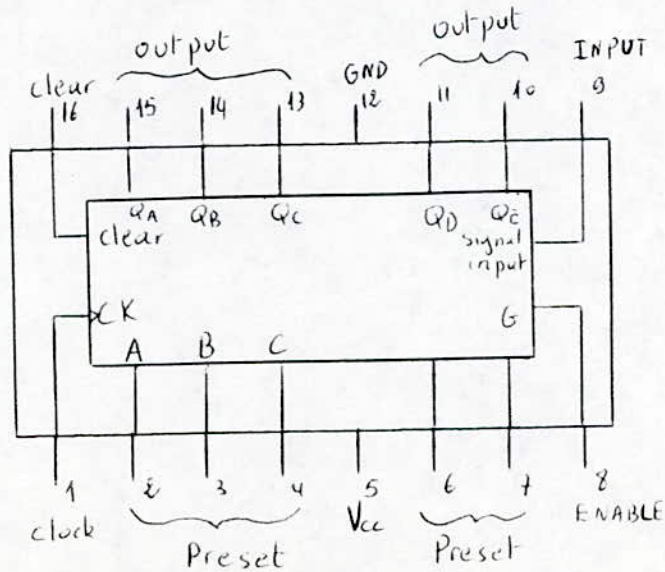
) Le circuit intégré NE 555



) Le circuits intégré CD 4030



) Le circuit intégré SN 7496



III.4. Modélisation du conduit vocal par un filtre en treillis :

La méthode d'Itakura-SAITO consiste à modéliser le conduit vocal par un filtre en treillis de coefficients  $K_i$ , dont la forme la plus courante est celle dite à deux multiplieurs.

C'est un filtre à  $P$  étages ( $P$  est l'ordre du filtre) dans chaque étage, on réalise deux multiplications des signaux par les coefficients de réflexions appropriés.

Considérons maintenant un étage de filtre ( voir figure 3.8.) ; il se présente sous forme d'un quadripôle avec les signaux  $X^+$  et  $X^-$  à gauche,  $Y^+$  et  $Y^-$  à droite ; les signes + et - en indice se réfèrent aux signaux direct et rétrograde respectivement.

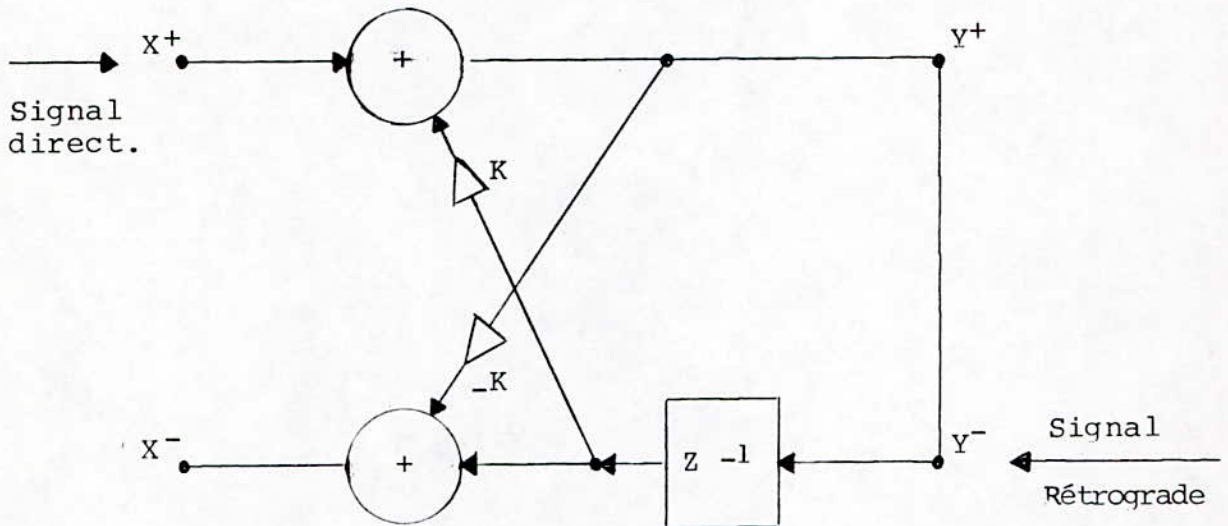


Figure (3.8) : Étage élémentaire du filtre en treillis

Les équations de cet étage s'établissent simplement :

- Parcours direct :  $Y^+ = X^+ + KZ^{-1} Y^-$  ;

Soit encore :  $X^+ = Y^+ - KZ^{-1} Y^-$  ;

- Parcours rétrograde :

$$X^- = -KY^+ + Z^{-1} Y^-;$$

Equations que l'on peut écrire matriciellement :

$$\begin{bmatrix} X^+ \\ X^- \end{bmatrix} = \begin{bmatrix} 1 & -KZ^{-1} \\ -K & Z^{-1} \end{bmatrix} \begin{bmatrix} Y^+ \\ Y^- \end{bmatrix}$$

Dans un mono-étage, les bornes  $Y^+$  et  $Y^-$  sont connectées ensemble et donnent le signal de sortie  $S$ , tandis que par la borne  $X^+$  pénètre le signal d'entrée  $E$  ; l'équation matricielle précédente devient :

$$\begin{bmatrix} E \\ X^- \end{bmatrix} = \begin{bmatrix} 1 & -KZ^{-1} \\ -K & Z^{-1} \end{bmatrix} \begin{bmatrix} S \\ S \end{bmatrix}$$

Soit :  $E = (1 - KZ^{-1}) S$

Où :  $\frac{S}{E} = \frac{1}{1 - KZ^{-1}}$

On trouve ici la fonction de transfert du circuit de synthèse par prédiction linéaire :

$$\frac{S(Z)}{E(Z)} = \frac{1}{1 - \sum_{k=1}^P a(k) Z^{-1}} = \frac{1}{1 - a(1) Z^{-1}} \quad \text{lorsque } P = 1$$

Prenons le cas où l'on a deux étages dans le filtre  
Voir figure (3.9).

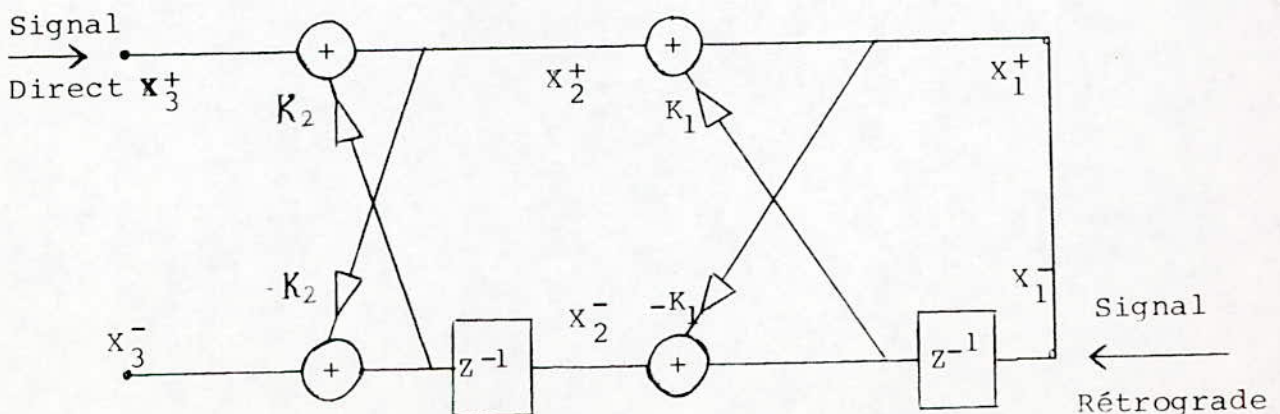


Figure (3.9) : Filtre en treillis à deux étages.

Leur équation est :

$$\text{- Filtre à gauche : } \begin{bmatrix} X_3^+ \\ X_3^- \end{bmatrix} = \begin{bmatrix} 1 & -K_2 z^{-1} \\ -K_2 & z^{-1} \end{bmatrix} \begin{bmatrix} X_2^+ \\ X_2^- \end{bmatrix}$$

$$\text{- Filtre de droite : } \begin{bmatrix} X_2^+ \\ X_2^- \end{bmatrix} = \begin{bmatrix} 1 & -k_1 z^{-1} \\ -K_1 & z^{-1} \end{bmatrix} \begin{bmatrix} X_1^+ \\ X_1^- \end{bmatrix}$$

Le signal d'entrée E pénètre par  $X_3^+$  ; on recueille aux bornes (connectées entre elles)  $X_1^+ - X_1^-$  ; le signal de sortie S.

$$\text{D'où : } \begin{bmatrix} E \\ X_3^- \end{bmatrix} = \begin{bmatrix} 1 & -K_2 z^{-1} \\ -K_2 & z^{-1} \end{bmatrix} \begin{bmatrix} 1 & -K_1 z^{-1} \\ -K_1 & z^{-1} \end{bmatrix} \begin{bmatrix} S \\ S \end{bmatrix}$$

$$\text{Et : } \frac{S}{E} = \frac{1}{1 + K_1 (K_2 - 1) z^{-1} - K_2 z^{-2}}$$

Qui est identifiable à la fonction de transfert :

$$\frac{S(Z)}{E(Z)} = \frac{1}{1 - \sum_{k=1}^P a(k) z^{-k}}$$

Lorsque :  $P = 2$  ;  $a(1) = -k_1 (K_2 - 1)$

$$\text{et } a(2) = K_2$$

Avec un filtre à trois étages, on a de même :

$$\frac{S}{E} = \frac{1}{1 - \sum_{k=1}^P a(k) z^{-k}}$$



Avec :

$$a(1) = - \left[ K_2 K_3 + K_1 (K_2 - 1) \right]$$
$$a(2) = - \left[ K_1 K_3 (1 - K_2) - K_2 \right]$$
$$a(3) = K_3$$

D'une façon générale, l'équation du filtre en échelle à deux multiplieurs d'ordre P s'écrit :

$$\frac{S}{E} = \frac{1}{1 - \sum_{k=1}^P a(k) z^{-k}}$$

Remarque : Il existe d'autres configurations, comme le filtre à un multiplieur et celui à quatre multiplieurs.

### III.5. Détection du voisement et mesure de la période du fondamental dans les cas d'une séquence sonore :

L'oreille humaine est très sensible aux variations du paramètre "Fo". Donc le naturel de la parole synthétique dépend beaucoup de la justesse et de la finesse de l'analyse de "Fo", mais sa détection et sa mesure posent de sérieux problèmes car le signal de parole ne présente pas de périodicité réelle (forme de l'onde d'excitation non constante).

Plusieurs chercheurs se sont penchés sur ce problème mais jusqu'à présent ils n'ont aboutis à aucune solution pleinement satisfaisante.

Parmi les méthodes développées comme exemple :

- La méthode du cepstrum.
- Les méthodes à base d'autocorrelation d'ITAKURA, de SAITO et de MORTEL.
- Méthode d'intercorrelation avec une fonction peigne.

#### III.5.1. Détection du voisement et mesure de la mélodie par la technique simplifiée du filtre inverse :

Pour détecter le voisement et mesurer le fondamental dans le cas d'une séquence sonore notre choix s'est porté sur l'algorithme de SIFT (Markel 1972).

Cette méthode est basée sur le calcul de la fonction d'autocorrelation.

Or ce dernier présente l'inconvénient suivant :

Le signal de parole est un produit de convolution entre le signal quasi-périodique de la source sonore et la réponse impulsionnelle du conduit vocal ( EL MALAWANY ).

Les formants du conduit vocal ont des largeurs de bandes suffisamment étroites ( 50 à 80 HZ ) pour produire des oscillations d'amplitudes élevées dans la fonction d'autocorrelation, or il peut y avoir interférences entre ces oscillations et la composante représentant la période de mélodie, ceci peut conduire à une estimation erronée de cette dernière, car de telles interférences sont principalement causées par le premier formant ( $F_1$ ).

D'où la nécessité d'éliminer la réponse du conduit vocal avant le calcul de la fonction d'autocorrelation.

La prédiction linéaire est le moyen le plus simple pour l'affaiblissement des résonances du conduit vocal (par l'application d'un filtrage inverse).

La convolution entre le signal  $\hat{S}(n)$  obtenu après échantillon et filtrage du signal de parole et le filtrage inverse permet d'obtenir le signal d'erreur  $e(t)$ .

On peut estimer que le signal d'erreur pour les sons non voisés est considéré comme un bruit blanc gaussien de valeur moyenne nulle. Ainsi théoriquement quand la séquence du signal analysé devient grande, la fonction d'autocorrelation  $R(n)$  tend vers zéro pour  $n$  différent de zéro.

Pour une longueur finie de la séquence, on peut déterminer un seuil tel que avec une certaine probabilité et pour un certain intervalle de confiance ; aucun échantillon de la fonction d'autocorrelation ( excepté celui de l'origine ) ne dépasse la valeur de ce seuil ( EL MALAWANY ).

Ce seuil  $\alpha$  qui satisfait la relation suivante :

$$\Pr ( r \leq \alpha ) = 0,01 \beta$$

Est obtenu à partir d'un graphe de la fonction  $\alpha = f(N)$  pour des intervalles de confiance différents.

( Voir figure 3.10 ).

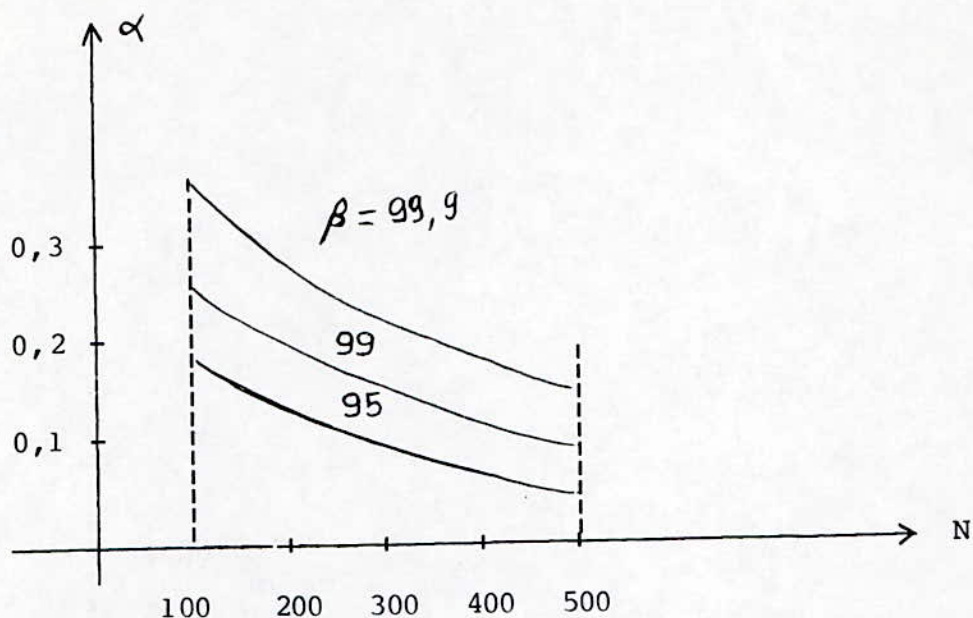


Figure (3.10) : Courbes du seuil  $\alpha$  en fonction du nombre  $N$  d'échantillons du signal pour différentes valeurs de l'intervalle de confiance  $\beta$ .

$N$  : Représente le nombre d'échantillon du signal.

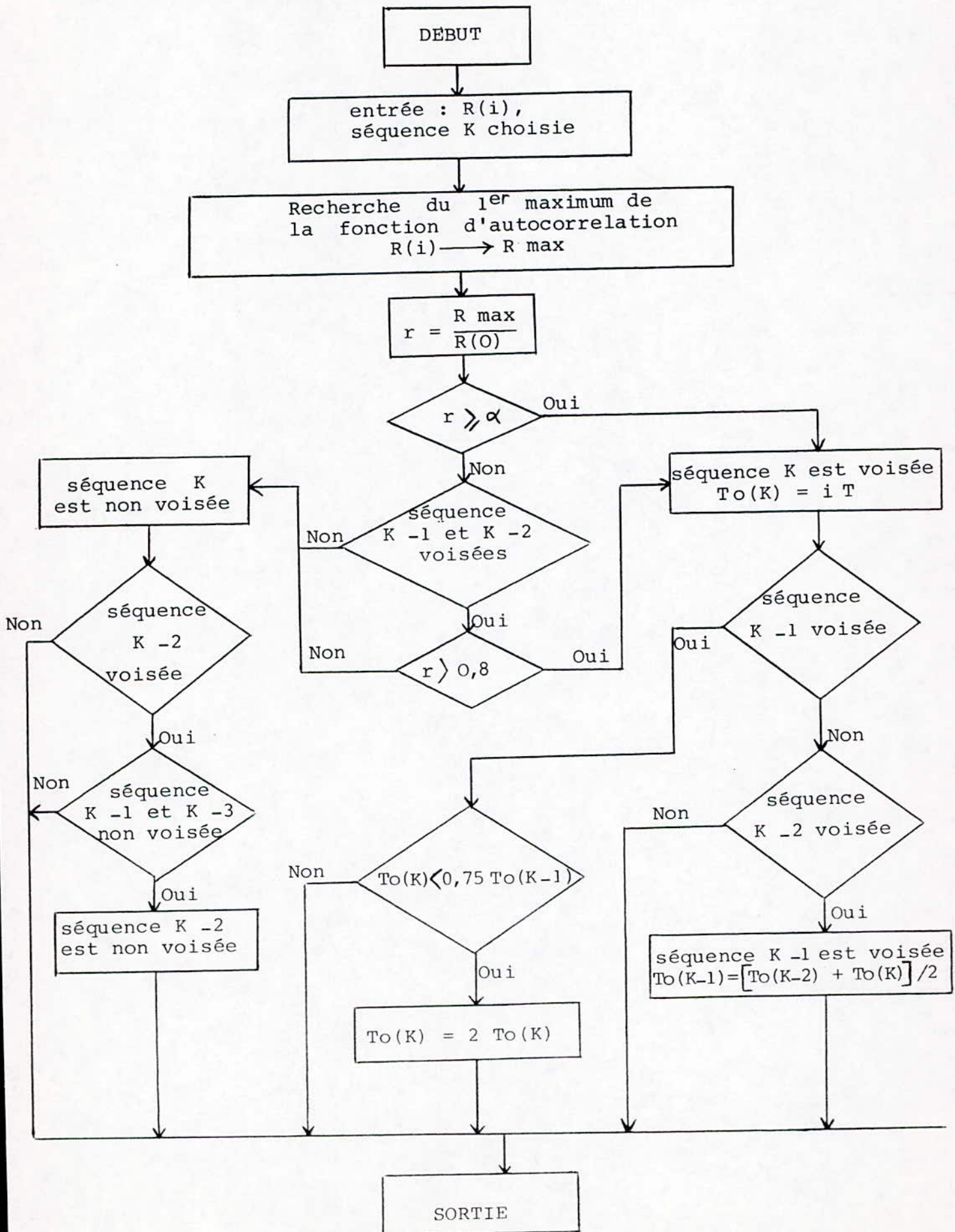
$r = \frac{R(n)}{R(0)}$  avec  $R(n)$  le premier maximum de la fonction d'autocorrelation à partir de  $R(0)$ .

Le graphe de  $\alpha = f(N)$  est valable pour un bruit blanc gaussien ; mais le signal d'erreurs pour les sons non voisés peut être différent d'un bruit gaussien.

Donc afin de ne commettre d'erreur dans la décision voisée / non voisée, on est conduit à relever ce seuil .

Dans le cas d'une fin de segment de parole voisée le seuil  $\alpha$  est pondéré par un facteur 0,8 ; car dans <sup>ce cas</sup> la période du fondamental varie dans une large proportion, par conséquent les périodes du signal sont alors faiblement corrélés, ce qui réduit la valeur de maximum de la fonction d'autocorrelation ( ce maximum correspond à une valeur moyenne de la période de mélodie).

Organigramme de la décision voisé / non voisé.



Commentaire de l'organigramme de la décision voisée/non-voisée :

- Initialisation de l'algorithme avec la valeur de  $r$  et du seuil  $\alpha$ .
- Recherche du premier maximum de la fonction d'autocorrelation.
- On fait un test :

Premier\_cas :  $r$  est supérieur ou égal à  $\alpha$

La séquence  $K$  est dite voisée et la période de mélodie est

$$T_0(K) = n T_e \quad (0 \leq n \leq N-1)$$

avec  $T_e$  : période d'échantillonnage.

Cette décision permet de corriger deux erreurs possibles :

- \*) Si le cadre précédent,  $K-1$  a été déclaré non-voisé à tort, cette erreur peut être corrigée dans la mesure où les cadres  $K-2$  et  $K$  sont voisés, par conséquent, on relève l'anomalie et le cadre  $K-1$  est déclaré voisé avec une période de mélodie qui est la moyenne de celles des deux cadres  $K-1$  et  $K$ .
- \*) Si la période  $T_0(K)$  correspond au premier formant et non à la période du fondamental.

Cette erreur peut être détectée en comparant  $T_0(K)$  à la valeur  $T_0(K-1)$  de la séquence précédente déclarée voisée. Comme la différence entre les périodes de deux séquences consécutives ne dépasse 20%, le test  $T_0(K) < 0,75 T_0(K-1)$  permet de détecter l'erreur et de corriger. La correction sera faite en prenant le double de  $T_0(K)$

Deuxième\_cas :  $r$  est inférieur à  $\alpha$

Si les séquences précédentes  $K$  sont voisées le seuil est pondéré par le facteur 0,8. Si  $r > 0,8$  la séquence  $K$  est déclarée voisée.

La séquence  $K$  est non-voisée dans deux cas :

- Quand  $r$  est inférieur à  $0,8\alpha$ .
- Quand les séquences  $K_{-1}$  et  $K_{-2}$  sont non-voisées.

Pour terminer si les séquences  $K$ ,  $K_{-1}$  et  $K_{-3}$  sont non-voisées alors la séquence  $K_{-2}$  est non-voisée.

## Chpitre IV : LES SYNTHETISEURS A FORMANTS

### IV.1. Introduction :

Le fonctionnement de tous les systhèmes de synthèse quelque soit leur spécificité, peut être décomposé en deux étapes :

- Elaboration, à partir de l'entrée, d'une suite de vecteurs qui represente l'évolution au cours du temps des differents paramètres de commande d'un synthétiseur, chaque vecteur étant constitué par l'ensemble des valeurs des paramètres du synthétiseur, pour une fenêtre temporelle donnée.

- Synthèse du signal de parole à partir des paramètres élaborés au cours de la première étape.

On distingue donc le synthétiseur proprement dit, c'est à dire le modèle acoustique, et la méthode d'élaboration de ses commandes, c'est à dire la méthode de synthèse.

### IV.2. Définition des formants :

Le signal issu de la source vocale est modifié lors de la traversée des cavités du conduit vacal ; ces cavités jouent le rôle de caisse de résonance et dans le spectre sonore, vont renforcer les composantes voisines de leurs fréquences propres. Les maximums de la courbe de réponse en fréquence du conduit vocal sont les "formants" (voir figure 4.1) ; la fréquence du premier formant se situe entre 200 et 900 HZ ; celle du second 500 à 2500 HZ ; le troisième se situe entre 1000 et 3500 HZ. Des formants d'ordre supérieur existent, mais leur rôle sur le plan perceptif est limité, ils contribuent à caractériser la voix.

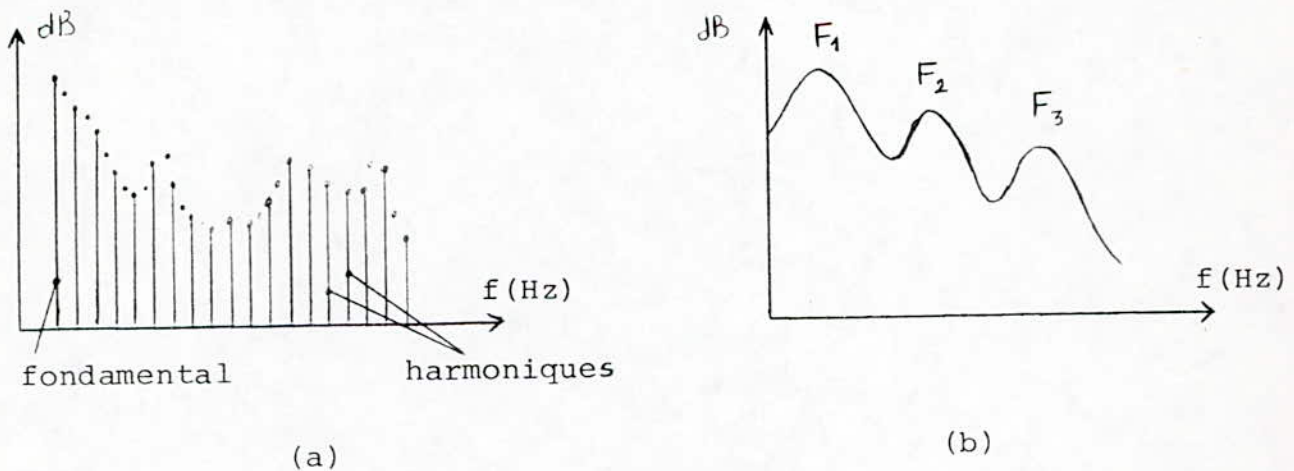


Figure 4.1. :

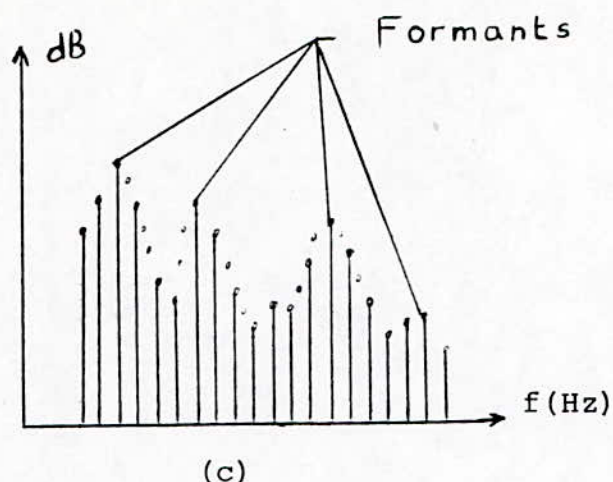


Figure 4.1. : Le signal sonore dont le spectre est en (a), traverse le conduit vocal, dont la courbe de réponse est en (b). On obtient le signal vocal en (c) ; les résonances observées sont les formants.

Lors de la production d'une voyelle, le conduit vocal se déforme très peu et ses fréquences de résonance, donc les fréquences des formants sont stables.

En revanche, la forme du conduit vocal évolue rapidement dans le temps lors de la production de certaines consonnes, les fréquences des formants changent et les transitions sont représentatives de la consonne prononcée.

Les formants des sons "s" ; "ch" ; "f" obtenus à partir d'une source de bruit, se situent vers les hautes fréquences. Au contraire, les consonnes nasales sont caractérisées par un affaiblissement des formants aux fréquences élevées.

Remarque : Pour un même individu, les formants d'une même voyelle se déplacent légèrement ; ils se déplacent énormément et de façon aléatoire lorsque l'exercice d'élocution est répété par un nombre croissant d'individus. De plus, les formants glissent vers l'aigu lorsqu'on passe de l'homme à la femme, puis à l'enfant.



#### IV.3. Les résonateurs :

##### 3.1. Définition de la résonance :

La résonance est la propriété qu'a un système oscillant plus volontier à une certaine fréquence qu'à toutes les autres. Cette fréquence privilégiée est sa fréquence naturelle de vibration appelée : "fréquence de résonance".

Tout système ayant cette propriété est dit résonateur. Il présente une impédance réactive somme d'une inductance et d'une capacitance.

Il existe deux types de résonateurs :

- Les resonateurs mécaniques : qui n'ont pas de lieu avec la phonation.

- Les résonateurs acoustiques : jouent un rôle très important dans la production de la parole.

##### 3.2. Propriétés des résonateurs acoustiques :

Un résonateur est dit parfait s'il ne vibre qu'a une seule fréquence  $f_0$  (voir figure 4.2a)

En général, un résonateur n'est pas parfait.

Il vibre pour une bande de fréquences plus ou moins étroite (voir figure 4.2b). Cette bande est appelée bande passante du "résonateur filtre". La fréquence centrale correspond à une amplitude maximale de vibration. Elle présentera alors la fréquence de résonance du résonateur.

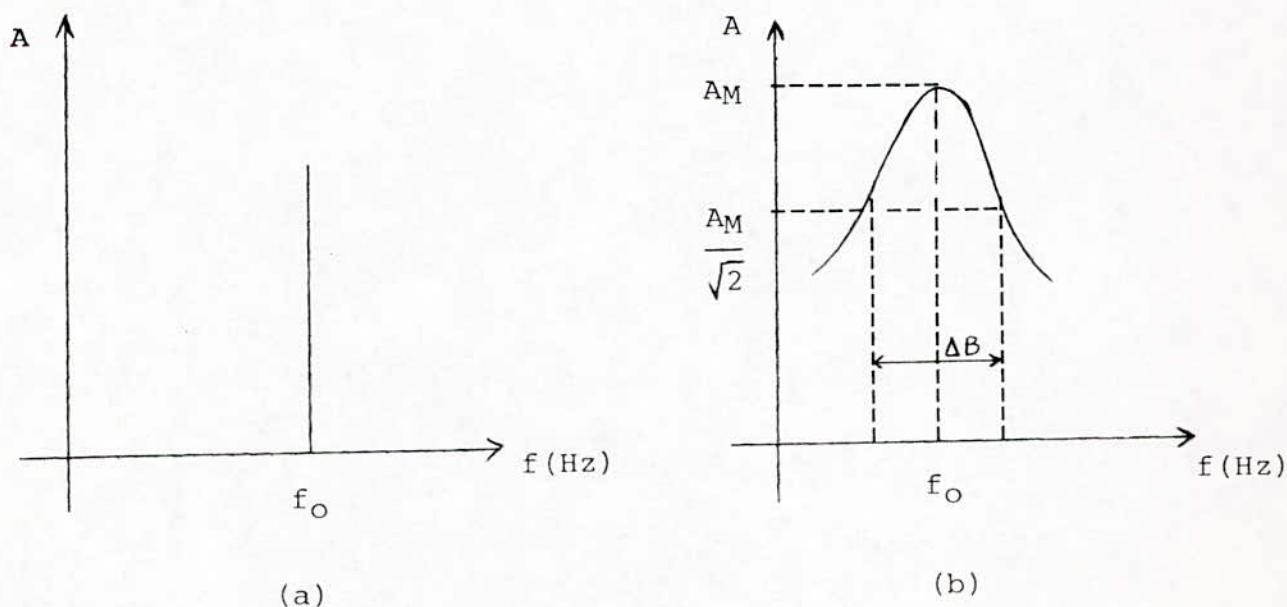


Figure 4.2. : Réponse en fréquence d'un résonateur acoustique.  
(a) parfait. ; (b) non parfait.

### 3.3. Application en filtre sonore :

Un son complexe est un mélange d'harmoniques (s'il est périodique). Alors si on le présente à l'entrée d'un résonateur, il sera filtré. Seuls les harmoniques dont les fréquences sont voisines de celle qui caractérise ce résonateur, seront transmises.

En outre, ce dernier produira une amplification de l'harmonique ayant une fréquence égale à la fréquence de résonance.

Donc, un résonateur peut jouer non seulement un rôle de "filtre" mais aussi "d'amplificateur". Il en est ainsi des différents résonateurs du conduit vocal.

### IV.4. Les synthétiseurs à formants :

Le principe de base du synthétiseur à formant est une simulation des caractéristiques acoustiques du signal de parole, ou plus exactement une simulation de la fonction de transfert du conduit vocal considéré comme l'ensemble "excitateur + résonateur". Dans le synthétiseur à formants, on essaye de reconstituer cette fonction à l'aide d'un assemblage de filtres passe-bas résonnants du second ordre qui simulent les formants.

La paramètre du synthétiseur (à réactualiser toutes les 20 ms environ) sont alors : la commande voisée/non-voisée, la fréquence centrale, la largeur de bande et l'énergie de chacun des formants.

Il existe deux façons principales de connecter entre plusieurs filtres de formants :

- Structure serie.
- Structure parallèle.

#### 4.1. Structure serie :

Les filtres sont placés en cascade (voir figure 4.3) le filtre  $F_2$  ne peut renforcer que le signal délivré par le filtre  $F_1$ .

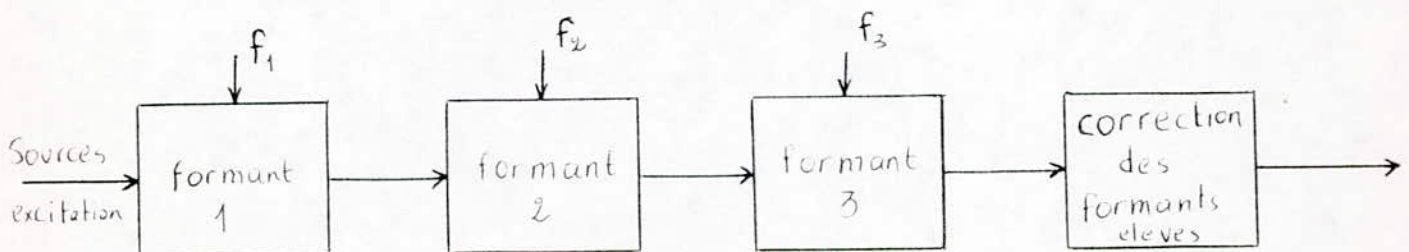


Figure 4.3 : Synthétiseur serie à formants.

La fonction de transfert du conduit vocal s'écrit :

$$H(p) = \frac{w_1^2}{p^2 + b_1.p + w_1^2} \cdot \frac{w_2^2}{p^2 + b_2.p + w_2^2} \dots\dots\dots (4.1)$$

En théorie, le nombre de termes du produit est infini ; chacun d'eux produit un pic dans le spectre d'énergie (voir figure 4.4) : ce sont les "formants".

Considérons maintenant la modélisation d'un formant par la fonction de transfert :

$$H(p) = \frac{w_c^2}{p^2 + bp + w_c^2} \dots\dots\dots (4.2)$$

$w_c$  représente la fréquence nominale de résonance.

Posons  $p = jw$  ; le spectre énergétique est :

$$|H(jw)|^2 = \frac{w_c^4}{\left[ w^2 - \left( w_c^2 - \frac{b^2}{2} \right) \right]^2 + b^2 \left( w_c^2 - \frac{b^2}{4} \right)} \dots\dots\dots (4.3)$$

Il passe par un maximum ; lorsque  $w$  varie et atteint la valeur :

$$\left( w_c - \frac{b^2}{2} \right)^{1/2} \quad \text{pour laquelle le premier terme du dénominateur s'annule.}$$

En général la largeur de bande  $b$  est petite devant la valeur de la fréquence centrale  $w_c$ , de sorte qu'une approximation du pic est  $w = w_c$  et son amplitude est alors égale à  $w_c/b$ .

A 3dB au-dessous du pic, l'énergie vaut :

$$|H(jw_{3dB})|^2 = \frac{1}{2} |H(jw_{max})|^2 = \frac{w_c^2}{2b^2} \dots\dots\dots (4.4)$$

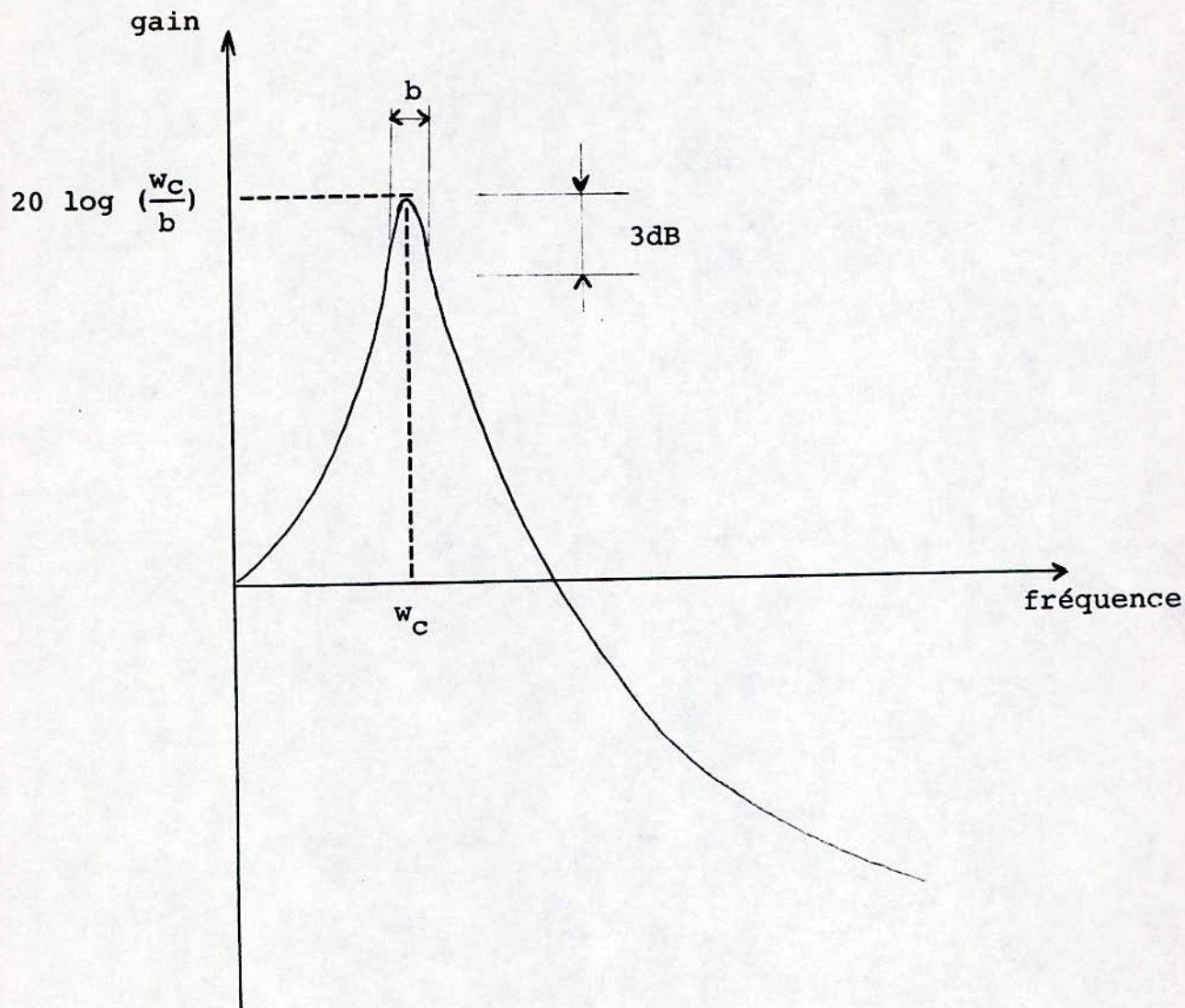


Figure (4.4) : Spectre énergétique d'un formant pris individuellement.

Et la valeur correspondante de la fréquence est, avec une bonne approximation.

$$w_{3dB} = w_c \pm \frac{b}{2} \dots\dots\dots (4.5)$$

La largeur de bande  $b$  d'un formant est quasiment constante quelle que soit la fréquence du formant.

4.2. Structure parallèle :

Les filtres sont tous alimentés par le même signal, leurs sorties sont additionnées. (Voir figure 4.5).

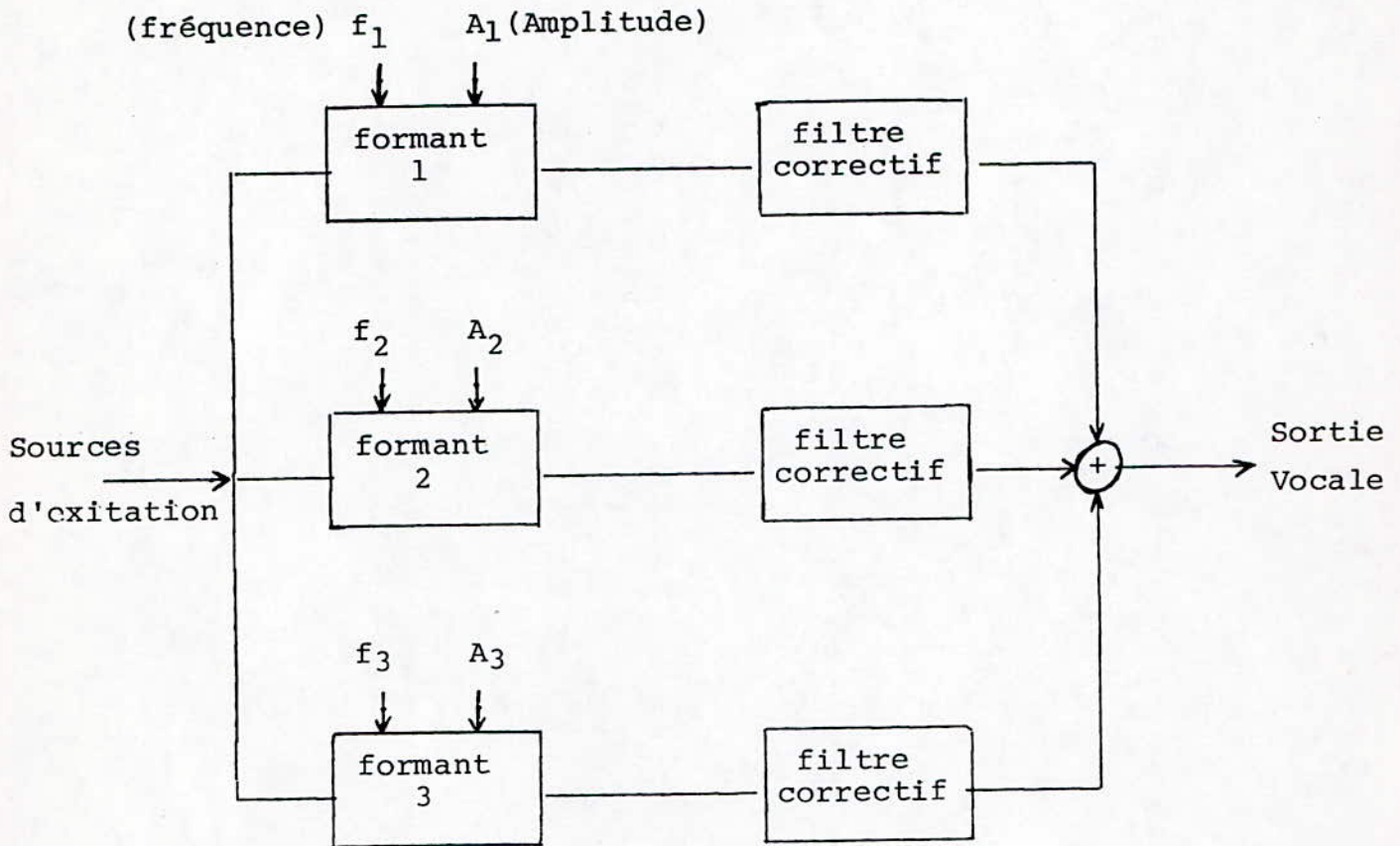


Figure 4.5. : Synthétiseur à formants.

On associe à chaque filtre de formant, un filtre de correction à faible gain afin d'éviter que de grandes variations d'amplitude se produisent.

La fonction de transfert de deux filtres du second ordre mis en parallèle est :

$$H(p) = \frac{A_1 \omega_1^2}{p^2 + b_1 p + \omega_1^2} + \frac{A_2 \omega_2^2}{p^2 + b_2 p + \omega_2^2} \dots \dots \dots (4.6)$$

IV.5. Comparaison entre les deux structures :

Le premier problème important, lorsque l'on détermine une structure de synthétiseur à formants, est de choisir entre structure série et structure parallèle. Ces deux structures s'opposent sur plusieurs points.

D'abord la structure série est très bien adaptée à la production des voyelles non nasalisées, dans les hautes où la décroissance de leur spectre vers les hautes fréquences est naturellement simulée par la mise en cascade des filtres de formants (FANT, 1956). La production des voyelles est un peu moins bonne avec une structure, car les interactions entre formants voisins ne sont pas complètement inexistantes, malgré les précautions prises. Toutefois la qualité obtenue est satisfaisante.

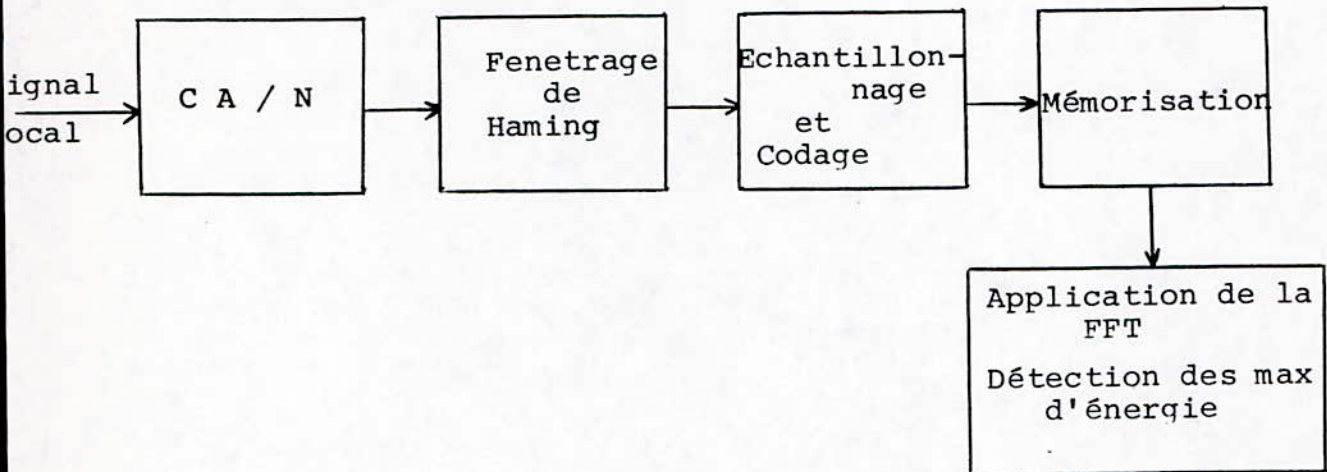
Une autre opposition réside dans le nombre des paramètres à définir :

Pour la structure série, chaque formant est défini par sa fréquence et sa bande passante, alors que pour la structure parallèle, il faut également spécifier le gain du formant. Cela pose des problèmes à l'analyse, car il y a un plus grand nombre de paramètres à élaborer, mais cela permet d'obtenir une souplesse d'ajustement inexistante avec la structure série. De plus, pour une structure parallèle, une erreur dans la détermination d'un formant n'affecte pas les autres formants et l'ensemble du spectre, contrairement à ce qui se passe structure série. (Holmes, 1973).

Enfin, dans le cas de la mise en cascade des filtres, les "bruits de calcul" liés à la précision finie des calculateurs ont tendance à se cumuler le long de la chaîne, tandis que statiquement, ils ont tendance à s'annuler par sommation au niveau de la sortie d'une structure parallèle.

Détermination des trois premiers formants et de la fréquence fondamentale à partir d'un signal sonore.

Schéma de principe :



- Le signal vocal est convertis en un signal numérique par un convertisseur analogique/numérique.
- Pour sélectionner la zone à étudier on applique la fenêtre de Haming qui est

$$Q(t) = 0,54 + 0,46 \cos \left( \pi \frac{t}{NT} \right)$$

- Cette dernière fonction a la grande partie de son énergie concentrée dans le lobe principal.

T : période d'échantillonnage.

N : le nombre d'échantillons.

- Le signal ainsi convertis est échantillonné puis codé.
- Une fois le transfert de toutes les données est effectué, le signal codé est stocké en mémoire.
- On observera sur l'écran le signal analysé.

A la fin on fait appelle au programme de la "FFT" qui est implanté dans le micro-ordinateur pour avoir le spectre du signal. Les maximums d'énergies nous donnent les formants correspondants. Comme exemple voir pages : 58, 59, et 60.

Deuxième Partie : PROGRAMMATION

Nous présentons ici trois programmes :

- ) Pour les deux premières on a utilisé le micro OLIVETTI M 24.
- ) Le troisième a été fait avec le TEKTRONIX 4052 à l'institut linguistique de BEN-AKNOUN.

Remarque :

Programme 1 : "Calcul des coefficients de prédiction pour la méthode d'autocorrelation" (Selon l'algorithme de DURBIN).

Programme 2 : "Detection du voisement et mesure du fondamental" (Méthode S.I.F.T.)

Programme 3 : "Mesure du fondamental et des trois premiers formants en utilisant la FFT".



```
2 REM *** CALCUL DES COEFFICIENTS DE PREDICTIONS A(I)***
4 REM *** PAR LA METHODE D'AUTOCORRELATION***
6 PRINT "CALCUL DES COEFFICIENTS PAR LA METHODE D'AUTOCORRELATION"
10 PRINT "ENTREZ LA VALEUR DE N"
20 INPUT N
30 PRINT "ENTREZ LA VALEUR DE P"
40 INPUT P
50 DIM S(N)
72 OPEN "I",#1, "c:DATA"
74 FOR I=1 TO N
76 INPUT #1, S(I)
77 NEXT
78 PRINT
90 DIM R(P)
100 R(0)=0
110 FOR J=1 TO N
120 H=S(J)^2
130 R(0)=R(0)+H
140 NEXT J
150 PRINT "R(0)=";R(0)
160 H=0
170 FOR I=1 TO P
180 R1=0
190 FOR J=1 TO N-I
200 H=S(J)*S(J+1)
210 R1=R1+H
230 NEXT J
240 R(I)=R1
250 PRINT "R(";I;")=";R(I)
260 NEXT I
265 REM RECHERCHE DES COEFFICIENTS DE REFLEXION K(P)
270 DIM E(P), K(P), A(P,P)
280 E=0
290 K=0
300 A=0
310 E(0)=R(0)
320 LIST
333 A(1,1)=K(1)
340 PRINT "K(1)=";K(1);"A(1,1)=";A(1,1);"E(0)=";E(0)
350 E(1)=(1-K(1)^2)*E(0)
360 PRINT "E(1)=";E(1)
370 FOR I=2 TO P
380 T=0
390 FOR J=1 TO I-1
400 T=T+A(J,I-1)*R(I-J)
410 NEXT J
420 K(I)=(R(I)-T)/E(I-1)
430 PRINT "K(";I;")=";K(I)
```

```
435 REM RECHERCHE DES COEFFICIENTS DE PREDICTION A(J,I)
440 A(I,I)=K(I)
450 FOR J=1 TO I-1
460 A(J,I)=A(J,I-1)-K(I)*A(I-J,I-1)
470 PRINT "A(";J;",";I;")=";A(J,I)
480 NEXT J
485 REM RECHERCHE DE L'ERREUR QUADRATIQUE E(P)
490 E(I)=(1-K(I)^2)*E(I-1)
500 PRINT "E(";I;")=";E(I)
510 NEXT I
515 PRINT "VOULEZ VOUS ESSAYER D'AUTRES VALEURS?"
520 PRINT "oui ou non"
530 INPUT M$
540 IF M$="OUI" THEN GOTO 10
545 IF M$="oui" THEN GOTO 10
547 IF M$="NON" GOTO 560
548 GOTO 515
560 END
```

```
10 REM      ***DETECTION DU VOISEMENT ***
20 REM      *** MESURE DU FONDAMENTAL ***
40 REM *** LECTURE DES COEFFICIENTS D'AUTOCORRELATION ***
50 REM      *** A PARTIR D'UN FICHER ***
60 PRINT "ENTRER LA VALEUR DU SEUIL ALPHA"
70 INPUT ALPHA
80 PRINT "ENTRER LE NOMBRE DE TRAME K"
90 INPUT N
100 PRINT "ENTRER LE NOMBRE DE PREDICTIONS P"
110 INPUT P
120 PRINT "ENTRER LA PERIODE D'ECHANTILLONNAGE T"
130 INPUT T
131 K=1
132 REM *** OUVERTURE DU FICHER DE DONNEES ***
140 OPEN "I",#1,"AUTOCOR"
150 FOR I=0 TO P
160 INPUT #1 , R(I)
170 NEXT
180 GOSUB 540
200 RP=(RMAX / R(0))
210 IF RP >= ALPHA GOTO 370
220 IF K<=2 GOTO 700
230 REM **CARACTERISTIQUE DE LA SEQUENCE K **
240 REM**IF F(K)=1 SEQUENCE VOISEE **
250 REM **IF F(K)=0 SEQUENCE NON VOISEE **
260 IF F(K-1) = 1 AND F(K-2) = 1 THEN GOTO 280
270 GOTO 290
280 IF RP>=ALPHA*.8 GOTO 370
290 PRINT "LA SEQUENCE",K,"EST NON VOISEE"
300 IF F(K-2)=1 GOTO 330
310 GOTO 630
320 IF K=3 GOTO 640
330 IF F(K-1)=0 AND F(K-3)=0 GOTO 350
340 GOTO 630
350 PRINT "LA SEQUENCE",K-2,"EST NON VOISEE"
360 GOTO 630
370 PRINT "LA SEQUENCE",K,"EST VOISEE"
380 TO(K)=J*T
390 REM *** MESURE DU FONDAMENTAL ***
400 PRINT "LE FONDAMENTAL EST TO(",K,")=" ,TO(K)
410 F(K)=1
420 IF K<=2 GOTO 640
430 IF F(K-1)=1 GOTO 460
440 IF F(K-2)=1 GOTO 500
450 GOTO 630
460 IF TO(K)<.75*TO(K-1) THEN 480
470 ELSE 630
480 TO(K)=2*TO(K)
490 GOTO 630
500 PRINT "LA SEQUENCE ",K-1 , "EST VOISEE"
510 TO(K-1)=(TO(K-2)+TO(K))/2
520 PRINT "TO(",K-1,")=" ,TO(K-1)
530 GOTO 630
```

```
535 REM *** RECHERCHE DU PREMIER MAXIMUM A PARTIR DE R(0) ***
540 J=1
550 RMAX=R(J)
560 FOR I=J+1 TO P
570 IF R(J)<R(I) GOTO 600
580 NEXT
590 RETURN
600 RMAX=R(I)
610 J=I
620 GOTO 560
630 IF K=N THEN GOTO 660
640 K=K+1
650 GOTO 150
660 END
700 PRINT "LA SEQUENCE",K,"EST NON VOISEE"
710 GOTO 640
```

PROGRAMME D'ANALYSE SPECTRALE FFT

```
100 INIT
102 ON SRQ THEN 105
105 PAGE
106 PRINT "JJJ "
107 PRINT
110 POLL A1 ,A2,A3
120 DIM A$(5500),B$(6),A(1024),M9(513),M8(513)
      B(1024),C(256),A5(1024)
130 DIM Y2(20)
140 A=0
150 PRINT "JJ TRANSFERT DE DONNEES";
160 I=1
170 R8=100
180 M=1
190 PRINT @2 : "ACCES LEFTS; ASCII; CURVE"
200 INPUT @2:A$
210 A(1)=VAL(A$)
220 P=0
230 E=100
240 E9=0
250 FOR I=1 TO 1015
260 P=POS(A$,"",P+1)
270 B$=SEG(A$,P,6)
280 A(I+1)= VAL(B$)
290 NEXT I
295 PRINT "GII TRANSFERT TERMINE"
296 PRINT "JJ ENTRER LE NOM DU LOGATOME:II";
297 INPUT L$
300 PRINT "J BASE DE TEMPS DE L'OSCILLO E:II";
310 INPUT B9
320 F9=1/(10*B9)
330 F5=1016/(B9*10)
340 PRINT "J FREQUENCE D'ECHANTILLONNAGE:II";F5
350 PRINT "J FREQUENCE MAX DE VISUALISATION DU SPECTRE:I";
360 INPUT F8
370 PRINT "J SPECTRE EN ECHELLE (LIN) OU (LOG):II";
380 INPUT Y$
390 PRINT "J TABLE TRACANTE (OUI=O, NON=N):I";
400 INPUT Q$
410 PRINT "J TABLE TRACANTE (OUI=O, NON=N):II";
420 INPUT T$
430 PAGE
440 S=SUM(A)
450 B=A
460 A=A-A/1024
470 VIEWPORT 10,130,0,45
480 WINDOW 0,130,0,100
485 MOVE 5,90
486 PRINT "LOGATOME:L".L$;"I"
490 MOVE 95,90
500 PRINT "SIGNAL ANALYSE"
510 WINDOW 1,E, _4*E, 4*E
520 AXIS 0,0,1,-4*E
530 AXIS 0,0,0,4*E
540 0,0,0,4*E+1
550 MOVE 1,A(1)
560 FOR J=2 TO I
570 DRAW J, A(J)
580 NEXT J
590 IF T$ <> "0" THEN 720
595 VIEWPORT 0,150,0,45
```

```
600 WINDOW 0,50,0,100
605 PRINT @1,17:1.2,2
607 MOVE @1,5,90
608 PRINT @1:"LOGATOME :[ ";L$;,"J"
610 MOVE @1:125,90
620 PRINT @1:"SIGNAL ANALYSE"
630 WINDOW 1,1,-4+E
640 AXIS @1:0,0,1,-4+E
650 AXIS @1:0,0,0,4+E
660 AXIS @1:0,0,1,4+E+1
670 MOVE @1:1,A(1)
680 FOR J=2 TO I
690 DRAW @1:J,J(A)
700 NEXT J
710 AS=A
720 CALL "TAPER" ,A,0.25
730 CALL "FFT",A
740 CALL "POLAR",A,M9,M8
750 CALL "MAX" M9,M7,18
760 CALL "MIN" M9,M6,18
770 IF Q$="0" THEN 870
780 IF Y$="LIN" THEN 870
790 U9=1 TO 513
800 G= M9(U9)/M7
810 M9(U9)=LGT(G)*10
820 NEXT U9
830 E9=10
840 C6=0
850 Y$="LIN"
860 GO TO 750
870 F7=F8/F9
880 IF F7<512 THEN 910
890 PRINT "GGG"
900 F7=512
910 F6=(F7-1)/10
920 VIEWPORT 10,130,55,100
930 WINDOW 0,130,0,100
940 MOVE 95,90
950 PRINT "SPECTRE DU SIGNAL"
960 IF Q$="0" THEN 190
970 MOVE 95,82
980 PRINT "(SUR";F7;" PIONTS)"
990 X9=M6-(M7*1,5-M6)/10
1000 IF E9=10 THEN 1030
1010 E9=(M7-M6)/5
1020 C6=M6
1030 WINDOW 1,F7,X9,M7
1040 AXIS F6,E9,1,C6
1050 FOR J=1 TO F7
1055 MOVE J,M6
1060 DRAW J,M9(J)
1070 NEXT J
1080 WINDOW 1,F7,0,100
1090 Z9=-F6
1100 FOR K=0 TO 9
1110 Z9=Z9+F6
1120 X$=STR(F8/10*K)
1130 MOVE Z9,1
1140 PRINT X$
1150 NEXT K
1160 IF E9< > 10 THEN 1280
1170 WINDOW 1,F7,X9,M7
1180 R8=1
```

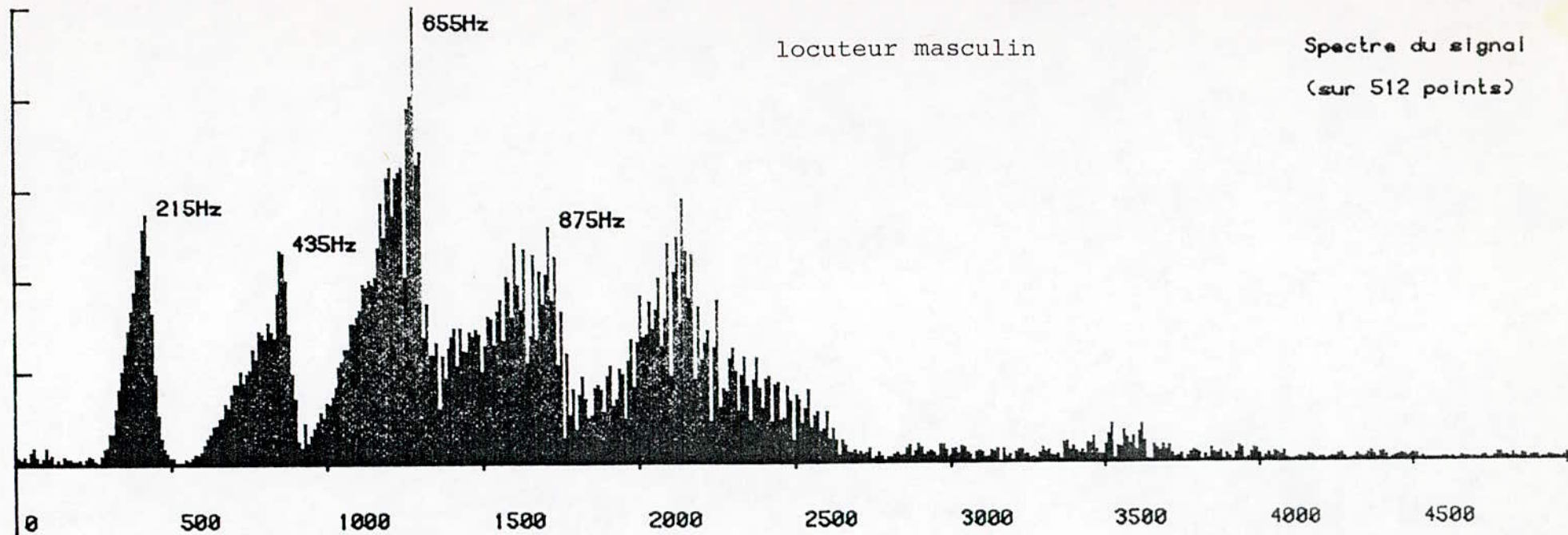
```
1190 AXIS 0,E9,F7,M7*1.2
1200 AXIS F6,0,1-F6,M6
1210 VIEWPORT 1,10,55,100
1220 WINDOW 1,F7,X9,M7
1230 FOR W6=0 TO M6 STEP-10
1250 MOVE 5,W6
1260 PRINT W6
1270 NEXT W6
1280 GO TO 2890
1290 IF T#< >'0' THEN 1750
1300 A=A5
1310 VIEWPORT 10,150,55,100
1320 WINDOW 0,150,0,100
1330 MOVE @1:125,90
1340 PRINT @1:'SPECTRE DU SIGNAL'
1350 IF Q#="0" THEN 2410
1360 MOVE @1: 125,82
1370 PRINT @1:'(SUR";F7;POINTS)'
1380 X9=M6-(M7*1.5-M6)/10
1390 IF E9=10 THEN 1420
1400 E9=(M7-M6)/5
1410 C6=M6
1420 WINDOW 1,F7,X9,M7
1430 AXIS @1,F6,E9,1,C6
1440 FOR J=1 TO F7
1445 MOVE @1:J, M6
1450 DRAW @1:J,M9(J)
1460 NEXT J
1470 WINDOW :1,F7,0,100
1480 Z9=-F6
1490 FOR K=0 TO 9
1500 Z9=Z9+F6
1510 X#=STR(F8/10*K)
1520 MOVE @1:Z9,1
1530 PRINT @1:X#
1540 NEXT K
1550 IF E9 < > 10 THEN 1670
1560 WINDOW 1,F7,X9,M7
1570 R8=1
1580 AXIS @1:0,E9,F7,M7*1.2
1590 AXIS @1:F6,0,1-F6,M6
1600 VIEWPORT 1,10,55,100
1610 WINDOW 1,F7,X9,M7
1620 FOR W6=0 TO M6 SEP-10
1630 MOVE @1:5,W6
1640 PRINT @1:W6
1660 NEXT W6
1670 WINDOW 1,F7,X9,M7
1680 FOR L=1 TO J7
1690 S6=0
1700 IF M9(Y2(L)) <M7> THEN 1720
1710 S6=M7*0.05
1720 MOVE @1:Y2(L),M9(Y2(L))-S6
1730 PRINT @1:F(L):'HZ'
1740 NEXT L
1750 INPUT R#
1760 PAGE
1770 PRINT 'J DEBUT DU PROGRAMME (OUI=0;NON=N):';
1780 INPUT R#
1790 IF R#< >'0' THEN 1820
1800 A=B
1810 GO TO 370
1820 PAGE
```

```
1830 PRINT "J AFFICHAGE DES FORMANTS (OUI=O;NON=N) : ";
1840 INPUT K$
1850 IF K$="N" THEN 1900
1860 PRINT "J FREQUENCE FONDAMENTALE : "; F(1); "HZ"
1870 PRINT "J PREMIER FORMANT : "; F(2); "HZ"
1880 PRINT "J DEUXIEME FORMANT : "; F(3); "HZ"
1890 PRINT "J TROISIEME FORMANT : "; F(4); "HZ"
1900 END
1910 SET DEGREES
1920 T9=10*B9
1930 D9=T9/1016*256
1940 F9=1/D9
1950 F7=F8/F9
1960 F6=(F7-1)/10
1970 U=0
1980 C1=0
1990 C2=0
2000 FOR W=0 TO 750 STEP 50
2010 U=0
2020 T=0
2030 FOR W1=1+W TO 256+W
2040 U=U+1
2050 C(U)=B(W1)
2060 NEXT W1
2070 S=SUM(C)
2080 C=C-S/256
2090 DIM P8(129),P9(129)
2100 CALL "TAPER",C,0,25
2110 CALL "FFT",C
2120 CALL "POLAR",C,P9,P8
2130 CALL "MAX",P9,P7,18
2140 CALL "MIN",P9,P6,18
2150 C1=C1+0.5
2160 D=C1*COS(60)
2170 C2=C2+1.4
2180 H=C2*SIN(60)
2190 VIEWPORT 20+D,100+D,50+H,100+H
2200 IF W>0 THEN 2340
2210 WINDOW 1,F7,M6-(M7-M6)/16,M7
2220 AXIS F6,0,1-F6,M6
2230 Z9=-F6+1
2240 FOR K=0 TO 9
2250 Z9=Z9+F6
2260 IF Z9<1 THEN 2290
2270 Z7=2
2280 GO TO 2300
2290 Z7=1
2300 MOVE Z9-Z7,M6-(M7-M6)/10
2310 X$=STR(F8/10*K)
2320 PRINT X$
2330 NEXT K
2340 MOVE 1,P9(1)
2350 FOR U=2 TO 128
2360 (9)
2360 DRAW U,P9(U)
2370 NEXT U
2380 NEXT W
2390 IF T$="O" THEN 1310
2400 END
2410 SET DEGREES
2420 T9=10*B9
2430 D9=T9/1016*256
2440 F9=1/D9
```



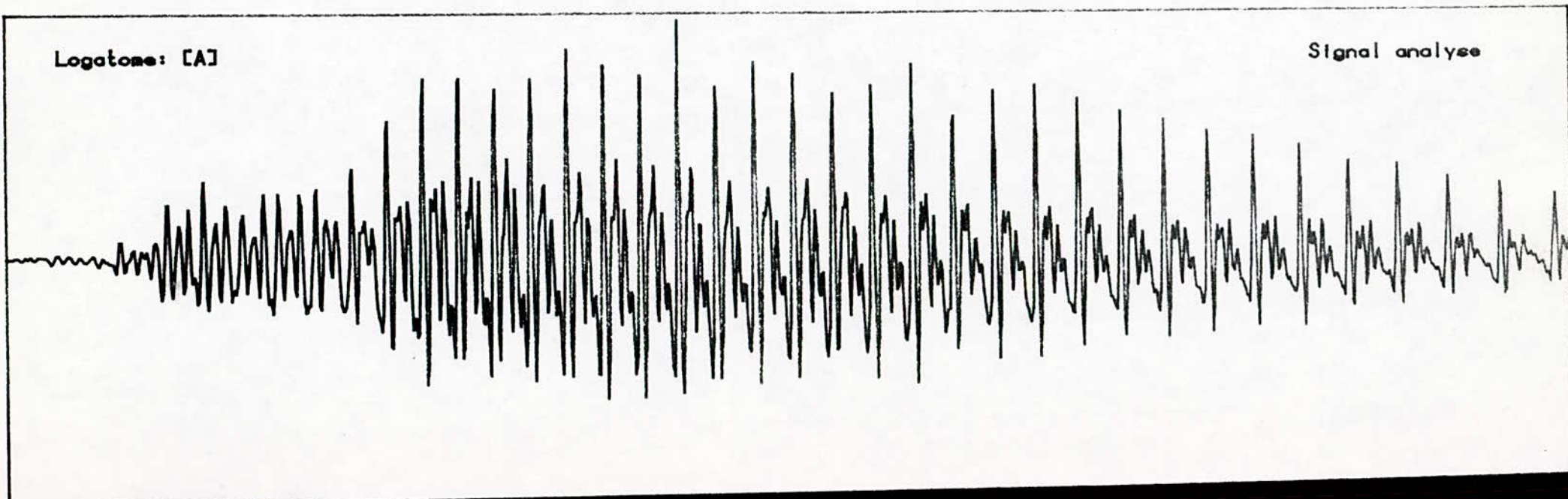
```
2450 F7=F8/F9
2460 F6=(F7-1)/10
2470 U=
2470 U=0
2480 C1=0
2490 C2=0
2500 FOR W=0 TO 750 STEP 50
2510 U=0
2520 T=0
2530 FOR W1=1+W TO 256+W
2540 U=U+1
2550 C(U)=B(W1)
2560 NEXT W1
2570 S=SUM(C)
2580 C=C-S/256
2590 CALL 'TAPER',C ,0,25
2600 CALL ''FFT' ',C
2610 CALL 'POLAR',C,P9,P8
2620 CALL 'MAX',P7, P9,18
2630 CALL 'MIN',P9,P6,18
2640 C1=C1+0.5
2650 D=C1*ACOS(60)
2660 C2=C2+1.4
2670 H=C2*SIN(60)
2680 VIEWPORT 20+D,100+D,50+H,100+H
2690 IF W>0 THEN 2830
2700 WINDOW 1,F7,M6-(M7-M6)/10,M7
2710 AXIS @1:F6,0,1-F6,M6
2720 Z9=-F6+1
2730 FOR K=1 TO 9
2740 Z9=Z9+F6
2750 IF Z9<=1 THEN 2780
2760 Z7=2
2770 GO TO 2790
2780 Z7=1
2790 MOVE @1:Z9-Z7,M6-(M7-M6)/10
2800 X#=STR(F8/10*K)
2810 PRINT @1:X#
2820 NEXT K
2830 MOVE @1,P9<1>
2840 FOR U=2 TO 128
2850 DRAW @1:U,P9(U)
2860 NEXT U
2870 NEXT W
2880 END
2890 WINDOW 1,F7,X9,M7
2900 IF E9< >10 THEN 2920
2910 VIEWPORT 10,130 ,55,100
2920 WINDOW 1,F7,X9,M7
2930 DIM(20)
2930 DIM F(20)
2940 J7=0
2950 F=0
2960 FOR I7=1 TO F7
2970 MOVE I7 ,M9(17)
2980 RDRAW 1,1*RB
2990 RDRAW -2,0
3000 RDRAW 1,-1*RB
3010 RDRAW 0,4*RB
3020 INPUT R#
3030 IF R#< >'S' THEN 3070
3040 J7=J7+1
3050 F(J7)=(I7-1)*F9
```

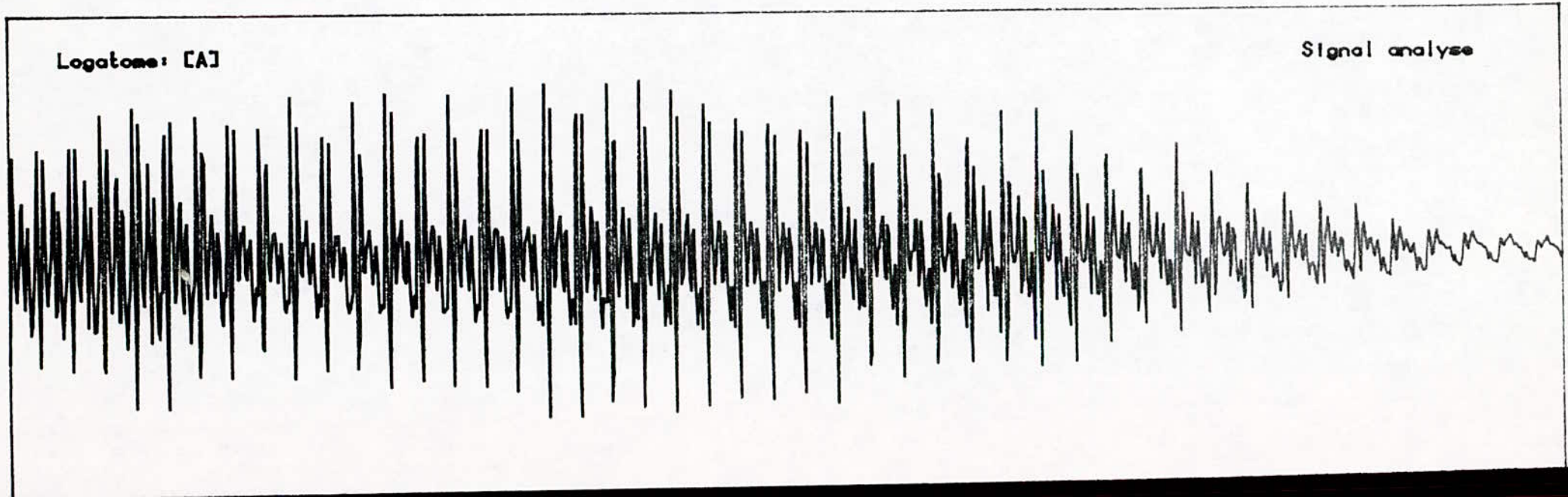
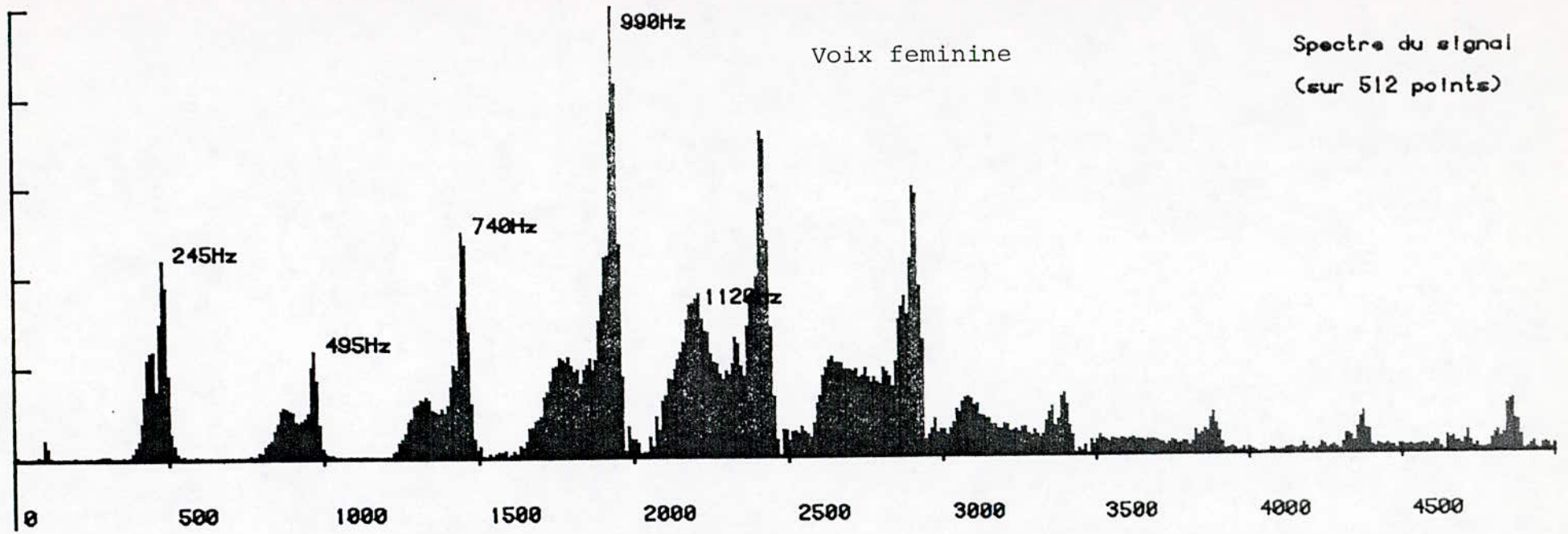
```
3060 Y2(J7)=I7
3070 IF R4="F" THEN 3090
3080 NEXT I7
3060 Y2(J7)=I7
3090 FOR L=1 TO J7
3100 MOVE Y2(L),M9(Y2(L))-5
3110 PRINT F(L); "HZ"
3120 NEXT L
3130 GO TO 1290
```

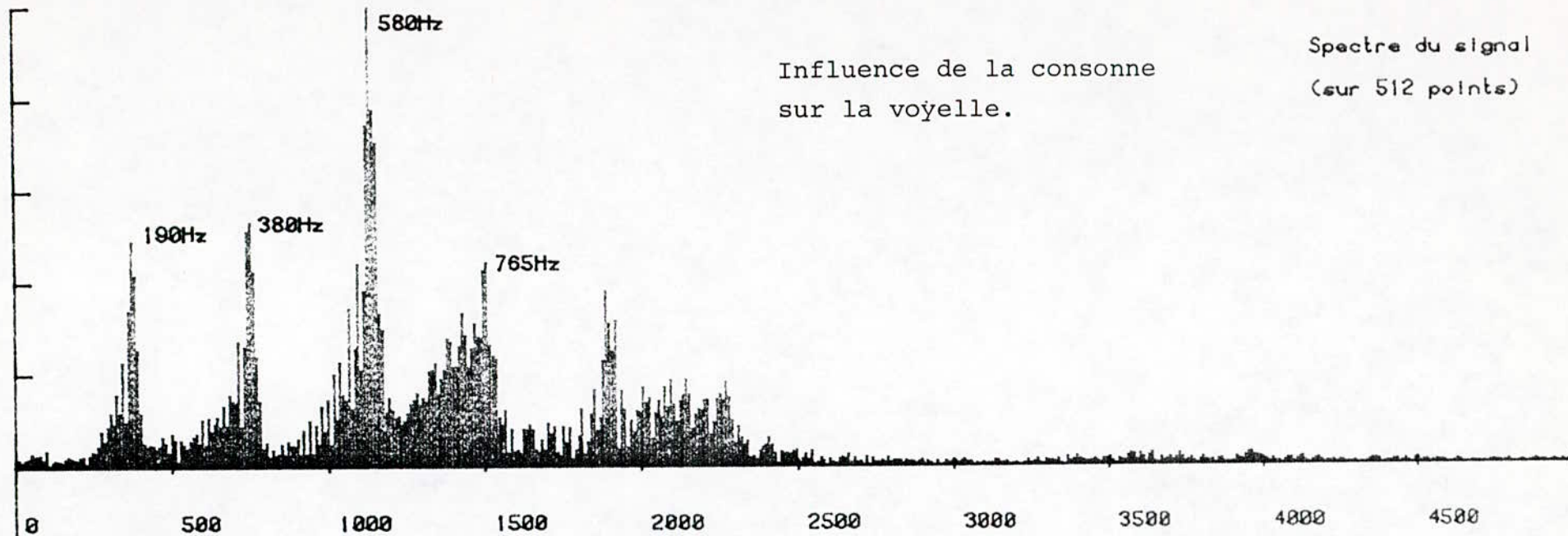


Logatome: [A]

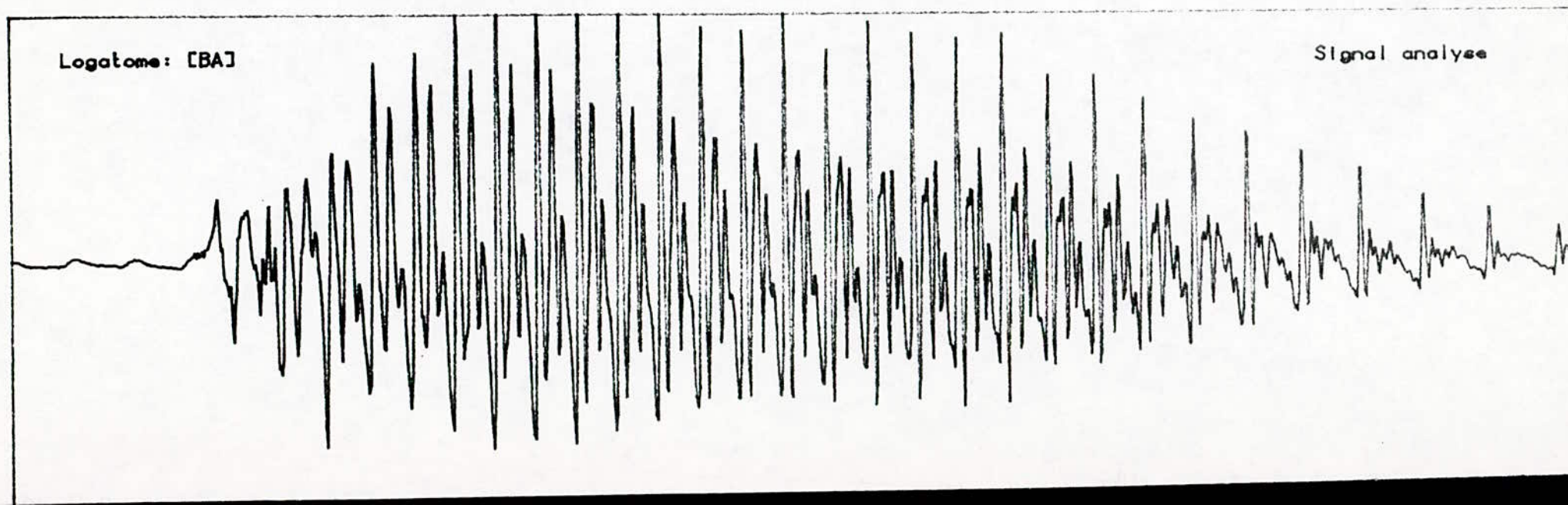
Signal analyse







- 60 -



Conclusion :

Ce travail nous a permis d'élargir nos connaissances dans le domaine du traitement du signal de la parole qui est actuellement un nouveau foyer de recherche.

L'objectif du travail présenté dans ce mémoire était la mise en place de bonne qualité passant par l'élaboration d'une méthode efficace, basée sur une représentation très simplifiée du mécanisme de production de la parole qui consiste à appliquer des vibrations périodiques ou une source de bruit au conduit vocal modelisé par un filtre en treillis, suivant que les sons sont voisés ou non voisés.

Les paramètres de commandes obtenus par la méthode d'autocorrelation selon l'algorithme de Durbin sont :

- Les coefficients de réflexion de chaque trame.
- La valeur du fondamental d'excitation.
- La valeur du gain.

On a présenté aussi à la fin une autre technique de production de la parole qui est : la synthèse à formants.

La voix artificielle devient une partie de plus en plus importante et potentiellement indispensable de l'interface homme/machine. La synthèse électronique de la parole est une technologie jeune mais à évolution rapide.

Dans peu de temps, le réveil électronique vous décrira le programme de votre journée. Il sera aussi courant d'entendre l'ordinateur de votre voiture vous raconter ses problèmes mécaniques pendant que vous allez au travail.

