

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT ELECTRONIQUE

المدرسة الوطنية المتعددة التقنيات  
BIBLIOTHEQUE — المكتبة  
Ecole Nationale Polytechnique

PROJET DE FIN D'ETUDES

S U J E T

IDENTIFICATION  
DU  
LOCUTEUR

Proposé par :  
B. BOUSSEKSOU

Etudié par :  
K. AIT ALI  
H. MERRAD

Dirigé par :  
B. BOUSSEKSOU

PROMOTION :  
Janvier 1986

THESE D'INGENIORAT D'ETAT  
EN ELECTRONIQUE,

présentée à

l'Ecole Nationale Polytechnique

par

K. AIT-ALI & H. MERRAD

المدرسة الوطنية المتعددة التقنيات  
BIBLIOTHEQUE — المكتبة  
Ecole Nationale Polytechnique

IDENTIFICATION

DU

LOCUTEUR

proposée et dirigée

par

B. BOUSSEKSOU

Janvier 1986

## DEDICACES

---

المدرسة الوطنية المتعددة التقنيات  
BIBLIOTHEQUE — المكتبة  
Ecole Nationale Polytechnique

A mon père qui m'a tout donné.  
A ma mère qui a veillé pour moi.  
A toute ma famille.  
A tous(tes) mes amis(es).  
Et à tous ceux qui me sont chers.

**Kamel**

A la mémoire de mon père.  
A la mémoire de ma mère.  
A mon oncle et à toute ma famille.  
A tous(tes) mes amis(es).

**Hamoud**

A NOTRE AMITIE

Kamel - Hamoud



## REMERCIEMENTS

---

المدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

Nous tenons à remercier vivement notre promoteur Mr B. BOUSSEKSOU, pour son aide et ses conseils tout au long de l'élaboration de notre projet.

Nous tenons à remercier aussi Melle MOUSSAOUI et Mr BENIDIR pour nous avoir fourni une documentation qui a été très utile pour notre travail.

Nous remercions également Mr A. CHALA et Mr BEN HOUHOU pour leurs aides sur le plan informatique qui nous ont été très précieuses.

Qu'il nous soit permis d'exprimer ici toute notre reconnaissance à tous ceux qui ont contribué à notre formation.



TABLE DES MATIERES



PAGES

INTRODUCTION. ....	1
CHAPITRE I: TRAITS DISTINCTIFS DU LOCUTEUR.	
I.1 Introduction. ....	2
I.2 Production de la parole. ....	2
I.3 Mélodie . ....	5
I.4 Le timbre. ....	5
I.5 Articulation. ....	6
I.6 Conclusion. ....	6
CHAPITRE II: ANALYSE DE LA PAROLE.	
II.1 Introduction. ....	7
II.2 Méthode des sonagrammes. ....	7
II.3 Analyse par LPC. ....	7
II.4 Analyse temporelle. ....	12
II.5 Analyse cepstrale. ....	14
II.6 Analyse spectrale. ....	20
CHAPITRE III: CLASSIFICATION.	
III.1 Introduction. ....	22
III.2 Classification de formes par la distance minimale. ....	23
III.3 Types de classification. ....	26
III.4 Recherche des classes. ....	29
III.5 Etude détaillée des algorithmes. ....	32
III.5.1 Un simple algorithme de classification. ...	32
III.5.2 L'algorithme de la distance MAXIMIN. ....	33
III.5.3 L'algorithme des K-Moyennes . ....	36
III.5.4 L'algorithme ISODATA. ....	38
III.6 Reconnaissance de formes non supervisée. ....	42
III.7 Conclusion. ....	42

CHAPITRE IV: RECONNAISSANCE DU LOCUTEUR.

IV.1 Systèmes de reconnaissance du locuteur. ....	43
IV.2 Système de vérification du locuteur. ....	45
IV.3 Alignement temporel. ....	48
IV.4 Estimation du pitch et des formants. ....	53
IV.5 Systèmes d'identification du locuteur. ....	56

ANNEXE .

- Programme de l'algorithme des k-moyennes .
- Programme de l'algorithme ISODATA .

BIBLIOGRAPHIE.





## INTRODUCTION



La reconnaissance du locuteur est le problème dual de reconnaissance de la parole : il ne s'agit pas de reconnaître ce qui est dit, mais la personne qui le dit .

Dans sa généralité, ce problème nécessite la prise en compte de nombreux éléments ( timbre, hauteur de la voix, accent, habitudes linguistiques, tournures de phrases . . . ) permettant d'identifier le locuteur par sa voix .

La reconnaissance du locuteur, nécessite une phase d'apprentissage. Le locuteur prononce son mot de passe qui est traité pour mettre en relief les caractéristiques de sa voix qui seront par la suite mémorisées .

Lors de la reconnaissance, le locuteur prononce le mot de passe ( qui représente le test ) et le système compare cette prononciation avec la référence .

Dans le cas de la vérification, le locuteur est accepté si la ressemblance est suffisante, sinon, il est rejeté .

Tandis que dans le cas de l'identification, le système compare la voix du locuteur à toutes les références et lui attribue l'identité qui correspond à la référence la plus proche .

Notre travail consiste essentiellement en le problème de reconnaissance de formes . Ces formes représentent les paramètres obtenus à partir des différentes méthodes d'analyse discutées dans le chapitre II . Ces paramètres sont les caractéristiques de la voix du locuteur présentées dans le chapitre I .

Dans le chapitre IV, nous exploitons tout ce qui a été décrit dans les trois chapitres précédents pour l'étude du système d'identification du locuteur .

Enfin, nous présentons en annexe deux programmes écrits en Basic correspondant aux algorithmes de classification cités dans le chapitre III .



CHAPITRE : I

TRAITS DISTINCTIFS DU LOCUTEUR.

## CHAPITRE : I

### TRAITS DISTINCTIFS DU LOCUTEUR

#### I.1 INTRODUCTION

Lors de la réalisation d'un système de reconnaissance du locuteur, nous nous trouvons face à un problème primordial qui est le choix des paramètres . Celui-ci réside en l'extraction du signal vocal des traits caractéristiques signifiants ( porteurs d'une quantité d'information suffisante sur l'identité du locuteur ). D'où la possibilité d'identifier un locuteur par sa voix .

Dans ce qui suit, nous allons parler brièvement des traits distinctifs du locuteur les plus importants .

#### I.2 PRODUCTION DE LA PAROLE

L'appareil phonatoire humain est constitué d'un organe respiratoire ( source d'énergie ) , des cordes vocales ( jouant le rôle d'oscillateur ) , et des cavités bucales et nasales ( cavités résonantes ) .

La parole, en tant que phénomène physique, résulte de l'excitation du conduit vocal par deux types de sources . Ce canal peut être considéré comme un tube acoustique partant des cordes vocales ( région du larynx ) jusqu'aux lèvres .

Le tube constitué par les cavités nasales, branché en dérivation sur le tube acoustique n'agit que lors de la production des sons nasalisés, par l'ouverture du vélum .

Les deux types d'excitations du canal vocal sont :

- excitation par les cordes vocales .
- excitation par des bruits .

##### I.2.1 LES SONS VOISÉS

La production des sons voisés ( voyelles, semi-voyelles, nasales ) exige une source d'excitation par mise en vibration des cordes vocales . La



pharynx nasal

palais mou

pharynx bucal

épiglotte

pharynx

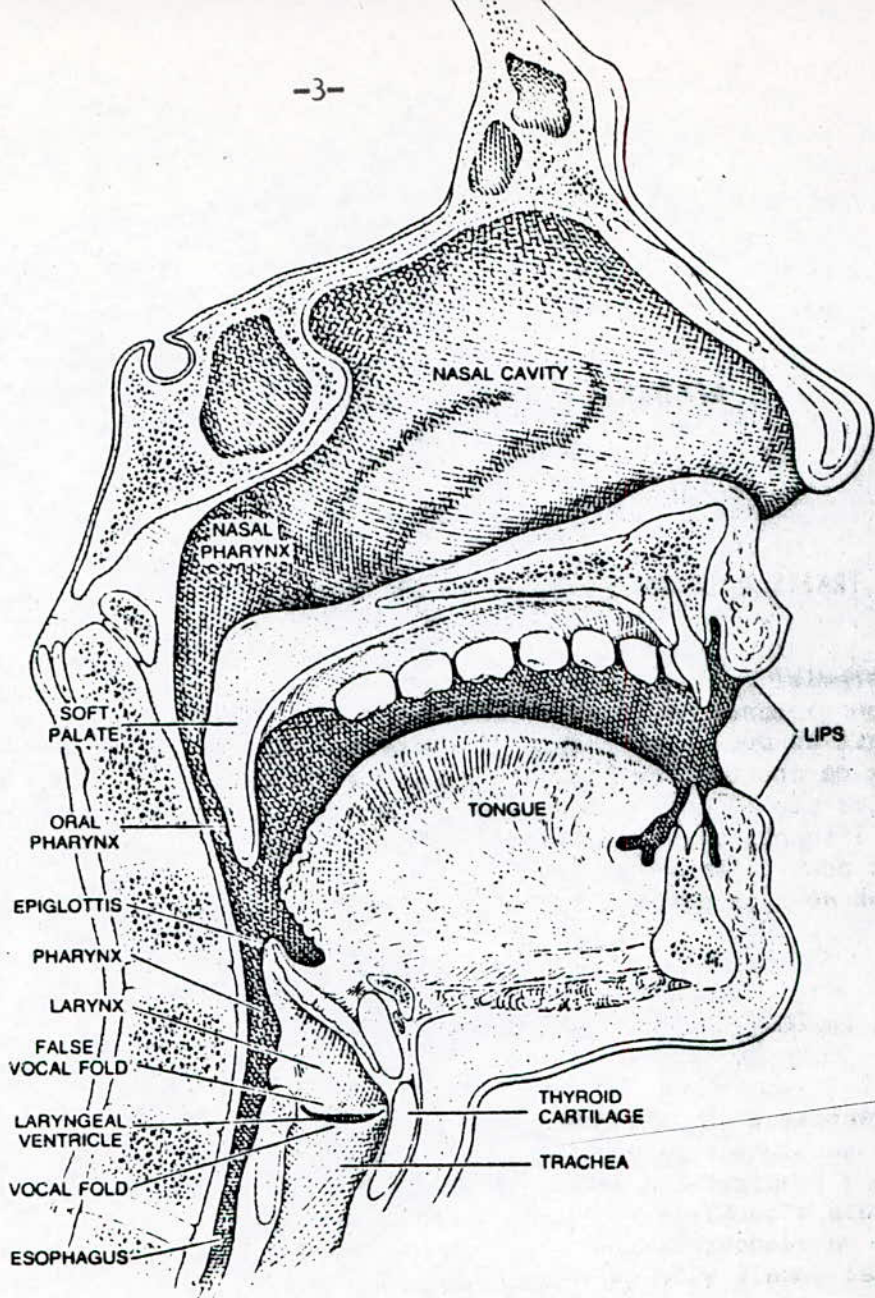
larynx

fausse corde vocale

ventricule laryngé

cordes vocales

oesophage



cavité nasale

lèvres

langue

cartilage thyroïde

trachée

cartilage thyroïde

THYROID CARTILAGE

VOCAL FOLD

corde vocale

ARYTENOID CARTILAGE

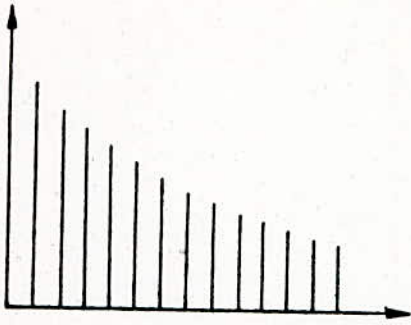
cartilage arythén

Fig.1.1

coupe de l'appareil vocal  
coupe du larynx et cordes vocales

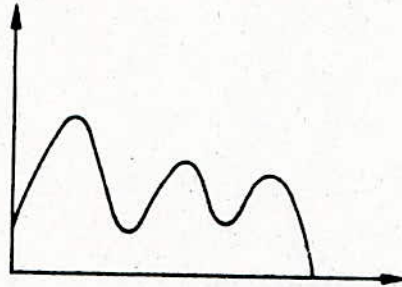


Sons voisés



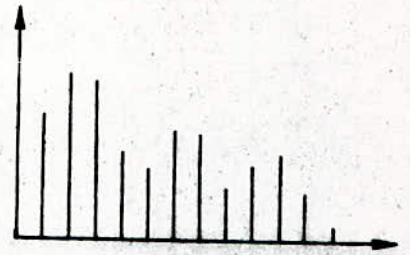
Spectre de la source

+



canal

→



signal

Sons non voisés

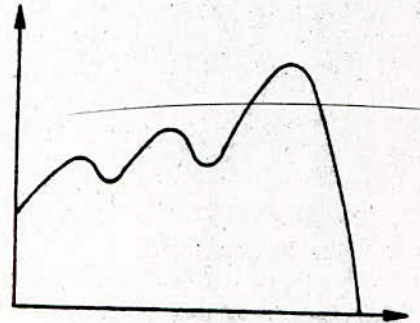
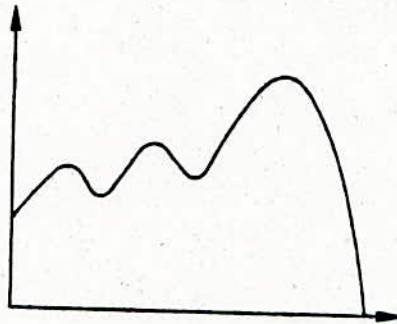
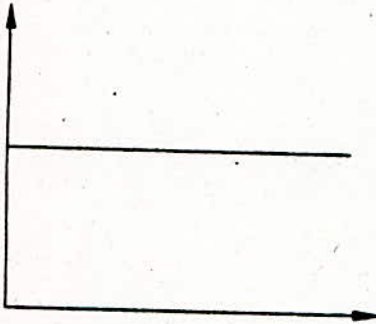


Fig.1.2

glotte s'ouvre et se ferme périodiquement à la fréquence  $F_0$  appelée pitch qui correspond à la première raie du spectre de ce signal . Cette fréquence varie de 70 à 150 Hz pour les hommes, 150 à 300 Hz pour les femmes et peut dépasser 400 Hz pour les enfants .

### I.2.2 LES SONS NON VOISES

Le second mode d'excitation est obtenu par le passage de l'air en un point de resserement du conduit vocal produisant un bruit blanc (Gaussien) Ce type d'excitation peut aussi se produire par des bruits d'occlusions ou de plosions provoqués par la fermeture ou l'ouverture des lèvres, ou des chocs de la langue contre le palais .

Notons que certaines consonnes présentent simultanément les deux modes d'excitations ( exp: /v/,/z/,/j/ et /b/,/d/,/g/ )

### I.3 LA MELODIE

La mélodie est la variation de la fréquence fondamentale  $F_0$  (pitch) au cours du temps .

Elle porte une information sur l'identité du locuteur .

### I.4 LE TIMBRE

Le timbre, qui n'est perceptible que pour les sons voisés, représente la caractéristique première de la voix d'un locuteur . Il est presque totalement exprimé par le spectre du signal . Ce spectre est le produit du spectre de la source et de celui du canal vocal ( représentant la fonction de transfert du canal dont les pôles sont appelés formants ) car le signal de la parole est le résultat de la convolution du signal glottique ( excitation ) et la réponse impulsionnelle du conduit vocal .

CHAPITRE: II

ANALYSE DE LA PAROLE.



## I.5 ARTICULATION

Ce type de traits ne concerne que l'activité musculaire du locuteur . En prononçant une phrase, le locuteur produit des phonèmes enchainés de façon continue, en reliant les parties stables du signal par des zones de transition. Ces transitions varient suivant le locuteur .

La durée du silence précédant l'explosion dans les plosives peut servir aussi de support pour l'identification .

## I.6 CONCLUSION

Ces paramètres distinctifs qui vont servir à l'identification du locuteur doivent présenter les qualités suivantes :

- Porter l'information utile sur l'identité du locuteur .
- Etre stables dans le temps et faciles à mesurer .
- Apparaître fréquemment dans la parole et varier peu suivant l'environnement .
- Ne pas être imitable .

---

CHAPITRE: II

ANALYSE DE LA PAROLE.

## CHAPITRE :II

### ANALYSE DE LA PAROLE

#### II.1 INTRODUCTION

L'analyse est une étape importante dans le traitement du signal de la parole en vue d'une réalisation d'un système de reconnaissance du locuteur.

Le but de cette analyse est l'extraction des paramètres représentant les traits distinctifs (voisement, mélodie, timbre, articulation, etc ...). Il existe plusieurs types d'analyse que nous essayerons de développer dans les paragraphes suivants.

#### II.2 METHODE DES SONAGRAMMES

Le sonagramme constitue la représentation de l'évolution spectrale d'un signal de la parole continu pendant un intervalle de temps sur une bande de papier, par noircissement proportionnel à l'énergie du spectre à l'instant et la fréquence considérés (Fig.2.1).

Les premières études faites sur la reconnaissance du locuteur concernaient l'identification par examen visuel de sonagrammes.

#### II.3 ANALYSE PAR (LPC)

La prédiction linéaire est une technique d'analyse de la parole considérée comme étant l'une des plus puissantes. Elle permet une estimation des paramètres de base de la parole (formant, pitch, spectre, etc ...) dont le critère d'optimisation est de minimiser l'erreur quadratique moyenne.

Les statisticiens utilisent depuis un certain temps cette méthode sous le nom d'analyse autorégressive.

Cette méthode consiste en la recherche d'un modèle se rapprochant le plus possible du signal original.



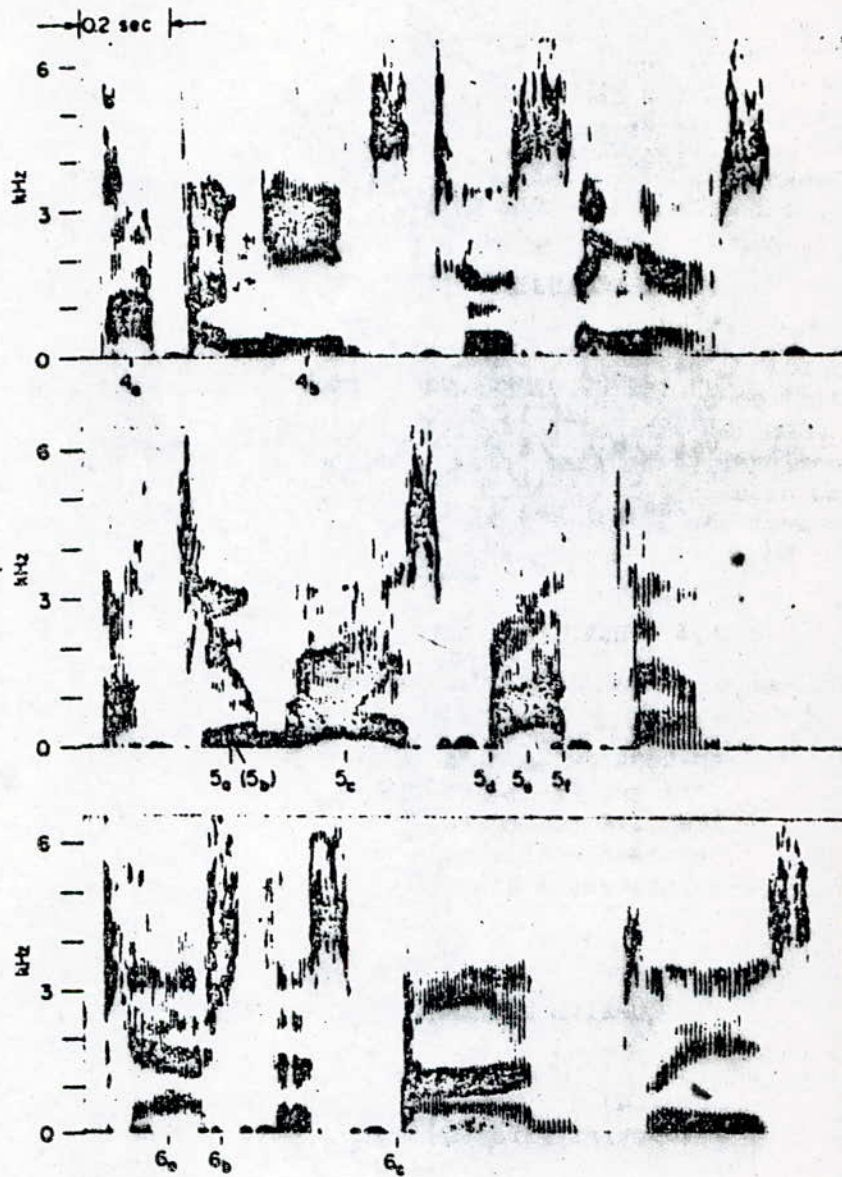


Fig.2.1

L'échelle horizontale représente les temps, l'échelle verticale les fréquences, et le noircissement l'énergie.

### II.3.1 PRINCIPE DE L'ANALYSE PAR LPC

L'idée de cette méthode d'analyse est fondée sur l'hypothèse selon laquelle, un échantillon de parole est estimé par une combinaison linéaire des échantillons qui le précèdent .

Ceci est alors illustré par la relation suivante :

avec

$$\hat{S}(n) = \sum_{k=1}^P a_k S(n-k) \quad (2.3.1)$$

P: l'ordre du prédicteur  
 $a_k$  : coeff de la prédiction

Considérons le modèle discret de la production de la parole décrit dans la ( Fig.2.2 ) pour mieux comprendre le principe de l'analyse par la (Lpc).

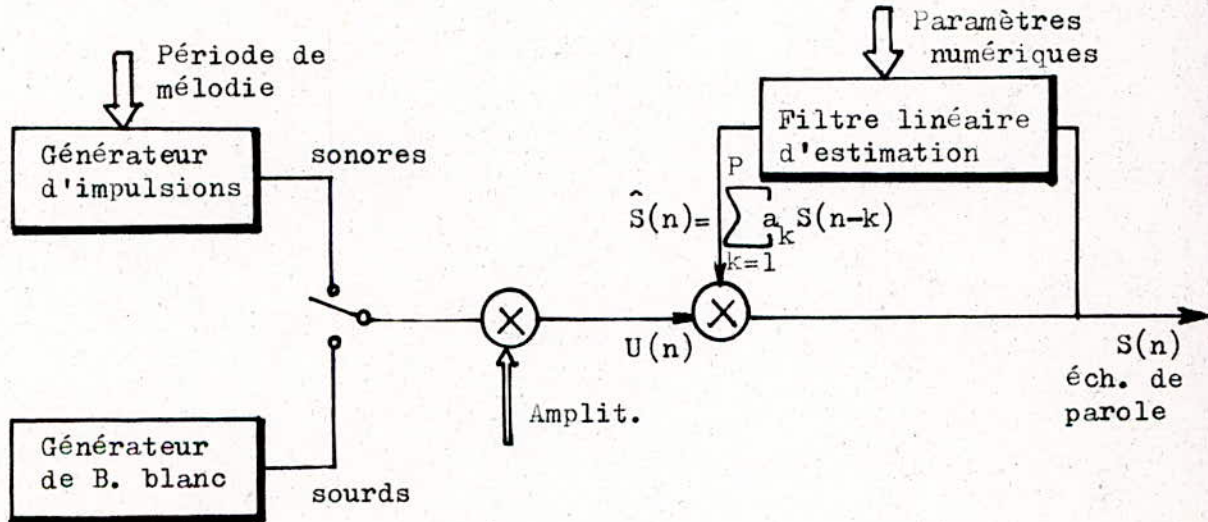


Fig.2.2 Modèle numérique de la production de la parole .

L'erreur de prédiction entre le signal et son estimation est alors :

$$\begin{aligned} e(n) &= S(n) - \hat{S}(n) \\ &= S(n) - \sum_{k=1}^P a_k S(n-k) \end{aligned}$$

En appliquant la transformée en Z à cette équation nous obtenons :

$$E(z) = S(z) - \sum_{k=1}^P a_k z^{-k} S(z) = S(z) \left[ 1 - \sum_{k=1}^P a_k z^{-k} \right]$$



ou  $E(z)$  et  $S(z)$  sont les transformées en  $Z$  respectivement de  $e(n)$  et  $S(n)$ .  
La fonction de transfert du filtre prédicteur est donnée par :

$$G(z) = \frac{A}{1 - \sum_{k=1}^P a_k z^{-k}} \quad A: \text{facteur de gain}$$

On remarque que cette fonction de transfert ne possède que des pôles .  
 $G(z)$  représente la fonction de transfert du conduit vocal .

### II.3.2 DETERMINATION DES COEFFICIENTS DE LA LPC

Pour cette méthode d'analyse le problème consiste à déterminer les  $P$  coefficients  $a_k$  afin d'optimiser l'estimation du signal  $S(n)$  .  
Le principe d'optimisation est la minimisation de l'erreur quadratique .  
On utilise alors le critère des moindres carrés .  
Cela revient donc à minimiser la quantité définie par :

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

Cette quantité est minimale si :

$$\frac{\partial E}{\partial a_k} = 0 \iff \sum_{k=1}^P a_k \left[ \sum_n S(n-k) S(n-i) \right] = \sum_n S(n-i) S(n) \quad (\alpha)$$

Pour la résolution de ce système linéaire d'équations, on utilise deux principales méthodes qui sont :

- Méthode de covariance.
- Méthode d'autocorrélation.

#### A) Méthode de covariance

Cette méthode est fondée sur les hypothèses suivantes :

- 1 - Le signal est défini pour  $P+N$  échantillons consécutifs , ou  $P$  est l'ordre du prédicteur et  $N$  la taille de l'échantillon estimé .
- 2 - Un échantillon est prédit par les  $P$  échantillons qui le précèdent.  
Ceci est valable pour les  $N$  échantillons successifs .
- 3 - L'erreur quadratique totale est minimisée pour l'ensemble des  $N$  échantillons consécutifs .



Le système ( $\alpha$ ) peut se mettre sous la forme :

$$\sum_{k=1}^P a_k C(i,k) = C(i,0) \quad 1 \leq i \leq P$$

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

Nous obtenons alors un système de P équations à P inconnues qui nous permettra de déterminer les coefficients  $a_k$   $1 \leq k \leq P$

Ecrivons ce système sous forme matricielle

$$\begin{bmatrix} C(1,1) & C(1,2) & \dots & C(1,P) \\ C(2,1) & C(2,2) & \dots & C(2,P) \\ \vdots & \vdots & \ddots & \vdots \\ C(P,1) & C(P,2) & \dots & C(P,P) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} C(1,0) \\ C(2,0) \\ \vdots \\ C(1,0) \end{bmatrix}$$

### B) Méthode d'autocorrélation

Citons les hypothèses sur lesquelles est fondée cette méthode :

- 1 - Le signal est nul à l'extérieur de l'intervalle considéré. Ceci est réalisé à l'aide d'un fenêtrage temporel de largeur N .
- 2 - Pour tout le temps ( $n \in ]-\infty, +\infty[$ ), chaque échantillon est prédit par les P échantillons qui le précèdent .
- 3 - La minimisation de l'erreur quadratique entre le signal fenêtré et le modèle se fait de  $-\infty$  à  $+\infty$  .

De ces hypothèses considérées nous aboutissons aux "Equations Normales " d'autocorrélation et cela à partir du système ( $\alpha$ )

$$\sum_{k=1}^P a_k R_{|i-k|} = R_i \quad i=1,2,\dots,P$$

avec  $R_i = \sum_{n=0}^{N-1-|i|} S(n) S(n+|i|)$

$$R(i) = R(-i)$$

Ecrivons le système sous forme matricielle :

$$\begin{bmatrix} R_0 & R_1 & \dots & R_{P-1} \\ R_1 & R_0 & \dots & R_{P-2} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ R_{P-1} & \dots & \dots & R_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_P \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \cdot \\ \cdot \\ R_P \end{bmatrix}$$

### II.3.3 CONCLUSION

L'analyse par prédiction linéaire a bien des applications dans le domaine de la parole . A titre d'exemple on peut citer les systèmes de reconnaissance du locuteur :

Si plusieurs locuteurs prononcent la même phrase, l'analyse par LPC nous permet de correspondre à chacun des locuteurs un vecteur de référence dont les composantes sont des paramètres caractéristiques de sa voix .

Lorsque l'un des locuteurs parmi les précédents prononce cette phrase, on compare alors le vecteur obtenu à tous les vecteurs de référence, le locuteur identifié sera donc celui dont le vecteur de référence est le plus proche du vecteur obtenu au sens d'une certaine distance .

L'analyse par LPC donne de bon résultats pour le calcul des paramètres de la parole .

La méthode de covariance est une méthode générale utilisée sans aucune restriction. Le seul problème réside en la stabilité du modèle .

Par contre la méthode d'autocorrélation, donne une stabilité du modèle, mais ne donne pas des résultats précis à cause de la troncature du signal original .

### II.4 ANALYSE TEMPORELLE

Cette méthode d'analyse met en évidence l'évolution temporelle de certains paramètres tels que la fréquence du fondamental  $f_0$ , les formants et certains coefficients de la LPC .

A titre d'exemple on peut utiliser cette méthode pour caractériser le fait suivant : Lors de la production d'une plosive, la fermeture du conduit vocal est mieux caractérisée par son aspect temporel que par son spectre .

Nous citons quelques méthodes d'analyse temporelle .



#### II.4.1 MESURE DE L'ENERGIE

Pour un signal échantillonné  $x(n)$ , l'énergie est donnée par la relation suivante :

$$E = \sum_{-\infty}^{+\infty} x^2(n)$$

En général, les intervalles de temps considérés sont très courts (environ 20 ms) pour lesquels le signal de la parole est stationnaire. Ceci est réalisé à l'aide d'une fenêtre temporelle  $h(m)$ , dont le but est d'ignorer le plus possible les échantillons éloignés dans le temps. L'énergie  $E$  donnée par la relation suivante est considérée comme étant un paramètre simple et utilise en vue de la détection des silences ou de la séparation des syllabes.

$$E = \sum [h(m) x(n+m)]^2$$

#### II.4.2 METHODE DE PASSAGE PAR ZERO

Le signal de la parole s'annule ou change de signe à des instants où la répartition dans le temps est liée à des caractéristiques spectrales. Pour un signal discret il suffit de vérifier l'inégalité suivante :

$$x(n) x(n-1) < 0$$

Ceci est utilisé pour mesurer la fréquence du fondamental et pour la détermination des formants.

Cette méthode présente certains avantages que nous citons ci-après :

- Cette méthode est facile à mettre en œuvre et rapide.
- Elle est indépendante de l'information correspondant à l'amplitude (car on s'intéresse seulement au signe) du signal.
- Permet la séparation des fricatives non voisées des phonèmes voisés.

Cette méthode d'analyse ne permet donc pas de déterminer un modèle du conduit vocal.

#### II.4.3 METHODE D'AUTOCORRELATION

Pour un signal échantillonné, la relation de définition est donnée par:



On considère que le signal de la parole est stationnaire sur un intervalle court ( 15 à 20 ms ) pour pouvoir définir alors la fonction d'autocorrélation à court terme par la relation suivante :

$$R_N(m) = \frac{1}{N} \sum_{i=0}^{N-|m|-1} X(i) X(i-m)$$

$$R_N(m) = 0 \quad \text{pour } m \geq N \quad m = 0, \pm 1, \dots, \pm(N-1)$$

La méthode d'autocorrélation dans le domaine temporel permet de chercher les lois de périodicité du signal de la parole .

Cette fonction permet aussi de déterminer les coefficients de la LPC .

## II.5 ANALYSE CEPSTRALE

Le but de cette analyse est de pouvoir séparer dans un espace particulier deux signaux combinés par convolution ( ou par multiplication ) .

L'analyse des systèmes non linéaires n'est pas aussi simple que celle des systèmes linéaires représentés par des relations mathématiques simples .

Le traitement homomorphique basé sur la superposition généralisée nous permet une analyse plus aisée de ces systèmes non linéaires .

### II.5.1 SUPERPOSITION GENERALISEE

Un système linéaire est défini par :

$$L [ a U_1(k) + b U_2(k) ] = a L [ U_1(k) ] + b L [ U_2(k) ] \quad (2.5.1)$$

Le filtrage linéaire est utilisé pour séparer deux signaux dont les bandes de fréquences sont disjointes, en utilisant un filtre passe-bande. Mais ceci n'est obtenu que dans le cas où les deux signaux sont combinés par addition .

Supposons maintenant que nous sommes dans le cas d'un signal obtenu par convolution ou par multiplication .

$$\text{Exp: } U(k) = \sum_{l=-\infty}^{+\infty} U_1(l) U_2(k-l)$$

ou

$$U(k) = U_1(k) \cdot U_2(k)$$

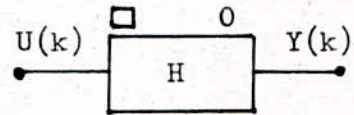
Le problème posé étant l'extraction du signal utile (désiré) .

Généralisons la relation (2.5.1)

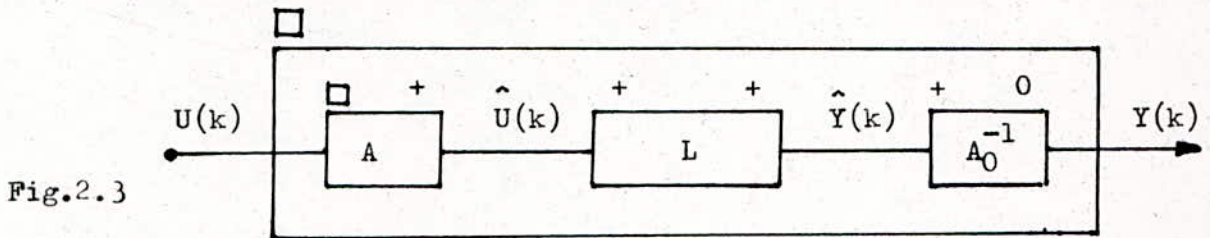
$$\begin{aligned} H \begin{bmatrix} U_1(k) \\ 0 \end{bmatrix} &= H \begin{bmatrix} U_1(k) \\ 0 \end{bmatrix} \quad 0 \quad H \begin{bmatrix} U_2(k) \\ 0 \end{bmatrix} \\ H \begin{bmatrix} C \\ 0 \end{bmatrix} &= C \quad 0 \quad H \begin{bmatrix} U_1(k) \\ 0 \end{bmatrix} \end{aligned}$$

H: étant l'opérateur du système, représente une transformation linéaire entre espaces vectoriels . Les systèmes représentés par H sont appelés homomorphiques .

Soit la représentation suivante , celle d'un système homomorphique H avec la loi d'entrée  $\square$  et de sortie  $\circ$  :



D'après OPPEHEIM, nous pouvons décomposer ce système en trois autres en série, dont celui du milieu correspond à un système linéaire conventionnel, c'est la forme dite canonique des systèmes homomorphiques (Fig.2.3) .



Dans ce qui suit nous allons considérer les systèmes homomorphiques convolutifs et multiplicatifs qui donnent une bonne illustration de cette représentation .

### A) Système multiplicatif

La forme générale des signaux existant à l'entrée de ce système est :

$$U(k) = [U_1(k)]^{a_1} \cdot [U_2(k)]^{a_2}$$

avec  $a_1, a_2$  : Constantes arbitraires .

Notre but est de trouver un opérateur A satisfaisant la relation suivante :

$$A \left[ [U_1(k)]^{a_1} \cdot [U_2(k)]^{a_2} \right] = a_1 A [U_1(k)] + a_2 A [U_2(k)]$$

Nous remarquons que la fonction logarithme répond à ceci .

$$\text{Log} \left[ [U_1(k)]^{a_1} \cdot [U_2(k)]^{a_2} \right] = a_1 \text{Log} [U_1(k)] + a_2 \text{Log} [U_2(k)]$$



Comme le signal de la parole est bipolaire (  $>0$  et  $<0$  ) on utilise alors la fonction logarithme complexe

$$U(k) = |U(k)| e^{j \text{Arg } U(k)}$$

ou

$$\text{Log } U(k) = \text{Log } |U(k)| + j \text{Arg } U(k)$$

La forme canonique d'un tel système est donnée par la (Fig.2.4)

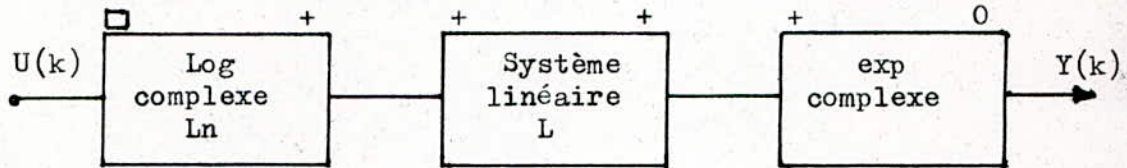


Fig.2.4

### B) Système convolutif

Un tel système présente à son entrée des signaux du type:

$$U(k) = \sum_{-\infty}^{+\infty} U_1(l) \cdot U_2(k-l) = U_1(k) * U_2(k)$$

Le but est de trouver D étant l'opérateur de déconvolution tel que:

$$D [U(k)] = D [U_1(k) * U_2(k)] = D [U_1(k)] + D [U_2(k)]$$

$$U'(k) = U_1'(k) + U_2'(k)$$

$U'(k)$ ,  $U_1'(k)$  et  $U_2'(k)$  sont les résultats de déconvolution respectivement des fonctions  $U(k)$ ,  $U_1(k)$  et  $U_2(k)$ .

Nous savons que mathématiquement la transformée en Z permet de passer de:

$$U(k) = U_1(k) * U_2(k)$$

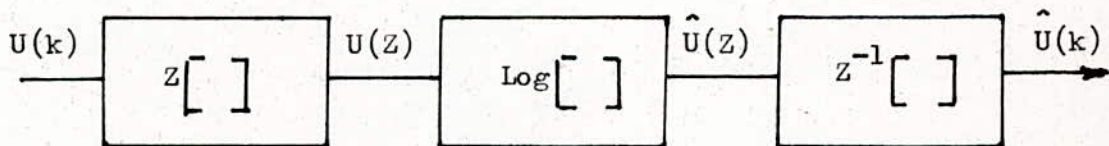
à

$$U(z) = U_1(z) \cdot U_2(z)$$

La fonction Log complexe nous permet le passage du produit à la somme .

$$\text{Log } U(z) = \text{Log } U_1(z) + \text{Log } U_2(z)$$

D'ou le système de déconvolution homomorphique:





## II.5.2 LE CEPSTRE

C'est la transformation de Fourier inverse du logarithme du spectre du signal, le cepstre d'un signal est défini par :

$$\hat{X}(k) = \frac{-1}{j2\pi k} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{X(f)} \frac{d(X(f))}{df} \exp(j2\pi kf) df, \quad k \neq 0$$

$$\hat{X}(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{Log } X(f) df$$

### Propriétés du cepstre

Soit une fonction de transfert sous sa forme générale :

$$X(f) = A \frac{\prod_{n=1}^{Z_i} (1 - \alpha_n e^{-j2\pi f}) \cdot \prod_{n=1}^{Z_e} (1 - \beta_n e^{j2\pi f})}{\prod_{n=1}^{P_i} (1 - \gamma_n e^{-j2\pi f}) \cdot \prod_{n=1}^{P_e} (1 - \delta_n e^{j2\pi f})}$$

avec  $A > 0$ , et  $|\alpha_n|$ ;  $|\beta_n|$ ;  $|\gamma_n|$  et  $|\delta_n| < 1$

D'où les coefficients cepstraux seront donnés par :

$$\hat{X}(k) = \begin{cases} - \sum_{n=1}^{Z_i} \frac{\alpha_n^k}{k} + \sum_{n=1}^{Z_e} \frac{\beta_n^k}{k} & k > 0 \\ \sum_{n=1}^{P_i} \frac{\beta_n^k}{k} + \sum_{n=1}^{P_e} \frac{\delta_n^k}{k} & k < 0 \end{cases}$$

$$\hat{X}(0) = \text{Log } A$$

Considérons la causalité du système :

$$X(k) = 0 \quad \text{pour } k < 0 \quad \text{d'où } \hat{X}(k) = 0 \quad \text{pour } k < 0$$

Ceci se traduit par le fait que la fonction de transfert n'a ni pôle, ni zéro en dehors du cercle unité ( donc nous considérons seulement l'intérieur du cercle unité c-à-d la stabilité du système ), d'où :

$$X(f) = \frac{\prod_{n=1}^{Z_i} (1 - \alpha_n e^{-j2\pi f})}{\prod_{n=1}^{P_i} (1 - \gamma_n e^{-j2\pi f})}$$

Echelle MEL des coefficients cepstraux

Nous remarquons que les coefficients cepstraux décroissent en  $1/k$  d'où la nécessité d'un nombre réduit de ceux-ci pour caractériser le signal .

L'échelle spectrale dans le calcul du cepstre est linéaire en fréquence . D'après les études faites sur la perception de l'oreille, il semble indiqué que celle-ci est sensible à une échelle logarithmique de la fréquence . Donc les informations pertinentes se trouvent réparties de façon exponentielle dans les zones éloignées .

L'échelle logarithmique tente d'éviter de donner un même poids à des zones de fréquences qui n'ont pas la même densité d'informations .

Pour cela on utilise une échelle quasi-logarithmique appelée : "échelle MEL", linéaire sur le premier KHz et logarithmique au delà .

Les coefficients cepstraux sont obtenus à l'aide d'un vocodeur à canaux (Fig. 1) composé de 25 filtres passe-bande de formes triangulaires . Les dix premiers ont une largeur uniforme et les autres ont une largeur qui suit une progression géométrique ( de raison supérieure à 1 car les filtres ont des largeurs croissantes exponentiellement ) .

Nous pouvons alors obtenir les coefficients cepstraux dans l'échelle MEL à parti de l'équation suivante :

$$MFCC (i) = \frac{1}{NF} \sum \text{Log } E(k) \cdot \text{Cos } n(k - i) \frac{\pi}{NF}$$

$$1 \leq i \leq M$$

$n$  : Rang du coefficient cepstral .

$k$  : Numéro du filtre .

$NF$  : Nombre de filtres triangulaires ( en général ,  $NF = 25$  ) .

$M$  : Nombre de coefficients cepstraux .

$E(k)$  : Réponse en énergie des filtres triangulaires (Fig.a) .



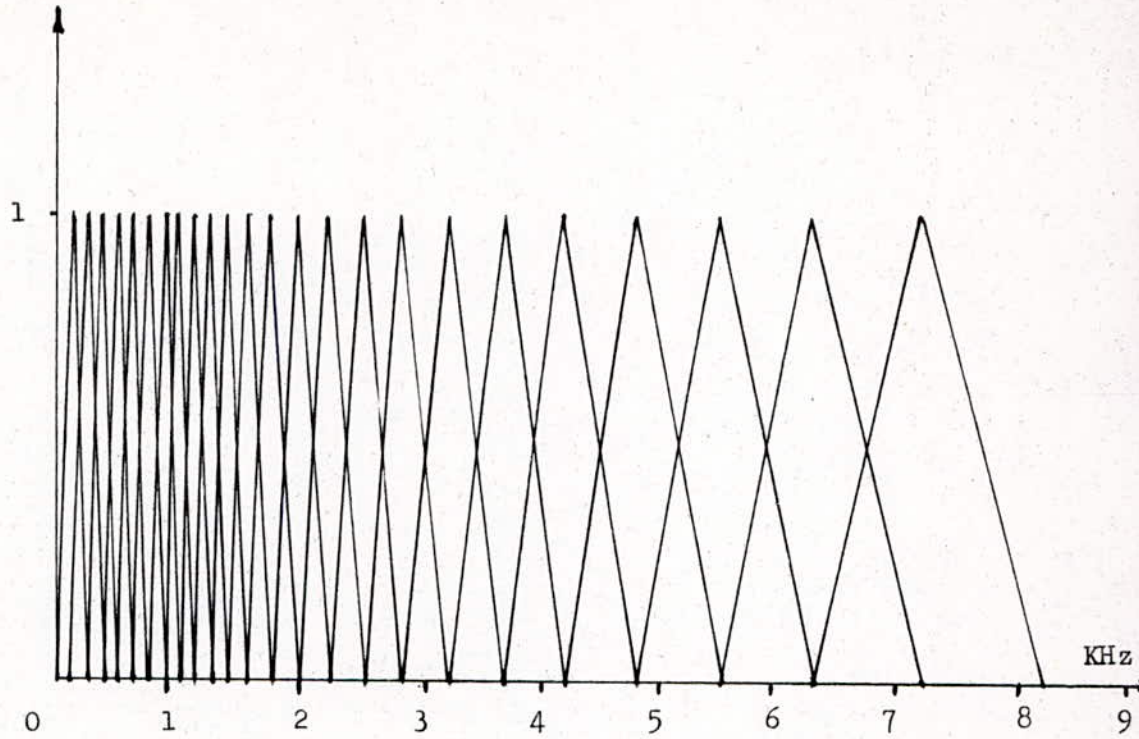


Fig.a  
Bancs de filtres triangulaires .

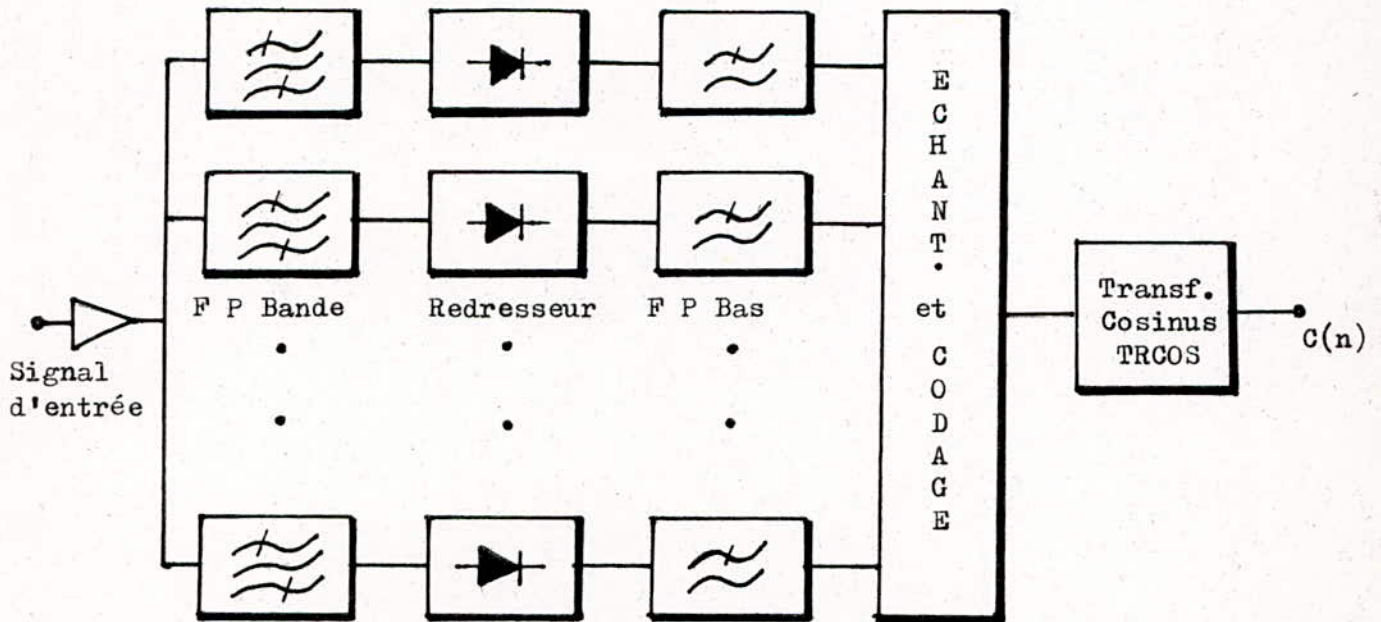


Fig.b  
Obtention des coefficients cepstraux MEL par un vocodeur à canaux .



## II.6 ANALYSE SPECTRALE

Les méthodes d'analyse spectrale ont été et sont toujours largement utilisées pour l'analyse du signal de parole en vue d'obtenir son spectre de puissance à court-terme .

Parmi ces méthodes d'analyse spectrale du type numérique on trouve la transformée de Fourier discrète (TFD) dont le calcul est facilité par les algorithmes de la FFT [11] .

### II.6.1 TRANSFORMEE DE FOURIER DISCRETE (TFD)

La TFD est la représentation fréquentielle des suites temporelles périodiques, elle possède trois caractères fondamentaux :

- 1- Le signal est une suite périodique finie .
- 2- Son spectre est également une suite périodique finie .
- 3- Les périodes du signal et les périodes du spectre contiennent le même nombre d'échantillons .

La relation de définition de la TFD est donnée sur une période par :

$$X(n) = \sum_{k=0}^{N-1} x(k) e^{-j2\pi \frac{nk}{N}} \quad n=0,1,\dots,N-1 \quad (2.6.1)$$

Remarque:

On peut remarquer que la TFD n'est autre que la transformée en Z évaluée aux points  $0,1,\dots,N-1$  avec  $Z = \exp(-j2\pi \frac{n}{N})$

Ecriture matricielle

Posons  $W = \exp(-j2\pi \frac{1}{N})$

La relation (2.6.1) peut s'écrire sous la forme matricielle suivante :

$$\begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} = \begin{bmatrix} W^0 & W^0 & \dots & W^0 \\ W^0 & W^1 & \dots & W^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ W^0 & W^{N-1} & \dots & W^{(N-1)^2} \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}$$

ou encore

$$\vec{X}(n) = [W] \vec{x}(k)$$

## II.6.2 TRANSFORMEE DE FOURIER RAPIDE (FFT)

La FFT est un ensemble d'algorithmes permettant d'accélérer le calcul des TFD .

Nous citons à titre d'exemple deux algorithmes qui sont [13] :

- L'algorithme de SANDER .
- L'algorithme de COOLEY.



CHAPITRE: III

CLASSIFICATION

## CHAPITRE : III

### CLASSIFICATION

#### III.1 INTRODUCTION

Les paramètres obtenus à partir des méthodes d'analyse discutées dans le chapitre précédent vont servir de données pour la classification en vue de la reconnaissance du locuteur .

La classification est un ensemble d'algorithmes permettant de définir des classes ( qui doivent recouvrir tout notre espace ) à partir d'un ensemble d'éléments composant un espace totalement inconnu .

La première étape de la classification dite d'apprentissage permet de choisir un représentant pour chaque classe , ces représentants de classes constitueront les références .

La deuxième étape dite de reconnaissance permet d'identifier un élément inconnu en le rangeant dans l'une des classes .

Dans ce chapitre, nous commençons l'étude des classificateurs de formes en considérant l'une des plus simples et intuitives approches de ce problème: le concept de classification de formes par les fonctions de distance .

La motivation pour l'utilisation des fonctions de décision comme instrument de classification suit naturellement le fait que la méthode la plus évidente à l'établissement de la mesure de similarité entre les vecteurs formes ( constitués par les paramètres d'analyse ) , que nous considérons aussi comme des points dans l'espace euclidien , est en déterminant leur proximité .



### III.2 CLASSIFICATION DE FORMES PAR LA DISTANCE MINIMALE

La classification par les fonctions de distance est l'un des premiers concepts dans la reconnaissance automatique des formes . Cette simple technique de classification est un instrument efficace pour la solution des problèmes dont les classes de formes exhibent un degré de variabilité raisonnablement limité .

En premier lieu nous considérons des classes caractérisées par des formes d'un seul prototype . Les concepts résultants sont étendus au cas des multiprototypes . Enfin, les propriétés générales de classification de cette approche sont discutées, et les limites des performances sur la classification sont établies .

#### III.2.1 MONOPROTOTYPES

Dans certaines situations, les formes de chaque classe tendent à bien se grouper autour d'une forme typique ou représentative de cette classe . Ceci a lieu, lorsque la variabilité de formes est élevée .

Un exemple typique est le problème de lecture de chèques bancaires par machine . Les caractères des chèques sont généralement écrits en encre magnétique pour faciliter la mesure du processus. Dans une telle situation, la mesure des vecteurs (formes) résultant de chaque classe de caractères sera à peu près identique puisque les mêmes caractères dans différents chèques sont identiques pour toutes les propositions pratiques. Sous ces conditions, les classificateurs à distance minimale peuvent constituer une approche très efficace pour le problème de classification .

Considérons  $M$  classes de formes et supposons qu'elles seront représentées par des formes prototypes  $Z_1, Z_2, \dots, Z_M$  .

La distance euclidienne entre un vecteur forme arbitraire  $X$  et le  $i^e$  prototype est donnée par :

$$D_i = \|X - Z_i\| = \sqrt{(X - Z_i)^t (X - Z_i)} \quad (3.2.1)$$

Le classificateur à distance minimale calcule la distance d'une forme  $X$  d'une classification inconnue du prototype de chaque classe, et l'affecte à la classe la plus proche . En d'autres termes ,  $X$  est affecté à une

classe  $w_i$  si :

$$D_i < D_j \quad \text{pour tout } j \neq i$$

L'équation (3.2.1) peut être exprimée en une forme plus convenable, élevons au carré les deux membres de l'équation donnée :

$$\begin{aligned} D_i^2 &= \|X - Z_i\|^2 = (X - Z_i)^t (X - Z_i) \\ &= XX^t - 2 X^t Z_i + Z_i^t Z_i \\ &= XX^t - 2 \left( X^t Z_i - \frac{1}{2} Z_i^t Z_i \right) \end{aligned} \quad (3.2.2)$$

Chercher le minimum de  $D_i^2$  est équivalent à chercher celui de  $D_i$  puisque toutes les distances sont positives .

De l'équation (3.2.2), cependant, puisque le terme  $XX^t$  est indépendant de  $i$  dans  $D_i^2$  avec  $i=1, \dots, M$  . Chercher le minimum de  $D_i^2$  revient à chercher le maximum de  $\left( X^t Z_i - \frac{1}{2} Z_i^t Z_i \right)$  .

Par conséquent on peut définir les fonctions de décision comme étant :

$$d_i(X) = X^t Z_i - \frac{1}{2} Z_i^t Z_i \quad i=1, \dots, M \quad (3.2.3)$$

où la forme  $X$  est affectée à la classe  $w_i$  si :

$$d_i(X) > d_j(X) \quad \text{pour tout } j \neq i$$

On remarque que  $d_i(X)$  est une fonction de décision linéaire; elle l'est si :

$$Z_{ij} \quad j=1, 2, \dots, n$$

sont les composantes de  $Z_i$  et posons

$$\begin{aligned} w_{ij} &= Z_{ij} \quad j=1, 2, \dots, n \\ w_{i \ n+1} &= -\frac{1}{2} Z_i^t Z_i \end{aligned} \quad \text{et} \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ 1 \end{bmatrix} \quad (3.2.4)$$

Alors on peut exprimer l'équation (3.2.3) sous la forme linéaire familière :

$$d_i(X) = W_i X \quad i=1, 2, \dots, M \quad (3.2.5)$$

où  $W_i = (w_{i1}, w_{i2}, \dots, w_{in}, w_{i \ n+1})$



On montre que pour la décision linéaire, la surface séparant chaque paire de prototypes  $Z_i$  et  $Z_j$  est un hyperplan qui est perpendiculaire à la ligne joignant les deux points .

La limite de décision pour un exemple de deux classes dont chacune est caractérisée par un seul prototype est illustrée par la (Fig. 3.1)

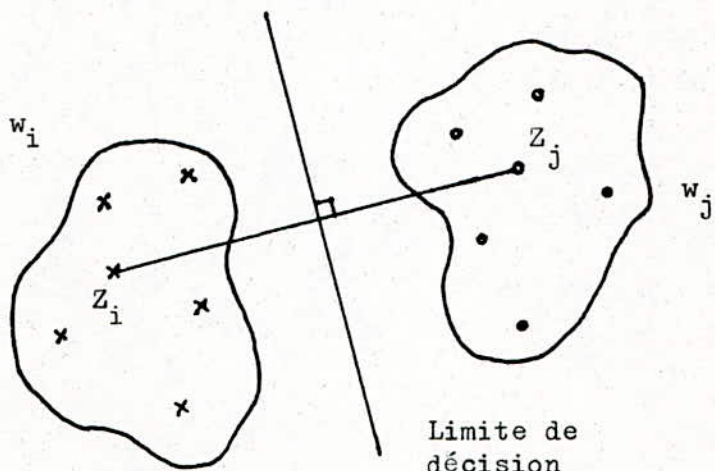


Fig.3.1  
Limite de décision  
pour deux classes  
caractérisée par un  
seul prototype .

### III.2.2 MULTIPROTOTYPES

Supposons qu'au lieu d'être représentée par une seule forme prototype, chaque classe est caractérisée par plusieurs prototypes, telle que, chaque forme de la classe  $w_i$  tend à se grouper autour de l'un des prototype  $Z_i^1, Z_i^2, \dots, Z_i^{N_i}$ , où  $N_i$  est le nombre de prototypes dans la  $i^e$  classe de formes .

Sous cette condition nous pouvons alors désigner un classificateur similaire à celui cité dans le paragraphe précédent .

La fonction de distance entre une forme arbitraire  $X$  et la classe  $w_i$  est notée par:

$$D_i = \min_l \|X - Z_i^l\| \quad l = 1, 2, \dots, N_i$$

$D_i$  représente alors la plus petite des distance entre  $X$  et chaque prototype de  $w_i$  .

La forme inconnue  $X$  est classée dans  $w_i$  si:

$$D_i < D_j \quad \begin{array}{l} i = 1, 2, \dots, M \\ \text{pour tout } i \neq j \end{array} \quad M: \text{ nombre de classes.}$$

Comme dans le paragraphe précédent on aboutit à la fonction de décision:

$$d_i(X) = \max_l (X^t Z_i^l) - \frac{1}{2} (Z_i^l)^t Z_i^l$$

$$d_i(X) = \max_l d_i^l(X) \quad \begin{matrix} l = 1, 2, \dots, N_i \\ i = 1, 2, \dots, M \end{matrix}$$

où  $d_i^l(x)$  est donné par :

$$d_i^l(X) = w_{i1}^l x_1 + w_{i2}^l x_2 + \dots + w_{in}^l x_n + w_{i, n+1}^l = W_i^l \cdot X$$

X est placé dans  $w_i$  si:

$$d_i(X) > d_j(X) \quad \text{pour tout } i \neq j$$

La (Fig. 3.2) nous donne un exemple sur les limites de décision qui sont représentées par les bissectrices perpendiculaires aux lignes joignant les prototypes des classes  $w_i$  et  $w_j$ , chacune étant représentée par deux prototypes.

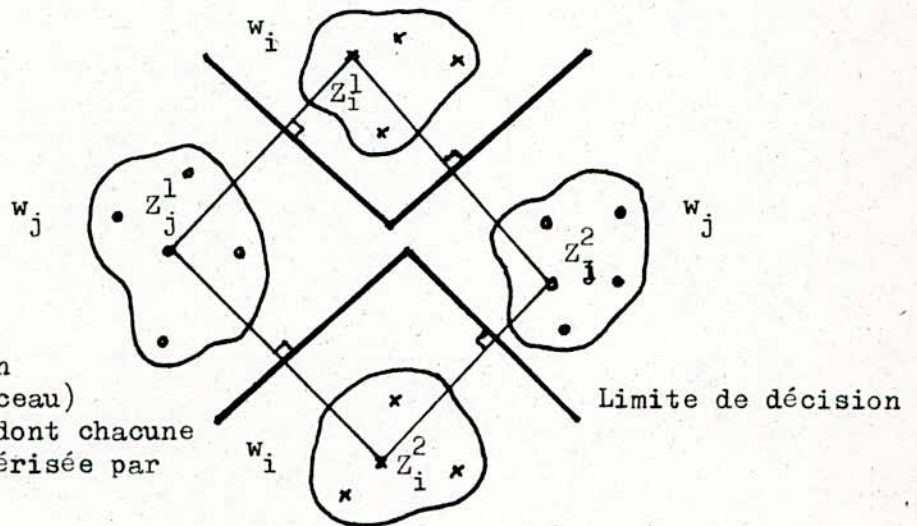


Fig.3.2  
Limites de décision  
(linéaires par morceau)  
pour deux classes dont chacune  
d'elles est caractérisée par  
deux prototypes .

### III.3 TYPES DE CLASSIFICATION

Il existe plusieurs techniques de classification qui sont réparties en deux groupes: les techniques supervisées et non supervisées selon le cas où nous possédons ou non les échantillons à traiter .

Si une modification des paramètres (ou de la structure de l'algorithme de classification) optimisant ces algorithmes se produit, la classification est dite adaptative, exemple l'algorithme ISODATA qui sera détaillé plus loin .



### III.3.1 CLASSIFICATION SUPERVISEE

Les méthodes supervisées peuvent être divisées en deux groupes:

- Celles, utilisant les hypothèses statistiques .
- Celles qui utilisent les hypothèses géométriques .

#### A) Méthodes statistiques

Parmi les concepts statistiques que font intervenir ces méthodes, citons la loi de distribution multinormale (Laplace - Gauss) .

Dans le cas où on ne peut pas faire cette hypothèse, on peut utiliser alors, une méthode statistique ne faisant pas appel aux lois de probabilité . Nous aboutissons alors à la méthode des K plus proches voisins et ses variantes .

#### B) Méthodes géométriques

Ces méthodes sont plus utilisées car elles sont moins coûteuses que les méthodes statistiques en temps de calcul .

Soit  $d_i(X, Y_j)$  la distance entre l'échantillon X et le  $j^e$  élément de la classe  $C_i$  .

On définit alors une mesure de similarité S entre X et la classe  $C_i$  :

$$S(X, C_i) = \frac{1}{p} \sum d_i^2(X, Y_j)$$

Où P est le nombre d'échantillons dans la classe  $C_i$  . On affecte alors X à la classe dont il est le plus proche .

### III.3.2 CLASSIFICATION NON SUPERVISEE

Dans ce cas le nombre et la nature des classes sont ignorés .

Il s'agit donc de grouper les vecteurs formes en sous ensembles .

Citons les principales méthodes pour ce type de classification .

#### A) Méthode des centroïdes

On va illustrer cette méthode à travers l'exemple suivant:

Supposons que nous avons à déterminer trois classes, choisissons alors trois formes quelconques de notre espace . Ces dernières serviront de premières estimations des centroïdes de ces trois classes, on associe alors à chaque

centroïde les éléments qui lui sont les plus proches . On a ainsi divisé notre espace en trois domaines . On détermine ensuite les centres de gravité de ces trois domaines, puis on classe à nouveau les formes . Ce qui donne trois nouveaux domaines .

Au bout d'un certain nombre d'itérations nous obtenons une stationnarité du processus : les trois domaines ainsi que leurs centres cessent d'évoluer . On a ainsi isolé les trois classes .

### **B) Méthode de groupement en chaîne**

Cette méthode ne suppose à priori aucune connaissance sur le nombre de classes . Dans le cas où les amas sont assez denses et séparés on prend les formes une à une dans un ordre quelconque, la première étant prise comme représentant de la première classe, on calcule ensuite la distance la séparant de la deuxième forme . Si cette distance excède un certain seuil fixé au préalable, on commence alors une deuxième classe, sinon la deuxième forme est affectée à la première classe . On procède de la même manière jusqu'à l'épuisement de toutes les formes .

Cette méthode sera détaillée dans le paragraphe (III.5.1)

### **C) Méthode utilisant une mesure de similitude fondée sur le nombre de voisins communs**

Etant donnés  $N$  échantillons  $X_1, X_2, \dots, X_N$  . On établit une liste des  $K$  plus proches voisins pour chaque échantillon ( chaque échantillon étant lui-même son plus proche voisin d'ordre zéro ) :

Les  $K$  échantillons sont rangés par ordre de distance croissante à  $X_i$  . On considère tous les couples  $(X_i, X_j)$ , on regarde d'abord si  $X_i$  figure parmi les  $K$  plus proches de  $X_j$  et réciproquement .

Si cette double condition est vérifiée et si en plus  $X_i$  et  $X_j$  ont en commun un nombre  $K_s$  de voisins ( $K_s \leq K$ ) alors  $X_i$  et  $X_j$  appartiennent à la même classe.

Le nombre minimal  $K_s$  de voisins communs correspond à un seuil de similitude . Les classes ainsi obtenues peuvent être groupées à leur tour en super classes jusqu'à ce que le nombre et la densité des groupements obtenus conviennent à l'application envisagée .



### III.4 RECHERCHE DES CLASSES

A partir des paragraphes précédents, il est évident que le moyen de déterminer les prototypes caractéristiques (centres de groupes) ayant un ensemble de données, joue un rôle central dans la désignation des classificateurs de formes basés sur le concept de la distance minimale .

Au début il est bien de mentionner, cependant, que la recherche de groupe est une expérience très bien orientée . Dans le sens que la performance des algorithmes donnés n'est pas seulement dépendante du type des données à analyser, mais aussi fortement influencée par la mesure de similarité des formes choisies et la méthode utilisée pour identifier les groupes .

Ces concepts, qui seront discutés dans les paragraphes suivants, fournissent aussi un fondement pour les systèmes de reconnaissance de formes non supervisée comme discuté dans le paragraphe (III.6) .

#### III.4.1 MESURE DE SIMILARITE

Pour définir un groupement de données, il est nécessaire en premier lieu de définir une mesure de similarité qui établira une règle pour assigner des formes au domaine du centre du groupement particulier .

Dans les paragraphes précédents nous avons considéré la distance euclidienne entre deux formes X et Z .

$$D = \| X - Z \| \quad (3.4.1)$$

comme mesure de leur similarité - plus la distance est petite plus la similarité est grande . Tous les algorithmes développés dans ce chapitre seront basés sur ce concept . Il y a, cependant, d'autres importantes mesures de distance qui sont parfois utiles . Citons à titre d'exemple, la distance de MAHALANOBIS de X à m

$$D = ( X - m )^t C^{-1} ( X - m ) \quad (3.4.2)$$

est une mesure de similarité pratique quand les propriétés statistiques sont explicitement considérées . Dans l'équation (3.4.2), C représente la matrice de covariance de la population de formes, m étant le vecteur moyen et X représente une forme variable .

Distance de MINKOWSKI :

$$d(X,Y) = |\alpha_i - \alpha'_i|$$

Distance de TCHEBYTCHEV :

$$d(X,Y) = \text{Max} |\alpha_i - \alpha'_i|$$

avec

$$X = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

$$Y = (\alpha'_1, \alpha'_2, \dots, \alpha'_n)$$

Les mesures de similarité n'ont pas besoin d'être restreintes à la mesure de distance . pour exemple, citons la mesure de similarité suivante :

$$S(X,Z) = \frac{X^t Z}{\|X\| \|Z\|} \quad (3.4.3)$$

qui est le cosinus de l'angle entre les vecteurs X et Z . Cette fonction est maximale quand X et Z sont orientés dans la même direction en respectant l'origine . Cette mesure de similarité est utile quand les régions de groupement tendent à se grouper le long d'axes principaux comme le montre la (Fig. 3.3) .

On voit sur cette figure, par exemple que X est plus similaire à la forme Z<sub>1</sub> qu'à Z<sub>2</sub> puisque

$$S(X,Z_1) > S(X,Z_2)$$

Cette mesure de similarité dépend de certaines conditions telle que la séparation suffisante entre les régions de groupement tout en respectant l'origine des coordonnées .

$$S(X,Z_1) = \text{Cos } \theta_1 = \frac{X^t Z_1}{\|X\| \|Z_1\|}$$

$$S(X,Z_2) = \text{Cos } \theta_2 = \frac{X^t Z_2}{\|X\| \|Z_2\|}$$

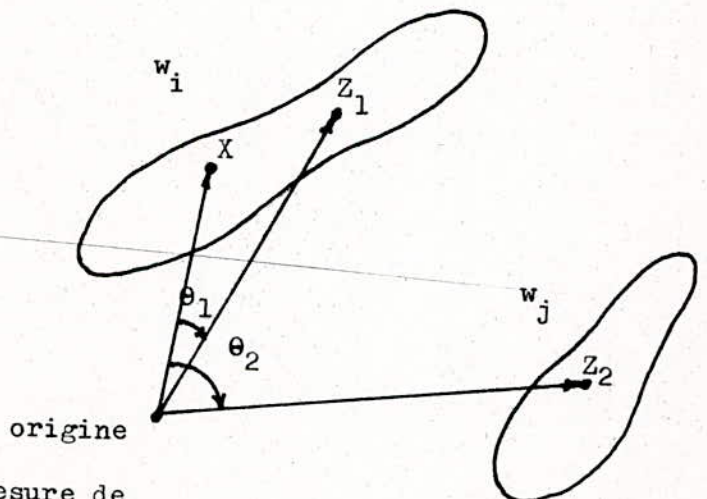


Fig.3.3  
illustration d'un mesure de similarité .



Dans ce qui va suivre, la discussion sera basée sur la mesure de similarité euclidienne donnée par l'équation (3.4.1) pour sa simple interprétation en termes du concept familier de proximité .

### III.4.2 CRITERE DE GROUPEMENT

Après l'adoption d'une mesure de similarité de formes, nous sommes toujours en face du problème de spécification d'une procédure pour partitionner les données en groupes .

Le critère de groupement utilisé peut représenter un schéma heuristique, où il peut être basé sur la minimisation (ou maximisation) d'un certain indice de performance .

L'approche heuristique est guidée par l'intuition et l'expérience . Elle consiste en la spécification en groupe de règles qui exploitent le choix de la mesure de similarité dans le but d'assigner les formes à un groupe . Cependant, puisque la proximité (mesure de la distance euclidienne) de deux formes est une mesure de similarité relative, d'où la nécessité d'établir un seuil pour une similarité acceptable dans la recherche des groupes .

Les algorithmes discutés dans les paragraphes (III.5.1) et (III.5.2) illustrent cette approche .

L'indice de performance est guidé par le développement d'une procédure qui minimisera ou maximisera l'indice de performance choisi . L'un des indices les plus souvent utilisé est la somme des erreurs quadratiques donnée par :

$$J = \sum_{j=1}^J \sum_{X \in S_j} \|X - m_j\|^2 \quad (3.4.4)$$

Où  $N_c$  est le nombre de classes ,  $S_j$  est la  $j^e$  classe et

$$m_j = \frac{1}{N_j} \sum_{X \in S_j} X \quad (3.4.5)$$

est le vecteur moyen de  $S_j$  ,  $N_j$  représente le nombre d'échantillons dans  $S_j$  . Un algorithme basé sur cet indice de performance sera discuté dans le paragraphe (III.5.3) .

L'algorithme ISODATA discuté dans le paragraphe (III.5.4) est basé sur la combinaison des deux approches précédentes .

### III.5 ETUDE DETAILLEE DES ALGORITHMES

#### III.5.1 UN SIMPLE ALGORITHME DE CLASSIFICATION

Supposons qu'on a un ensemble de  $N$  échantillons de formes  $X_1, X_2, \dots, X_N$  . Admettons que le premier centre de groupe  $Z_1$  est égal à quelques échantillons de formes, pour convenance on peut choisir  $Z_1 = X_1$  et fixons un seuil arbitraire non négatif  $T$  . Ensuite, nous calculons la distance  $D_{21}$  de  $X_2$  à  $Z_1$  ( voir équation (3.2.1) ) . Si cette distance dépasse  $T$  , un nouveau centre de groupe ,  $Z_2 = X_2$  est commencé . Autrement, on affecte  $X_2$  au domaine du centre de groupe  $Z_1$  .

Supposons que  $D_{21} > T$  alors  $Z_2$  est établi . Dans la marche suivante, la distance  $D_{31}$  et  $D_{32}$  de  $X_3$  à  $Z_1$  et  $Z_2$  sont calculées . Si les deux distances  $D_{32}$  et  $D_{31}$  sont supérieures à  $T$ , un nouveau centre de groupe  $Z_3 = X_3$  est créé . Autrement, on affecte  $X_3$  au domaine du centre de groupe qui est le plus proche . De la même manière, la distance de chaque nouvelle forme à chaque centre de groupe établi, est calculée et comparée au seuil, et un nouveau centre de groupe est créé si toutes les distances dépassent  $T$ .

Autrement, la forme est assignée au domaine du centre de groupe qui est le plus proche .

Les résultats des procédures acquises dépendent du premier centre de groupe choisi, la valeur de  $T$  et naturellement, les propriétés géométriques des données . Ces effets sont illustrés sur la (Fig.3.4 ) , on a trois arrangements du centre de groupe différents qui étaient obtenus pour les mêmes données simplement en variant  $T$  et le point de départ .

Bien que ces algorithmes ont des limitations, ils constituent une méthode simple et rapide pour établir des propriétés approximatives pour un exemple de données . Ainsi, puisqu'il est requiert seulement un passage à travers l'ensemble pour établir le domaine du groupe pour une valeur de  $T$  . En pratique, cependant, cette procédure requiert normalement des expérimentations avec différentes valeurs du seuil et différents points de départ dans le but d'obtenir un aperçu utile de la distribution géométrique des échantillons .



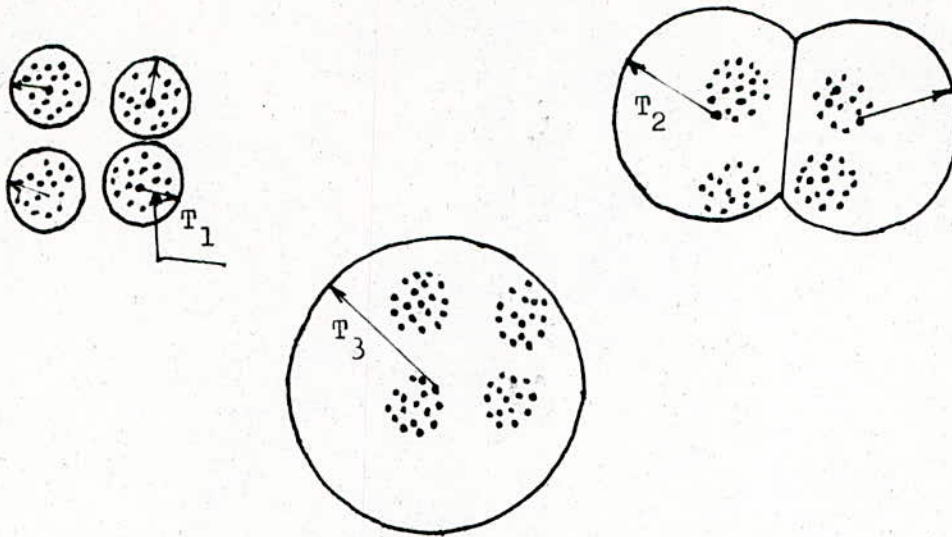


Fig. 3.4

Effets du seuil et des points de départ dans le schéma de la recherche des groupes d'échantillons .

Puisque les échantillons sont généralement de dimension élevée, qui empêche l'interprétation visuelle des résultats, cette information est obtenue premièrement par comparaison de la distance entre les centres de groupes et les nombres d'échantillons contenus dans les domaines résultants après chaque passage à travers les données . Les points les plus proches et les plus loins dans chaque domaine du centre de groupe et de la variance de ces domaines sont aussi des paramètres utiles . L'information ainsi obtenue dans chaque passage à travers les données peut être utilisée pour avantager la sélection de la nouvelle valeur de  $T$  et le point de départ pour le prochain passage . Cette procédure peut être attendue pour donner des résultats utiles dans des situations où les caractéristiques des données sont raisonnablement bien séparées en respectant les valeurs choisies du seuil .

### III.5.2 L'ALGORITHME DE LA DISTANCE MAXIMIN

L'algorithme de la distance maximin ( maximum - minimum ) est une autre procédure heuristique simple basée sur le concept de la distance euclidienne . Cet algorithme est principalement similaire au schéma présenté

dans le paragraphe(III.5.1) , avec l'exception qu'il identifie en premier lieu les régions de groupement qui sont loins l'une de l'autre . Cette procédure est bien illustrée par un exemple .

Considérons les dix échantillons bidimensionnels comme le montre la (Fig. 3.5.a) . Il est désiré d'utiliser l'algorithme de la distance MAXIMIN pour avoir une idée sur le nombre de domaines de groupes présents . Pour l'explication de cette algorithme, nous considérons la table de la ( Fig. 3.5.b ) . Une table contient les échantillons; l'autre, la liste des assignements des centres de groupes développés durant l'exécution de l'algorithme . La dernière table est initialement vide . Alors, dans le premier pas nous laissons arbitrairement  $X_1$  devenir le premier centre de groupe, désigné par  $Z_1$  dans la ( Fig. 3.5.b ) . Les nombres au dessus des flèches dans cette figure indiquent les pas dans lesquels les assignements des groupes ont pris place .

Aprés, nous déterminons l'échantillon le plus loin de  $X_1$ , qui dans ce cas est  $X_6$ , et appelons-le le centre  $Z_2$  .

Dans le troisième pas on calcule la distance entre chaque échantillon restant à  $Z_1$  et  $Z_2$  .

Pour chaque paire de ces calculs nous gardons les distances minimales .Ainsi, nous sélectionnons la distance maximale de ces minimales . Si cette distance est une fraction appréciable de la distance entre  $Z_1$  et  $Z_2$  ( on dit, au moins la moitié de cette distance) , on appelle l'échantillon correspondant le centre de groupe  $Z_3$  .

Autrement l'algorithme est terminé . Si nous utilisons ce critère dans le présent exemple, il est facilement vu par vérification que l'échantillon  $X_7$  devient le centre  $Z_3$  . Dans le pas suivant nous calculons la distance de chacun de ces trois centres de groupes établis aux formes restantes et gardons le min de chaque groupe des trois distances . Si cette distance est une fraction appréciable des distances max typiques précédentes, l'échantillon correspondant devient le centre  $Z_4$  . Une mesure utile des distances typiques précédentes est la moyenne de ces distances . Si ce critère est utilisé dans le présent exemple, et nous exigeons que la nouvelle distance max soit au moins égale à la moitié de la moyenne, on voit à partir de la (Fig.3.5.a) que la nouvelle distance maximale, qui arrive à être la distance entre  $X_1$  et  $X_3$  , ne satisfera pas cette condition . L'algorithme est donc terminé



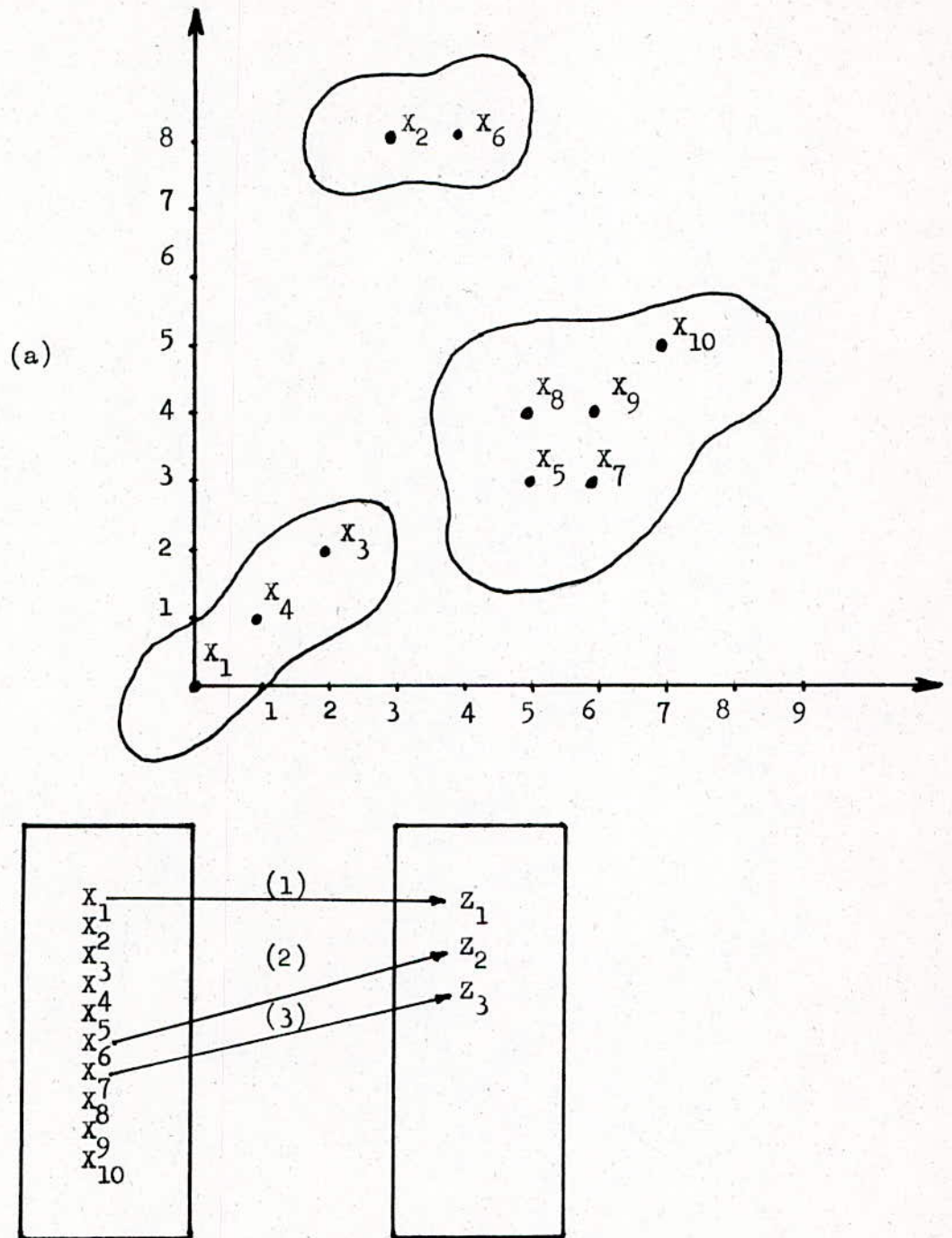


Fig. 3.5  
(a) Les formes d'échantillons utilisées dans l'illustration de l'algorithme de la distance MAXIMIN .  
(b) Echantillons et tables de correspondance .

à cette étape .

Dans le cas général, la procédure ci-dessus est répétée jusqu'à la nouvelle distance maximale à une étape particulière qui n'arrive pas à satisfaire la condition de création d'un nouveau centre de groupe .

Dans ce simple exemple nous avons obtenu trois centres de groupes  $X_1, X_2, X_7$  .

Pour assigner les échantillons restants au domaine de ces trois centres , on assigne simplement chaque échantillon au centre de groupe le plus proche . Donc, de la (Fig. 3.5.a ) on obtient les domaines de groupes  $( X_1 , X_3 , X_4 )$  ,  $( X_2 , X_6 )$  et  $( X_5 , X_7 , X_8 , X_9 , X_{10} )$  . Ces résultats concordent avec les classes auxquelles nous pourrions intuitivement s'attendre à avoir à partir de ces données . Pour obtenir une bonne représentation du centre pour chaque classe, on peut prendre la moyenne des échantillons pour chaque ensemble, en utilisant l'équation ( 3.4.5 ) . Ces moyennes peuvent alors être utilisées comme des nouveaux centres de classes .

### III.5.3 L'ALGORITHME DES K-MOYENNES

Les algorithmes tirés des paragraphes ( III.5.1) et (III.5.2) sont fondamentalement des procédures intuitives . L'algorithme présenté ci-dessus, dans l'autre main, est basé sur la minimisation de l'indice de performance qui est défini comme étant la somme des carrés de distances de tous les points dans un domaine de groupe du centre . Cette procédure, souvent appelée l'algorithme des k-moyennes, se compose des étapes suivantes :

Etape 1 :

Choisir K centres de groupes initiaux  $Z_1(1), Z_2(1), \dots, Z_K(1)$  . Ils sont arbitrairement et généralement sélectionnés comme étant les K premiers échantillons de l'ensemble d'échantillons donnés .

Etape 2 :

A la  $K^e$  étape itérative, on distribue les échantillons X entre les K domaines de groupes, utilisant la relation:

$$X \in S_j(k) \quad \text{si} \quad \| X - Z_j(k) \| < \| X - Z_i(k) \| \quad (3.5.6)$$

pour tout  $i \neq j \quad i=1, 2, \dots, K$



où  $S_j(k)$  dénote l'ensemble des échantillons dont le centre de groupe est  $Z_j(k)$ . Attachés à l'expression (3.5.6) sont résolues arbitrairement.

**Etape 3 :**

A partir des résultats de l'étape 2, on calcule les nouveaux centres  $Z_j(k+1)$ ,  $j=1,2,\dots,K$ , telle que la somme des distances de tous les points dans  $S_j(k)$  du nouveau centre est minimisée. Autrement dit, le nouveau centre de groupe  $Z_j(k+1)$  est calculé ainsi que l'indice de performance.

$$J_j = \sum_{X \in S_j(k)} \|X - Z_j(k+1)\|^2 \quad j=1,2,\dots,K$$

est minimisé.  $Z_j(k+1)$  qui minimise cet indice de performance est simplement la moyenne des échantillons de  $S_j(k)$ . Donc le nouveau centre de groupe est donné par:

$$Z_j(k+1) = \frac{1}{N_j} \sum_{X \in S_j(k)} X \quad j=1,2,\dots,K$$

Où  $N_j$  est le nombre d'échantillons dans  $S_j(k)$ . Le nom "K - Moyennes" est évidemment "tiré" de la manière dont les centres de classes sont séquentiellement renouvelés.

**Etape 4 :**

Si  $Z_j(k+1) = Z_j(k)$  pour  $j=1,2,\dots,K$ , l'algorithme a convergé et la procédure est terminée. Autrement, revenir à l'étape 2.

Le maintien de l'algorithme des K - Moyennes est influencé par le nombre de centres de classes spécifié, le choix des centres de groupes initiaux, l'ordre dans lequel les échantillons sont pris, et bien sûr, les propriétés géométriques des données. Bien qu'il n'existe pas de preuve générale de convergence pour cet algorithme, il peut être attendu de donner des résultats acceptables lorsque les données sont relativement loins l'une de l'autre. Dans la plupart des cas pratiques, l'application de cet algorithme exigera expérimentalement avec plusieurs valeurs de K aussi bien que les différents choix de la configuration du début.

### III.5.4 L'ALGORITHME ISODATA

L'algorithme ISODATA présenté dans ce paragraphe est principalement similaire à la procédure des K-Moyennes dans le sens que les centres de classes sont itérativement déterminés par les moyennes des échantillons . A la différence du dernier algorithme, cependant, ISODATA représente un ensemble absolument compréhensif d'une addition de procédures heuristiques qui étaient incorporées dans un schéma interactif .

Le mot " heuristique " pourrait être gardé clairement dans l'esprit comme le lecteur progresse à travers la discussion suivante, puisque plusieurs étapes qui pourront être présentées ont été incorporées dans l'algorithme comme le résultat de l'expérience .

Avant d'exécuter l'algorithme, il est nécessaire de spécifier un ensemble  $N_c$  de centres de classes initiaux ,  $Z_1, Z_2, \dots, Z_{N_c}$  . Cet ensemble, qui n'a pas besoin d'être nécessairement égal en nombre aux centres de classes désiré, peut être formé par sélection des échantillons de l'ensemble des données .

Pour un ensemble de  $N$  échantillons  $X_1, X_2, \dots, X_N$  , ISODATA consiste en les étapes principales suivantes:

#### Etape 1 :

Spécification des paramètres du processus suivants :

$K$ : Nombre de centres de classes désiré .

$\theta_N$ : Un paramètre auquel est comparé le nombre d'échantillons dans un domaine de groupe .

$\theta_s$  : Paramètre de déviation standard .

$\theta_c$  : Paramètre de groupement .

$L$  : Nombre max de paires de centres de groupes qui peuvent être groupés .

$I$  : Nombre d'itérations admises .

#### Etape 2 :

Distribuons les  $N$  échantillons entre les présents centres de groupes, en utilisant la relation suivante:

$$X \in S_j , \text{ si } \|X - Z_j\| < \|X - Z_i\| \\ i=1,2,\dots,N_c , i \neq j$$



pour tout  $X$  dans l'ensemble des échantillons . Dans cette relation,  $S_j$  représente le sous-ensemble des échantillons assignés au centre de groupe  $Z_j$  .

**Etape 3 :**

Ecartons les sous-ensembles d'échantillons dont les nombres sont inférieurs à  $\theta_N$ , tel que, pour chaque  $j$ ,  $N_j < \theta_N$  , écartons  $S_j$  et réduisons  $N_c$  de 1 .

**Etape 4 :**

Cherchons chaque centre  $Z_j$  ,  $j=1,2,\dots,N_c$  en égalisant celui-ci à la moyenne des échantillons de l'ensemble correspondant  $S_j$  ,

$$Z_j = \frac{1}{N_j} \sum_{X \in S_j(k)} X \quad j=1,2,\dots,N_c$$

où  $N_j$  est le nombre d'échantillons dans  $S_j$  .

**Etape 5 :**

Calculons la distance moyenne  $\bar{D}_j$  des échantillons dans  $S_j$  à partir de son centre , utilisant la relation :

$$\bar{D}_j = \frac{1}{N_j} \sum_{X \in S_j(k)} \|X - Z_j\| \quad , \quad j=1,2,\dots,N_c$$

**Etape 6 :**

Calculons la distance moyenne générale des échantillons à partir de leurs centres de groupes respectifs, utilisant la relation :

$$\bar{D} = \frac{1}{N} \sum_{j=1}^{N_c} N_j \bar{D}_j$$

**Etape 7 :**

- a) Si cette itération est la dernière, faisons  $\theta_c = 0$  et allons à l'étape 11 .
- b) Si  $N_c \leq \frac{K}{2}$  , aller à l'étape 8 .
- c) Si le nombre d'itérations est pair ou si  $N_c \geq 2K$  aller à l'étape 11 .  
Autrement, continuer .

**Etape 8 :**

Trouver le vecteur de déviation standard

$$\sigma_j = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{pj})^t \text{ pour chaque sous-ensemble}$$

d'échantillons, utilisant la relation :

$$\sigma_{ij} = \sqrt{\frac{1}{N_j} \sum_{X \in S_j(k)} (X_{ik} - Z_{ij})^2} \quad \begin{matrix} i=1,2,\dots,P \\ j=1,2,\dots,N_c \end{matrix}$$

où P est la dimension de l'échantillon,  $X_{ik}$  est la  $i^e$  composante du  $k^e$  échantillon dans  $S_j$ ,  $Z_{ij}$  est la  $i^e$  composante de  $Z_j$ , et  $N_j$  est le nombre d'échantillons dans  $S_j$ . Chaque composante de  $\sigma_j$  représente la déviation standard des échantillons dans  $S_j$  le long d'un axe de coordonnées principal.

**Etape 9 :**

Trouver la composante max de chaque  $\sigma_j$   $j=1,2,\dots,N_c$  et appelons la :  $\sigma_j \text{ max}$

**Etape 10 :**

Si pour chaque  $\sigma_j \text{ max}$   $j=1,2,\dots,N_c$ , on a :

$$\sigma_j \text{ max} > \theta_s \text{ et}$$

$$a) \bar{D}_j > \bar{D} \quad \text{et} \quad N_j > 2(\theta_N + 1)$$

ou

$$b) N_c \leq \frac{K}{2}$$

alors divisons  $Z_j$  en deux nouveaux centres  $Z_j^+$  et  $Z_j^-$ , effacer  $Z_j$ , et augmentons  $N_c$  de 1.

Le centre de groupe  $Z_j^+$  est formé en additionnant une quantité donnée  $\gamma_j$  à la composante de  $Z_j$  qui correspond à la valeur de la composante max de  $\sigma_j$ ;  $Z_j^-$  est obtenu en retranchant  $\gamma_j$  de la même composante de  $Z_j$ ,

$$\text{avec } \gamma_j = k \sigma_j \text{ max}, \quad 0 < k \leq 1$$

Si la division a eu lieu dans cette étape, revenir à l'étape 2, sinon, continuer.



**Etape 11 :**

Calculer les distances par paires  $D_{ij}$  pour tous les centres .

$$D_{ij} = \| Z_i - Z_j \| \quad , \quad \begin{matrix} i=1,2,\dots,N_c-1 \\ j=1, \dots, N_c \end{matrix}$$

**Etape 12 :**

Comparons les distances  $D_{ij}$  au paramètre  $\theta_c$  .

Arrangeons les L plus petites distances qui sont inférieures à  $\theta_c$  dans l'ordre croissant :

$$\left[ D_{i_1 j_1}, D_{i_2 j_2}, \dots, D_{i_L j_L} \right]$$

où  $D_{i_1 j_1} < D_{i_2 j_2} < \dots < D_{i_L j_L}$  et L est le nombre max de paires de centres qui peuvent être associés ensemble . Le processus de groupement est discuté dans l'étape suivante .

**Etape 13 :**

Avec chaque distance  $D_{i_l j_l}$  est associée une paire de centres  $Z_{i_l}$  et  $Z_{j_l}$  . Commençons par la plus petite de ces distances, exécutant l'opération de groupement en paire qui s'accorde avec la règle suivante :

pour  $l = 1, 2, \dots, L$  , si ni  $Z_{i_l}$  ni  $Z_{j_l}$  n'ont été utilisés dans le groupement dans cette itération, confondons ces deux centres en utilisant la relation suivante :

$$Z_l^* = \frac{1}{N_{i_l} + N_{j_l}} \left[ N_{i_l} (Z_{i_l}) + N_{j_l} (Z_{j_l}) \right]$$

effaçons  $Z_{i_l}$  et  $Z_{j_l}$  , et diminuons  $N_c$  de 1 .

**Etape 14 :**

Si celle-ci est la dernière itération, l'algorithme est terminé . Autrement revenir à l'étape 1 si des paramètres du processus exigent un changement, ou revenir à l'étape 2 si les paramètres restent inchangés pour l'étape suivante .

Une itération est comptée chaque fois que la procédure retourne à l'étape 1 ou 2 .

### III.6 RECONNAISSANCE DE FORMES NON SUPERVISEE

La recherche des classes peut être vue comme un problème dans la reconnaissance de formes non supervisée .

Le problème de l'étude non supervisée peut être rangé comme celui de l'identification des classes dans l'ensemble de formes donné .

L'application des algorithmes de la recherche des classes à l'étude non supervisée est, en principe loyale . Etant donné un ensemble de formes  $( X_1, X_2, \dots, X_N )$  d'une classification inconnue . Ces formes peuvent être soumises à l'un ou plusieurs algorithmes discutés précédemment dans l'effort d'identifier les centres de classes représentatifs . Les domaines de groupes résultants peuvent être interprétés comme des classes de formes différentes .

Les centres de classes identifiés dans l'étude non supervisée peuvent être utilisés comme la base pour un classificateur à distance minimale . Donc, il est clair que la désignation d'un système de reconnaissance de formes non supervisé exigera une grande intuition et un grand nombre d'expérimentations .

### III.7 CONCLUSION

L'important dans ce chapitre sont les principaux concepts soulignant l'idée des classificateurs de formes basés sur la distance .

Deux approches sont mises en évidence dans le développement des algorithmes présentés dans ce chapitre, il s'agit des approches, heuristique et celle basée sur l'indice de performance .



CHAPITRE : IV

RECONNAISSANCE DU LOCUTEUR .

---

## CHAPITRE: IV

### RECONNAISSANCE DU LOCUTEUR

#### IV.1 SYSTEMES DE RECONNAISSANCE DU LOCUTEUR

Dans la reconnaissance du locuteur, les techniques du traitement numérique sont souvent la première étape dans ce qui est essentiellement un problème de reconnaissance de forme . Ceci est schématisé sur la figure 4.1 .

Comme vu dans cette figure, une représentation ( vecteur forme ) du signal de la parole est obtenue en utilisant les techniques du traitement numérique de la parole qui conservent les caractéristiques du signal de la parole qui ont un rapport avec l'identité du locuteur . Les formes résultantes sont comparées aux formes de référence précédemment et une décision subséquente et logique est utilisée pour faire un choix parmi les alternatives valables .

Il existe deux types distincts de reconnaissance du locuteur - vérification du locuteur et identification du locuteur .

Pour la vérification du locuteur une identité est prétendue par l'utilisateur, et la décision exigée du système de vérification est strictement binaire; i.e , accepter ou rejeter l'identité prétendue . Dans le but de faire cette décision, un ensemble de caractéristiques désigné pour retenir l'information essentielle concernant l'identité du locuteur est mesurée par un ou plusieurs essais de ce dernier, et les mesures résultantes sont comparées ( souvent en utilisant une mesure de comparaison non linéaire ) à un ensemble de formes de références stocké pour le locuteur prétendu . Donc pour la vérification du locuteur seule une comparaison singulière entre l'ensemble ( les ensembles ) des mesures , et la forme de référence exige de faire la décision finale pour accepter ou rejeter l'identité prétendue .

Généralement une mesure de distance entre les mesures données et la distribution des références stockées est calculée . Basé sur des coûts relatifs pour faire les deux types d'erreurs possibles ( i.e , accepter un imposteur et rejeter le vrai locuteur ), un seuil approprié est fixé sur la fonction de distance, . Il est facilement montré que la probabilité de faire les deux types d'erreurs décrits précédemment est essentiellement



indépendant de  $N$ , le nombre de formes références stockées dans le système, puisque les formes références pour tous les autres locuteurs vont former une distribution stable qui caractérise tous les locuteurs. Énoncé dans plusieurs termes mathématiques, si nous notons la distribution de probabilité pour un vecteur de mesure  $X$  pour le  $i^e$  locuteur comme  $P_i(X)$ , alors une règle de décision simple pour la vérification du locuteur pourrait être de la forme

- vérification du  $i^e$  locuteur si:  $P_i(X) > C_i P_{av}(X)$

- rejet du  $i^e$  locuteur si:  $P_i(X) < C_i P_{av}(X)$

où  $C_i$  est une constante pour le  $i^e$  locuteur qui détermine les probabilités d'erreurs pour le  $i^e$  locuteur,  $P_{av}(X)$  est la moyenne (sur tous les locuteurs dans l'ensemble) de la distribution de probabilité pour la mesure  $X$ . En variant la constante  $C_i$ , un contrôle simple de l'erreur de confusion entre les deux types d'erreurs est facilement obtenu.

Le problème d'identification du locuteur diffère significativement du problème de vérification du locuteur. Dans ce cas le système est demandé pour faire une identification absolue parmi les  $N$  locuteurs d'une population. Donc au lieu d'une comparaison singulière entre un ensemble de mesures et une forme référence stockée,  $N$  comparaisons complètes sont exigées. La règle de décision pour tels systèmes est de la forme :

Choisissons le  $i^e$  locuteur tel que  $P_i(X) > P_j(X)$   
 $j=1,2,\dots,N$   
 $i \neq j$

i.e, choisissons le locuteur avec la probabilité d'erreur absolue minimale. Dans ce cas il paraît raisonnable que comme la population test devient très grande, la probabilité d'erreur doit tendre vers 1 puisque un nombre infini de distributions ne peut pas rester clair dans un espace de paramètres finis. i.e, il devient de plus en plus probable que deux ou plusieurs locuteurs dans un ensemble, peuvent avoir des distributions de mesures qui sont extrêmement proches  $P_i \approx P_j$  l'une de l'autre. Sous ces circonstances, une identification du locuteur fiable devient essentiellement impossible. Il peut être vu à partir de la discussion précédente qu'il existe un grand nombre de similarités, aussi bien que des différences, entre les systèmes

d'identification et de vérification du locuteur . Chaque processus exige du locuteur de prononcer une ou plusieurs phrases test, fait quelques mesures sur ces phrases test, et donc calcule une (ou plusieurs ) fonction de distance entre le **vecteur** de mesure et celui de la référence . Donc, en termes d'aspect du traitement du signal de ces deux problèmes, les méthodes sont bien similaires .

Les différences majeures se produisent dans la décision logique .

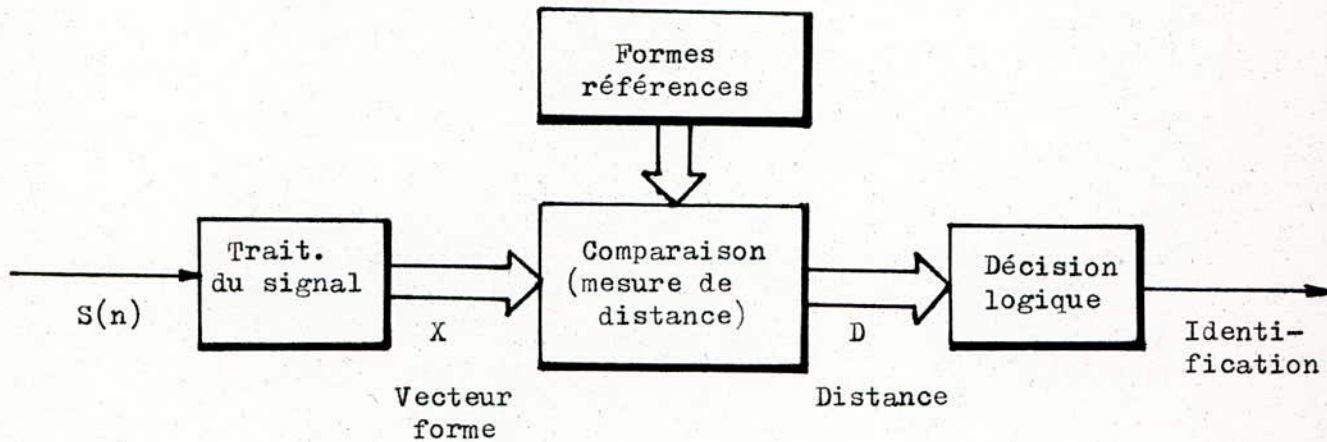


Fig.4.1 :  
Représentation générale de reconnaissance du locuteur .

#### IV.2 SYSTEME DE VERIFICATION DU LOCUTEUR

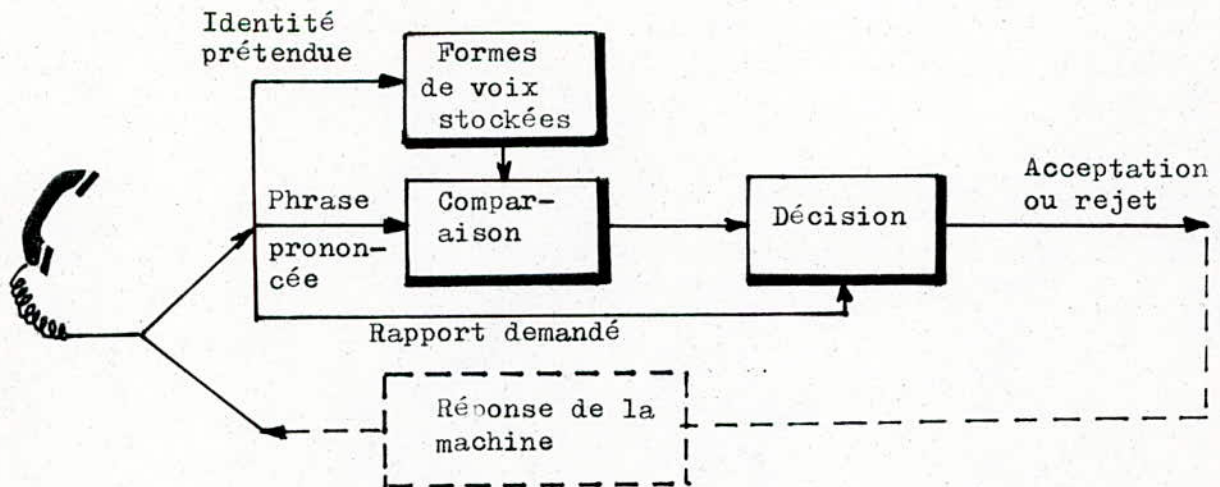


Fig.4.2 :  
Diagramme d'un système de vérification du locuteur  
( d'après Rosenberg )



La figure 4.2 montre un diagramme en bloc d'un système digital de vérification du locuteur développé par Roseberg . La personne souhaitée d'être vérifiée entre en premier son identité prétendue, et alors sur demande du système, prononce sa phrase de vérification, et demande ce qu'elle a à faire dans le cas où elle est acceptée . Par exemple, cette demande peut être l'accès à une information privilégiée d'une banque, etc .

La phrase prononcée est traitée pour obtenir une forme qui est comparée aux autres formes de référence stockées pour l'identité prétendue et donc une décision d'accepter ou de rejeter l'utilisateur est faite .

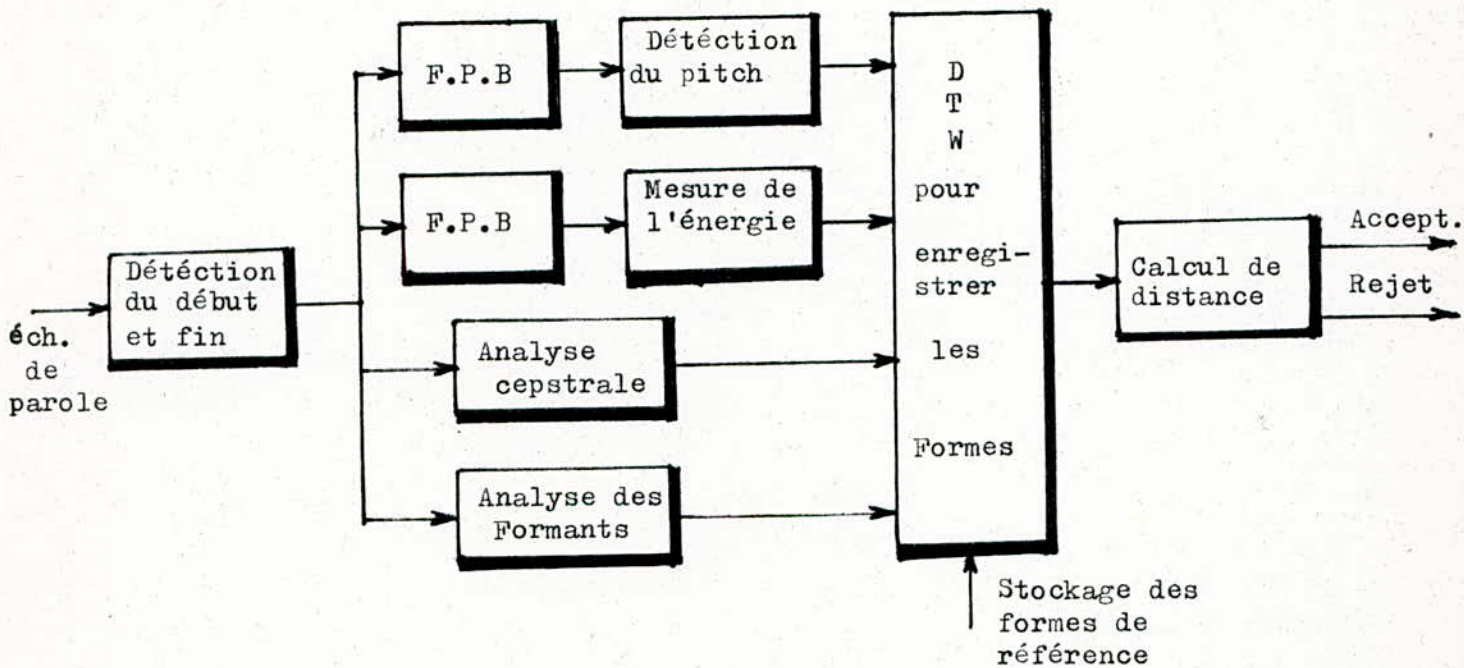


Fig.4.3 :

Aspects du traitement du signal pour le système de vérification du locuteur .

La figure 4.3 montre un diagramme en bloc des parties traitement du signal de vérification du locuteur . L'échantillon de voix qui se produit

dans un intervalle de temps présélectionné doit être exactement détecté . Ceci étant le travail du système de détection .

Une fois que le début et la fin du mot sont trouvés, des séries de mesures et estimations des paramètres sont faites pour fournir les formes qui représentent la voix .

En particulier, un détecteur du pitch est utilisé pour mesurer le contour du pitch de la voix, une mesure d'énergie est faite pour donner les contours d'énergie, l'analyse cepstrale est utilisée pour donner les contours des paramètres cepstraux, et finalement une estimation de la localisation des formants est faite . ( Tous les paramètres montrés dans la figure 4.3 sont parfois utilisés dans le système de vérification ; cependant, on réduit l'ensemble de mesures aux mesures du pitch et de l'intensité, vu la grande qualité de calcul exigée pour l'analyse cepstrale ou l'estimation des formants.) .

La figure 4.4 montre les diagrammes des trois premiers formants avec le pitch et l'intensité pour une certaine version de la phrase test : " We were away a year ago " prononcée par des locuteur mâles .

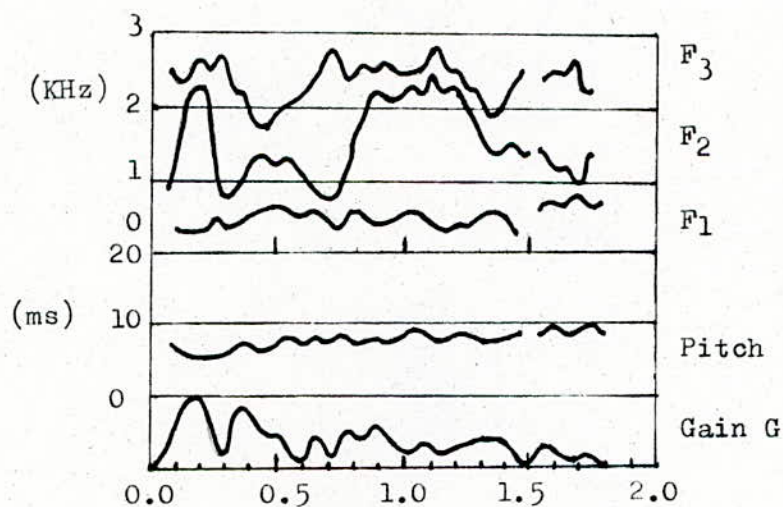


Fig.4.4  
Diagramme des trois premiers Formants, Pitch et intensité pour la vérification du locuteur  
( d'après Lummis.)



Après que la représentation paramétrique désirée soit calculée, il est nécessaire de la comparer aux formes de référence correspondantes pour le locuteur dont l'identité est prétendue . On n'a pas le droit de comparer les formes temporelles comme le pitch, l'intensité et les variations des formants, puisque un locuteur n'est généralement pas capable de prononcer la phrase de vérification précisément avec la même vitesse pour différentes répétitions . La solution à cette difficulté a été une déformation non linéaire de l'échelle du temps pour les formes à l'entrée pour obtenir la meilleure correspondance possible entre les formes de référence stockées du locuteur prétendu, et les formes mesurées pour son échantillon de parole .

Le processus de la déformation temporelle est extrêmement important et il a été utilisé dans diverses applications du traitement de la parole .

#### IV.3 ALIGNEMENT TEMPOREL

Le processus de la déformation temporelle est comme illustré dans la figure 4.5 .

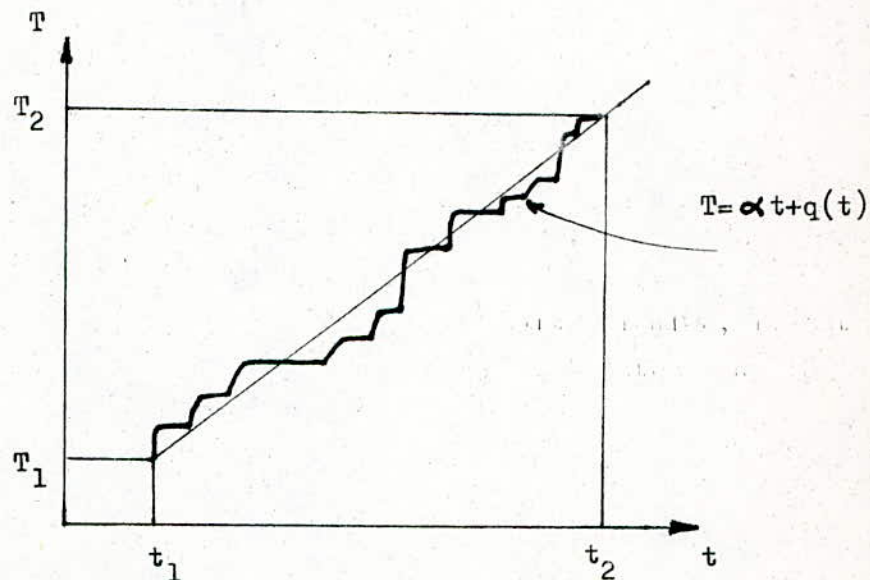


Fig.4.5 illustration de l'alignement (déformation) temporel .

L'idée est de déformer (aligner ) l'échelle temporelle  $t$  de la référence pour que les résultats significatifs dans quelques contours des mesures  $a(t)$  s'alignent avec les mêmes résultats significatifs dans les contours de référence  $r(t)$  . La fonction de déformation est de la forme :

$$T = \alpha t + q(t) \quad (4.3.1)$$

où  $q(t)$  est la fonction de déformation non linéaire,  $\alpha$  représente la moyenne répandue de la fonction de déformation temporelle . Notons que l'élimination de  $q(t)$  correspond à une simple modification linéaire . Les conditions aux limites sont imposées pour assurer l'alignement des points de début et fin de l'échantillon et de la référence . Ces conditions aux limites sont :

$$T_1 = \alpha t_1 + q(t_1) \quad (4.3.2 a)$$

$$T_2 = \alpha t_2 + q(t_2) \quad (4.3.2 b)$$

Le problème qui reste est de choisir la fonction  $q(t)$  et la constante  $\alpha$  dans le but d'améliorer l'alignement des contours de mesures . Une très simple approche est de définir une fonction de  $t$  linéaire par morceaux avec un nombre fini de points le long de l'axe des temps où la pente de  $q(t)$  change . Les points et les pentes de  $q(t)$  ( aussi bien que la pente  $\alpha$  ) sont donc déterminés par le type de descente raides d'optimisation dans lequel, soit la mesure de distance entre les contours du test et de la référence, ou une mesure de corrélation est utilisée pour contrôler la direction de la recherche d'optimisation .

Une solution simple et très rapide pour le problème d'alignement temporel est d'utiliser la méthode de la programmation dynamique pour un choix optimal des contraintes de la fonction de déformation .

En imposant la condition de continuité de la fonction de déformation, plutôt qu'imposer une approximation linéaire par morceaux, il est plus intéressant de déterminer la fonction de déformation optimale appropriée pour un ensemble de contours . Considérons l'exécution de l'alignement temporel approprié pour une paire de contours qui sont échantillonnés à un ensemble de points discrets . Soient les points dans le contour mesuré classés :  $n=1,2,\dots,N$  et soient les points dans le contour de référence classés :  $m=1,2,\dots,M$  , nous espérons donc choisir une fonction de déformation temporelle  $W$  , telle que :

$$m = W(n) \quad (4.3.3)$$



Les conditions aux limites sur  $W(n)$  sont :

$$\begin{cases} W(1) = 1 & \text{points de départ .} \\ W(N) = M & \text{points de la fin .} \end{cases} \quad (4.3.4)$$

Si une déformation linéaire était utilisée, alors la fonction de déformation temporelle  $W$  pourrait être de la forme :

$$W(n) = \left[ \left( \frac{M - 1}{N - 1} \right) (n - 1) + 1 \right] \quad (4.3.5)$$

Pour une fonction de déformation non linéaire plus générale, nous devons considérer une stratégie pour le début au point  $n = 1$ ,  $m = 1$  et un traitement à travers la grille discrète pour finir au point  $n = N$ ,  $m = M$  comme exigé par les conditions aux limites. Pour limiter le degré de non linéarité de la fonction de déformation, il est plus raisonnable d'imposer la condition de continuité souple : la fonction de déformation  $W$  ne peut changer par plus de deux points de la grille pour quelques indices  $n$ . Plus formellement nous exigeons :

$$\begin{aligned} W(n + 1) - W(n) &= 0, 1, 2 & \text{si } W(n) &= W(n - 1) \\ &= 1, 2 & \text{si } W(n) &= W(n - 1) \end{aligned} \quad (4.3.6)$$

Donc la pente de la fonction de déformation est soit 0, 1 ou 2 si au précédent indice de grille l'indice de déformation change, ou 1, 2 si au précédent indice de la grille l'indice de déformation reste constant.

La mesure de similarité (ou mesure de distance) est utilisée pour déterminer le chemin de la fonction de déformation qui localement minimise la distance totale maximale, sujet des contraintes de continuité de l'équation (4.3.6).

L'exemple de la figure(4.6) montre le domaine des coordonnées possibles de la grille  $(n, m)$  et une fonction typique de déformation  $W(n)$  ( la ligne dans la grille ) pour l'alignement de 20 points de référence ( $N = 20$ ) à 15 points de test ( $M = 15$ ) de la parole. A cause des contraintes de continuité, la fonction de déformation doit être dans le parallélogramme montré sur cette figure.

La technique de l'alignement temporel est appliquée pour la reconnaissance du locuteur et pour la reconnaissance de la parole . Comme exemple la figure (4.7) ( due à Rosemberg ) montre un ensemble de contours d'intensité d'une référence et d'un test de parole, avant et après l'alignement temporel non linéaire .

Les étapes finales dans le processus de vérification de la figure (4.2) sont le calcul de quelques mesures de distance, et donc de comparer la distance au seuil choisi .

La simple mesure de la distance du contour est une somme normalisée de carrés; pour le  $j^e$  contour, la distance  $d_j$  devrait être de la forme :

$$d_j = \sum_i \left[ \frac{[a_{js}(i) - a_{jr}(i)]}{a_j(i)} \right]^2 \quad (4.3.7)$$

où  $a_{js}(i)$  est la valeur du  $j^e$  contour de mesure au temps  $i$  ,  $a_{jr}(i)$  est la valeur de la  $j^e$  mesure du contour de référence au temps  $i$  ,  $a_j(i)$  est déviation standard de la  $j^e$  mesure au temps  $i$  .

La fonction de distance totale  $D$  , est généralement une somme pondérée des carrés, i.e,

$$D = \sum_j w_j d_j \quad (4.3.8)$$

où  $w_j$  est le  $j^e$  poids, choisi sur la base de l'efficacité de la  $j^e$  mesure vérifiant le locuteur .

Pour un tel système, le taux d'erreur égal ( i.e, quand les probabilités d'une fausse rejection, et ou d'une fausse acceptation sont égales ) utilisant des imitateurs professionnels est autour de 4.1 % , dans le cas d'une parole téléphonique on atteint 7% , avec le pitch et l'intensité .

L'addition de mesures plus sophistiquées, telles que les formants, ou les paramètres cepstraux pourrait réduire significativement ce taux d'erreur.



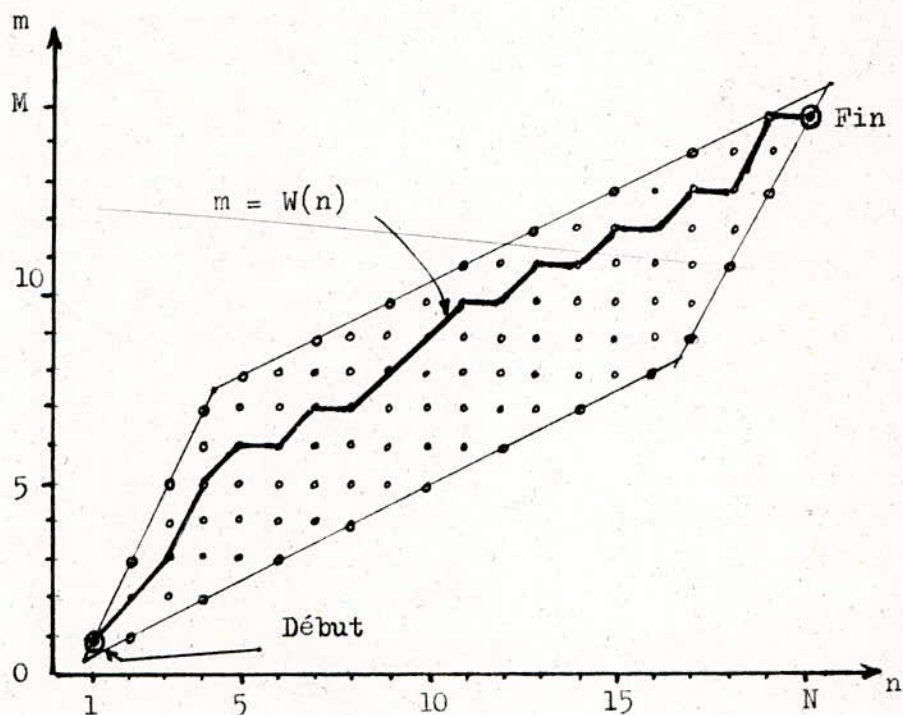


Fig.4.6  
Un exemple d'une fonction de déformation  
typique (d'après ITAKURA) .

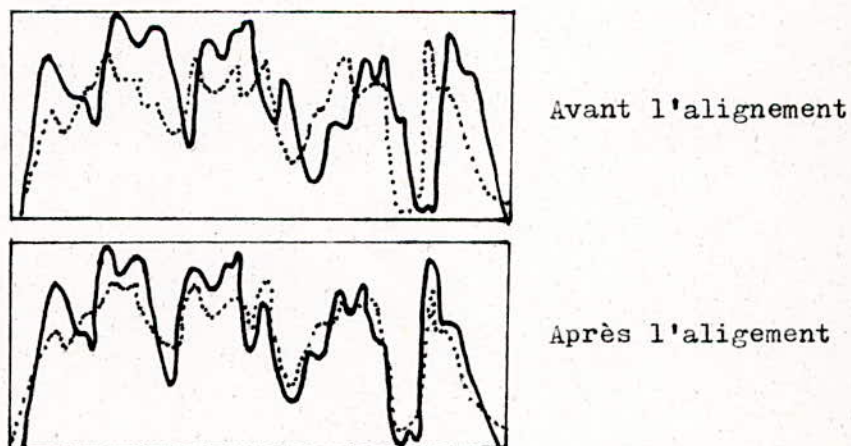


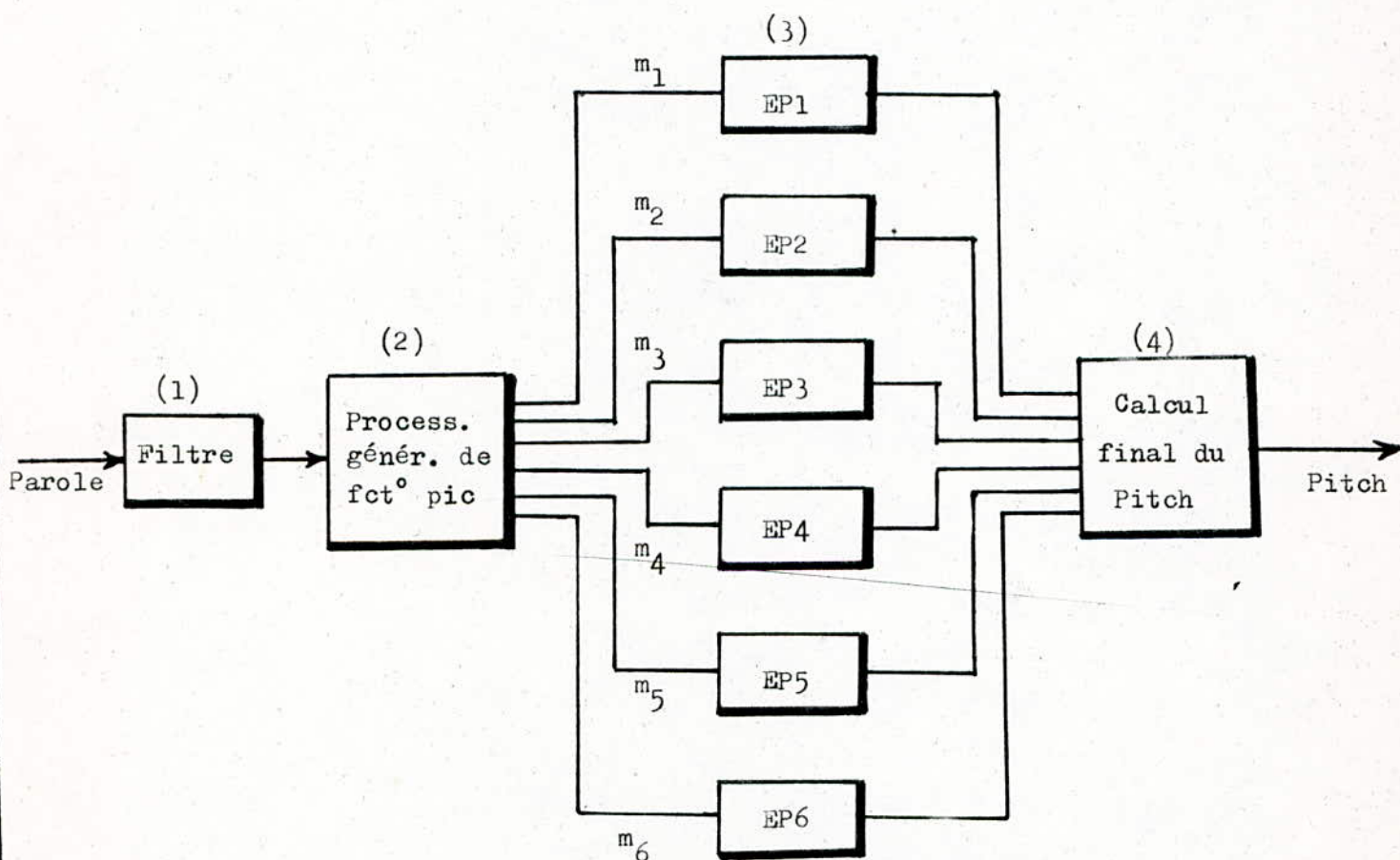
Fig.4.7  
Un exemple des effets de l'alignement temporel  
sur le contour de l'intensité de la parole.  
( d'après Rosenberg.)

#### IV.4 ESTIMATIONS DU PITCH ET DES FORMANTS

##### IV.4.1 ESTIMATION DU PITCH PAR LA METHODE DU TRAITEMENT PARALLELE

L'idée de base dans le traitement parallèle est qu'une amélioration de la précision pourrait être obtenue en combinant convenablement les sorties de plusieurs estimateurs élémentaires du pitch [6].

Un algorithme a été utilisé pour l'estimation finale du pitch basée sur l'utilisation de six estimateurs élémentaires ( Fig.4.8 ).



Six estimateurs individuels du pitch

Fig.4.8 :

Un diagramme en bloc de l'algorithme de l'estimation du pitch .



Cet algorithme peut être convenablement divisé en quatre étapes comme le montre la figure 4.8 .

- 1- Filtrage du signal de parole .
- 2- Génération de six fonctions de pics du signal de parole traité .
- 3- Six estimateurs de pitch simples et identiques, travaillant sur l'une de ces fonctions .
- 4- Calcul du pitch final, basé sur l'examen des résultats à partir de chaque estimateur .

Pour améliorer cet algorithme, deux modifications peuvent être introduites :

- 1- La première modification a pour but d'accélérer l'exécution et de limiter la valeur de la fréquence fondamentale à 220 Hz ( donc la voix masculine est la plus convenable pour l'analyse ) .
- 2- La deuxième modification conduit à une réalisation hardware appréciable, plus simple que celle de l'algorithme original et donne une fréquence fondamentale de la voix en dessous de 300Hz .

Cet algorithme ainsi que ses deux modifications ont été trouvés précis et efficaces dans les tests sur la parole réelle et synthétique .

La vitesse de calcul de cet algorithme ( et ses modifications ) est plus rapide que celle des techniques utilisant le cepstre ou la fonction d'autocorrelation .

#### IV.4.2 ESTIMATION DES FORMANTS

La partie inférieure de l'échelle temporelle du cepstre correspond essentiellement aux conduit vocal, pulsation glottale et information transmise, tandis que la partie supérieure est due à la l'excitation .

Les formants peuvent être estimés en localisant les pics dans le  $\text{Log}(\text{spectre})$  . Celui-ci est obtenu à partir du cepstre dans lequel la partie supérieure de l'échelle temporelle est éliminée ( lissage , élimination de l'information correspondant au pitch ) , voir (Fig.4.9) [20] .

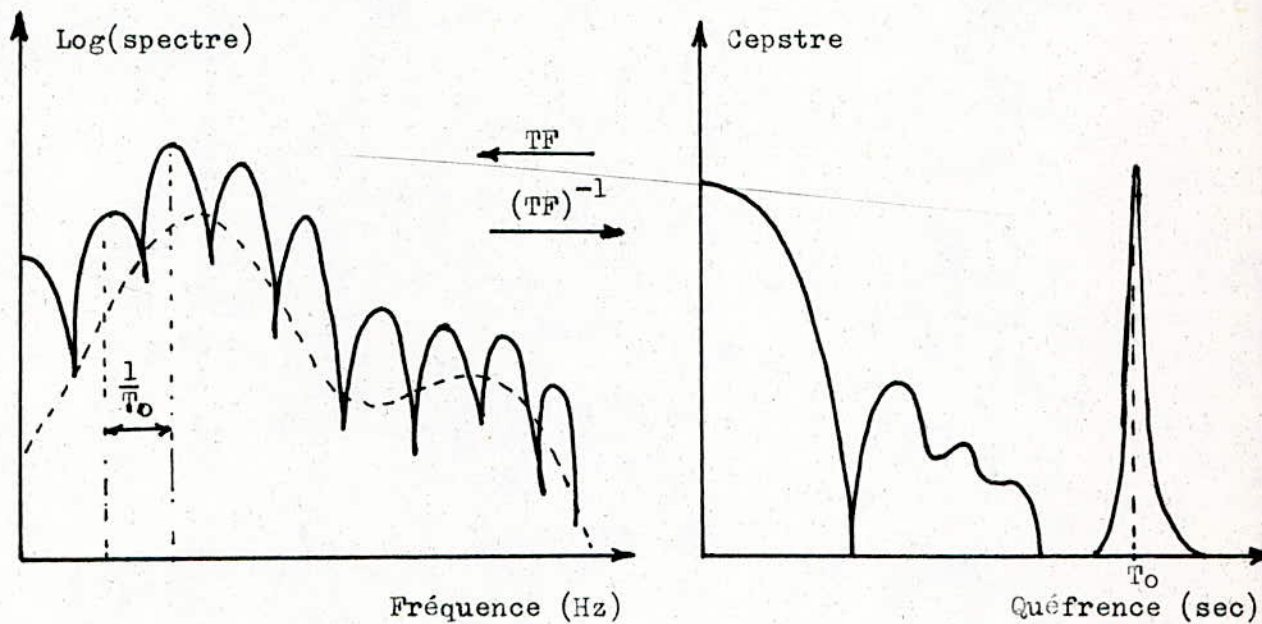


Fig.4.9 [9]

Cepstre d'un signal . A la période du fondamental de la voix,  $T_0$  , correspond dans le cepstre un pic à la quéfreence  $T_0$  .

Les basses quéfreences correspondent à la structure formantique

Notons aussi que les formants peuvent être estimés par plusieurs méthodes telles que la LPC et la méthode de passage par zéro du signal vocal .



#### IV.5 SYSTEMES D'IDENTIFICATION DU LOCUTEUR

Il y a beaucoup de similarités entre les problèmes de vérification et d'identification du locuteur .

En termes d'aspects de traitement du signal, les deux problèmes sont à peu près traités de la même manière dont la plus grande partie du traitement montrée par la fig.4.3 est appliquée à l'identification aussi bien qu'à la vérification du locuteur . Les différences principales dans le traitement nécessitent le choix des paramètres utilisés pour les mesures des distances, et la nécessité de mesurer N distances ( pour l'identification ) au lieu d'une seule . La décision finale pour l'identification du locuteur est de choisir le locuteur dont les formes de référence sont les plus proches en terme de distance à celle du test, tandis que pour la vérification du locuteur, la décision finale est un choix binaire : accepter ou rejeter l'identité prétendue basée sur la comparaison de la distance à un seuil de distance convenablement choisi .

Bien qu'une mesure de distance classique simple du type donné dans l'équation (4.3.8) est généralement appropriée pour la plupart des systèmes de vérification du locuteur, une mesure de distance plus sophistiquée et robuste était utilisée par Atal et les autres pour les systèmes d'identification du locuteur . Rappelons que l'objet du calcul de la distance est de fournir une mesure de similarité entre la forme de référence et celle du test .

La mesure de distance utilisée par Atal peut être comme suit : Soit X un vecteur forme à L dimensions représentant la forme d'entrée, dans lequel la  $k^e$  composante de X est la  $k^e$  mesure .

Il est pris que la fonction de densité de probabilité commune aux mesures pour le  $i^e$  locuteur est une distribution gaussienne multidimensionnelle avec moyenne  $m_i$  et matrice de covariance  $W_i$  . Donc, la fonction de densité gaussienne à L dimension pour X est donnée par:

$$g_i(X) = \frac{1}{\sqrt{2\pi |W_i|}} \exp \left[ -\frac{1}{2}(X-m_i)^t W_i^{-1}(X-m_i) \right] \quad (4.5.9)$$

Où  $W_i^{-1}$  est la matrice inverse de la matrice  $W_i$  ( en prenant  $W_i$  non singulière ),  $|W_i|$  est le déterminant de  $W_i$ , et  $t$  indique la transposé d'un vecteur, la règle de décision qui minimise la probabilité d'erreur énonce que le vecteur de mesure  $X$  pourrait être assigné à la classe  $i$  si:

$$P_i g_i(X) \geq P_j g_j(X) \quad \text{pour tout } i \neq j \quad (4.5.10)$$

où  $P_i$  est la probabilité à priori que  $X$  appartienne à la classe  $i$ . Puisque  $\ln y$  est une fonction monotone et croissante de la variable  $y$ , la règle de décision de l'équation (4.5.10) peut être considérablement simplifiée et réécrite comme:

Décidons la classe  $i$  si

$$d_i(X) = \frac{1}{2} (X - m_i)^t W_i^{-1} (X - m_i) + \frac{1}{2} \ln |W_i| - \ln P_i \leq d_j(X) \quad (4.5.11)$$

Les deux derniers termes de la partie droite de l'équation (4.5.11) ne dépendent pas du vecteur de mesure  $X$  et peuvent être considérés comme un terme constant représentant un biais envers la  $i^e$  classe.

Comme un problème pratique il a été trouvé que la plupart des cas d'intérêt, la règle de décision qui inclut le biais ne fournit pas un avantage significatif sur une règle de décision basée exclusivement sur le premier terme de la partie droite de l'équation (4.5.11). Donc, au lieu de la règle de décision de l'équation (4.5.11), la mesure de distance  $\hat{d}_i$  est définie comme:

$$\hat{d}_i = (X - m_i)^t W_i^{-1} (X - m_i) \quad (4.5.12)$$

est généralement utilisée et l'indice  $i$  est choisi tel que  $\hat{d}_i$  est minimisé pour tout  $i$ .

La règle de décision exige le calcul du vecteur moyen  $m_i$  et de la matrice de covariance  $W_i$  pour chaque classe  $i$  dans l'ensemble de décision. Pour un ensemble formé de  $N_i$  vecteurs de mesure avec les composantes  $X_1(n)$  appartenant à la classe  $i$ , la moyenne et la matrice de covariance sont donnés par les relations :



$$m_i = \frac{1}{N_i} \sum_{n=1}^{N_i} X_i^{(n)} \quad (4.5.13)$$

$$W_i = \frac{1}{N_i} \sum_{n=1}^{N_i} X_i^{(n)} X_i^t(n) - m_i m_i^t \quad (4.5.14)$$

L'hypothèse de la distribution normale pour les mesures peuvent être justifiée à partir de plusieurs considérations .

En premier, pour la correction de la règle de décision , il est suffisant que la distribution soit normale au centre de sa gamme - une propriété souvent trouvée pour être vraie dans le cas des mesures physiques .

Un avantage majeur de la distance métrique de l'équation (4.5.12) est que la distance  $\hat{d}_i$  est invariante en respectant une certaine transformation linéaire.

Une seconde propriété importante de cette distance métrique est qu'elle est particulièrement efficace dans la comparaison des composantes des différents vecteurs .

En utilisant la règle de décision de l'équation (4.5.12) Atal a examiné l'efficacité des différentes représentations paramétriques du signal de la parole en terme de problème d'identification du locuteur. Il utilise une population de 10 locuteurs, chacun prononce la même phrase test six fois . Chaque voix est divisée en 40 segments égaux donc fournissent une déformation temporelle . La longueur moyenne de chaque segment est autour de 50 ms. La précision de l'identification du locuteur est testée en sélectionnant une phrase pour chaque locuteur comme test et les cinq autres restantes pour références pour chaque locuteur .

Après étude et comparaison des résultats obtenus à partir des différentes représentations paramétriques pour un tel segment de parole, la représentation cepstrale donne de meilleurs précisions que les autres .

En combinant plusieurs trames pour obtenir un vecteur forme de dimension élevée, de hautes précisions peuvent être atteintes . La figure (4.10) montre les variations de la précision de l'identification atteinte en utilisant les mesures cepstrales comme fonction de la durée totale de la l'intervalle de parole dans l'établissement du calcul de distance . Il peut être vu dans cette figure que les précisions d'identification ont atteint 95 % pour cet ensemble limité de locuteurs pour des échantillons

de parole de durée de l'ordre de 0.5 seconde .

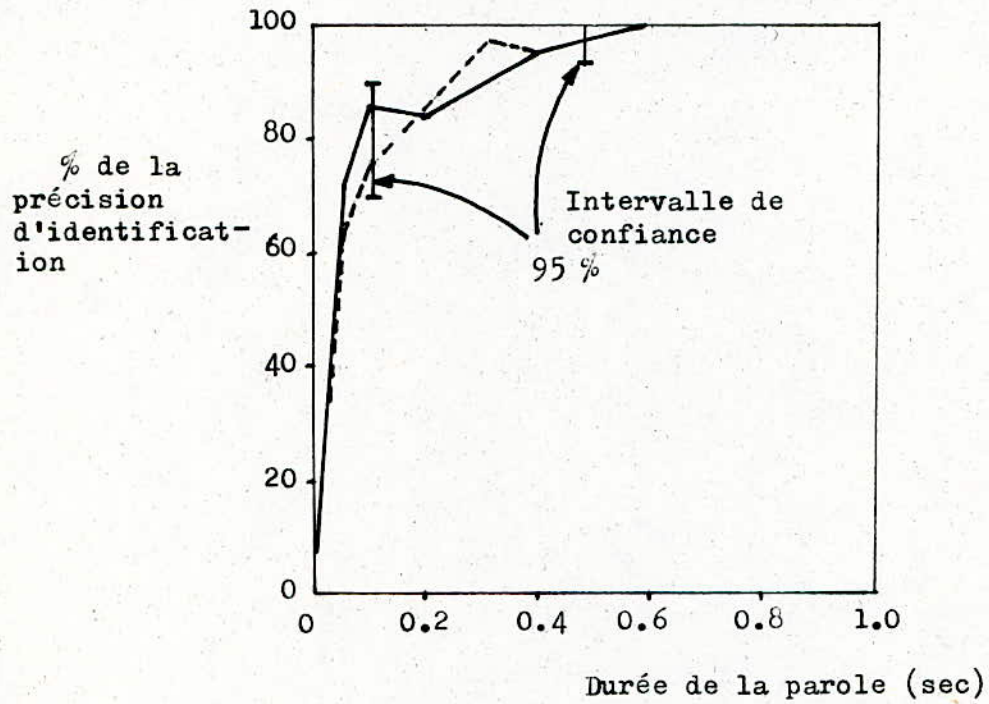


Fig.4.10  
Précision d'identification du locuteur ( utilisant  
les paramètres cepstraux ) comme une fonction de  
la durée de la parole .  
( d'après Atal. )



## CONCLUSION

Actuellement, la communication parlée est à la une dans la recherche, l'identification du locuteur qui en fait partie a ses applications dans le domaine judiciaire ( identifier un criminel par exemple) et peut aussi servir pour avoir accès à des informations secrètes ( accès à la salle des ordinateurs de Texas Instruments aux USA par exemple ) .

Dans ce problème d'identification, le premier souci est de déterminer les paramètres ( invariance interlocuteur ) caractérisant la voix d'un locuteur, Une question peut se poser : ne pourrait-il pas exister des invariants qui nous sont inconnus?, car jusqu'à présent, seuls les aspects temporels, fréquentiels et énergétiques ont été utilisés . Cela peut être l'objet d'une étude plus approfondie .

Le travail que nous avons effectué nous a permis d'élargir nos connaissances dans le domaine de la parole . Nous avons aussi approché de plus près l'algorithmique et la programmation .

ANNEXE :

PROGRAMMES .



```

5      PRINT"          *****
10     PRINT"          *      PROGRAMME DES K-MOYENNES      *
15     PRINT"          *****
20     INPUT "N";N
30     INPUT "K";K
40     INPUT "P";P
50     DIM TABLE(N,P) : DIM CENT(K,P) : DIM CLASSE(K,N)
60     DIM VECT(K,P)
70     FOR I=1 TO N
80     FOR J=1 TO P
90     INPUT TABLE(I,J)
100    NEXT J
110   NEXT I
120   FOR I=1 TO K
130   FOR J=1 TO P
140   CENT(I,J)=TABLE(I,J)
150   NEXT J
160   NEXT I
170   U=1 : CPT=0
180   WHILE U<=K
190   F=1 : L=1
200   WHILE L<=K
201   IF U=L THEN L=L+1
202   IF L>K GOTO 330
210   I=1
220   WHILE I<=N
230   N1=0 : N2=0
240   FOR M=1 TO P
250   N1=((TABLE(I,M)-CENT(U,M))^2)+N1
260   N2=((TABLE(I,M)-CENT(L,M))^2)+N2
270   NEXT M
280   N1=N1^.5
290   N2=N2^.5
300   IF N1-N2<0 THEN CLASSE(U,F)=I : F=F+1
310   I=I+1
320   WEND
330   L=L+1
340   WEND
350   FOR T=1 TO P
360   VECT(U,T)=CENT(U,T)
370   NEXT T
380   TROUV=0
381   J=1
390   WHILE J<=P
400   FOR S=1 TO F
410   CL=CLASSE(U,S)
420   CENT(U,J)=TABLE(CL,J)+CENT(U,J)
425   NEXT S
430   CENT(U,J)=CENT(U,J)/F
460   IF VECT(U,J)<>CENT(U,J) THEN TROUV=1

```

```
461   J=J+1
470   WEND
480   IF TROUV<>1 THEN CPT=CPT+1
490   U=U+1
500   WEND
501   IF CPT<>K GOTO 170
510   FOR I=1 TO K
520   PRINT "CLASSE NUMERO";I
530   J=1
540   WHILE CLASSE(I,J)<>0 AND J<=N-K+1
550   PRINT "ECHANTILLON";CLASSE(I,J)
560   FOR L=1 TO P
570   PRINT TABLE(CLASSE(I,J),L)
580   NEXT L
590   J=J+1
600   WEND
610   NEXT I
620   FOR I=1 TO K
630   PRINT "CENTRE";I
640   FOR J=1 TO P
650   PRINT CENT(I,J)
660   NEXT J
670   NEXT I
680   END
```



```

5      REM                                     *****
6      REM                                     *   PROGRAMME ISODATA   *
7      REM                                     *****
10     INPUT "N";N
20     INPUT "KI";KI
30     INPUT "P";P
40     INPUT "NC";NC
50     INPUT "LI";LI
60     INPUT "IT";IT
70     INPUT "TETAN";TETAN
80     INPUT "TETAS";TETAS
90     INPUT "TETAC";TETAC
91     IF NC>KI THEN NI=NC
92     NI=KI
93     INPUT "LI"
94     IF LI>NI(NI-1)/2 THEN GOTO 93
100    DIM MOBS(P,N); DIM MCLAS(N,NI); DIM NCLAS(NI); DIM IST(3,NI(NI-1))
110    DIM X1(P); DIM X2(P); DIM NCENT(P,NI); DIM VECT(NI)
120    DIM SIGMA(P,NI); DIM SIMAX(NI); DIM L(NI); DIM LST(NI)
121    FOR I=1 TO N
122    FOR J=1 TO P
123    INPUT MOBS(J,I)
124    NEXT J
125    NEXT I
130    CPT1=0 ;CPT3=1
150    DB=0
160    FOR J=1 TO NI
161    LL=1
162    FOR I=1 TO NI
170    IF I=J THEN GOTO 270
180    FOR K=1 TO N
190    FOR L=1 TO P
200    X1(L)=MOBS(L,K)-MOBS(L,J)
210    X2(L)=MOBS(L,K)-MOBS(L,I)
220    X1=X1+(X1(L))^2
230    X2=X2+(X2(L))^2
240    NEXT L
250    IF X1-X2<0 THEN MCLAS(LL,J)=K;LL=LL+1
260    NEXT K
270    NEXT I
280    NCLAS(J)=LL
290    IF NCLAS(J)>=TETAN THEN NI=NI-1;GOTO 300
291    REM
300    REM                                     CALCUL DU NOUVEAU CENTRE
301    REM                                     -----
310    FOR I=1 TO P
320    NCENT(I,J)=0
330    FOR L=1 TO NCLAS(J)
340    K=MCLAS(I,J)
350    NCENT(I,J)=NCENT(I,J)+MOBS(I,K)

```

```

360 NEXT L
370 NCENT(I,J)=NCENT(I,J)/NCLAS(J)
380 NEXT I
381 REM
390 REM          CALCUL DE DJB
391 REM          -----
400 DJB=0
410 FOR L=1 TO NCLAS(J)
420 K=MCLAS(L,J)
430 FOR I=1 TO P
440 X1(I)=MOBS(I,K)-NCENT(I,J)
450 X1=X1+(X1(I))^2
460 NEXT I
470 DJB(J)=DJB(J)+SQR(X1)
480 NEXT L
490 DJB(J)=DJB(J)/NCLAS(J)
491 REM
500 REM          CALCUL DE DB
501 REM          -----
510 DB=DB+(DJB(J)*NCLAS(J))
520 NEXT J
530 DB=DB/N
540 IF CPT1=IT THEN TETAC=0; GOTO 990
550 IF NC<=KI/2 THEN GOTO 590
560 Q=CPT1/2
570 IF 2*Q-CPT1=0 THEN GOTO 990
580 IF NC>=2*K THEN GOTO 990
590 FOR J=1 TO NI
600 FOR I=1 TO P
610 SIGMA(I,J)=0
620 FOR L=1 TO NCLAS(J)
630 K=MCLAS(L,J)
640 SIGMA(I,J)=SIGMA(I,J)+(MOBS(I,K)-NCENT(I,J))^2
650 NEXT L
660 SIGMA(I,J)=SIGMA(I,J)/NCLAS(J)
670 SIGMA(I,J)=SQR(SIGMA(I,J))
680 NEXT I
690 NEXT J
691 REM
700 REM          TROUVER LA COMPOSANTE MAX DE CHAQUE SIGMA(J)
701 REM          -----
710 CPT=1
720 FOR J=1 TO NI
730 FOR I=1 TO P
740 MAX=SIGMA(I,J)
750 IF MAX<SIGMA(I+1,J) THEN MAX=SIGMA(I+1,J);CPT=CPT+1;L(J)=CPT
760 NEXT I
770 SIGMA(J)=MAX
780 NEXT J
781 REM
790 REM          DIVISION DES CENTRES
791 REM          -----

```



```

800 FOR J=1 TO NI
810 IF SIMAX(J)<=TETAS THEN GOTO 980
820 IF DJB(J)<=DB THEN GOTO 850
830 IF NCLAS(J)<=2*(TETAN+1) THEN GOTO 850
840 GOTO 860
850 IF NC>KI/2 THEN GOTO 990
860 FOR L=1 TO P
870 NCENT(L,NI+1)=NCENT(L,J)
880 NEXT L
890 NCENT(L(J),NI+1)=NCENT(L(J),J)-.5*(SIGMA(J))
900 NCENT(L(J),J)=NCENT(L(J),J)+.5*(SIGMA(J))
910 NI=NI+1
920 VECT(CPT3)=J
930 CPT3=CPT3+1
940 VECT(CPT3)=NI
950 CPT3=CPT3+1
960 CPT1=CPT1+1
970 GOTO 150
980 NEXT J
981 REM
990 REM          CALCUL DES DISTANCES PAR PAIRE DE CENTRES
991 REM          -----
1000 FOR I=1 TO NI-1
1010 FOR J=I+1 TO NI
1020 V=0
1030 FOR L=1 TO P
1040 V=V+(CENT(I,L)-CENT(J,L))^2
1050 NEXT L
1060 IF V>=TETAC THEN GOTO 1110
1070 DIST(1,K)=V
1080 DIST(2,K)=I
1090 DIST(3,K)=J
1100 K=K+1
1110 NEXT J
1120 NEXT I
1130 FOR M=1 TO K-1
1140 FOR L=M+1 TO K
1150 IF DIST(1,M)<=DIST(1,L) THEN GOTO 1190
1160 Z1=DIST(1,M):Z2=DIST(2,M):Z3=DIST(3,M)
1170 DIST(1,M)=DIST(1,L):DIST(2,M)=DIST(2,L):DIST(3,M)=DIST(3,L)
1180 DIST(1,L)=Z1:DIST(2,L)=Z2:DIST(3,L)=Z3
1190 NEXT L
1200 NEXT M
1210 FOR U=1 TO LI
1220 BOOL=1
1230 V=1
1240 WHILE VECT(V)<>0 AND BOOL=1
1250 IF DIST(2,U)=VECT(V) THEN BOOL=0
1260 IF DIST(3,U)=VECT(V) THEN BOOL=0
1270 V=V+1
1280 WEND
1290 IF BOOL<>1 THEN GOTO 1370
1300 FOR I=1 TO P
1310 T=DIST(2,U): F=DIST(3,U)
1320 NCENT(I,T)=(NCLAS(T)*NCENT(I,T)+NCLA(F)*NCENT(I,F))

```

```
1330 NCENT(I,T)=NCENT(I,T)/(NCLAS(T)+NCLAS(F))
1340 NCENT(I,F)=0
1350 NEXT I
1360 NI=NI-1
1361 VECT(V)=0
1370 NEXT U
1372 I=1
1373 FOR V=1 TO NI
1374 IF VECT(V)=0 THEN LIT(I)=V: I=I+1
1375 NEXT V
1376 I=1
1377 FOR V=1 TO NI
1378 IF V>=LIT(I) THEN VECT(V)=VECT(V+1):I=I+1
1379 NEXT V
1380 IF CPT1<>IT GOTO 150
1390 FOR I=1 TO KI
1400 PRINT "CLASSE NUMERO";I
1410 J=1
1420 WHILE MCLAS(J,I)<>0 AND J<=N-NI+1
1430 PRINT "ECHANTILLON";MCLAS(J,I)
1440 FOR L=1 TO P
1450 PRINT MOBS(L,MCLAS(J,I))
1460 NEXT L
1470 J=J+1
1480 WEND
1490 NEXT I
1500 FOR I=1 TO KI
1510 PRINT "CENTRE";I
1520 FOR J=1 TO P
1530 PRINT NCENT(J,I)
1540 NEXT J
1550 NEXT I
1560 END
```



-0 BIBLIOGRAPHIE 0-

- [1]- P.Y ARQUES (1982)  
" Décision en traitement du signal." MASSON
- [2]- M. BELLANGER (1981)  
" Traitement numérique du signal." MASSON
- [3]- B. BOUSSEKSOU (1983)  
" Reconnaissance automatique de la parole par la méthode globale. Application aux particularités linguistiques de l'arabe standard."  
Thèse de magister. Université d'ALGER.
- [4]- A. FELDBAUM (1972)  
" Principes théoriques des systèmes asservis optimaux."  
MIR. MOSCOU
- [5]- P. FONDANECHÉ, P. GILBERTAS (1981)  
" Filtres numériques, principes et réalisations."  
MASSON
- [6]- B. GOLD, L. RABINER (1969)  
"Parallel processing techniques for estimating pitch periods of speech in the time domain."  
JASA
- [7]- Y. GRENIER (1977)  
" Identification du locuteur et adaptation au locuteur d'un système de reconnaissance phonémique."  
Thèse de Docteur-Ingénieur. ENST
- [8]- M. GUERTI (1983)  
" Contribution à la synthèse de la parole en arabe standard."  
Thèse de magister. Université d'ALGER
- [9]- J. GUIBERT (1979)  
" La parole. Compréhension et synthèse par les ordinateurs."  
PUF
- [10]- T. TOU JULIUS, RAFAEL C. GONZALEZ (1974)  
" Pattern recognition principles."
- [11]- M. KUNT (1981)  
" Traitement numérique des signaux." DUNOD

- [12]- J.S LIENARD (1977)  
 " Les processus de la communication parlée. Introduction  
 à l'analyse et à la synthèse de la parole." MASSON
- [13]- J. LIFERMAN (1981)  
 " Les méthodes rapides de transformation du signal."  
 MASSON
- [14]- J. MARIANI  
 " Traitement automatique de la parole: perspective  
 industrielle."  
 Support de cours de formation permanente. ENST
- [15]- J.D MARKEL, A.H GRAY (1976)  
 " Linear prediction of speech." SPRINGER-VERLAG
- [16]- J. MAX (1981)  
 " Méthodes et techniques de traitement du signal et  
 applications aux mesures physiques." Tome 1 MASSON
- [17]- A. MENACER (1985)  
 " Reconnaissance de la parole en mode multilocuteur  
 par des méthodes globales (mots isolés)."  
 Thèse de Docteur-Ingénieur. Université de RENNES
- [18]- B. PICINBOMO (1980)  
 " Eléments de théorie du signal." DUNOD
- [19]- A.E ROSENBERG  
 " Automatic speaker verification."  
 IEEE Avril 1976
- [20]- RW SCHAFER , L.R RABINER (1978)  
 " Digital processing of speech signals."  
 PRENTICE-HALL
- [21]- T. SUZUKI (1980)  
 " Etat des recherches en reconnaissance du locuteur."  
 Analyse bibliographique. CNET
- [22]- COURS DU LEA CNET