

وزارة التربية الوطنية  
MINISTERE DE L'EDUCATION NATIONALE

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT

*Electronique*

المدرسة الوطنية المتعددة التقنيات  
BIBLIOTHEQUE — المكتبة  
Ecole Nationale Polytechnique

# PROJET DE FIN D'ETUDES

SUJET

*Implémentation d'un synthétiseur à  
formants du type KLATT*

Proposé par :

*Melle M. GUERTI*

Etudié par :

*L. ZAÏDI*

*L. REMMACHE*

Dirigé par

*Melle M. GUERTI*

PROMOTION

*Juillet 1993*

بسم الله الرحمن الرحيم

“قل إن صلاتي ونسكي ومحياي ومماتي

لله رب العالمين لا شريك له ، وبذلك

أمرت وأنا أول المسلمين“

سورة الأنعام الآية ( ١٦٢ ، ١٦٣ )

- DEDPACCP -

المدرسة الوطنية المتعددة التقنيات  
BIBLIOTHEQUE — المكتبة  
Ecole Nationale Polytechnique

à la mémoire de ma mère,

à mon père,

à mes frères et sœurs,

à toute ma famille,

à tous mes amis et particulièrement RAËD, ZOËËR, MOËAMÈD,

NASSIM,

L. HAÏD

أ) ma mère,  
أ) mon père,  
أ) mon frère,  
أ) Nadia.

R. haïd.

-DÉDICATION-

À ma mère, ma chérie  
À mon père,  
À mon frère, mon frère.

À tous les hommes, femmes et enfants qui ont cru et qui croient  
toujours, du fond de leur cœur, au projet Islamique,

Hommes, Femmes, enfants qui furent un jour, qui sont aujourd'hui  
et qui seront demain encore plus forts,

voilà à qui je dédie cet humble mémoire et à qui je dédierai  
toute ma personnalité, tout mon savoir, toute mon honnêteté, tout  
mon Amour d'"ALLAH".

En particulier à Redha, le plus grand des hommes que j'ai connus,  
par sa bravoure, sa gentillesse et son Amour de "Dieu"; je t'aime  
Redha sur la parole de "Dieu",

à Yassine, Madjid, Brahim, Khalil, Abdelhak, Falim, Mohamed, Abdelkader  
et tant d'autres, tant d'autres ... Ce sont tous des Hommes.

À la Femme musulmane, saine d'esprit, à qui l'éducation de nos  
enfants revient; merci beaucoup "Femmes", merci beaucoup "maman".

À Nadia qui, j'espère (inchallah), sera comme fut Khadidja,  
Aïcha et tant d'autres femmes musulmanes.

Un serviteur d'"ALLAH".

R. Laid

## REMERCIEMENTS

Nous tenons à exprimer nos remerciements à notre promoteur, Mademoiselle M.GUERTI, pour son aide efficace, et ses précieux conseils tout au long de l'élaboration de ce mémoire.

Nous adressons également nos remerciements à Monsieur HALIMI pour son aide précieuse.

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

L.ZAIDI

L.REMMACHE

INTRODUCTION

I-GENERALITES SUR LE SIGNAL VOCAL

INTRODUCTION

1-L'APPAREIL PHONATOIRE HUMAIN.

1-1 PARTIE SUBGLOTTIQUE

1-2 PARTIE GLOTTIQUE (le larynx)

1-3 PARTIE SUPRAGLOTTIQUE

2-LE MECANISME DE PRODUCTION DE LA PAROLE

2-1 LES SONS VOISES OU SONORES

2-2 LES SONS NON VOISES OU SORDS

2-2-1 les consonnes occlusives

2-2-2 les consonnes fricatives

2-2-3 les consonnes sonnantes

3-LES DIFFERENTES TECHNIQUES D'ANALYSE-SYNTHESE

3-1 LES TECHNIQUES D'ANALYSE

3-1-1 Les méthodes de traitement du signal

a-transformation de FOURIER

b-transformée en Z

c-l'analyse cepstrale

d-méthode d'analyse indirecte de la parole

e-technique d'analyse par prédiction linéaire

3-1-2 Les méthodes d'extraction des paramètres

a-extraction de l'énergie

b-extraction des formants

c-extraction du fondamental

3-2 LES TECHNIQUES DE SYNTHÈSE

3-2-1 La synthèse à formants

3-2-2 La synthèse à prédiction linéaire

CONCLUSION

II-LES SYNTHÉTISSEURS DE LA PAROLE

INTRODUCTION

1-STRUCTURE GÉNÉRALE D'UN SYNTHÉTISSEUR

2-LES SYNTHÉTISSEURS A CANAUX

3-LES SYNTHÉTISSEURS ARTICULATOIRE

4-LES SYNTHÉTISSEURS A PREDICTION LINEAIRE

5-LES SYNTHÉTISSEURS A FORMANTS

5-1 LE SYNTHÉTISSEUR DE RABINER :structure serie

5-2 LE SYNTHÉTISSEUR DE HOLMES :structure parallèle

5-3 COMPARAISON DES STRUCTURES PARALLELES ET SERIE

6-COMPARAISON DES DIFFERENTS TYPES DE SYNTHÉTISSEURS

CONCLUSION

III-LE SYNTHÉTISSEUR DE KLATT

INTRODUCTION

1-OBJECTIFS DE REALISATION DE CE SYNTHÉTISSEUR

2-SIMULATION DU SYNTHÉTISSEUR

3-LES DIFFERENTS OUTILS UTILISES

3-1 TAUX D'ECHANTILLONNAGE

3-2 TAUX DE REGENERATION DES PARAMETRES

3-3 FILTRE NUMERIQUE

3-4 ANTIRESONATEUR NUMERIQUE (antiformants)

3-5 FILTRE PASSE-BAS

#### 4-DESCRIPTION DE LA STRUCTURE DU SYNTHETISEUR

##### 4-1 LES SOURCES DE SONS

###### 4-1-1 Source voisée

a-voisement normal

b-voisement quasi-sinusoidale

###### 4-1-2 Source de bruit

a-source fricative

b-source d'aspiration.

##### 4-2 LE SYSTEME DE FILTRAGE

###### 4-2-1 Le modèle serie du conduit vocal

###### 4-2-2 Le modèle parallèle du conduit vocal

#### 5-CARACTERISTIQUES DE RAYONNEMENT

#### 6-IMPLEMENTATION DU SYNTHETISEUR

##### 6-1 DESCRIPTION DU PROGRAMME PRINCIPAL

##### 6-2 DESCRIPTION DU SOUS-PROGRAMME PARCOE

##### 6-3 DESCRIPTION DU SOUS-PROGRAMME COEWAU

#### 7-RESULTATS GRAPHIQUES

#### CONCLUSIONS GENERALES

## INTRODUCTION

Les développements de l'électronique et de l'informatique ont permis la mise en oeuvre d'un grand nombre d'applications dans le domaine de la parole : nous pouvons citer le dialogue homme-machine , la plongée sous-marine , l'aide aux handicapés ou les petites applications en particulier dans les jeux vidéo.

Les recherches en synthèse de la parole se font à l'aide de synthétiseurs qui essaient de reproduire le signal acoustique de la parole.

L'un d'eux , le synthétiseur à formants , est un modèle acoustique du système phonatoire humain. L'intérêt de la synthèse à formants est qu'elle est particulièrement bien adaptée à une interprétation acoustique de ses paramètres de commande (formants, fréquence fondamentale etc...) , ce qui en fait un outil extrêmement intéressant en recherche.

L'objectif principal de notre travail , présenté dans ce mémoire , est justement l'étude d'une simulation d'un synthétiseur à formants et plus précisément un synthétiseur à formants du type KLATT.

Dans un premier chapitre , nous avons regroupé un certain nombre de rappels théoriques sur lesquels s'appuient la suite de ce travail à savoir , le fonctionnement de l'appareil phonatoire humain , les techniques d'analyse-synthèse nécessaires pour extraire les différents paramètres de commande du synthétiseur.

Le deuxième chapitre expose les différentes structures de synthétiseurs à formants et enfin la description de quelques synthétiseurs série et parallèle déjà mis en oeuvre dans des laboratoires de recherche.

Le troisième et dernier chapitre est consacré à l'étude du synthétiseur de KLATT du type hybride (série + parallèle). Nous verrons comment ce dernier a essayé d'allier les avantages des structures série et parallèle. Les différents modèles de la source vocale sont exposés afin de justifier le choix d'une source adéquate.

Enfin, une simulation basée sur des caractéristiques citées par KLATT, sera exposée en vue d'apprécier la validité des résultats graphiques obtenus pour quelques exemples de voyelles et consonnes du français en comparaison avec des signaux naturels.

## CHAPITRE I

### GENERALITES SUR LE SIGNAL

#### VOCAL

#### INTRODUCTION

Le rôle de ce chapitre est d'exposer les éléments essentiels sur lesquels s'appuie la suite de ce projet à savoir, une description de l'appareil phonatoire humain (qui nous permettra de nous initier au mécanisme de production de la parole ). Nous terminerons par la présentation de quelques techniques d'analyse-synthèse.

## 1 L'APPAREIL PHONATOIRE HUMAIN.

La parole est en fait un produit de la respiration .  
Lorsque l'air se dégage des poumons et ne rencontre aucun obstacle ,il n'y a pas d'émission de sons ,il y a simplement respiration.

Les obstacles provoquent des vibrations longitudinales ,des ondes acoustiques .En contrôlant l'introduction des obstacles (la langue par exemple) ,nous pouvons modifier les sons.

La parole est une forme "sophistiquée" de ce contrôle.

Il nous paraît donc utile de faire un bref rappel du système vocal humain au niveau anatomo-physiologique afin de mieux comprendre les phénomènes qui engendrent la production de la parole ( figure 1.1 ).<sup>(1)</sup> On distingue trois parties essentielles.

### 1-1 PARTIE SUBGLOTTIQUE.

Elle est constituée des poumons et de la trachée artère;elle est à l'origine de la parole par envoi d'air vers les parties glottiques .Elle constitue ainsi la source d'énergie de la parole .

### 1-2 PARTIE GLOTTIQUE ( LE LARYNX ).

Cette partie se présente sous forme de cartilages articulés , de ligaments ,de muscles et de muqueuses qui surmontent la trachée artère .A la base du larynx se trouvent les cordes vocales ,sorte de lèvres placées symétriquement à gauche et à droite de la ligne médiane du larynx .

Lors de l'émission de sons voisés ,les cordes vocales vibrent principalement dans un plan horizontal ,obeissant ainsi à un phénomène d'oscillations-relaxation .

### 1-3 PARTIE SUPRAGLOTTIQUE.

---

-(1): [2].

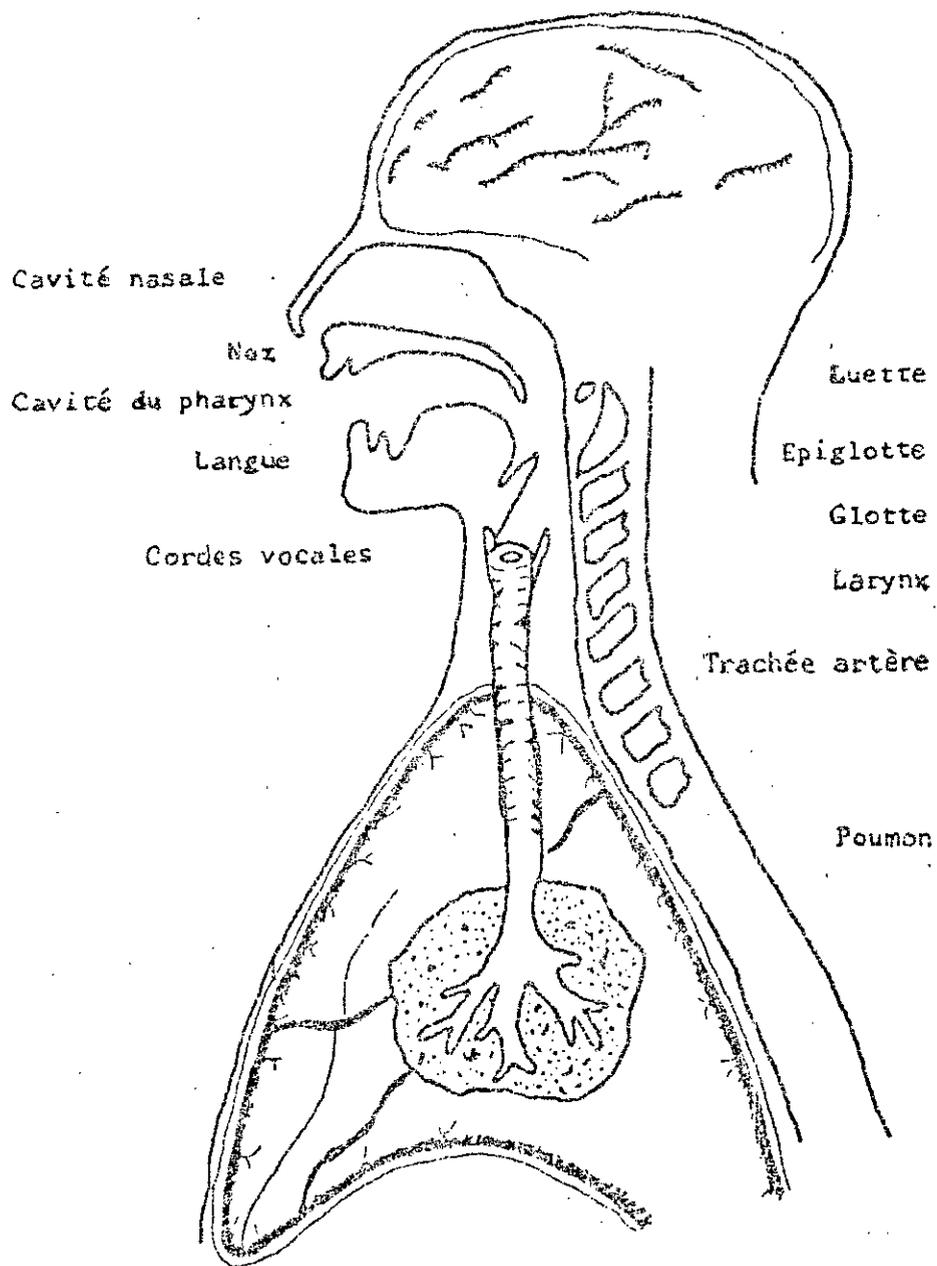


Figure 1-1 : Le Système Phonatoire humain

C'est tout ce qui vient au dessus du larynx .  
On distingue :

- le conduit oral,

il se compose du pharynx (cavité située au-dessus du larynx )et de la cavité buccale .La dimension et la géométrie du conduit oral sont extrêmement variables grâce à la grande mobilité de la langue ;

- le conduit nasal,

il se compose des fosses nasales qui ne sont que deux cavités de forme fixe et dont la communication avec la cavité orale est commandée par le voile du palais ( on parle de couplage entre les cavités nasales et les cavités orales ).

## 2 LE MECANISME DE PRODUCTION DE LA PAROLE.

Il y'a deux sortes de sources qui peuvent être activées durant la production de la parole .

L'une implique des vibrations quasi-periodiques de quelques structures notamment les cordes vocales ;la vibration de ces dernières est appelée voisement ( d'autres structures telles les lèvres ,le bout de la langue ou la luvette peuvent être mis en mouvement ) .

Le second type de source implique la génération d'une turbulence de bruit par un flux d'air rapide au-delà d'une constriction étroite .Le bruit résultant est appelé "aspiration" si la constriction est localisée au niveau des cordes vocales ,par exemple durant la prononciation du /h/ .

Si par contre la constriction est localisée au-dessus du larynx comme c'est le cas du /s/ ,le bruit résultant est dit "frication".

Quand le voisement et la turbulence coexistent ,comme par exemple lors d'une fricative /z/ ,le bruit est modulé périodiquement en amplitude par les vibrations des cordes vocales .

## 2-1 LES SONS VOISES OU SONORES.

Pour ces types de sons ,le conduit vocal est excité par une source d'impulsions pseudo-périodique ( figure 1-2 )<sup>(1)</sup>. En effet,les cordes vocales oscillent suivant un phénomène de relaxation .Une simple explication mécanique du fonctionnement des cordes vocales peut être donnée sur la base d'une force exercée sur les cordes alternativement par la pression subglottique ( cordes vocales fermées ) et par la dépression exercée à la glotte par le passage de l'air .Cette dernière est due à l'effet de BERNOULLI .

La source à son tour va exciter le conduit que l'on peut considérer comme un tuyau à section variable dans lequel se propage une onde plane .Il se comporte comme une ligne de transmission ou un "filtre acoustique".

Dans le cas des voyelles orales ,la fonction de transfert de ce filtre se présente sous la forme d'un produit filtre passe-bas du second ordre .

Dans le cas des voyelles nasalisées ,le conduit nasal est couplé au conduit oral ,ce qui introduit des zéros dans la fonction de transfert ,appelés aussi antifonnants.

## 2-2 LES SONS NON VOISES OU SOURDS.

Les consonnes forment ,au plan acoustique ,une classe très hétérogène que l'on peut décomposer en trois ensembles ayant des caractéristiques distinctes .

---

(1): [1]

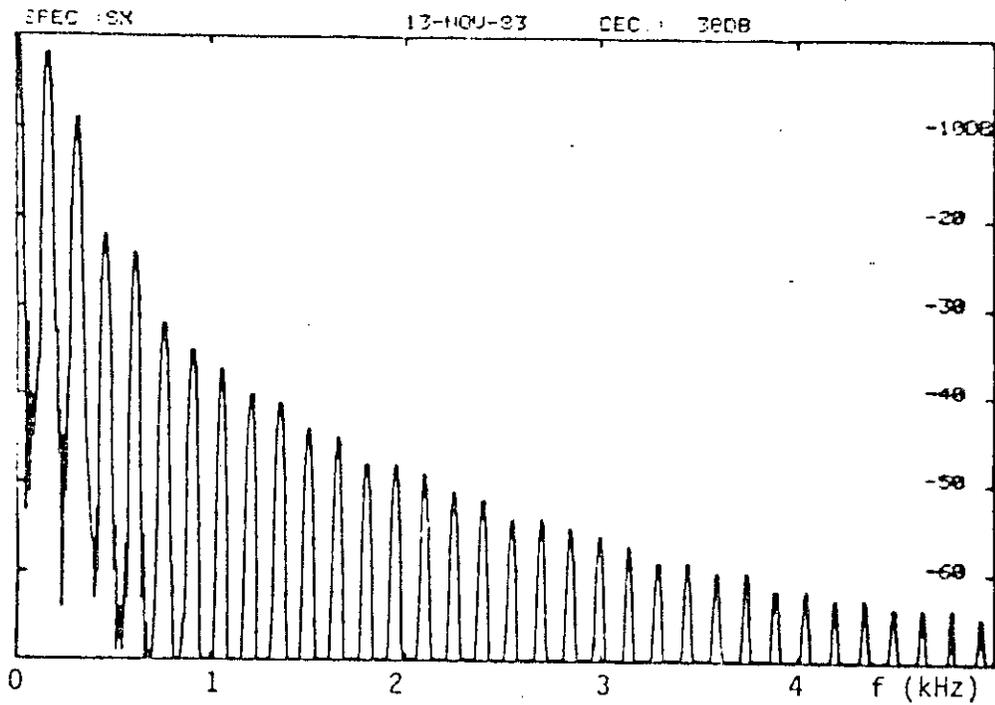
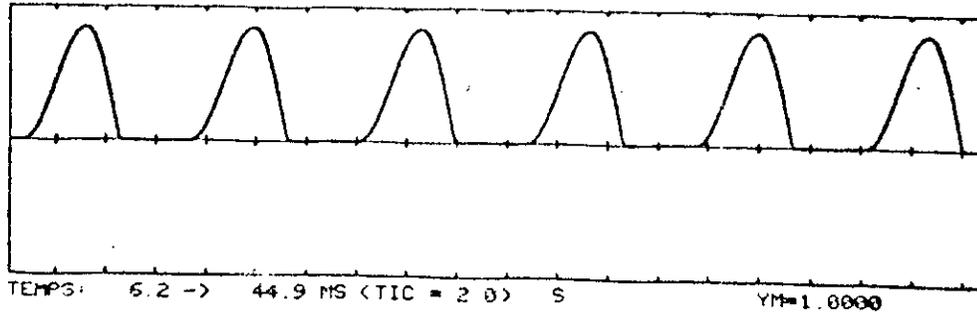


Figure 1-2 : L'onde de debit et son spectre

### 2-2-1 Les consonnes occlusives.

Les occlusives se caractérisent principalement par une fermeture complète d'une section du conduit oral, ce qui augmente la pression en arrière de cette occlusion. Ainsi une explosion se produit au moment de la brusque ouverture de l'occlusion.

Pendant la tenue, le signal est assez faible avec un spectre ne contenant que des fréquences allant jusqu'à 1KHZ.

Par contre pendant la phase d'explosion le spectre du signal contient du bruit. Les caractéristiques (spectre et amplitude) de bruit varient selon l'occlusion.

La figure 1-3<sup>(1)</sup> illustre un exemple avec les différentes régions expliquées ci-dessus.

### 2-2-2 Les consonnes fricatives.

Pour ces types de sons, une constriction est formée en un point du conduit oral où l'écoulement crée des turbulences qui donnent naissance à une source de bruit.

La source de bruit doit avoir un spectre approximativement plat, avec une distribution d'amplitude gaussienne (figure 1-3).

### 2-2-3 Les consonnes sonnantes.

Acoustiquement, les consonnes sonantes ( /l/, /r/, /m/, /n/ ) se caractérisent par une structure de formants.

Ce sont des consonnes à formants. Elles ne possèdent que peu ou pas de bruit. Intrinsèquement sonores, elles se dévoient au contact d'une consonne sourde.

Plusieurs sous classes peuvent être distinguées :

---

(1): [1]

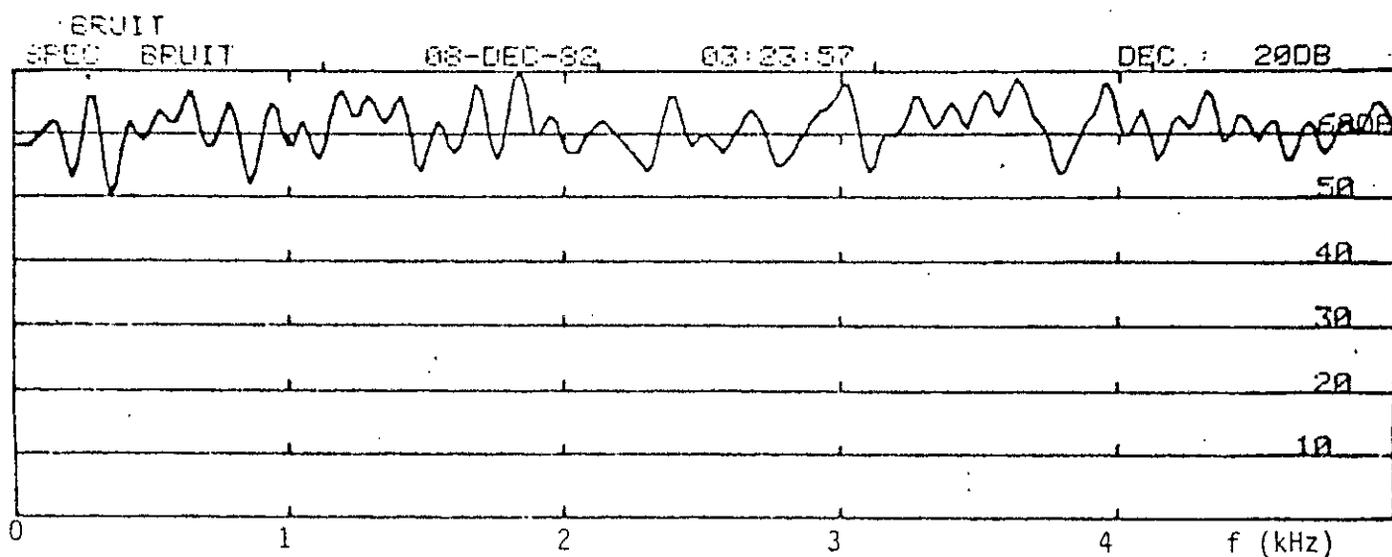
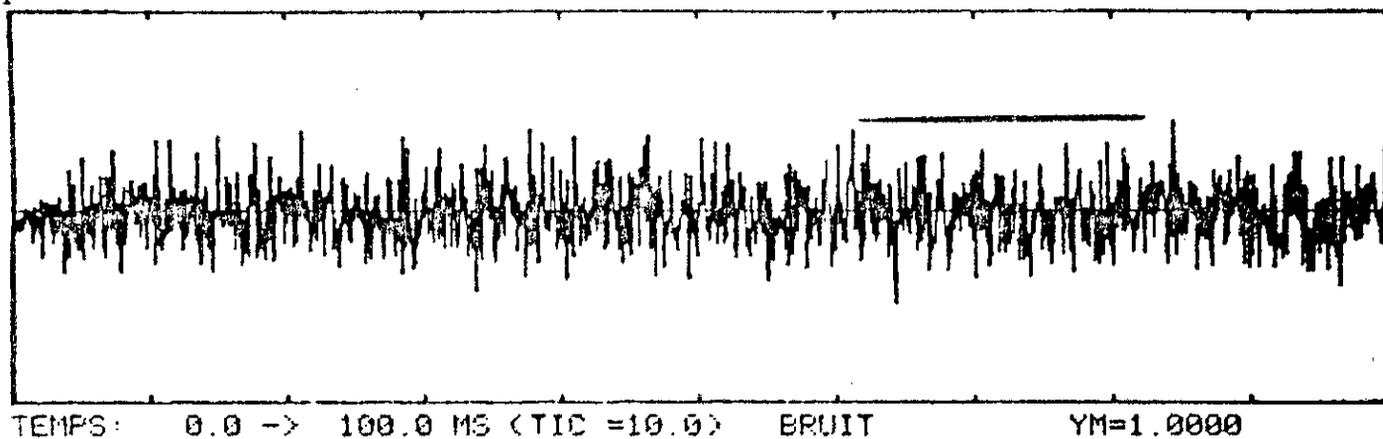


Figure 1-3: Le signal de bruit et son spectre

- les consonnes nasales ( /m/,/n/ ) ,produites avec le voile du palais abaissé et donc un couplage conduit oral - conduit nasal qui complique leur structure acoustique ;

- les consonnes liquides ( /l/,/r/ ) ,au nom traditionnel sans signification précise ,qui possèdent ou non des zéros dans leur fonction de transfert ;

- les semi-consonnes ( /j/,/w/ ) ,dont le nom générique est bien mal choisi ,tout comme son synonyme "semi-voyelle" ,et qui sont acoustiquement proches des voyelles .

### 3 LES DIFFERENTES TECHNIQUES D'ANALYSE-SYNTHESE.

#### 3-1 LES TECHNIQUES D'ANALYSE.

Le but poursuivi est de caractériser sur le plan acoustique les sons de parole et de pouvoir vérifier la validité des résultats par synthèse .Les paramètres fondamentaux à élaborer sont les suivants :

- énergie globale du signal pour une fenêtre temporelle donnée,
- spectre du signal sur une fenêtre temporelle et en fonction du spectre ,détermination des formants ,
- la fréquence fondamentale .

La parole est un signal extrêmement dense : sa transmission intégrale nécessite un débit d'information équivalent à 30 Kbits/s .

C'est un signal qui véhicule des informations ne concernant pas uniquement la signification objective du message ;il contient des données sur l'accent ,le rythme et l'intonation du locuteur.

Il est d'autre part très redondant ce qui lui permet d'être très résistant au bruit ambiant. Il faudrait disposer d'une mémoire d'une dizaine de méga-octets pour stocker quatre à cinq minutes de parole. La parole doit être codée afin de réduire son débit tout en conservant son intelligibilité.

Or qui dit codage pense ANALYSE préalable.

### 3-1-1 Les méthodes de traitement du signal. (1)

Dans notre cas, nous nous limiterons aux signaux échantillonnés car en ce qui concerne notre synthèse, elle se traite par ordinateur.

#### a - Transformation de FOURIER.

Un signal  $f(t)$  est présenté par une suite de  $N$  échantillons  $X(0), X(1), \dots, X(N-1)$ . Sa transformée de FOURIER discrète,  $F(\omega)$ , se présente sous la forme de  $N$  valeurs :

$$A(n) = \sum_{k=0}^{N-1} X(k) \cdot \exp(-2\pi jnk/N) \quad (1)$$

et la transformée inverse, celle qui permet de connaître les échantillons  $X$  à partir de la donnée des termes  $A$ , s'écrit

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} A(n) \cdot \exp(2\pi jnk/N) \quad (2)$$

#### b - Transformée en $Z$ .

On définit la transformée en  $Z$ ,  $X(Z)$ , de la séquence d'échantillons  $X(n)$  par :

---

(1): [3]

$$X(Z) = \sum_{n=-\infty}^{\infty} X(n) \cdot Z^{-n} \quad (3)$$

où  $Z^{-1} = \exp(-j2\pi fT)$  est un "multiplieur" réalisant en quelque sorte une pondération sur la somme des échantillons. Multiplions cette transformée par  $Z^{-1}$  pour obtenir une nouvelle transformée  $V(Z)$  :

$$V(Z) = Z^{-1} \cdot X(Z) = \sum_{n=-\infty}^{\infty} X(n) \cdot Z^{-n-1} \quad (4)$$

Ce qui peut également s'écrire :

$$V(Z) = \sum_{n=-\infty}^{\infty} X(n-1) \cdot Z^{-n} \quad (5)$$

Par conséquent

$$V(Z) = X(n-1) \quad (6)$$

On a ainsi réalisé un opérateur de retard qui est en fait un multiplieur dans le domaine de la transformée en  $Z$ . Un filtre pour la synthèse de la parole peut être représenté par

sa fonction de transfert en Z :

$$H(Z) = \frac{b_0 + b_1 z^{-1} + \dots + b_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}} \quad (7)$$

qui comporte p pôles et q zéros .C'est le cas par exemple en prédiction linéaire .

On utilise aussi une autre forme (parallèle)

$$H(Z) = \sum_{i=1}^j C_i \cdot \frac{1 + d_i z^{-1}}{1 + a_i z^{-1} + b_i z^{-2}} \quad (8)$$

Le filtre H se présente alors comme un ensemble de J cellules en parallèle .

Chaque cellule est constituée d'un filtre de premier ordre (un zéro) et d'une section de deuxième ordre (deux pôles) série avec un gain  $C_i$ .

Chaque cellule est de réalisation relativement simple .Le calcul des paramètres est aussi relativement peu complexe .

Enfin ceux-ci permettent de contrôler directement et de façon simple les caractéristiques perceptivement intéressantes de l'enveloppe du spectre :

- a et b déterminent la fréquence centrale d'une bande  $\Delta f$  de l'enveloppe du spectre et sa forme locale (maximum et largeur de bande) ;

- C contrôle l'amplitude globale de cette zone ;

- d permet d'incliner ,dans un sens ou dans l'autre , l'enveloppe dans cette zone .

c - L'analyse cepstrale

On pourrait tout simplement filtrer le signal d'entrée pour éliminer les ondulations non souhaitées du signal de parole . Ce qu'il y a c'est que le filtrage élimine tout ce qui a un caractère additif ,alors qu'ici ,on ne parle que de multiplication de fonctions d'onde .

Alors transformons les multiplications en additions grâce aux logarithmes et filtrons ce que nous voulons éliminer les composantes (de hautes fréquences) non désirées.

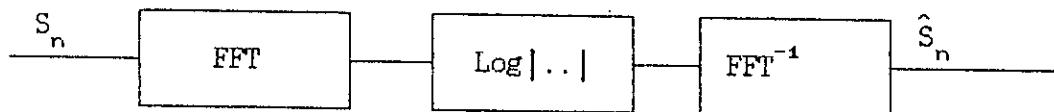


Figure 1-4 : schéma synoptique de l'analyse cepstrale.

Le cepstre est le spectre (ou transformée de FOURIER) du logarithme du spectre de fréquence .l'axe horizontal du cepstre porte une grandeur qui possède la dimension d'un temps :on l'appelle "quéfrence".

Il est à noter au passage que les signaux de hautes fréquences ont de faibles quéfrences .

Pour l'analyse des formants ,on élimine toutes les quéfrences dépassant un certain seuil et on identifie les pics du cepstre de plus faible quéfrence .

d - Méthodes d'analyse indirecte de la parole

On part d'un signal électrique image de la voix ,delivré par un microphone et on tente d'élaborer un signal électrique image de la fréquence de la mélodie .

Les analyseurs utilisent toutes les notions d'analyse spectrale du signal rayonné par les lèvres pour le traitement fréquentiel et l'analyse par autocorrélation pour le traitement temporel . L'analyse spectrale dans le domaine fréquentiel d'un signal  $f(t)$  variable dans le temps  $t$  est basée sur la détermination de sa transformée de FOURIER  $F(\omega)$

$$\int_{-\infty}^{+\infty} f(t).exp(-j\omega t).dt \quad (9)$$

Dans le cas où  $f(t)$  est periodique ,il apparaît par cette opération des raies dans le spectre qui correspondent à des fréquences multiples les unes des autres (harmoniques) et dont le plus petit multiple (fondamentale) est responsable de la sensation de hauteur lors de l'audition du signal  $f(t)$ .

On distingue ,parmi les analyseurs spectraux de mélodie ,les analyseurs à filtre passe-bas (un seul filtre linéaire) , celui-ci ayant pour objet d'éliminer du signal de parole les harmoniques, suivi d'un système de mesure ou de detection des périodes laryngiennes successives (évolution du fondamental).

Une analyse dans le domaine temporel consiste à chercher les lois de periodicité d'un signal  $f(t)$ . elle s'effectue par le calcul de la fonction d'autocorrélation

$$h(\tau) = \int_{-\infty}^{+\infty} f(t).f(t+\tau)dt \quad (10)$$

où l'on compare la valeur de la fonction à l'instant  $t$ , à la valeur de cette même fonction à l'instant  $t+\tau$ .

Toute oscillation dans la courbe représentative de la fonction d'autocorrélation indique la présence de motifs similaires dans le signal de parole analysé.

Toutes ces méthodes d'analyse nécessitent le recours à des ordinateurs pour la réalisation des calculs qui sont souvent longs et voraces en espace mémoire. Elles servent à la détermination du fondamental, des périodicités et des pics du signal vocal.

#### e - La technique d'analyse par prédiction linéaire

L'intérêt que présente l'analyse par prédiction linéaire est multiple :

i - c'est une technique d'analyse directe, car, une fois le critère d'approximation du signal réel par le modèle fixé, le résultat est obtenu par la résolution d'un système d'équations linéaires;

ii - elle offre la possibilité de nombreuses interprétations à partir des paramètres extraits: pôles, bandes passantes du système ect...

Le principe est fondé sur l'hypothèse selon laquelle un échantillon du signal de parole  $S(nT)$ , où  $T$  est la période d'échantillonnage, peut être "prédit" approximativement par une somme, pondérée linéairement, d'un certain nombre d'échantillons le précédant immédiatement.

La valeur approchée (prédite)  $\hat{S}(nT)$  du signal réel  $S(nT)$  s'écrit

$$\hat{S}(nT) = \sum_{k=1}^p a_k \cdot S(nT-kT) \quad (11)$$

Dans cette somme de "p" termes, les coefficients " $a_k$ " servent à la pondération des "p" signaux de parole recueillis ou

calculés aux instants " $nT-T, nT-2T, \dots, nT-pT$ ".

On dit que p est l'ordre de prédicteur linéaire. la différence  $S(nT) - \hat{S}(nT)$  est l'erreur  $e(nT)$  entre l'échantillon du signal de parole original et l'échantillon de signal prédit.

On établit des relations entre les coefficients  $a_k$  du prédicteur et la fonction d'autocorrélation du signal original pour minimiser l'erreur moyenne calculée sur l'ensemble des échantillons prédits.

Pratiquement, pour choisir les coefficients de prédiction linéaire, on calcule d'abord l'erreur quadratique moyenne "M" entre le signal à un instant donné et sa prédiction par des échantillons (en nombre supposé égal à p) évalués à de précédents instants:

$$M = \sum_n e(n)^2 = \sum_n \left[ S(n) - \sum_{k=1}^p a_k \cdot S(n-k) \right]^2 \quad (12)$$

$S(n)$  est la valeur prise par le signal de parole à l'instant  $nT$  et  $S_{n-k} = S(nT-kT)$  constitue les valeurs recueillies à l'un des instants précédents.

Il faut rendre M le plus petit possible donc minimiser M; il

existe pour cela des méthodes mathématiques : on calcule les dérivées partielles de M par rapport aux coefficients de pondérations  $a_k$  et l'on annule chacune d'entre-elles. On écrit donc pour l'une quelconques des dérivées partielles.

$$\frac{\partial M}{\partial a_j} = -2 \cdot \sum_n S(n-j) \left[ S(n) - \sum_{k=1}^p a_k \cdot S(n-k) \right] = 0 \quad (13)$$

ou encore d'une manière condensée , pour toutes les dérivées partielles:

$$\sum_{k=1}^p a_k \cdot \sum_n S(n-j) \cdot S(n-k) = \sum_n S(n) \cdot S(n-j) \quad (14)$$

avec  $j = \overline{1, p}$ .

On obtient une "équation matricielle" constituée d'un ensemble de "p" équations linéaires , avec p inconnues  $a_1, \dots, a_p$  que l'on résoud en inversant une matrice de dimensions  $p \times p$ .

Il existe d'autres méthodes fondées sur l'autocorrélation et la covariance. Pour la première , on écrit une relation d'autocorrélation sur une séquence infinie de valeurs  $X(n)$  d'un signal

$$\sum X(n-j) \cdot X(n-k) \quad (14)$$

### 3-1-2 Les méthodes d'extraction des paramètres

Après cet aperçu sur les différentes méthodes de traitement du signal , nous allons exposer un certain nombre de méthodes d'extraction de paramètres du signal de parole.

#### a - extraction de l'énergie

L'énergie du signal de l'échantillon n est donnée par :

$$E_n = \sum_m X^2(m) \cdot h(n-m) \quad (15)$$

où  $X(m)$  est la valeur de l'échantillon m.

$h(n)$  est une fonction représentant la fenêtre d'analyse.

La représentation de l'énergie varie selon la fenêtre choisie. Si on choisit une fenêtre  $h(n)$  très longue, on constate que l'énergie  $E_n$  montre un changement assez faible. Par contre, si on choisit une fenêtre très courte pour représenter les changements rapides du signal, on ne donne pas une représentation de l'énergie moyenne lissée.

#### b - extraction des formants

Un grand nombre d'algorithmes sont proposés dans la littérature et la plupart d'entre eux fonctionnent correctement dans le cas des sons voisés mais sont moins précis. Certes, des algorithmes simples permettent de rechercher des pics ; il suffit en effet de trouver l'échantillon numéro K tel que :

$$X(k) > X(k-1) \text{ et } X(k) > X(k+1) \quad (16)$$

Mais, l'algorithme ne détecte pas tous les formants, en particulier lorsque l'un est proche de l'autre. Il sert cependant à rejeter les pics qui ne sont pas des formants en expliquant certains critères de choix, comme une largeur de bande nécessairement inférieure à 500 HZ.

Une technique basée sur la transformée de FOURIER rapide (FFT) a été proposée par RABINER (RABINER et SHAFER:1970);celle-ci est automatique.

Le cepste du signal à analyser est calculé et filtré passe-bas afin d'éliminer les fréquences hautes correspondant à la source. La transformation de FOURIER est alors appliquée au résultat pour obtenir un spectre lissé. Ensuite, un algorithme de détection de pic calcule les trois premiers formants.

Les techniques qui utilisent la prédiction linéaire sont nombreuses et variées. L'une d'elle (MARKEL:1972) consiste à calculer les coefficients du filtre inverse en utilisant la prédiction linéaire. On obtient ensuite les spectres inverses du signal par la transformation de FOURIER et enfin les formants par détection de pics. Cette technique est assez rapide mais peu précise.

Parmi les techniques d'extraction de formants décrites ci-dessus aucune d'entre-elles ne fonctionnent pour les consonnes et plus généralement pour les sons non voisés.

Pour la détermination des formants des consonnes fricatives, elle présente un problème particulier, puisque les formants de ces consonnes possèdent souvent des bandes passantes larges, ce qui conduit à un spectre global peu différencié.

Parmi les techniques qui existent pour ce cas, on peut citer la plus essentielle qui est celle qui consiste à utiliser l'analyse par prédiction linéaire, suivie d'une extraction de pôles. Cette technique nous permet de suivre l'évolution des formants, avec une moins bonne précision en fréquence.

#### c - extraction du fondamental

Il existe un grand nombre d'études et de techniques qui traitent le problème du fondamental ou FO. Nous allons juste citer une technique que nous jugeons intéressante: c'est la technique

par "intercorrelation avec une fonction peigne"(MARTIN:1981).

Cette technique est basée sur la recherche d'une structure harmonique dans le spectre d'amplitude. Elle consiste à calculer la fonction peigne pour différentes valeurs de la distance entre les dents du peigne: le maximum de cette fonction a lieu pour une distance entre deux dents consécutives égales à la périodicité du signal à analyser.

### 3-2 LES TECHNIQUES DE SYNTHÈSE (1)

La synthèse nous permet de mieux comprendre le processus de production de la parole ainsi que de déterminer les caractéristiques et les indices acoustiques des sons. Il existe essentiellement deux techniques principales de synthèse à savoir la prédiction linéaire et la synthèse par formants.

La prédiction linéaire a déjà été traitée en analyse. En exploitant la synthèse par formant, nous comprendrons mieux la philosophie de la synthèse.

#### 3-2-1 la synthèse à formants

La synthèse à formants cherche à reproduire les caractéristiques acoustiques du signal de parole en simulant la fonction de transfert du conduit vocal et les différentes sources d'excitations la stimulant.

La fonction de transfert qui modélise un formant a la forme suivante:

$$H(p) = \frac{\omega_c^2}{p^2 + b.p + \omega_c^2} \quad (17)$$

$\omega_c$  représente la fréquence nominale de résonance (exprimée en rad/s) et b est la bande passante de la courbe

---

(1): [3]

de résonance (figure 1-5).

Posons  $p = j\omega$  et calculons le spectre énergétique ,soit:

$$\left| H(j\omega) \right|^2 = \frac{\omega_c^4}{\left[ \omega^2 - (\omega_c^2 - b^2/2) \right]^2 + b^2 (\omega_c^2 - b^2/4)} \quad (18)$$

Il passe par un maximum ,lorsque  $\omega$  varie et atteint la valeur  $(\omega_c^2 - b^2/2)^{1/2}$  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  $\omega_c$  ,de sorte qu'une bonne approximation du pic soit  $\omega = \omega_c$  et son amplitude est alors égale à  $\omega_c/b$ .

Lorsque la fréquence est supérieure au pic ,l'énergie varie comme  $1/\omega^4$  .

Sur une octave ,c'est à dire pour deux sons dont l'un a une fréquence double de l'autre ,le niveau d'énergie est divisé par un facteur 16 ,ce qui s'exprime couramment en decibels par  $10 \cdot \log(16) = 12$  dB/octave.

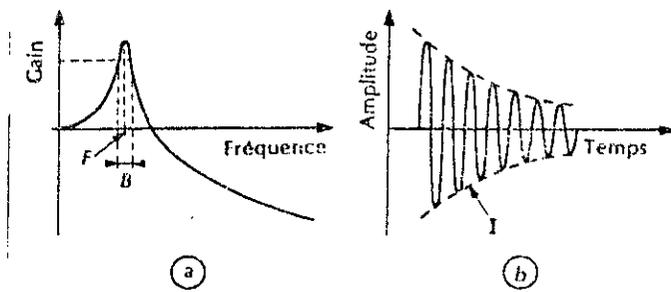
A 3 dB au dessous du pic ,l'énergie vaut:

$$\left| H(j\omega_{3dB}) \right|^2 = \frac{1}{2} \left| H(j\omega_{max}) \right|^2 = \frac{\omega_c^2}{2b^2} \quad (19)$$

et la valeur correspondante de la fréquence est ,avec bonne approximation:

$$\omega_{3dB} = \omega_c \pm b/2 \quad (20)$$

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



$B$  largeur de bande à 3 dB  
 $F$  fréquence centrale

Figure 1-5 : Aspect fréquentiel et temporel d'un formant pris individuellement.

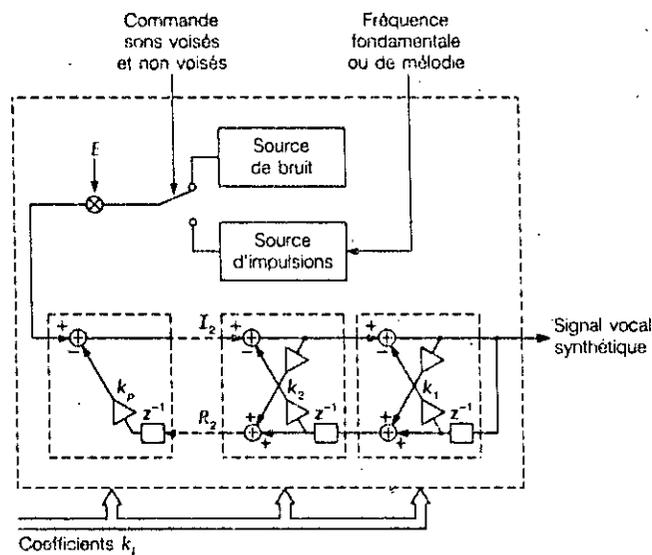


Figure 1-6 : Principe du filtre récursif dans la prédiction linéaire

### 3-2-2 La synthèse par prédiction linéaire

Puisque cette technique a déjà été traitée en analyse , nous connaissons donc sa structure et sommes en mesure de donner les principes sur lesquels elle est fondée.

Connaissant les coefficients de prédiction , la fréquence fondamentale et la nature de l'excitation , la synthèse se fait au moyen d'un filtre récursif (figure 1-6), c'est à dire un système dont la sortie réagit sur l'entrée , avec un retard égal à un pas d'échantillonnage.

#### CONCLUSION

Ce chapitre nous a permis de connaître une certaine terminologie et à travers elle de comprendre certains phénomènes se produisant au cours de la synthèse (naturelle ou artificielle) de la parole et qui nous sera d'une aide précieuse pour la poursuite de ce document à savoir la synthèse de la parole et les principaux synthétiseurs.

Il est à noter que les synthétiseurs qui seront exposés au chapitre suivant sont basés sur des techniques d'analyse déjà vues auparavant.

## CHAPITRE II

# LES SYNTHÉTISEURS DE LA PAROLE

### INTRODUCTION

Dans ce chapitre , nous examinerons les appareils permettant de reproduire la parole à partir de signaux déduits de l'analyse du message dits "synthétiseurs de parole".

Quatre grandes catégories sont à considérer selon le mode de reproduction du spectre instantané: les synthétiseurs à canaux , les synthétiseurs articulatoires , les synthétiseurs à prédiction linéaire et les synthétiseurs à formants.

Les trois premiers seront décrits brièvement contrairement au quatrième qui va être traité en détail et sera illustré par deux exemples de réalisation (série et parallèle).

## 1 STRUCTURE GENERALE D'UN SYNTHETISEUR. (1)

D'une façon générale ,un synthétiseur de parole est constitué de sources d'excitation qui sont des sources d'impulsions pour les sons voisés ,une source de bruit pour les sons non-voisés et un simulateur de conduit vocal (figure 2-1).

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

Le modèle peut être réalisé par un nombre plus ou moins important de filtres électriques dont les caractéristiques reproduisent les propriétés résonantes du conduit vocal (synthétiseurs à formants) ,ou encore dont les valeurs des gains sont adaptées au spectre de la parole (synthétiseurs à canaux).

Le modèle est également réalisable numériquement par un unique filtre numérique (synthétiseurs à prédiction linéaire).

Il existe un autre modèle dit "synthétiseur articulatoire", dont l'idée consiste à modéliser directement le conduit vocal lui-même et non plus ce qui y est généré.

## 2 LES SYNTHETISEURS A CANAUX.

Dans les appareils de ce type ,le spectre de la parole est divisé en un certain nombre de bandes de fréquences fixes (figure 2-2).

On excite un banc de filtres répartis sur la bande audio-fréquence à l'aide de l'une ou l'autre des sources d'excitation

---

(1): [3]

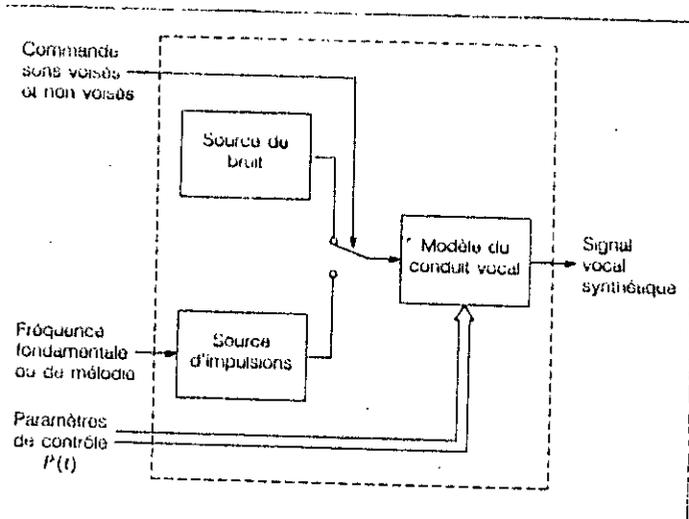


figure 2.1 : Principe du Synthétiseur de la parole

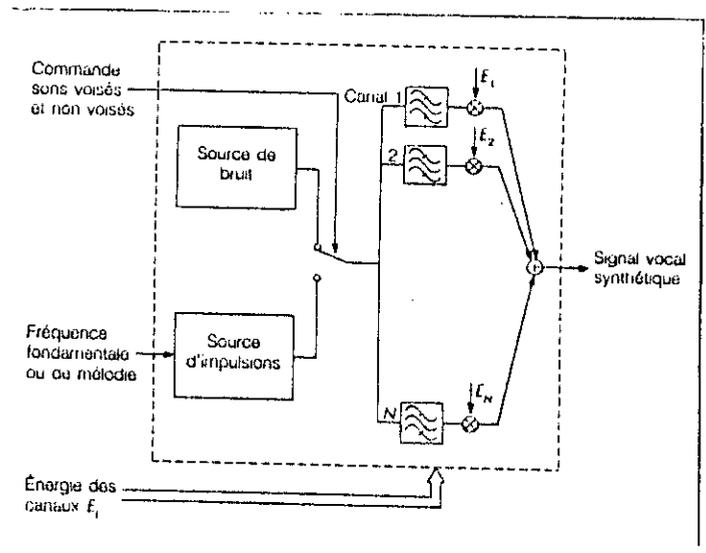


Figure 2-2 Forme canonique d'un synthétiseur à canaux

produisant des impulsions ou un bruit blanc , selon que les sons sont voisés ou non ; en outre , on dose l'énergie dans chaque canal ; également , nous faisons correspondre la fréquence des impulsions à celle de la mélodie.

Les paramètres de contrôle du synthétiseur pendant un certain laps de temps sont alors la commande "voisée/non voisée" , la fréquence du fondamental et l'énergie dans chaque canal.

Chaque portion de l'onde finale est donc construite par sommation des ondes partielles synthétisées par chacun des canaux.

### 3 LES SYNTHETISEURS ARTICULATOIRES.

Les synthétiseurs articulatoires ont tendances à simuler le fonctionnement acoustique du conduit vocal.

Un modèle articulatoire permet de reconstituer la forme du conduit vocal en fonction de la position des organes phonatoires (langue , machoire , lèvres) (figure 2-3).

Le signal vocal est ensuite calculé à l'aide d'une simulation mathématique de l'écoulement de l'air dans le canal vocal.

La complexité des calculs mis en jeu conduit actuellement soit à des réalisations très coûteuses et difficiles à mettre en oeuvre en temps réel , soit à des simplifications du modèle articulatoire dégradant ainsi la qualité de la synthèse.

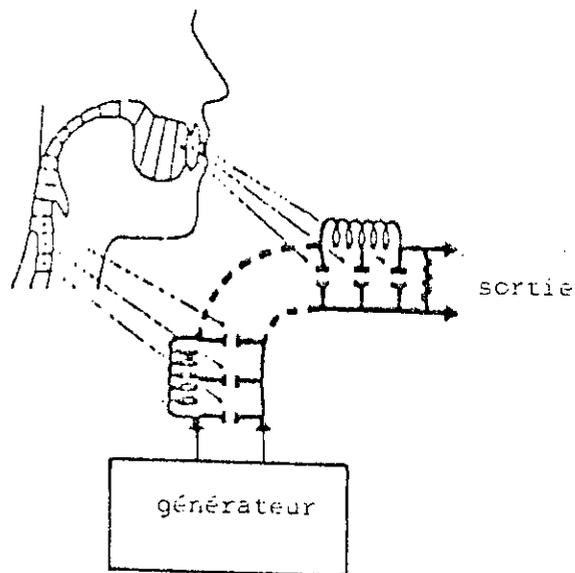


Figure : 2-3 Synthétiseur articulatoire

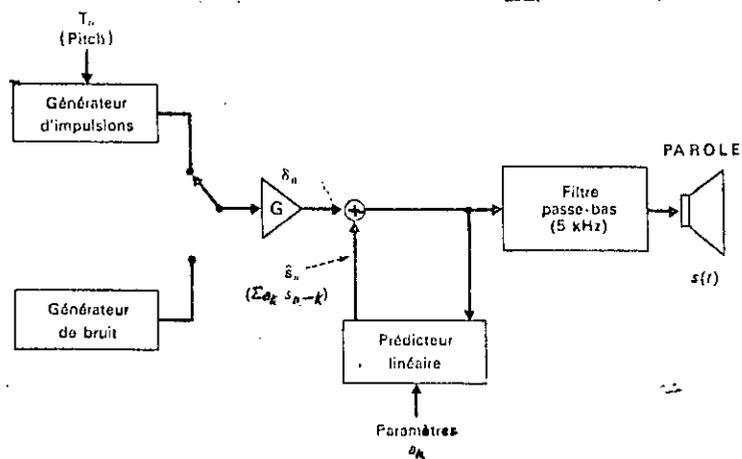


Figure : 2-4 Synthétiseur à prédiction linéaire

#### 4 LES SYNTHETISEURS A PREDICTION LINEAIRE. (1)

La figure 2-4 montre le principe de la réalisation d'un synthétiseur utilisant la prédiction linéaire. Les paramètres de contrôle fournis au synthétiseur sont:

- la période "T0" du fondamental ,
- un signal linéaire relatif au choix de l'excitation (voisée ou non-voisée) ,
- la valeur quadratique moyenne du signal échantillonné ,
- et les "p" coefficients de prédiction.

Le générateur d'impulsions fournit , au début de chaque période de pich , une impulsion unité .

Le générateur de bruit blanc produit un signal aléatoire.

C'est la commande linéaire qui contrôle la commutation de l'un ou de l'autre de ces deux dispositifs d'excitation .

L'amplitude du signal d'excitation est ajustée à l'aide de l'amplificateur "G". La valeur  $\hat{S}_n$  du signal à l'instant  $nT$  , prédite à partir des échantillons précédents , est combinée avec le signal d'excitation  $S_n$  lorsque ce dernier est présent ; on obtient ainsi le n-ième échantillon du signal synthétique. Les échantillons sont enfin soumis à un filtrage passe-bas.

#### 5 LES SYNTHETISEURS A FORMANTS. (2)

Le principe de base du synthétiseur à formants est une simulation des caractéristiques acoustiques du signal de parole , ou plus exactement une simulation de la fonction de transfert du conduit vocal considérée comme l'ensemble "excitateur+résonateur".

Dans le synthétiseur à formants , nous essayons de reconstituer cette fonction à l'aide d'un assemblage de filtres passe-bas

---

(1): [4]

(2): [3]

résonnants du second ordre qui simulent les formants.

Il existe deux façons principales de connecter plusieurs filtres de formants entre-eux: en série ou en parallèle.

#### S-1 LE SYNTHÉTISEUR DE RABINER: structure série (1)

La structure de ce synthétiseur est schématisée à la figure 2-5. Nous pouvons distinguer deux sources d'excitations et trois branches de filtrages connectées en parallèle.

La source voisée est une source d'impulsions mise en forme par un filtre passe-bas du second ordre. La source de bruit est un générateur de bruit blanc.

Le canal situé en haut de la figure est le plus important. Il comporte trois filtres de formants connectés en cascade, un réseau de nasalité et un réseau de rayonnement.

Le réseau de correction sert à représenter les effets des formants d'ordre supérieur qui ne sont pas simulés explicitement.

Le réseau de nasalité sert à introduire dans la fonction de transfert de ce canal une paire pôle/zéro qui représente les effets de nasalité lorsque le conduit nasal est couplé au conduit oral.

Enfin le réseau de rayonnement simule les effets de radiation dus à la tête. Ce canal sert à la production des sons voisés, sauf dans le cas où le commutateur "VA" permet l'attaque de ce canal par la source de bruit, pour les sons aspirés.

Le canal figuré en bas du schéma est celui du bruit: il est constitué d'une paire pôle/zéro et d'un filtre de mise en forme.

Une caractéristique importante de ce synthétiseur est la

---

(1): [1]

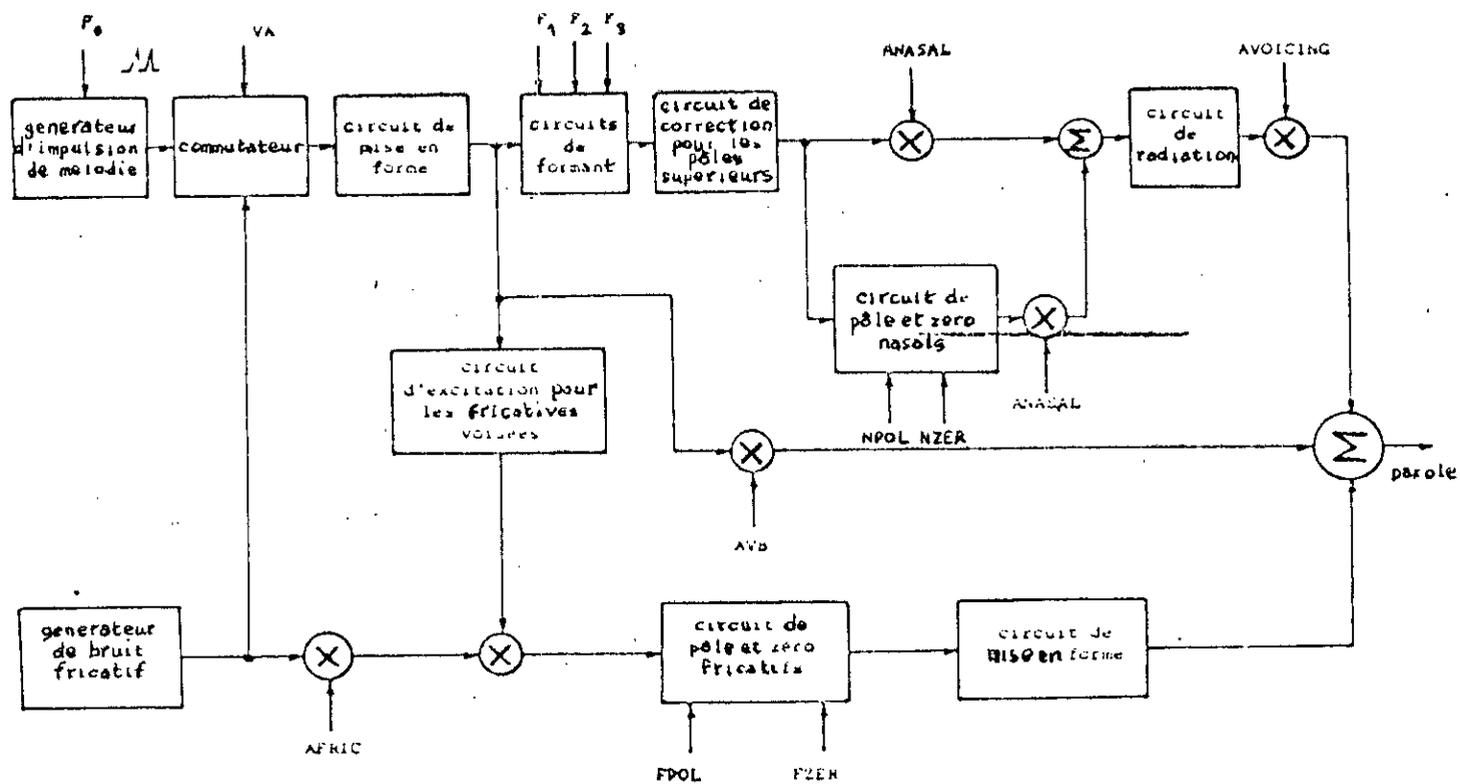


Figure 2-5 : Le Synthétiseur de Rabiner

modulation en amplitude du bruit , lors de la production des consonnes fricatives voisées (figure 2-6).

Le résonateur à la fréquence du premier formant "F1" est une approximation du premier ordre de la fonction de transfert du débit volumique au niveau de la constriction par rapport à la pression glottique.

La valeur " $V_{th}$ " correspond au seuil en-dessous duquel la génération du bruit de constriction n'a pas lieu.

Finalement , le bruit est modulé en synchronisme avec la source voisée par un signal qui présente approximativement le débit volumique au niveau de la constriction.

Ce système semble être une représentation assez fidèle de la réalité.

#### 5-2 LE SYNTHETISEUR DE HOLMES: structure parallèle (1)

Le synthétiseur contient six branches connectées en parallèle (figure 2-7). Les cinq premières sont constituées d'un amplificateur à gain variable et d'un filtre du second ordre passe-bas (filtre de formants) , le sixième contient simplement un amplificateur et un filtre passe-bande (3600-4000 HZ) , pour la partie haute du spectre des fricatives et des plosives.

Le filtre "FN" sert , en conjonction avec "F1", à simuler les effets de nasalisation en créant un pseudo-zéro entre "F1" et "F2". Les trois premiers formants sont variables mais "F4" est toujours fixe. Chaque branche contient un filtre de mise en forme spectrale qui sert à réduire l'interaction entre les formants.

Un mélange d'une source de bruit et d'une source de voisement peut attaquer les cinq branches , ce qui permet la production des fricatives voisées.

---

(1): [2]

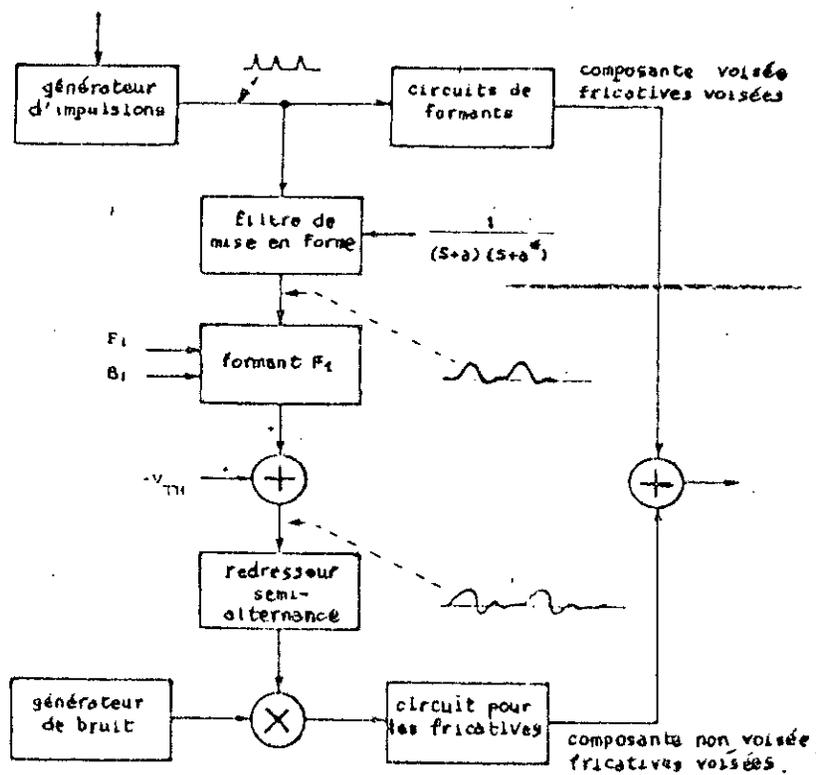


Figure : 2-6 Principe de la modulation dans le Synthétiseur de Rabiner

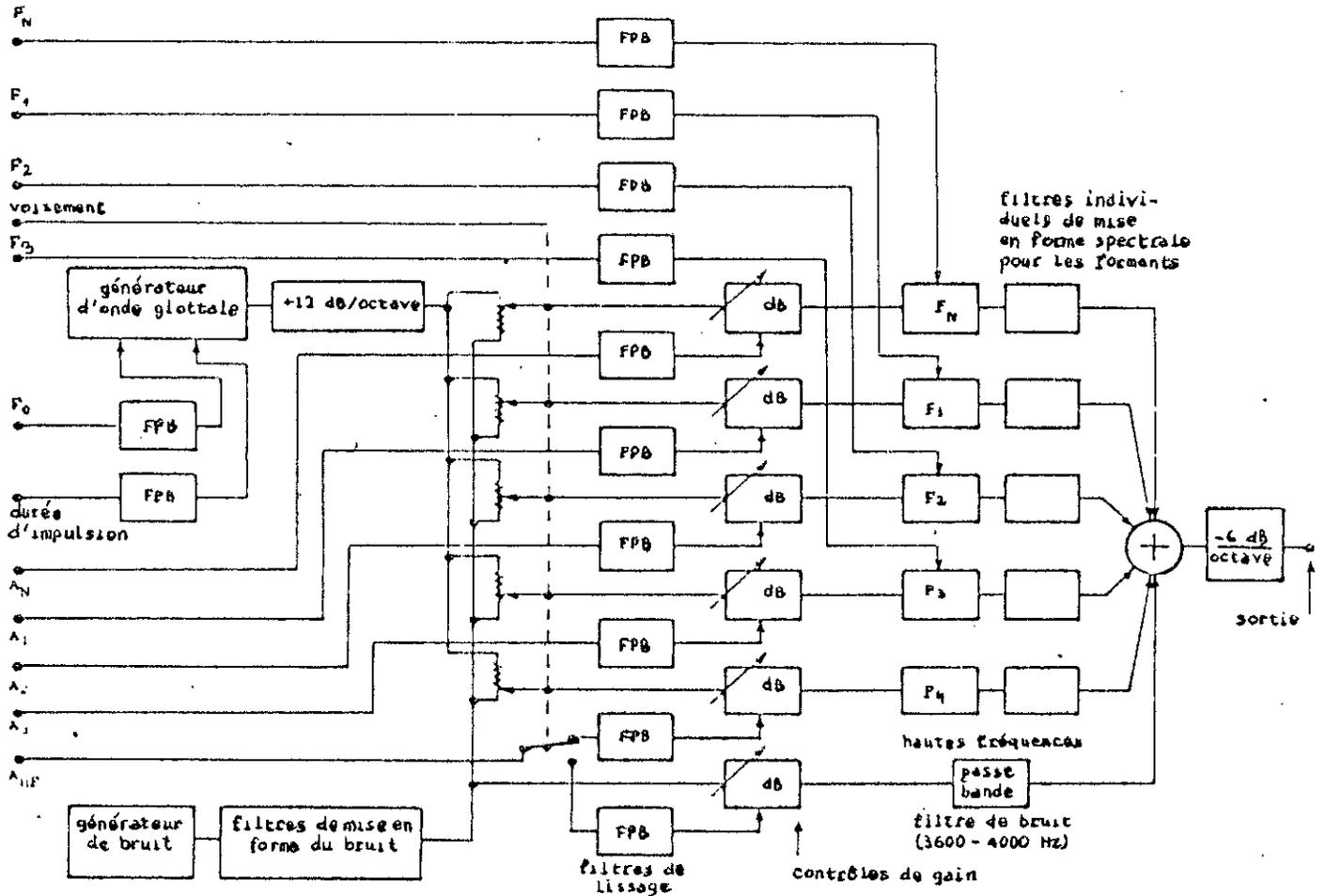


Figure : 247 Le Synthétiseur de Holmes

Par rapport au synthétiseur de RABINER , les différences se situent à plusieurs niveaux:

- la source de bruit n'est pas modulée par la source de voisement ,
- les circuits de mise en forme de chacune des branches ne sont pas précisés (ils sont fixes ou variables selon les sons).

### 5-3 COMPARAISON DES STRUCTURES PARALLELE ET SERIE

La structure serie est mieux adaptée à la production des sons voisés car elle reflète parfaitement la structure du système phonatoire (chaque pôle de la fonction de transfert traduit l'effet de l'une des cavités vocales sur le signal de la source vocale).

Par contre , la production des voyelles à l'aide d'une structure parallèle donne des résultats un peu moins bon car elle ne permet pas d'obtenir un spectre parfait à cause des interactions entre les formants voisins (ces interactions introduisent des zéros entre les formants).

Pour les sons non voisés , la structure serie est incapable de produire le spectre adéquat (spectre soit plat , soit croissant vers les hautes fréquences). Ce qui n'est pas le cas de la structure parallèle qui , grâce au réglage indépendant des gains , arrive à reproduire ces types de spectre.

En contre partie , la structure parallèle utilise plus de paramètres ; inconvénient qui pose des problèmes d'analyse mais qui permet d'obtenir une souplesse d'ajustement.

### 6 COMPARAISON DES DIFFERENTS TYPES DE SYNTHETISEURS.

Evoquons les avantages et les inconvénients respectifs des différents synthétiseurs décrits ci-dessus.

Les synthétiseurs à canaux présentent le gros avantage ,grâce à la représentation globale du signal de parole qu'ils opèrent , d'être commandés par des paramètres très facilement obtenus par analyse et de manière automatique.L'inconvénient lié à cette caractéristique est que ces paramètres sont difficiles à interpréter de manière directe ,surtout dans l'optique d'une étude acoustique du conduit vocal.

les synthétiseurs articulatoires présentent une très grande difficulté dans l'élaboration de leurs commandes ;de plus ,ils ne permettent pas d'agir de manière directe et précise sur les propriétés acoustiques du son.

Les synthétiseurs à prediction linéaire sont comparables aux synthétiseurs à canaux.

les synthétiseurs à formants sont particulièrement bien adaptés à une interprétation acoustique de leurs paramètres, ce qui en fait un outil intéressant en recherche.

D'autre part ,leur structure peut être aménagée afin de modéliser certaines caractéristiques particulières(modulation du bruit blanc pour les fricatives voisées).

les principaux inconvénients sont d'une part le nombre de paramètres qui est grand(de l'ordre d'une vingtaine) et d'autre part la difficulté à les obtenir.

Pour conclure ,dans le cadre d'une étude au niveau acoustique et au niveau perceptif des sons de parole ,le synthétiseur à formants est le mieux adapté.

## CONCLUSION

Après avoir exposé ,quoi que brièvement ,le fonctionnement de ces synthétiseurs et avoir mis en relief les avantages et inconvénients de chacun ,nous sommes en mesure d'affirmer que le synthétiseurs à formants est le mieux adapté ,remarque qui sera en fait le thème poursuivi au chapitre III.

## CHAPITRE III

### LE SYNTHETISEUR DE KLATT

#### INTRODUCTION

Ce chapitre présente en détail la conception d'un synthetiseur à formants du type KLATT; ce dernier génère la parole par une simple simulation utilisant un calculateur (personal-computer).

Le logiciel nécessaire pour le système est exposé. Enfin, nous allons montrer les performances de ce synthetiseur.

## 1 OBJECTIFS DE REALISATION DE CE SYNTHETISEUR.

L'un des buts recherché est de mettre en place un outil souple et efficace pour les études de perception. Il est indispensable que le modèle permette une bonne et simple maîtrise des paramètres acoustiques des sons produits. Plusieurs expériences de perception ont été élaborées sous le contrôle d'un calculateur et ,ainsi ,il y a des avantages à utiliser un simple ordinateur pour la generation de la parole comme étant une réponse à une excitation.

Les avantages du SOFTWARE sur la structure HARWARE sont nombreux. Le synthétiseur en SOFTWARE n'a pas besoin d'étalonnage répété; il est stable et le rapport signal sur bruit (RSB) peut être évalué comme on veut.

La configuration peut être facilement modifiée pour d'éventuelles idées nouvelles. Par exemple les voix de femmes et d'enfants peuvent être synthétisées par des modifications appropriées au niveau de l'onde de débit et du conduit vocal.

Les informations sont habituellement disponibles dans un ordinateur et peuvent être programmées pour le contrôle des paramètres ou la selection des portions du signal de sortie. Le spectre à court terme peut être cependant calculé et visualisé dans le but de comparer les signaux synthétiques et naturels de la parole.

## 2 SIMULATION DU SYNTHETISEUR

Le synthétiseur à formants de KLATT est une simulation commandée par calculateur et fonctionnant à la manière décrite en figure 3-1.

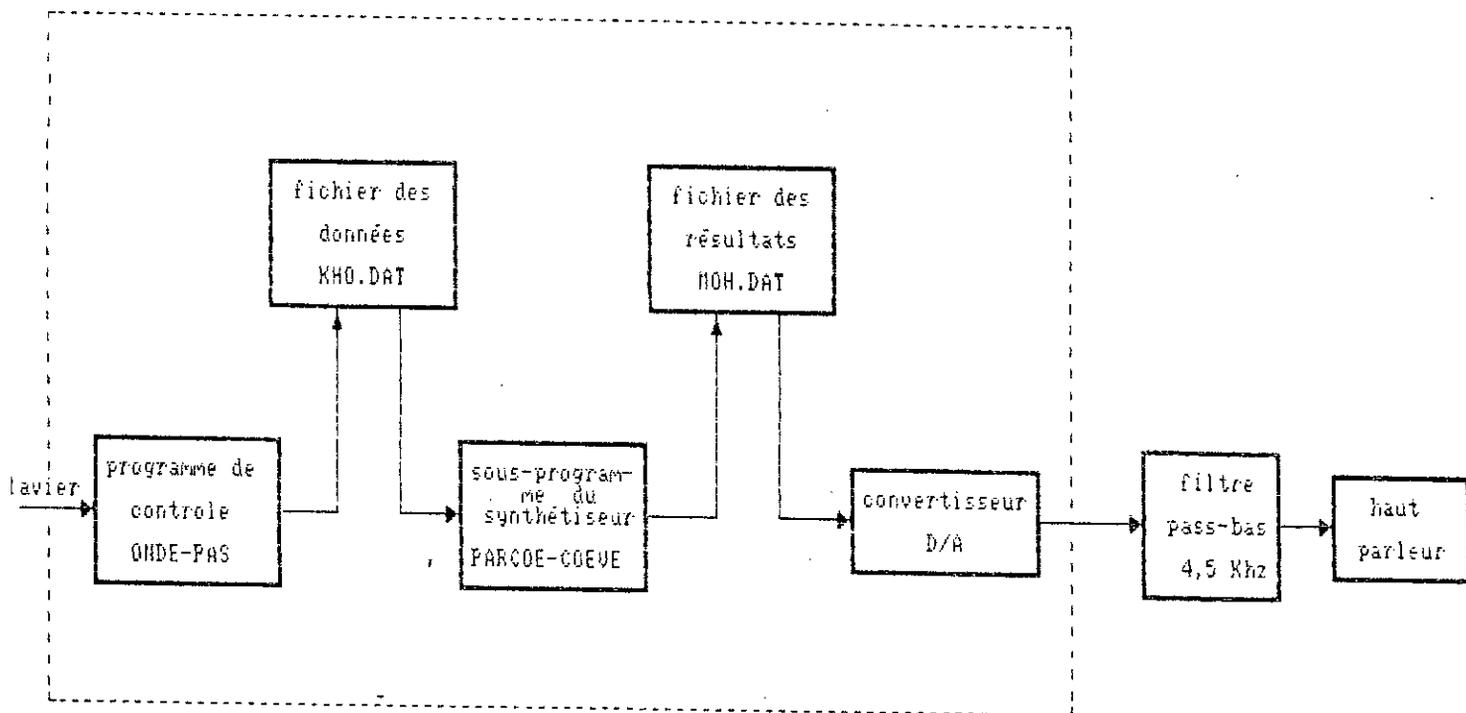


Figure 3-1 Simulation du synthetiseur.

KLATT présente un synthétiseur utilisant les variations de chaque paramètre de contrôle comme étant des fonctions du temps et ce , en utilisant un programme principal. Plus de vingt paramètres (table 1) peuvent varier de la sorte et servir d'entrées à deux autres sous-programmes.

Trois programmes semblables sont proposés en appendice. Les échantillons de l'onde de sortie sont calculés par le calculateur puis accumulés dans un fichier pour être ensuite normalement réécoutés à travers un CNA ,un filtre passe-bas et un haut-parleur.

### 3 LES DIFFERENTS OUTILS UTILISES.

Nous exposerons brièvement les différents outils utilisés et nécessaires à la compréhension du synthétiseur.

#### 3-1 TAUX D'ECHANTILLONNAGE

La plus grande énergie du signal de parole est comprise entre 80 et 8 KHZ. Quoi qu'il en soit ,des tests sur l'intelligibilité de la parole ont montré qu'il n'y avait pas de grandes variations sur cette intelligibilité si la fréquence de coupure du filtre passe-bande ne dépassait pas 5 KHZ. Ainsi donc ,un choix sur 5 KHZ (soit 10000 échantillons par seconde selon le critère de SHANON) servirait de taux d'échantillonnage au synthétiseur.

#### 3-2 TAUX DE REGENERATION DES PARAMETRES

Cette quantité représente la durée à la fin de laquelle sont régénérés les paramètres pour ainsi produire une suite de sons. Selon KLATT une valeur de 5 ms suffirait pour imiter la plus rapide des transitions de formants et les occlusives brèves.

N	variable ou constante	symbole	signification	valeur min	valeur max	valeur type
1	v	AV-dB	amplitude voisée	0	80	0
2	v	AF	amplitude fric.	0	80	0
3	v	AH	amplitude aspir.	0	80	0
4	v	AVS	amplitude q-sin.	0	80	0
5	v	AN	amplitude nasal	0	80	0
6	v	A1	amp. formant 1	0	80	0
7	v	A2	amp. formant 2	0	80	0
8	v	A3	amp. formant 3	0	80	0
9	v	A4	amp. formant 4	0	80	0
10	v	A5	amp. formant 5	0	80	0
11	v	A6	amp. formant 6	0	80	0
12	v	AB	amplitude schunt	0	80	0
13	v	F1-HZ	formant 1	150	900	450
14	v	F2	formant 2	500	2500	1450
15	v	F3	formant 3	1300	3500	2450
16	v	F4	formant 4	2500	4500	3300
17	v	F5	formant 5	3500	4900	3750
18	v	F6	formant 6	4000	4999	4900
19	c	FGP	fréquence RGP	0	600	0
20	c	FGZ	fréquence RGZ	0	5000	1500
21	c	FNP	fréquence RNP	200	500	250
22	c	FNZ	fréquence RNZ	200	700	250
23	v	B1-HZ	B. passante de F1	40	500	50
24	v	B2	B. passante de F2	40	500	70
25	v	B3	B. passante de F3	40	500	110
26	v	B4	B. passante de F4	100	500	250
27	v	B5	B. passante de F5	150	700	200
28	v	B6	B. passante de F6	200	2000	1000
29	c	BGP	B. passante de RGP	100	2000	100
30	c	BGZ	B. passante de RGZ	100	9000	6000
31	c	BNP	B. passante de RNP	50	500	100
32	c	BNZ	B. passante de RNZ	50	500	100
33	c	BGS	B. passante de RGS	100	1000	200
34	c	F0-HZ	fréquence fonda.	0	500	0
35	v	SW	interrupteur	0 (CASC)	1 (PARA)	0
36	v	NFC	nbre de filtres S	4	6	5

Table 1 : liste des paramètres de contrôle

Mais dans notre cas ,il n'est pas utile de mentionner cette quantité étant donné que nous avons effectué des essais sur des sons pris individuellement et sur une fenêtre temporelle donnée.

### 3-3 FILTRE NUMERIQUE (1)

Les circuits du synthétiseurs sont réalisés à l'aide de filtres numériques (simulés par ordinateur).

Ainsi ,si  $H(p)$  est la fonction de transfert d'un filtre analogique linéaire , $H(z)$  sera la transformée en  $z$  du filtre numérique de propriétés semblables.

Un résonateur numérique est une équation différentielle du second ordre et sa fonction de transfert est alors donnée par la relation équivalente à l'expression (8):

$$T(f) = \frac{A}{1 - B.Z^{-1} - C.Z^{-2}} \quad (21)$$

où  $Z = \exp(2j\pi fT)$  , $f$  est la fréquence variant entre 0 et 5 KHZ et  $1/T$  est la fréquence d'échantillonnage.

Par exemple ,pour  $f = 100$  HZ et une bande passante de 50 HZ ,nous pouvons tracer les deux fonctions de transfert numérique et analogique (figure 3-2).On constate qu'elles ne sont pas tout à fait identiques surtout aux fréquences hautes où la fonction de transfert analogique montre une atténuation par rapport à la fonction de transfert numérique.Cela est dû au fait que la transformation en  $Z$  introduit des distorsions par rapport à la fonction de transfert analogique mais qui ne sont pas ,heureusement ,trop importantes.

L'équation (21) peut ainsi être ramenée à l'expression suivante:

---

(1): [5]

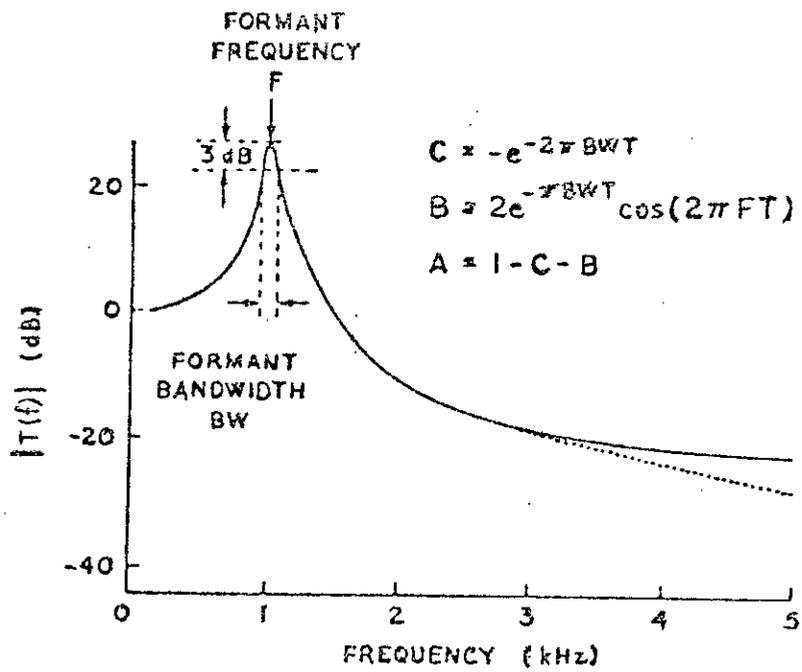


Figure : 3-2 Fonctions de transfert numérique-analogique

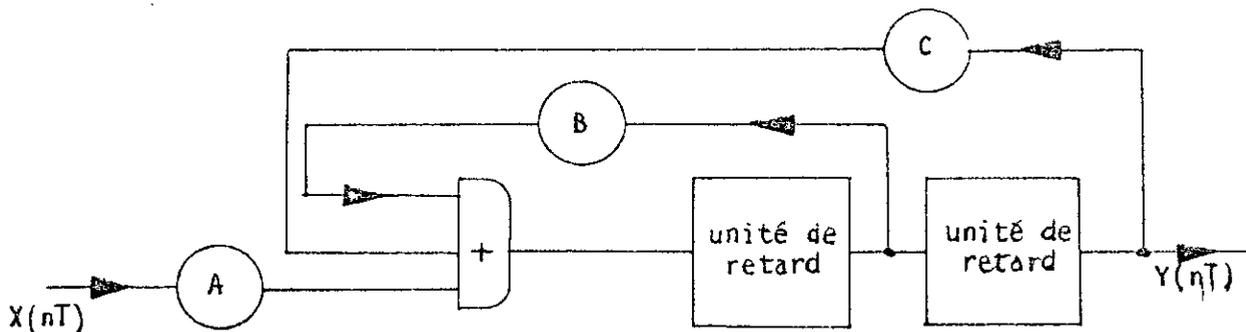


Figure 3-3 : Résonateur numérique

$$Y(nT) = A.X(nT) + B.Y(nT-T) + C.Y(nT-2T) \quad (22)$$

où  $X(nT)$  est la séquence d'entrée et  $Y(nT)$  est la séquence de sortie.

Les deux paramètres qui rendent compte des caractéristiques d'un circuit de formants sont la fréquence  $f$  et la bande passante  $BW$ . Les relations qui relient  $f$  et  $BW$  aux trois paramètres  $A, B$  et  $C$  sont les suivantes:

$$\begin{aligned} C &= -\exp(-2\pi.BW.T) \\ B &= 2\exp(-\pi.BW.T).\cos(2\pi.f.T) \\ A &= 1 - C - B \end{aligned} \quad (23)$$

Dans l'équation (22), l'échantillon  $Y(nT)$  peut être calculé de plusieurs façons, ce qui donne diverses possibilités de réalisations du filtre numérique.

Le choix entre ces structures est basé sur la précision et le temps de calcul. La figure 3-3 montre une réalisation qui donne une bonne précision.

### 3-4 ANTIRESONATEUR NUMERIQUE (ANTIFORMANTS) (1)

Il peut être réalisé par une mince modification de ces équations. La réponse fréquentielle d'un antiresonateur est l'image de la réponse dressée sur la figure 3-2 (remplacer dB par -dB).

Il est utilisé dans la synthèse pour modéliser le spectre de la source voisée et il est aussi utilisé pour simuler les effets du conduit nasal dans le modèle série de la fonction de transfert du conduit vocal.

La sortie d'un antiformant  $Y(nT)$  est associée à l'entrée  $X(nT)$  par l'équation:

---


$$(1): [5]$$

$$Y(nT) = A'.X(nT) + B'.X(nT-T) + C'.X(nT-2T) \quad (24)$$

où  $X(nT-T)$  et  $X(nT-2T)$  sont les deux valeurs préalables de  $X(nT)$  et les constantes  $A'$ ,  $B'$  et  $C'$  sont données par :

$$\begin{aligned} A' &= 1/A \\ B' &= -B/A \\ C' &= -C.A \end{aligned} \quad (25)$$

et  $A$ ,  $B$  et  $C$  sont obtenues par les équations (23) en inversant l'antiformant  $f$  et la bande  $BW$  à  $-3dB$ .

### 3-5 FILTRE PASSE-BAS

Pour un cas particulier, la fréquence  $f$  du filtre numérique peut être atténuée en passant par un filtre passe-bas ayant une atténuation nominale à  $-12$  dB/octave et à  $-3$  dB correspond la bande passante  $BW/2$ .

La source de voisement contient un filtre numérique "RGP" (figure 3-4) utilisant un tel filtre qui transforme l'impulsion glottique en une vibration ayant une onde et un spectre similaire à une voix normale.

Un second filtre numérique "RGS" est utilisé en passe-bas pour filtrer la voix normale pour reproduire l'onde glottique d'une manière quasi-sinusoidale (figure 3-5).

## 4 DESCRIPTION DE LA STRUCTURE DU SYNTHETISEUR (1)

La structure est schématisée en figure 3-4. On peut décomposer le synthétiseur en deux sous-ensembles: les sources de sons et le système de filtrage.

---

(1): [5]

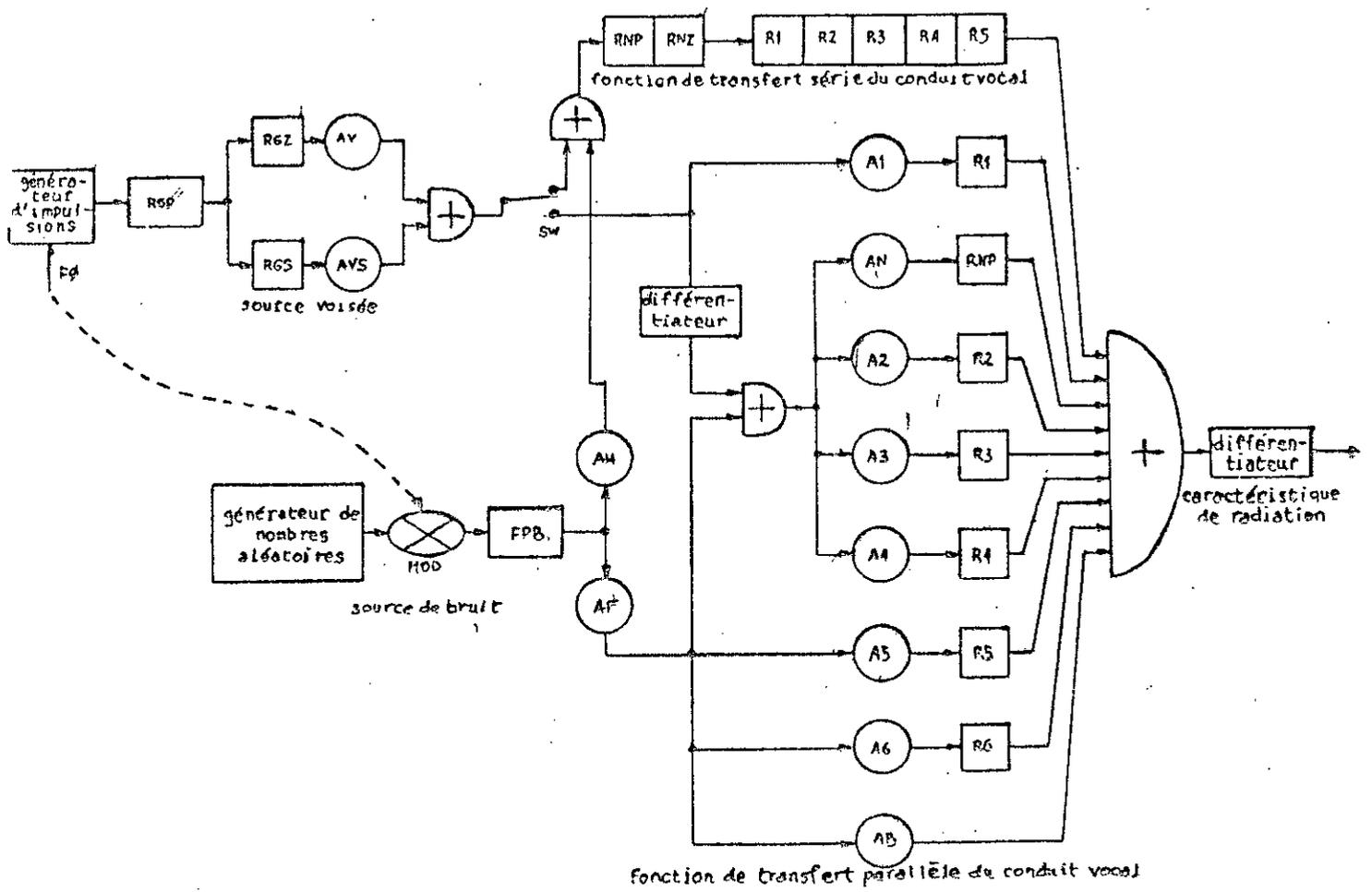


Figure : 3-4 Schéma bloc du Synthétiseur

#### 4-1 LES SOURCES DE SONS

##### 4-1-1 Source voisée

La structure de cette source est schématisée sur la gauche de la figure 3-4. Les paramètres variables de contrôle sont utilisés pour spécifier la fréquence fondamentale "FO", l'amplitude du voisement normal "AV" et l'amplitude du voisement quasi-sinusoidal "AVS".

L'élément de base est une source d'impulsions dont le spectre doit être plat jusqu'à 4 KHZ environ. Le train d'impulsions passe à travers un filtre passe-bas "RGP" pour produire une onde lisse qui ressemble à la vitesse volumique de l'onde glottique. La fréquence de résonance FGP est fixée à 0 HZ et BGP à 100 HZ.

Les impulsions ainsi filtrées ont un spectre dont la pente de décroissance est de 12 dB/octave au dessus de 50 HZ. Le signal obtenu à la sortie du filtre "RGP" est utilisé de deux manières: une pour générer un voisement normal et l'autre pour générer un voisement quasi-sinusoidal .

a -voisement normal (source AV).

L'onde générée par le filtre "RGP" n'a pas le même spectre de phase qu'une vibration glottique typique et ne contient pas de zéros dans le spectre du type qui apparaît souvent dans une voix naturelle. Ces différences sont très importantes au niveau de la perception .

L'antirésonateur "RGZ" est utilisé pour modifier les détails du modèle du spectre de la source.

L'onde et l'enveloppe spectrale ainsi générées sont illustrées en figure 3-5-a.

b - voisement quasi-sinusoidal (source AVS).

Cette onde est obtenue en filtrant "passe-bas" une impulsion par des filtres numériques "passe-bas" "RGP" et "RGS". La fréquence de contrôle de "RGS" est fixée à 0 pour produire un filtre passe-bas et "BGS" = 200 HZ détermine la fréquence de coupure au-delà de laquelle des harmoniques sont fortement atténuées (figure 3-5-b). (1)

#### 4-1-2 Sources de bruit.

La source de bruit pour les sons non-voisés est simulée dans le synthétiseur par un générateur de nombres pseudo-aléatoires, un modulateur, un filtre numérique passe-bas à -6 dB/octave et deux amplitudes de contrôle "AF" pour les sources fricatives et "AH" pour les sources d'aspirations.

##### a - source fricative

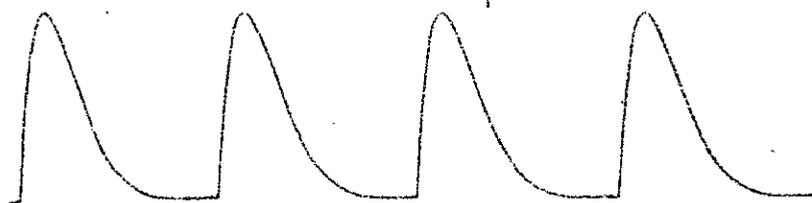
Le spectre de la source fricative est approximativement plat, et la distribution d'amplitude est gaussienne. Les signaux produits par le générateur ont une distribution uniforme d'amplitude entre les limites déterminées par la valeur du paramètre de contrôle "AF". Une distribution d'amplitude pseudo-gaussienne est obtenue dans le synthétiseur en sommant 16 des nombres produits par le générateur de nombres aléatoires.

La source de bruit attaque une fonction de transfert qui simule grossièrement l'impédance du conduit vocal vu de la source de bruit dans le cas réel.

Cette fonction de transfert est celle d'une intégration. L'intégrale est approximée par un filtre numérique passe-bas du premier ordre "LFF", illustré en figure 3-4. Les échantillons de sortie  $Y(nT)$  sont liés à  $X(nT)$  par l'équation:

---

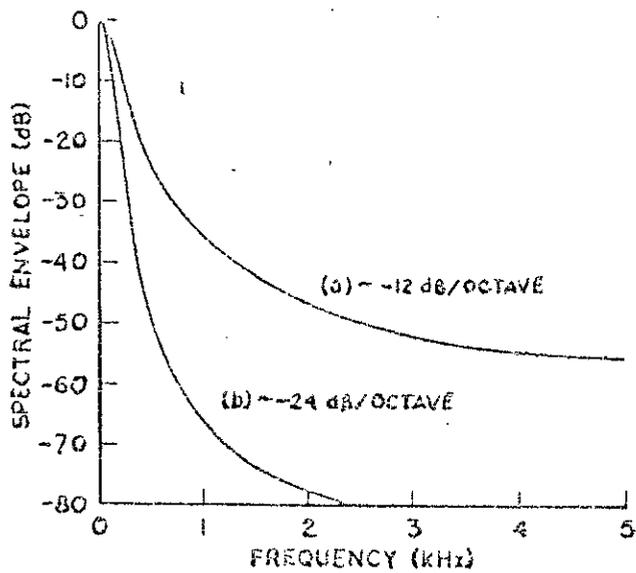
(1): [5]



(a) NORMAL VOICING WAVEFORM



(b) SMOOTHED VOICING WAVEFORM



(c) VOICING SOURCE MAGNITUDE SPECTRA

Figure : 3-5 (a) voisement normal  
 (b) voisement quasi-sinusoidal  
 (c) low spectre respectif

$$Y(nT) = X(nT) + Y(nT-T) \quad (26)$$

Un exemple de débit de bruit fricatif synthétisé est montré en figure 3-6.

Pour certains sons (par exemple les fricatives voisées), la source de bruit et la source voisée coexistent et le bruit est modulé en amplitude à la fréquence fondamentale "FO". Pour simplifier la réalisation, le bruit est modulé par un signal carré de fréquence "FO". Le degré de modulation est fixé à 50 % dans le synthétiseur. L'amplitude de bruit fricatif est déterminée par "AF"; une valeur de 60 dB générera une forte fricative tandis qu'une valeur nulle, éteint automatiquement la source fricative.

#### b - source d'aspiration

Cette source est essentiellement la même que la friction, exceptée qu'elle est générée dans le larynx. L'aspiration passe à travers le modèle série du conduit vocal. L'amplitude de cette source est déterminée par "AH", une valeur de 60 dB donne une forte aspiration alors qu'une valeur de zéro coupe la source d'aspiration.

#### 4-2 LE SYSTEME DE FILTRAGE

Le système de filtrage du synthétiseur se compose de deux branches : l'une série, représentant les propriétés de résonance du conduit vocal pour les sources voisées, et l'autre en parallèle représentant les propriétés de résonance du conduit vocal durant la production de fricatives (cependant, la branche parallèle peut aussi servir à simuler les sons voisés) (figure 3-4).

Un interrupteur SW permet la disjonction entre un fonctionnement "parallèle du synthétiseur" et un fonctionnement "série".

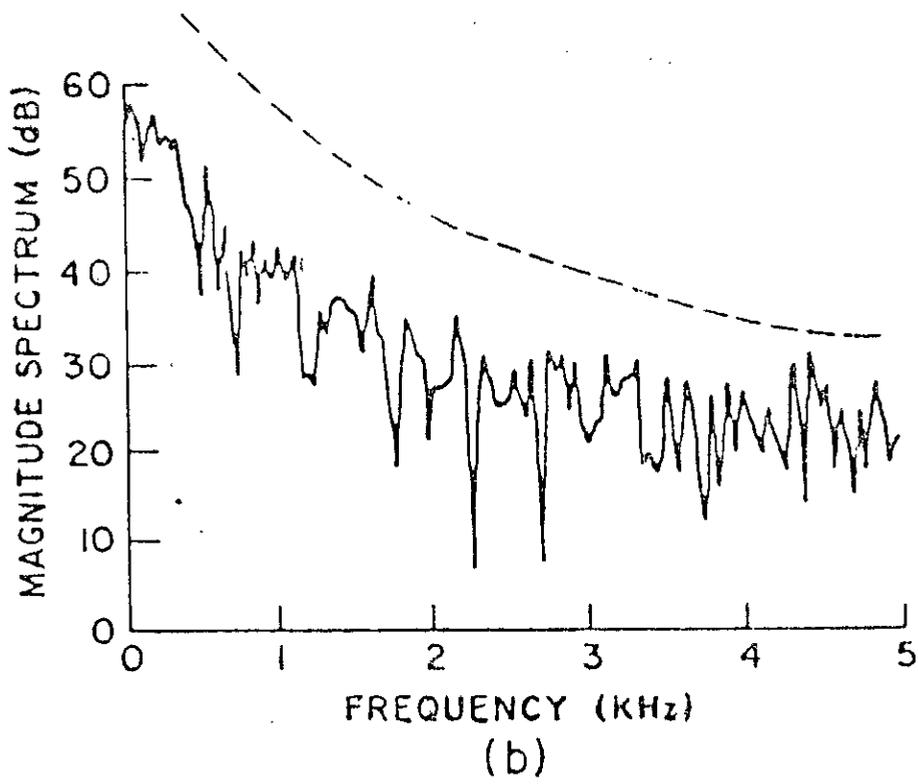
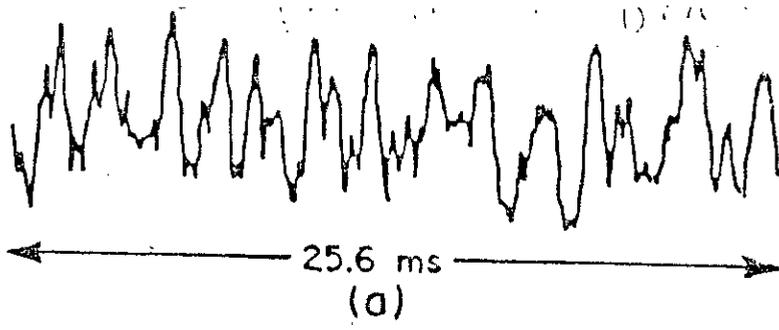


Figure 3-6 : Exemple d'une source fricature avec son spectre

Ce choix a été établi afin de pouvoir apprécier la qualité du son pour les deux configurations.

#### 4-2-1 Modèle série du conduit vocal (SW = 0)

Cette configuration illustre parfaitement les voyelles non-nasalisées et les voyelles nasalisées. Elle se distingue par l'association en série de cinq résonateurs "R1", "R2", "R3", "R4", "R5" et de deux autres résonateurs "RNP" et "RNZ" (les "Ri" et "RNP" ont même nombre de pôles).

Les résonateurs "R<sub>i</sub>" assurent la représentation formantique de chaque voyelle.

Les nasales, telles que /z/, /z̃/, /ɛ̃/, /ɛ̃̃/ sont assurées par les filtres "RNP" et "RNZ". "RNP" possède une fréquence de pôle nasal "FNP" et "RNZ" possède une fréquence de zéro nasal "FNZ". Si "FNP" = "FNZ", le phénomène de nasalisation n'a pas lieu et dont l'effet est la réduction du premier formant "F1" dû à la présence de cette paire pôle/zéro, justement, autour de "F1". Donc, par exemple, il suffit de fixer "FNP" à 270 HZ et de faire varier "FNZ".

Si nous ne considérons pas l'effet de "RNP" et "RNZ", la fonction de transfert de la configuration série sera:

$$T(f) = \prod_{n=1}^5 \frac{A(n)}{1 - B(n).z^{-1} - C(n).z^{-2}} \quad (27)$$

où "A(n)", "B(n)" et "C(n)" sont des constantes associées au n-ième formant et qui peuvent être calculées indépendamment d'après (23).

#### 4-2-2 Modèle parallèle du conduit vocal

##### configuration

Cette configuration est assurée lorsque  $SW = 1$  (figure 3-4). Il y a six résonateurs accompagnés chacun d'un contrôle d'amplitude. Le sixième formant a été ajouté à la branche pour la synthèse de très grandes fréquences de bruit tels que le /s/ ou le /z/ au lieu de déplacer "F5" vers les hautes fréquences.

La source de bruit attaque directement les filtres R2, R3, R4, R5 et R6. Une entrée directe (un shunt) du bruit est assurée par le contrôle de gain Ab car les fonctions de transfert pour le /f/ , /v/ , /p/ et /b/ ne contiennent pas de pics résonants remarquables.

Pour un essai de sons voisés avec cette configuration , le filtre R1 est attaqué par la source voisée tandis que R2, R3 et R4 , eux , par sa dérivée première (cette technique ayant pour but de supprimer l'énergie basse fréquence des formants R2, R3, R4 qui risquerait de perturber la réponse fréquentielle dans les régions du premier formant "F1").

En ce qui concerne RNP , il n'est pas vraiment indispensable , même pour la synthèse de sons nasalisés , vu que la fonction de transfert pour ce modèle peut elle-même introduire des zéros.

L'expression de la fonction de transfert est la somme de chaque fonction de transfert prise individuellement ; elle est semblable à l'expression (9).

#### 5 CARACTERISTIQUES DE RAYONNEMENT.

Le module , libélé "caracteristiques de rayonnement" en figure (3-4) , représente l'effet des exemples de directivités du son rayonnant au niveau des lèvres et de la tête comme étant une fonction de la fréquence.

La transformation est simulée dans le synthétiseur en prenant la première différence du débit "lèvres-nez":

$$P(nT) = U(nT) - U(nT-T). \quad (28)$$

Les caractéristiques de rayonnement ajoutent une montée graduelle dans le spectre total.

## 6 IMPLEMENTATION DU SYNTHETISEUR.

Le logiciel simulant le synthétiseur a été écrit en Turbo-Pascal vu la grande souplesse que présente ce langage.

Il se compose d'un programme principal ,permettant la gestion des paramètres de contrôle ,et de deux autres sous-programmes , l'un transformant les paramètres de contrôle en coefficients nécessaires au fonctionnement du synthétiseur et l'autre simulant les parties sources de sons et système de filtrage ,bref tout ce qui est calcul (l'initialisation du synthétiseur se fait au niveau de cette procédure).

### 6-1 DESCRIPTION DU PROGRAMME PRINCIPAL

Comme le montre son organigramme ,sa lecture n'est pas très difficile.

Au début ,la lecture de la durée de la fenêtre temporelle UTTDUR est obligatoire:cette dernière représente ,en milliseconde,la durée de l'articulation.Après quoi ,on ouvre deux fichiers ,l'un pour le fichier de données ,KHO.DAT ,composé de 36 paramètres et l'autre représentant le fichier résultat (onde de sortie) MOH.DAT.

Après la lecture de KHO.DAT ,l'appel des deux sous-programmes se fait automatiquement ,calculant ainsi l'onde de sortie et l'enregistreur dans le fichier MOH.DAT.

Ainsi à chaque nouveau vecteur ,KHO.DAT ,correspond une nouvelle sortie MOH.DAT qui pourra aisément être visualisée.

#### 6-2 DESCRIPTION DU SOUS-PROGRAMME "PARCOE"

Ce dernier transforme les paramètres lus à partir de KHO.DAT en coefficients.

En réalité ,ce sous-programme utilise lui-même un autre sous-programme qui ,pour un filtre numérique donné ,transforme le formant "F" et la bande passante "BW" à -3dB en trois coefficients A,B,C et une fonction transformant les contrôles de gains de l'unité "decibel" en une unité linéaire.

Nous avons noté en tout 45 coefficients correspondant à onze filtres numériques et douze contrôles de gains.

#### 6-3 DESCRIPTION DU SOUS-PROGRAMME "COEWAV"

C'est le sous-programme le plus long vu qu'il regroupe un nombre considérable d'opérations et surtout d'équations récurrentes toutes semblables aux expressions (22) et (24).

En outre ,c'est lui qui initialise le synthétiseur ,c'est-à-dire qu'il met à zéro toutes les périphéries du synthétiseur (entrée/sortie de chaque filtre numérique)..

Il se compose de deux boucles:

- une boucle principale représentant le nombre de périodes envisagées pour le calcul d'un son voisé ,
- une boucle secondaire représentant le nombre d'échantillons par période qui est directement lié au taux

d'échantillonnage et à la fréquence fondamentale qui dans notre cas a été prise à 120 HZ (soit  $10000/120 \approx 85$  échantillons par période).

La boucle principale commande la source d'impulsions et le calcul de la source de son et de la fonction de transfert se fait sur une période donnée.

L'interrupteur "SW" permet la disjonction entre un fonctionnement parallèle (SW = 1) du synthétiseur et un fonctionnement série (SW = 0).

Pratiquement, la modulation de la source de bruit par la source d'impulsions se fait de la manière suivante:

- on divise la source de bruit par deux lorsque la forme de l'onde à la sortie de RGP est négative ou nulle.

## 7 RESULTATS GRAPHIQUES.

Les exemples pris lors de l'implémentation du synthétiseur concernent les voyelles /e/ ,/i/ ,/a/ , les fricatives /f/ ,/s/ , et la plosive /p/. Il n'y a pas eu de choix bien précis, mais nous avons essayé d'avoir des résultats sur les différents sons existant en "français".

Bien sûr une extension sur les autres voyelles et consonnes peut être envisagée étant donné que le synthétiseur est habilité à modéliser toute forme de son.

La table 2 indique les différentes valeurs des paramètres de contrôle utilisés lors de la création de ces "articulations".

La fenêtre temporelle est de 100 ms, cela veut dire que la durée de l'articulation est de 0.1 s (il est à noter que cette valeur est fixée par l'utilisateur).

PARAMETRES	/a/	/e/	/i/	/s/	/f/	/p/
↓						
AV (dB)	60	60	60	-60	-60	-60
AF	-60	-60	-60	50	26	0
AH	-60	-60	-60	/	/	-60
AVS	-60	-60	-60	-60	-60	-60
AN	-60	-60	-60	-60	-60	-60
A1	/	/	/	0	0	0
A2	/	/	/	-5	35	2
A3	/	/	/	-5	34	2
A4	/	/	/	10	34	0
A5	/	/	/	10	-5	2
A6	/	/	/	0	0	0
AB	/	/	/	0	57	63
F1 (HZ)	750	370	250	320	501	100
F2	1244	2700	2900	3000	1300	1000
F3	2510	1900	2000	2530	2250	2181
F4	3030	3468	3400	3900	3159	3247
F5	3906	3496	3800	4600	4154	4363
F6	4900	4900	4900	4900	4900	4900
FGP	0	0	0	0	0	0
FGZ	1500	1500	1500	1500	1500	1500
FNP	900	900	900	900	900	900
FNZ	-900	-990	-900	-900	-900	-900
B1 (HZ)	75	55	55	200	200	150
B2	60	100	100	200	200	154
B3	84	74	50	200	200	225
B4	86	110	80	150	200	490
B5	109	132	148	150	223	460
B6	1000	1000	1000	1000	1000	1000
BGP	100	100	100	100	100	100
BGZ	6000	6000	6000	6000	6000	6000
BNP	500	500	500	500	500	500
BNZ	500	500	500	500	500	500
BGS	200	200	200	200	200	200
F0 (HZ)	120	120	120	120	120	120
SW	0	0	0	1	1	1
NFC	5	5	5	/	/	/

Table 2 : les différentes valeurs des paramètres de contrôle

Pour les sons voisés ,la fréquence fondamentale a été choisie à 120 HZ (à vous de trouver à quel locuteur elle appartient).

Le signal à la sortie de "RGP" a aussi été visualisé. (onde glottique).

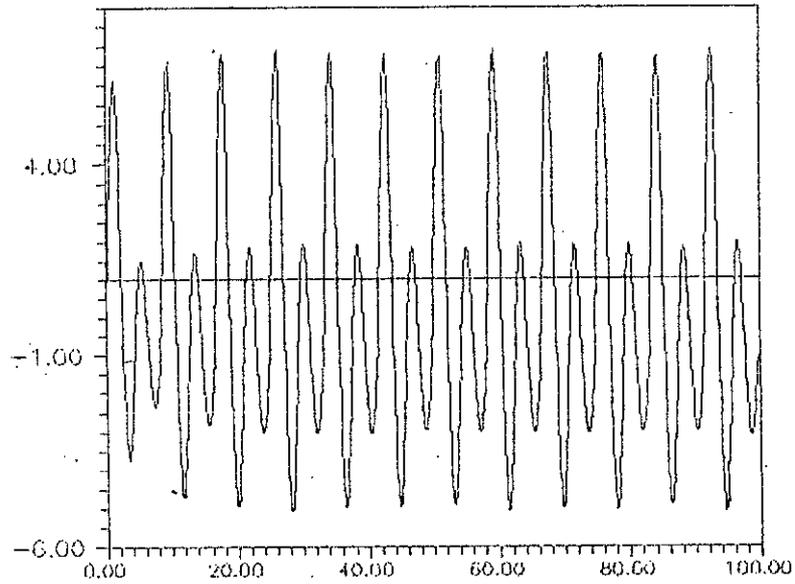
Le résultat graphique obtenu est très satisfaisant .Une simple visualisation graphique montre une périodicité de chacune des voyelles correspondant à la fréquence fondamentale 120 HZ ,point fondamental pour les sons voisés (figure 3-7 et 3-8).

ON voit que le /a/ présente plus d'harmoniques que le /e/ et qui lui même en présente plus que le /i/ ,chose évidente.

Les fricatives /s/ et /f/ représentent bien une réponse à un bruit (figure 3-9)

L'onde glottique est le signal à la sortie du filtre "RGP" dont les coefficients A,B et C ont été calculés à partir de la fréquence de résonance "FGP" = 0 HZ et la bande passante "BGP" valant 100 HZ (figure 3-10).

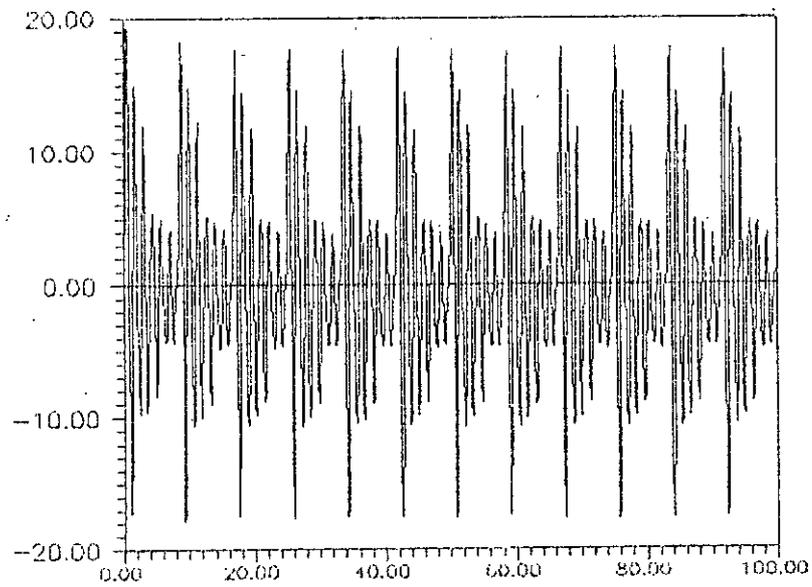
Amplitude



Temps(ms)

VOYELLE /i/

Amplitude



Temps(ms)

VOYELLE /a/

Figure: 3-7 Le signal temporel de /a/et/i/

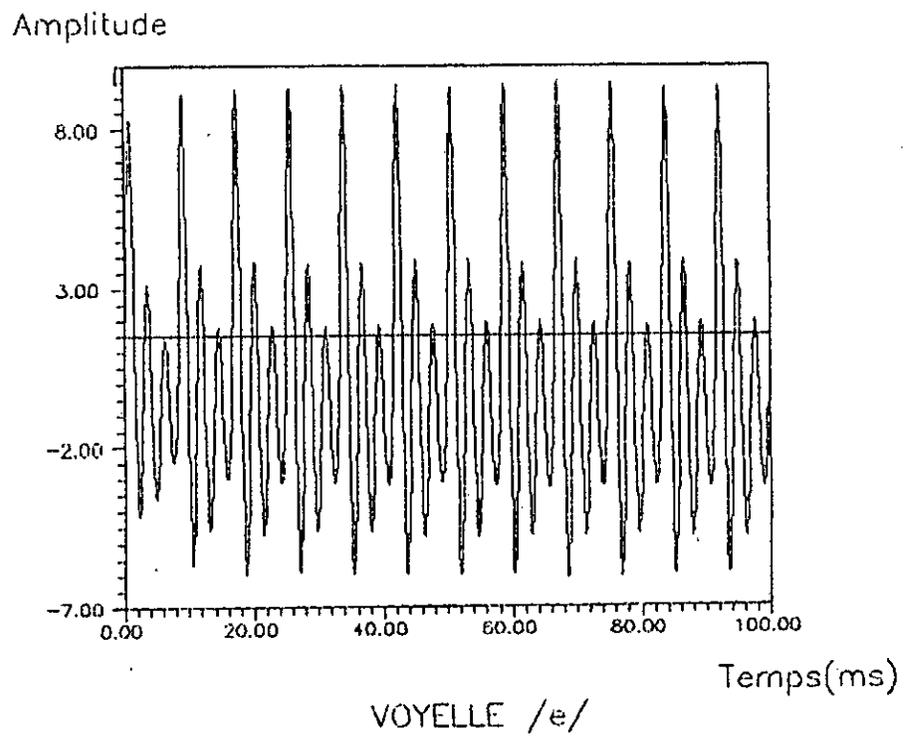
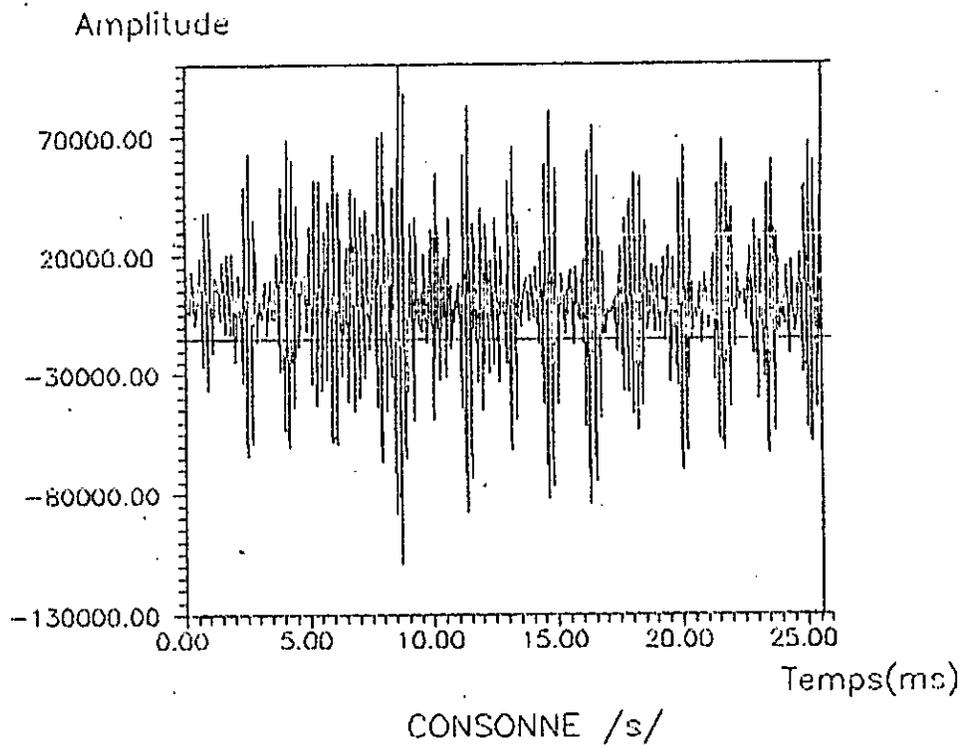


Figure: 3-8 Le signal temporel de /e/et/s/

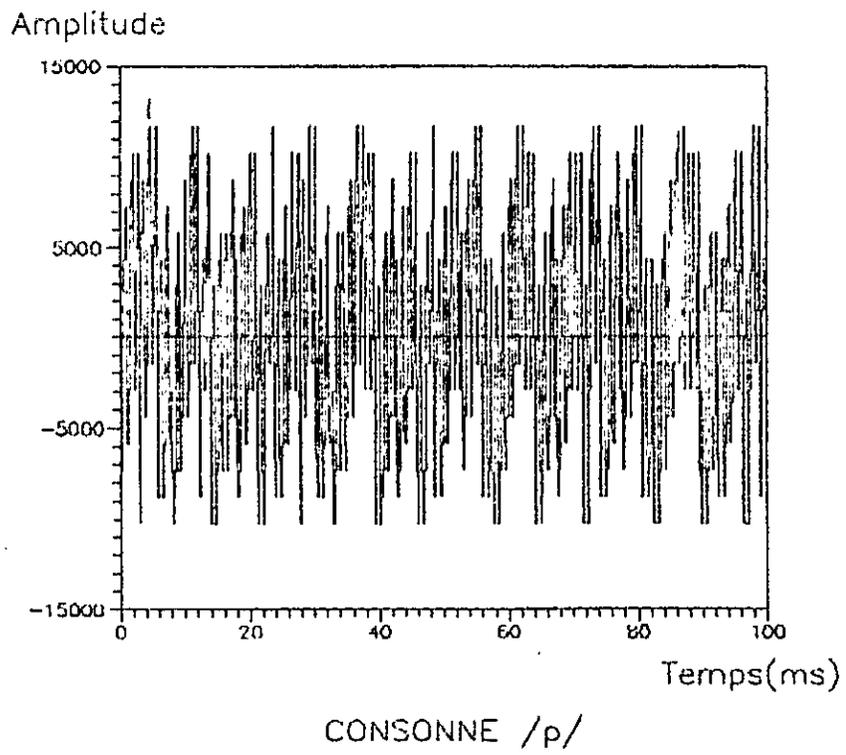
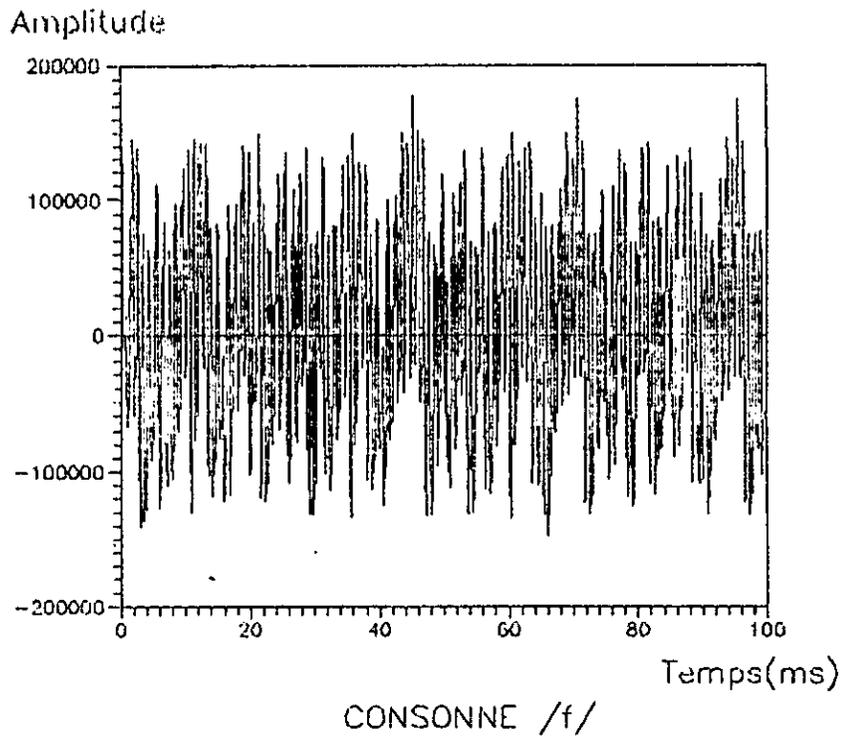


Figure: 3-9 Le signal temporel de /f/et/p/

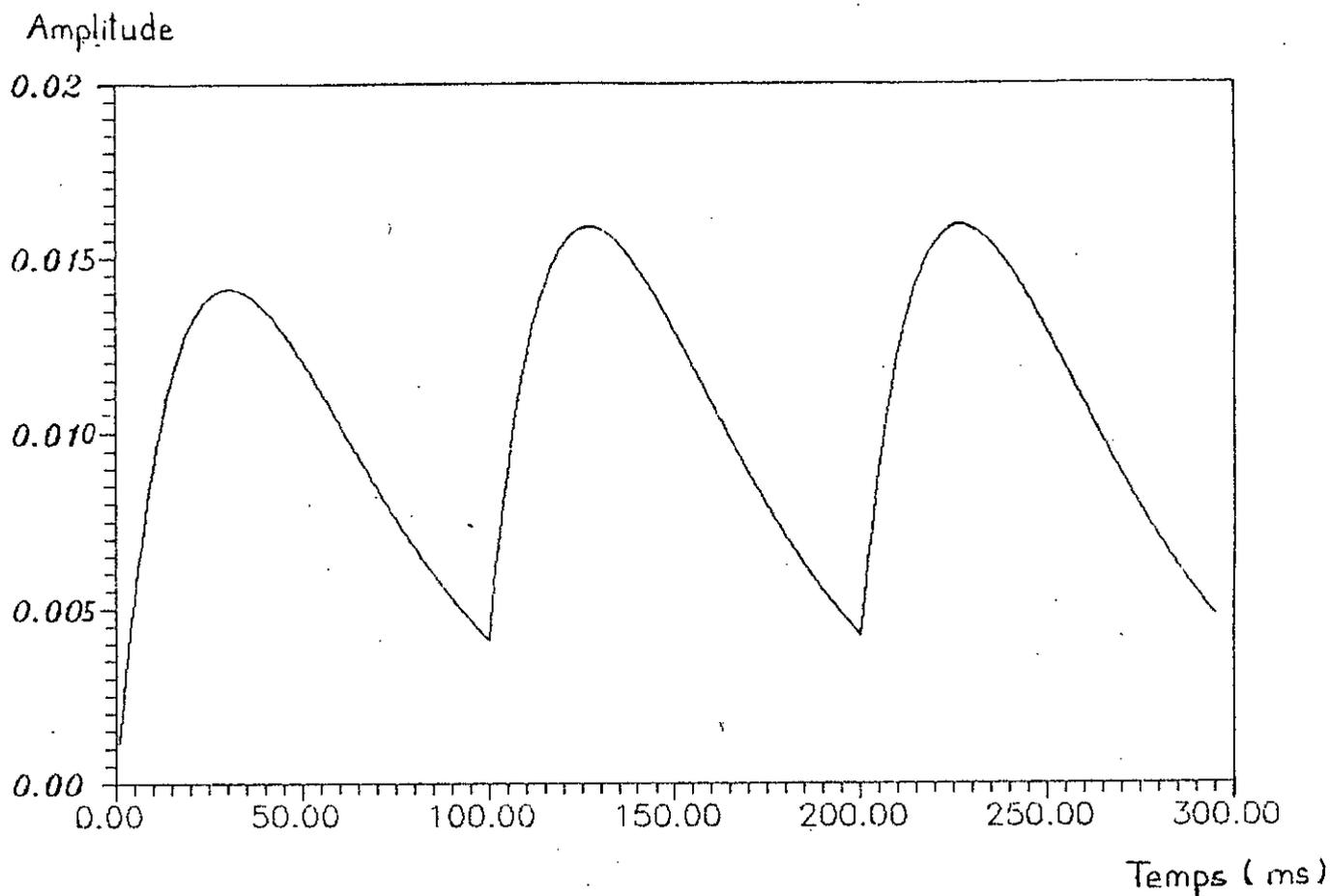


Figure 3\_10 : L'onde glottique pris sur trois périodes  
durée : 30 ms.

## CONCLUSIONS GENERALES

Notre étude a été consacrée d'une part à l'étude du synthétiseur à formants de KLATT ,et d'autre part à son implémentation au niveau du centre de calcul de l'Ecole Nationale Polytechnique d'Alger.

Nous avons d'abord effectué une étude sur les différents modèles de synthétiseurs à formants existants ,ce qui nous a permis d'entamer la structure hybride (serie-parallèle) de KLATT.

De plus ,nous avons exposé les différentes méthodes d'extraction des paramètres en discutant les limites ainsi que les avantages et les inconvénients de chacune des méthodes.

A travers l'implémentation du synthétiseur sur ordinateur , nous pouvons affirmer que le synthétiseur de KLATT est une simulation admirable de la synthèse de la parole ou de la synthèse par ordinateur.

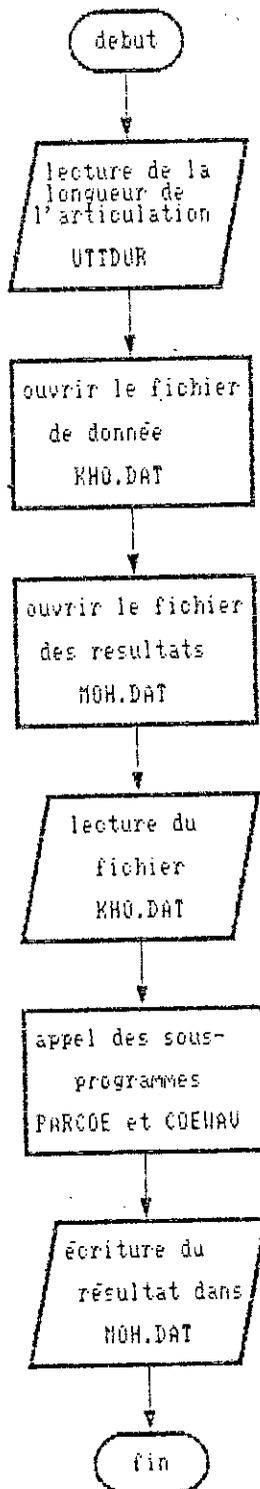
Quelques modifications nous paraissent déjà évidentes pour son application à d'autres langages "phonétiquement" plus simple tel que "l'arabe".

Une autre idée serait d'ajouter un logiciel permettant la conversion "orthographique-phonétique" c'est à dire qu'il suffirait d'écrire sur le visuel une phrase et nous aurons automatiquement la conversion en code phonétique et après traitement,nous aurons une réponse au niveau d'un haut-parleur accompagnant le système , chose qui est plutôt délicate étant donné que ceci est le thème de plusieurs recherches effectuées en Algérie (laboratoire de communication :C. D. T. A).

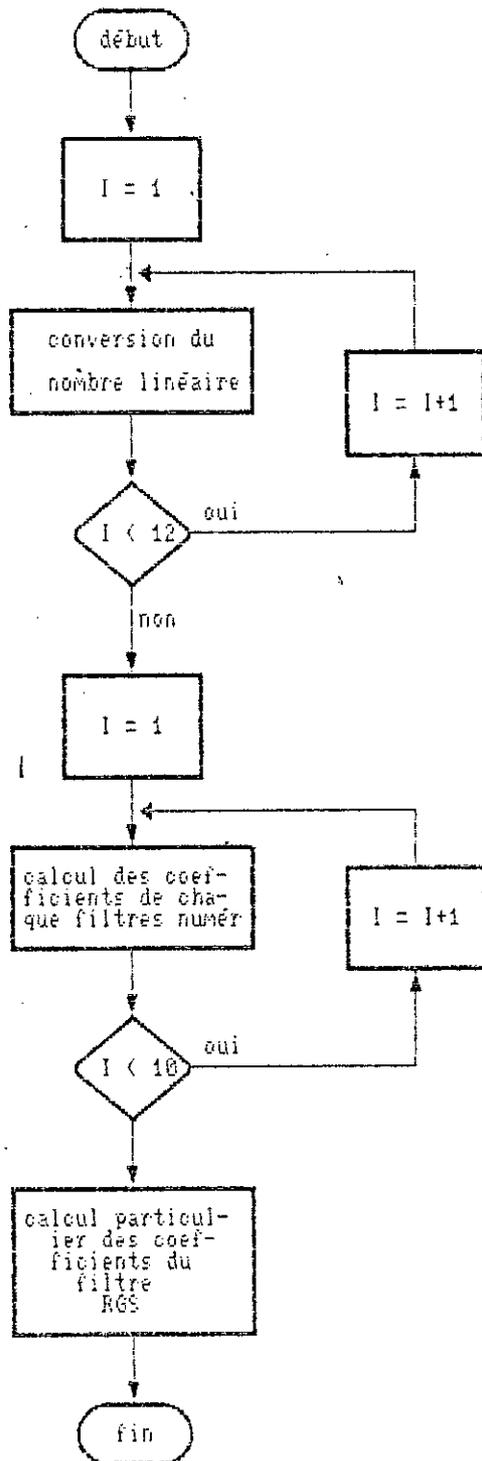
En bref ,voilà ce que fût notre travail basé énormément sur le modèle de KLATT qui simule parfaitement et avec une grande souplesse ,que ce soit pour l'anglais ou pour le français ,toutes les formes des sons existants.

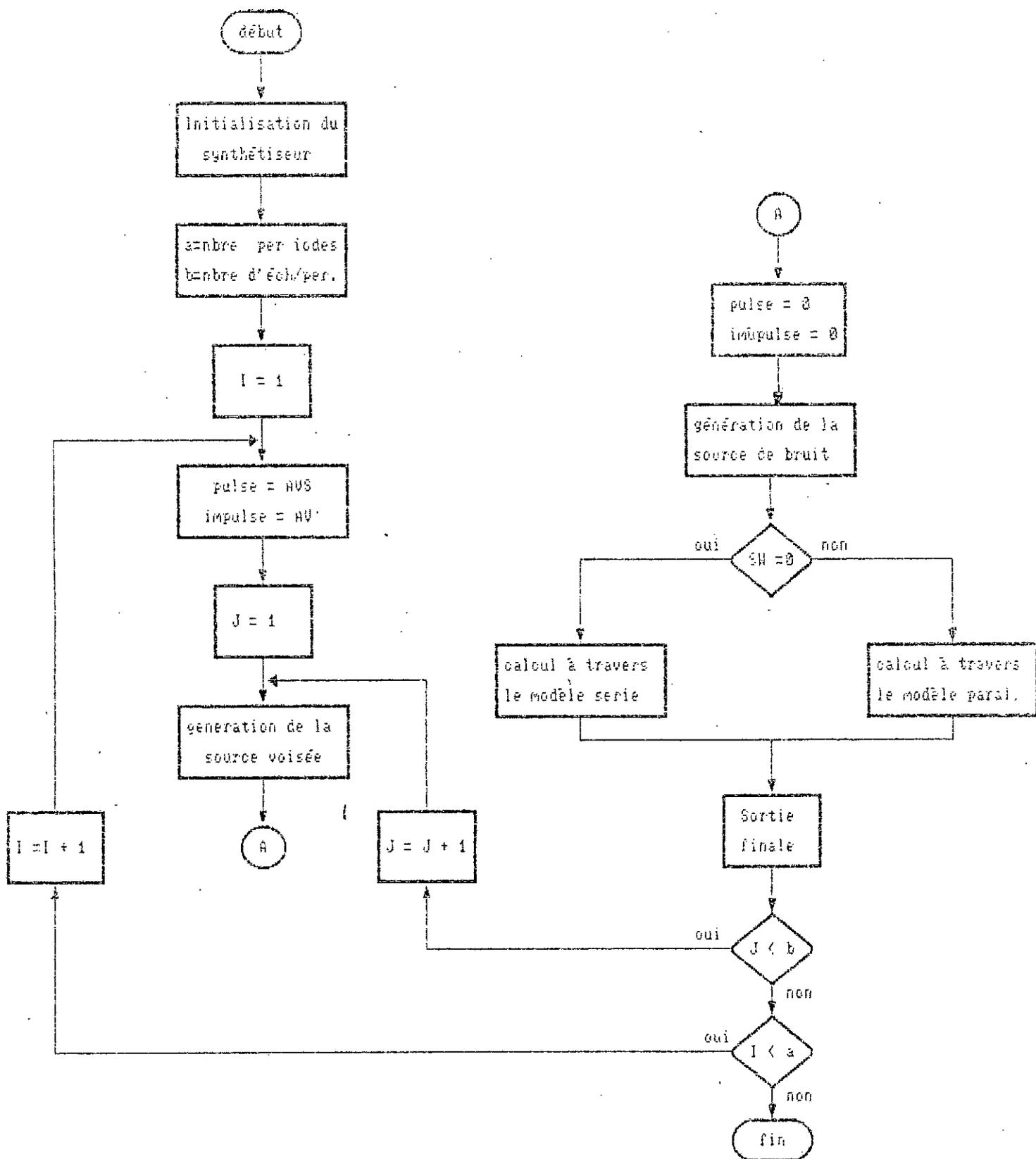
Enfin ,il est souhaité que ce sujet soit poursuivi en vue d'une application pratique et efficace dans la vie courante.

ANNEXE  
DES  
ORGANIGRAMMES  
ET DU  
PROGRAMME



- ORGANIGRAMME DU PROGRAMME PRINCIPAL -





-ORGANIGRAMME DU SOUS-PROGRAMME COENAU-

IMPLEMENTATION DU LOGICIEL

```
Program ESSAI2(input,output);
```

```
CONST Wsize=5000;
Type C=Array[1..50] of real;
      I=Array[1..39] of Integer;
Sortie=Array[1..Wsize] of Real;
Var
  Ipar :I;
  Opar :C;
  fichier,fich :text;
  j      :Integer;
  Uttdur:Integer;
  Sort:Sortie;
```

```
{*****}
{*****}
```

```
Function Getamp(Ndb:integer) : real;
```

```
Const
  Dtable :Array[1..11] of Real =
(1.8,1.6,1.43,1.26,1.12,1.0,0.89,0.792,0.702,0.623,0.555);
  Stable :Array[1..28] of real=
(65536,32768,16384,8192,4096,2048,
1024,512,256,128,64,32,16,8,4,2,1,0.5,0.25,0.125,
0.0625,0.0312,0.0156,0.0078,0.0039,0.00195,0.000975,0.000487);
Var b1,b2,b3:integer;
    x1:REAL;
    x2:real;
```

```
Begin
```

```
  b1 := Ndb ;
  getamp := 0 ;
  IF b1 > -72 Then
    Begin
      IF b1 > 96 then
        begin
          b1:=96;
          b2:=b1 div 6;
          b3:=b1-(6*b2);
          x1:=stable[17-b2];
          x2:=dtable[6-b3];
          getamp:=x1*x2;
        end
      else
        begin
          b2:=b1 div 6;
          b3:=b1-(6*b2);
          x1:=stable[17-b2];
          x2:=dtable[6-b3];
          getamp:=x1*x2;
        end;
    end;
end;
```

```
end;
```

```
{*****}
{*****}
```

```
Procedure SetABC(f,fb:integer;VAR a,b,c:Real);
```

```
const pi = 3.14159265;
      T = 0.0001 ;
```

```
var R : real;
```

```
begin
```

```

R := exp(-PI*T*FB) ;
C := -R*R ;
B := 2*R*COS(2*PI*T*F);
A := 1-B-C ;
IF F<0 then
Begin
A:=1/A ;
B:=-B/A ;
C:=-A*C ;
End;
End;
{*****}
{*****}

Procedure PARCOE(V1:I);
Begin
for j:=1 to 12 do Opar[j]:=Getamp(V1[j]);
for j:=1 to 10 do
Setabc(V1[12+j],V1[22+j],Opar[10+3*j],Opar[11+3*j],Opar[12+3*j]);
Setabc(0,V1[33],Opar[43],Opar[44],Opar[45]);
end;
{*****}
{*****}

Procedure COEWAV(C1:C;VAR IWAVE:sortie);
VAR
YPP,YPP1,YPP2,YZZ,YZZ1,YZZ2,YSS,YSS1,YSS2,YSS3,YSS4:real;
YSO,YSO1,YSO2,YP1,YP11,YP12,YZ1,YZ11,YZ12,Y1,Y11,Y12:real;
Y2,Y21,Y22,Y3,Y31,Y32,Y4,Y41,Y42,Y5,Y51,Y52,Y6,Y61,Y62:real;
YO,YO1,YO2:real;
YF,YF1,YF2,YFF,YFF1,YFF2,YFH,YFH1,YFH2:real;
YS1,YS11,YS12:real;
impulse,pulse:real;

VAR
noise:real;
I,k,a,b :integer;

begin
{initialisation de tous les vecteurs}
YPP1:=0;YPP2:=0;YZZ1:=0;YZZ2:=0;
YSS1:=0;YSS2:=0;YSS3:=0;YSS4:=0;YSO1:=0;YSO2:=0;YP11:=0;YP12:=0;
YZ11:=0;YZ12:=0;Y11:=0;Y12:=0;Y21:=0;Y22:=0;Y31:=0;Y32:=0;
Y41:=0;Y42:=0;Y51:=0;Y52:=0;Y61:=0;Y62:=0;YO1:=0;
YO2:=0;YF1:=0;YF2:=0;
YFF1:=0;YFF2:=0;YFH1:=0;YFH2:=0;

a:=trunc(uttdur*0.001*Ipar[34]);
b:=trunc(10000/Ipar[34]);
for k:=1 to (a+1) do
{Generation de la source voisée}
Begin
impulse:=C1[1];
pulse:=C1[4];
for i:=1 to b do
begin

{Resonateur RGP}
YPP:= C1[31]*impulse+ C1[32]*YPP1+ C1[33]*YPP2 ;
YPP2:=YPP1;
YPP1:=YPP;
impulse:=0;
{Resonateur RGZ}

```

```

    YZZ:= C1[34]*YPP+ C1[35]*YZZ1+ C1[36]*YZZ2;
    YZZ2:=YZZ1;
    YZZ1:=YPP;
        {Resonateur RGS}
    YSS:= C1[43]*pulse+ C1[44]*YSS1+ C1[45]*YSS2;
    YSS2:=YSS1;
    YSS1:=YSS;
    YSS:=C1[31]*YSS+ C1[32]*YSS3 +C1[33]*YSS4;
    YSS4:=YSS3;
    YSS3:=YSS;
    pulse:=0;
        {La source voisée est YSO}
    YSO:= YZZ + YSS ;
        {Generation du bruit blanc pseudo-Gaussien}
    randomize;
    noise:=random(16);
    noise:=noise -8 ;
        if YPP<=0 then
    noise:=noise/2;
    YF:=noise + YF1 ;
    YF1:=YF;
    YFH:=C1[3]*YF;
    YFF:=C1[2]*YF;
    {Fonction de transfert à travers le modèle serie}
        if Ipar[35] = 0 then
    begin
    {Resonateur R6}
    Y6:=YSO+YFH;
    if Ipar[39]=6 then
    begin
    {resonateur R6}
    Y6:= C1[28]*(YSO+YFH)+ C1[29]*Y61+ C1[30]*Y62;
    Y62:=Y61;
    Y61:=Y6;
    end
    else
    Y5:=Y6;
        if Ipar[39]=5 then
    begin
    {resonateur R5}
    Y5:= C1[25]*Y6+ C1[26]*Y51+ C1[27]*Y52;
    Y52:=Y51;
    Y51:=Y5;
    end
    else
    Y4:=Y5;
    Y4:= C1[22]*Y5+ C1[23]*Y41+ C1[24]*Y42;
    Y42:=Y41;
    Y41:=Y4;
        {Resonateur R3}

    Y3:= C1[19]*Y4+ C1[20]*Y31+ C1[21]*Y32;
    Y32:=Y31;
    Y31:=Y3;
        {Resonateur R2}

    Y2:= C1[16]*Y3+ C1[17]*Y21+C1[18]*Y22;
    Y22:=Y21;
    Y21:=Y2;
        {Resonateur R1}

    Y1:= C1[13]*Y2+ C1[14]*Y11+C1[15]*Y12;
    Y12:=Y11;
    Y11:=Y1;

```

```

(Resonateur RN2)
YZ1:=C1[40]*Y1+ C1[41]*YZ11+C1[42]*YZ12;
YZ12:=YZ11;
YZ11:=Y1;
  (Resonateur RNP)
YP1:=C1[37]*YZ1+ C1[38]*YP11+ C1[39]*YP12;
YP12:=YP11;
YP11:=YP1;
YO:=YP1;
  end
  else
  {Ipar[35]=1}
  {Fonction de transfert à travers le modèle PARALLELE}
  begin
  {Resonateur R1 attaqué par la source source voisée }
Y1:= C1[6]*C1[13]*YSO+ C1[14]*Y11+ C1[15]*Y12;
Y12:=Y11;
Y11:=Y1;
  {derivee de la source voisée}
YS1:=YSO - YSO1 ;
YSO1:=YSO;
  {Resonateur R2-R4 attaques par YS1 et YFF}
YP1:=C1[5]*C1[37]*(YS1+YFF)+ C1[38]*YP11+ C1[39]*YP12;
YP12:=YP11;
YP11:=YP1;

Y2:= C1[7]*C1[16]*(YS1+YFF)+ C1[17]*Y21+ C1[18]*Y22;
Y22:=Y21;
Y21:=Y2;

Y3:= C1[8]*C1[19]*(YS1+YFF)+ C1[20]*Y31+ C1[21]*Y32;
Y32:=Y31;
Y31:=Y3;

Y4:= C1[9]*C1[22]*(YS1+YFF)+ C1[23]*Y41+ C1[24]*Y42;
Y42:=Y41;
Y41:=Y4;

  {resonateur R5-R6 attaques par YFF}
Y5:= C1[10]*C1[25]*YFF+ C1[26]*Y51+ C1[27]*Y52;
Y52:=Y51;
Y51:=Y5;

Y6:= C1[11]*C1[28]*YFF+ C1[29]*Y61+ C1[30]*Y62;
Y62:=Y61;
Y61:=Y6;

  {sortie finale}
YO:=Y1-Y2+Y3-Y4+Y5-Y6+YP1-C1[12]*YFF;
end;

IWAVE[i+(k-1)*b]:=(YO-YO1);
YO1:=YO;
end;
end;
end;{ de la procedure }
{*****}
{*****}
begin
{sm+}
writeln('donnez la longueur de l'articulation');
read(uttdur);

```

```
read(uttdur);
Assign(fichier, 'd:\Moh.dat');
rewrite(fichier);
Assign(fich, 'd:\kho.dat ');
reset(fich);
for j:=1 to 39 do Readln(fich.Ipar[j]);
Parcoe(Ipar);
Coewav(Opar,sort);
  for j:=1 to (uttdur*10) do
    writeln(fichier,0.1*j, ' ', sort[j]:5:10);
  end.
```

## REFERENCES BIBLIOGRAPHIQUES

- [1] - AWAD. S (1983), "Synthétiseur à formants en temps réel : étude d'une source d'excitation élaborée, codage optimum des paramètres de commande".  
Thèse de Docteur-Ingenieur  
I. N. P Grenoble.
- [2] - BADIN. P (1983), "Analyse de la parole-Synthèse à formants : application à la synthèse des contours voisés du français. Discussion d'une méthode de détermination des indices acoustiques de la parole".  
Thèse de Docteur-Ingenieur  
I. N. P Grenoble.
- [3] - FERRITI. M et CINARE. F (1983), "Synthèse , reconnaissance de la parole".  
Edition TESTS.
- [4] - GUIBERT. J (1979), "la parole , compréhension et synthèse par les ordinateur".  
Edition PUF.
- [5] - KLATT. D. H (1980), "Software for a Cascade/Parallel formant synthesizer".  
J. Acoust. Soc. Am (JASA) 67 (971-996).
- [6] - LIENARD. J. S (1977), "Processus de la communication parlée, introduction à l'analyse et à la synthèse de la parole".  
Edition MASSON.
- [7] - RABINER L. R et SCHAFER R. W (1978), "Digital processing of speech signal".  
Prentice Hall, Englewood cliffs, New Jersey.

## REFERENCES BIBLIOGRAPHIQUES

- [1] - AWAD. S (1983), "Synthétiseur à formants en temps réel : étude d'une source d'excitation élaborée, codage optimum des paramètres de commande".  
Thèse de Docteur-Ingenieur  
I. N. P Grenoble.
- [2] - BADIN. P (1983), "Analyse de la parole-Synthèse à formants : application à la synthèse des contours voisés du français. Discussion d'une méthode de détermination des indices acoustiques de la parole".  
Thèse de Docteur-Ingenieur  
I. N. P Grenoble.
- [3] - FERRITI. M et CINARE. F (1983), "Synthèse , reconnaissance de la parole".  
Edition TESTS.
- [4] - GUIBERT. J (1979), "la parole , compréhension et synthèse par les ordinateur".  
Edition PUF.
- [5] - KLATT. D. H (1980), "Software for a Cascade/Parallel formant synthesizer".  
J. Acoust. Soc. Am (JASA) 67 (971-996).
- [6] - LIENARD. J. S (1977), "Processus de la communication parlée, introduction à l'analyse et à la synthèse de la parole".  
Edition MASSON.
- [7] - RABINER L. R et SCHAFER R. W (1978), "Digital processing of speech signal".  
Prentice Hall, Englewood cliffs, New Jersey.