الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

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

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT D'ELECTRONIQUE

PROJET DE FIN D'ETUDES

SUJET

Application de la classification à la reconnaissance des phonémes

Proposé par :
Mr. B. Bouseksou

Etudié par :
R. Belkessa
D. Hamoutene

Dirigé par :
Mr B. Bouseksou

PROMOTION : JUIN 1985

E.N.P. 10, Avenue Hacen Badi – EL-HARRACH – ALGER

D E D I C A C E S

à toute ma famille

R A B A H

à mes parents
à tous ceux qui me sont chers

D J A M E L

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

Nous tenons remércier vivement notre promoteur Mr BOUSEKSOU POUR toute l'aide qu'il nous a apportée durant la réalisation de ce projet.

Nous tenons à exprimer aussi notre reconnaissance à Mr AMENAS DE L'I.N.I pour son précieux concours.

Nous tenons à remércier également Mr FERRAH MENOUER pour les travaux de dactylographie.

S O M M A I R E

Table des figures

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

Depuis le développement des ordinateurs, la communication entre l'Homme et la machine n'est plus une utopie. En effet le mode de communication privilégié entre les hommes qu'est la parole a beaucoup préocupé les chercheurs.

La communication entre la machine et L'Homme pose des problémes trés différents suivant le sens dans lequel elle se fait :

- Dans le sens machine-Homme, il s'agit de synthétiser une voix qui soit à la fois intéllégible et le moins possible artificiélle.
- Dans le sens Homme-machine, il s'agit de permettre à la machine de reconnaitre le semantique de la parole. Or, la reconnaissance de ce contenu sémantique passe par un certain stades intérmédiaires correspendant aux divers niveaux d'information accessibles dans le de la parole

L'avénement de nouveaux moyens de calcul a incité les chércheurs à les utiliser pour le traitement numérique de la parole. Cet intérét ne s'est imfléchi, bien au contraire, des nouvelles méthodes de traitement du signal en général et du signal de la parole en particulier ne cessent de voir le jour.

Le champs de recherches sur la parole est vaste, il exige des connaissances multidisciplinaires.

Au début, l'intéret manifesté par les universitaires etait purement scientifiques, ensuite s'est ajouté à l'intérét économique de dizaines d'entreprises qui ont stimuler les recherches dans ce domaine. En effet, plusieurs entreprises commerciallisent actuellement des circuits intégrés spécialisés, des cartes de reconnaissances et de synthéses à usage général et enfin l'intérét stratégique des militaires.

Des resultats encourageants ont été obtenus dans le domaine de la synthése, cela est du au fait que le mode de production de la parole ainsi que les modes articulatoires combinés de tous les organes du conduit vocal sont assez bien compris.

Dans le domaine de la reconnaissance automatique de la parole, quoique les recherches s'intensifient, l'objectif final n'est pas encore atteint à savoir réaliser un systéme qui a les caractéris--tiques suivantes:

- Compréhension de la parole continue
- Vocabulaire de mots(avec syntaxe) d'assez grande taille
- Multilocuteur
- Un taux d'érreur(de sémantique) trés faible(moins de 10%)
- Réponse en temps réel

Un tel systéme n'éxiste pas actuellement car le processus par lequel l'être humain décode la parole en un ensemble de traits phonétiques n'est pas encore connu;.
En effet la pérception de la parole n'est pas seulement due à une extraction passive des traits phonétiques ,elle met en jeu un phénomème de mémorisation comprenant la génération d'hypothéses de synthése interne avec comparaison au signal reçu.
La pérception de la parole par l'être humain passe par trois phases:
1- une phhase d'aquisition (capteur)
2- la paramétrisation réalisée par l'oreille moyenne et l'oreille intérne.
3- la décision ou la reconnaissance éffectué par le systéme nérveux central.

Ces trois phases se retrouvent dans tout systéme de reconnaissance automatique de la parole.

Le problème de la rennaissance de la parole est en fait un problème de reconnaissance de forme. Il s'agit de trouver une méthode automatique qui partage un ensemble de données en sous ensembles ayant chacun un nom. Notre travail consiste à étudier et à élaborer un algorithme de classification automatique (algorithme des K - Moyennes) pour la reconnaissance des phonémes. Ces dernier permet après décodage léxical de reconnaitre des mots ou des séquences de mots.

CHAPITRE I

CHAPITRE -I-

ETUDE DE LA PAROLE

I-1)- INTRODUCTION.

" Nous allons parler succintement, dans ce qui suit ; de l'origine et de la nature de la parole, nous préciserons aussi certaines propriétés du signal de la parole.

1-2--)-- PRODUCTION DE LA PAROLE.

1-2 -1)-FONCTIONNEMENT DE L'APPAREIL VOCAL:

La parole, en tant que phénoméne physique, résulte de l'excitation du conduit vocal par deux types de sources sonores. Le conduit vocal est une suite de cavités qui servent de résonnateurs et suivant leurs formes, il apparait des resonnances à des fréquences variables appelés formants, nous distinguons la cavité pharyngale, la cavité buccale et enfin la cavité nasale. Cette dernière est soit inutilisée soit elle se met en dérivation sur la cavité buccale par l'abaissement du vélum. Cela se traduit par l'apparition d'antiformants sur le spectre des signaux (sons nasalisés).

La production des " sons voisés", comme les voyelles, fait intervenir une source d'implulsions périodiques constituée par l'ensemble poumons cordes vocales. Les sons " nons voisés", comme certaines consonnes, sont engendrés par une source de bruit, c'est à dire de signaux de forme aléatoire. Le bruit est produit par l'action du courant d'air issu des poumons sur les parois du conduit vocal, soit au niveau d'un resserrement, soit au niveau d'une fermeture totale de celui-ci.

I-2-2) E T U D E S P E C T R A L E

Les deux sources que nous avons decrites possedent des caractéristiques spectrales differentes.La source périodique, ou source vocale délivre des oscillations de "rélaxation" qui prennent naissance au niveau des cordes vocales.

La première raie du spectre de la source se trouve à la fréquence fondemmentale Fo (pitch) tandis les raies suivantes espacés de Fo sont des harmoniques . La fréquence fondamentale varie entre 70 et 150 Hz.pour les hommeset de 150 Hz à 300 Hz pour les femmes, et peut dépasser 400 Hz chez les enfants.

.../...

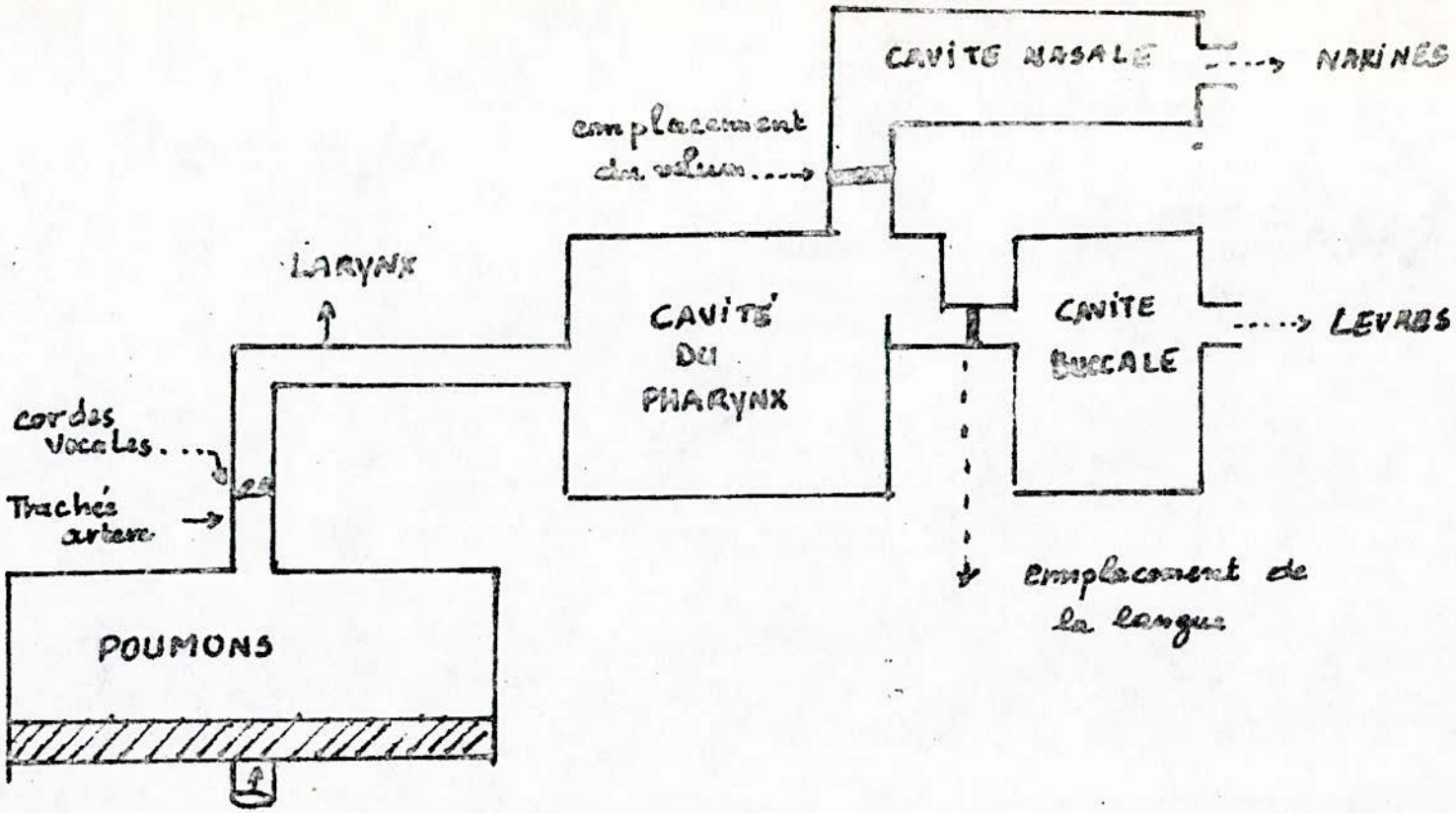


Fig (1) SCHÉMATISATION DE L'APPAREIL VOCAL

La source de bruit est localisée sur le parcours du conduit vocal et sa situation peut varier de 200 à 900 Hz, le second de 500 à 2500 Hz. Les fréquences des formants sont stables lors de la production d'une voyelle et dépendant de cette dernière. En revanche les fréquences des formants changent lors de la production de certaines consonnes car le conduit vocal évolue rapidement dans le temps. Une analyse spectrale de la parole, que l'on peut obtenir à l'aide d'un sonagraphe, met en évidence les évolutions des formants.

I-3) DECOMPOSITION DE LA PAROLE

La phonétique considére depuis longtemps l'existence de sons élémentaires appelés phonémes . Une trentaine de phonènes pourrait décrire la prononciation française, les phonèmes du français, au nombre de 36, sont classés en quatre groupes:

-Les voyelles : sons périodiques dont la fréquence du fondamentale défini la hauteur du son. La nature de la voyelle est déterminée par la forme des cavités.

-Les consonnes fricatives: caractérisées par un rétrécissement du passage de l'air. Ce retrecissement, qui a lieu dans certains endroit du conduit vocal, est équivalent à une source de bruit mise en forme par les cavités de conduit. Si en plus de cette source de bruit les cordes vocales interviennent, la consonne produite est dite voisée. Z/S , V/f, J/ ch.

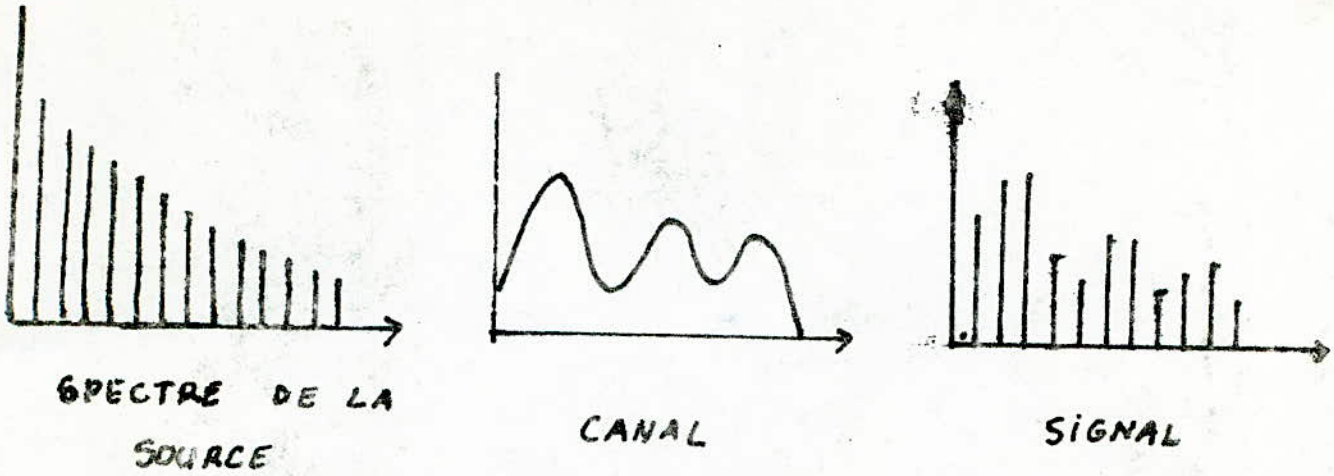
-Les consonnes plosives: sons purement transitoires, ce sont des consonnes momentanées qui supposent une occlusion complète suivie d'une ouverture brusque, semblable à une explosion. Il existe des couples de consonnes plosives semblables en ce qui concerne le caractère voisé ou non b/p, d/t, g/k.

-Les nasales; sons caractérisés par la mise en paralléle des fosses nasales sur le conduit vocal. "In", "an" sont des voyelles nasales, "m" "n" sont des consonnes nasales.

Enfin, il existe quelques cas particulières : les liquides (l) les vibrantes (V•) , les diphtongues.

.../....

Sons voises



SONS NON VOISES

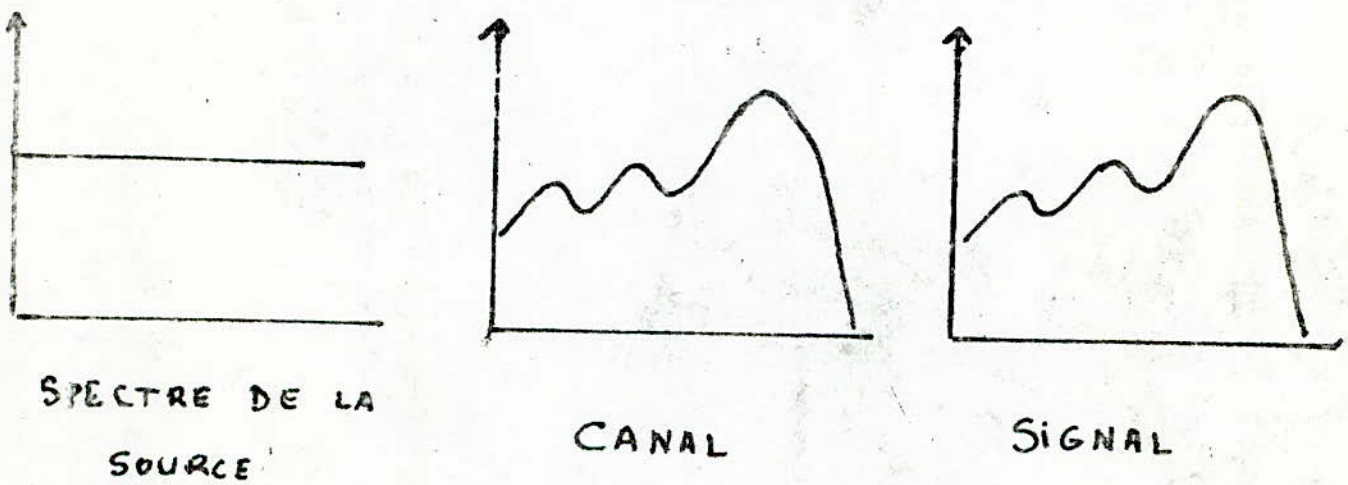


Fig (2)

Pour décrire phonétiquement une phrase parlée, il suffit de reconnaitre chacun des phonèmes qui la compose, le plus souvent, les systèmes de reconnaissance de la parole utilise la "segmentation" . Cette dernière consiste à découper le signale contenu en unités plus petites qui représenteraient des phonèmes. Au préalable, une segmentation séparant la parole du bruit est nécéssaire. Le problème dans ce cas, est de detecter le debut et la fin de la phrase. Cette tâche est aisée lorsque le rapport signal sur le bruit est de l'ordre de 60dB mais devient plus difficile lorsque le rapport est inferieur à 30 dB.

La méthode la plus courante, pour la segmentation, est celle qui utilise l'amplitude ou l'energie du signal.

LES SONS DU FRANCAIS

VOYELLES ORALES

PHONETIQUE	EXEMPLE	CODE MACHINE
[I]	HABIT	^I
[e]	CAFÉ	^E
[ɛ]	PROCÉS	AI
[a]	PAPA	Á
[ɑ]	VASE	Idem
[ɔ]	ROBE	AU
[o]	POT	'O
[u]	MOU	OU
[y]	TU	^Y
[ø]	PEU	EU
[œ]	PEUR	Idem
[ə]	LE	EE

VOYELLES NASALES

[ɛ̃]	PAIN	^N
[œ̃]	BRUN	Idem
[ɑ̃]	BLANC	AN
[ɔ̃]	BLOND	ON

SEMI-VOYELLES

[j]	YEUX	J-
[ɥ]	HUILE	U-
[w]	OUI	W-

CONSONNES

PHONETIQUE	EXEMPLE	CODE MACHINE
[p]	PERE	~P
[t]	TU	~T
[k]	CAS	~K
[b]	BON	~B
[d]	DIRE	~D
[g]	GOUT	~G
[f]	FEU	~F
[v]		~V
[s]	SORT	~S
[z]	ZERO	~Z
[ʃ]	CHOU	CH
[ʒ]	JOUR	ZZ
[ʁ]		~R
[l]		~L
[m]		~M
[n]		~N
[ɲ]	AGNEAU	Non Codé

* designent les consonnes voiés

[ɲ] est considéré comme décomposé en [n] + [j] (donc ^N + J-)

CHAPITRE II

CHAPITRE II

ANALYSE DU SIGNAL DE LA PAROLE EN VUE DE SA RECONNAISSANCE

II-1 INTRODUCTION

L'analyse acoustique est une partie importante dans le traitement que subit le signal de la parole pour pouvoir réaliser un système de reconnaissance de la parole.

Le but de cette analyse est d'extraire les paramétres pertiments ' (les énergies dans les bandes de fréquence, les coéfficients cepstraux.) Qui caractérisent au mieux ce signal qu'est la parole.

Il existe plusieurs types d'analyse dont les plus importantes sont
-analyse spectrale par la transformée de fourrier rapide (F.F.T) ou par vocodeurs
-Analyse par la prédiction linéaire
-analyse cepstrale

Enfin citons les méthodes temporelles tels que le nombre de passage du signal par zero, la mesure de l'énergie et la fonction d'auto-correllation à court terme.

L'étude de ces différentes méthodes permet de choisir celle qui donne les meilleurs performances le tableau ci-dessus indique le nombre de coefficients nécéssaires pour une bonne réstitution du signal de la parole pour chaq type de représentation.

.../...

types de coefficients	energie dans les bondes de frequence	coeff de la L.P.C	coeff du cepstre
nombre de coefficients	$\simeq$ 20	$\simeq$ 12	$\simeq$ 8
type de distance	Euclidienne	statistique	euclidienne

II-2 ANALYSE SPECTRALE

II-21 INTRODUCTION

L'analyse spéctrale d'un signal échantillonné (numérique) consiste à le décomposer en une serie infinie de sinusoides de fréquence f et d'amplitude X(f). Cette décomposition est obtenue par le transformée de fourrier discréte (T.F.D). le calcul de la T.F.D est facilitée par les algorithme de la F.F.T qui reduisent considérablement le temps de calcul.

L'analyse du spectre à court terme peut etre obtenue à l'aide d'un vocodeur à canaux, enfin, les vocodeurs à formant effectue eux aussi l'analyse spectrale et detecte la fréquence des formants.

II 22- ANALYSE DU SPECTRE A COURT TERME PAR LA F F T

transformée de fourrier discréte (T.F.D)

soit une suite de valeurs discrétes d'un signal z(k)

la T.F.D est une suite de valeurs définies par $x(K)$

$$X(n) = \sum_{k=0}^{N-1} x(k)\, e^{j\frac{2\pi n}{N}k} \qquad k = 0, 1, \ldots N-1$$

la T.F.D inverse existe et a pour expression

$$x(k) = \frac{1}{N} \sum_{n=0}^{N-1} X(n)\, e^{-j\frac{2\pi k}{N}} \qquad k = 0, 1, \ldots N-1$$

on peut l'écrire sous forme matricielle

on pose $W_N = e^{-j\frac{2\pi}{N}}$ et $W_N^{nk} = (W_N)^{nk}$

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

transformée de fourrier rapide (F.F.T)

si l'on considére la formule qui définit la T.F.D d'une suite de valeurs discrétes, il apparait qu'il convient de faire n² opérations (addition et multiplication) pour la calculér.

la F.F.T est une famille d'algorithmes qui permettent de réduire considérablement le nombre d'opération. En effet, l'algorithme de COOLEY-TUKEY nécéssite $N\log_2 N$ opérations, soit un gain en temps-calcul de $N/\log_2 N$ ceci est évidement interessant lorsque N est trés grand.

II-23 ANALYSE SPECTRALE PAR VOCODEUR.

1/ INTRODUCTION.

L'analyse spéctrale de la parole peut être obtenue par les vocodeurs à canaux ou à formants. Ces derniérs ont une fonction commune, la détéction et la mesure de la fréquence du fondamentale. Chaque vocodeur se comporte d'un analyseur et d'un systhètiseur placés de part et d'autre du canal de transmission.

2-VOCODEUR A CANAUX.

En anglais le mot vocoder(voice coder)signifie codage de la parole l'analyseur du vocodeur à canaux assure deux fonctions dinstinctes.

-L'analyse spectrale; elle est obtenue par une batterie de filtres passe-bas recouvrant tout le spectre de la parole (300 à 3000 Hz). au nombre de 12à32 canaux. Chaque canal est constitué par un filtre passe bande suivi d'un redresseur et d'un filtre passe-bas permettant le lissage de l'energie (fig 3).

.../...

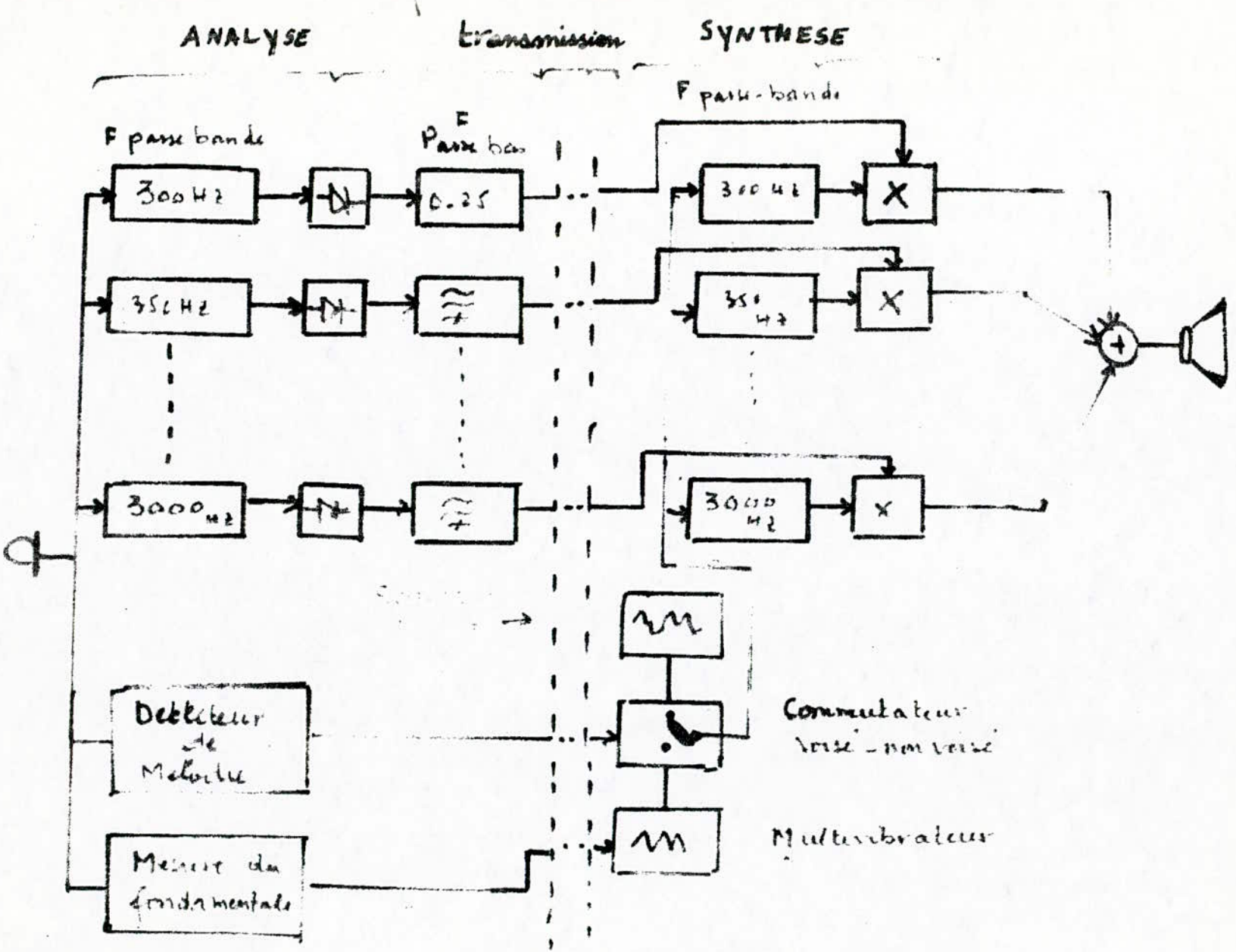


fig (3) VOCODEUR A CANAUX

-un détecteur de mélodie qui élabore des signaux à la fréquence du fondamentale. Ces signaux nous renseignent sur le caractére voisé ou non voisé du son analysé.

Le banc de filtres éféctue donc un échantillonnage du spectre de la parole . En effet à la sortie de chaque filtre on peut mesurer la valeur de l'énergie dans chaque bande de fréquence toutes les 20 ms.

Ces énergies constituent des paramètres que nous pouvons utiliser dans les systèmes de reconnaissance automatique de la parole en général.
Le synthétiseur permet de reconstituer le message parlé selon un processus inverse de celui utilisé dans l'analyse, la parole synthétisée est intelligible mais le naturel de la voix est dégradé.

3-LE VOCODEUR A. FORMANTS

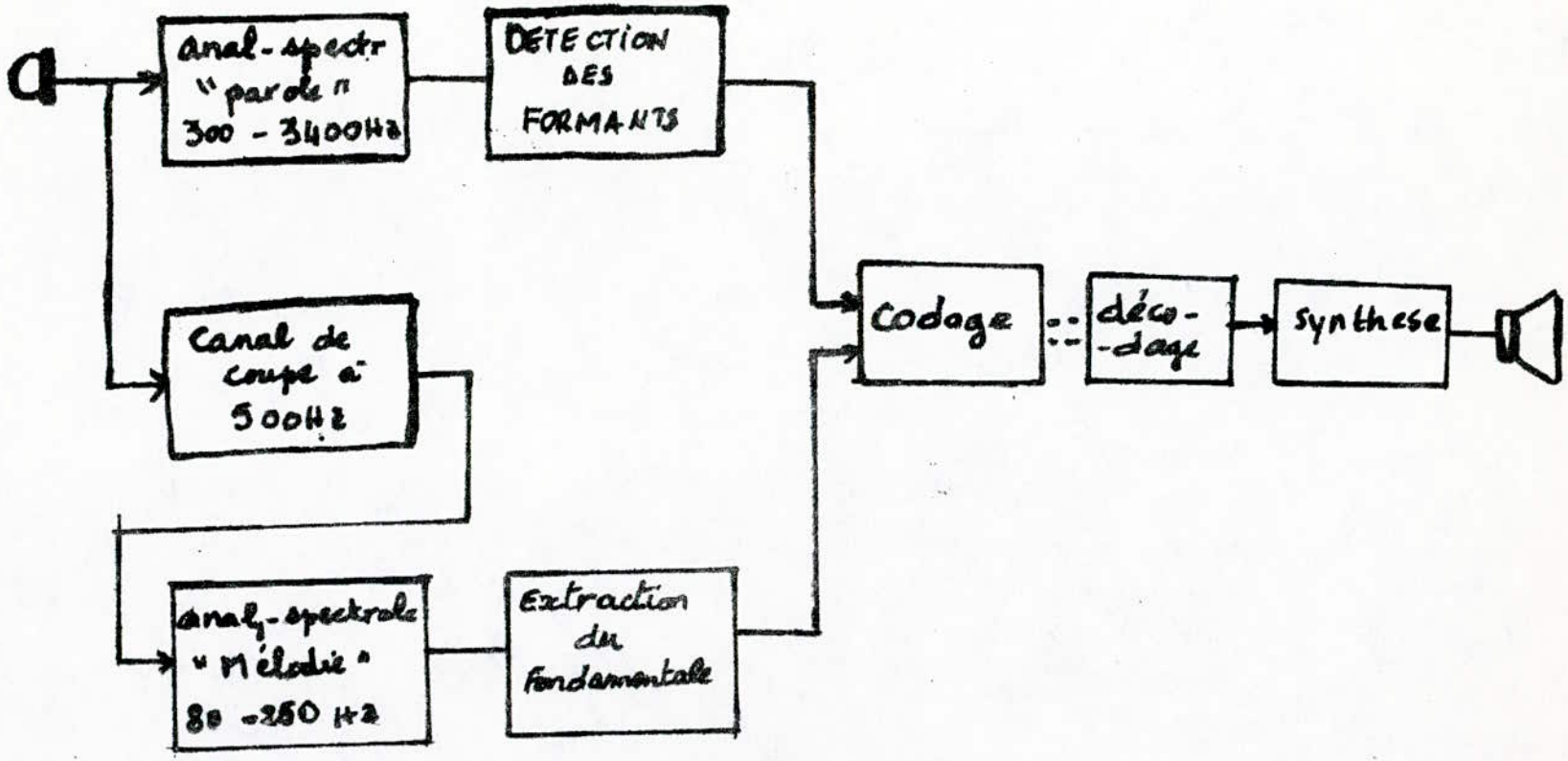
L'analyseur du vocodeur à formant effectue une analyse spectrale du signal identique à celle du vocodeur à canaux et détécte les formants (fig 4) le synthétiseur par contre est constitué d'un ensemble de filtres résonnants dont la courbe de réponse globale en fréquence reproduit celle de conduit vocal, une source de bruit et une source d'impulsions periodique attaque ce circuit.

On remarque une certaine analogie entre le mode de production de la parole humaine et celle produite par ce circuit, ce qui laisse prevoir une meilleur synthése de la parole que celle produite par le vocodeur à canaux .
En effet, ce dernier ne tient pas compte des caractéristiques du conduit vocal, ni des contraintes qu'il impose. Malheuresement l'analyse par vocodeur à formant reste difficile à effectuer pour deux raisons:
-L'apparition de nombreux parasites sur le spectre provoque des erreurs importantes
-Le premier maximum peut être confondu avec le pitch.

4 CONCLUSION

Si les vocodeurs à formants donnent de bons resultats, il reste que la détection automatique des formants et en temps réel est difficile, c'est la raison pour laquelle les vocodeurs à formant ne sont pas utilisés fréquemment dans la synthése de la parole.

.../...



Fig(4)

VOCODEUR A FORMANTS

II_3_ ANALYSE PAR LA PREDICTION LINEAIRE (L.P.C)

II_31_ INTRODUCTION

La prédiction linéaire est une téchnique d'estimation des paramétres de base de la parole (pitch,formants....)dont le critére d'op -misation est la minimisation de l'erreur quadratique moyenne.

L'analyse par la prédiction linéaire consiste à trouver un modéle à un signal original,ce modéle doit se rapprocher le plus de ce signal.

Chaque echantillon est éstimer par une fonction linéaire des echantillon qui le précéde immediatement.Ce qui se traduit par la relation suivante:

$$\hat{S}(n) \simeq \sum_{k=1}^{p} a_k S(n-k) \qquad (1)$$ a_k :coefficients de la L.P.C

Les avantages de la prédiction linéaire sont:

- une représentation du signal de la parole dans un espace de dimension réduite (de l'ordre de 10) .
- une estimation précise des paramétres de base de la parole.
- une vitesse relative de calcul de ces paramétres.

II_32 _ PRINCIPE DE L'ANALYSE PAR LA L.P.C

Pour mieux comprendre le principe de l'analyse par la L.P.C ,on fait a au modéle de la production de la parole par le systéme de la figure (5) Ce systéme est excité par une source dimpulsions ou de bruit blanc . D'une maniére générale on peut exprimer tout signal temporel en terme d modéle prédit et d'un signal d'erreur

on a $$S(n) = \sum_{k=1}^{p} a_k S(n-k) + \sigma e(n) = \hat{S}(n) + \sigma e(n) \qquad (2)$$

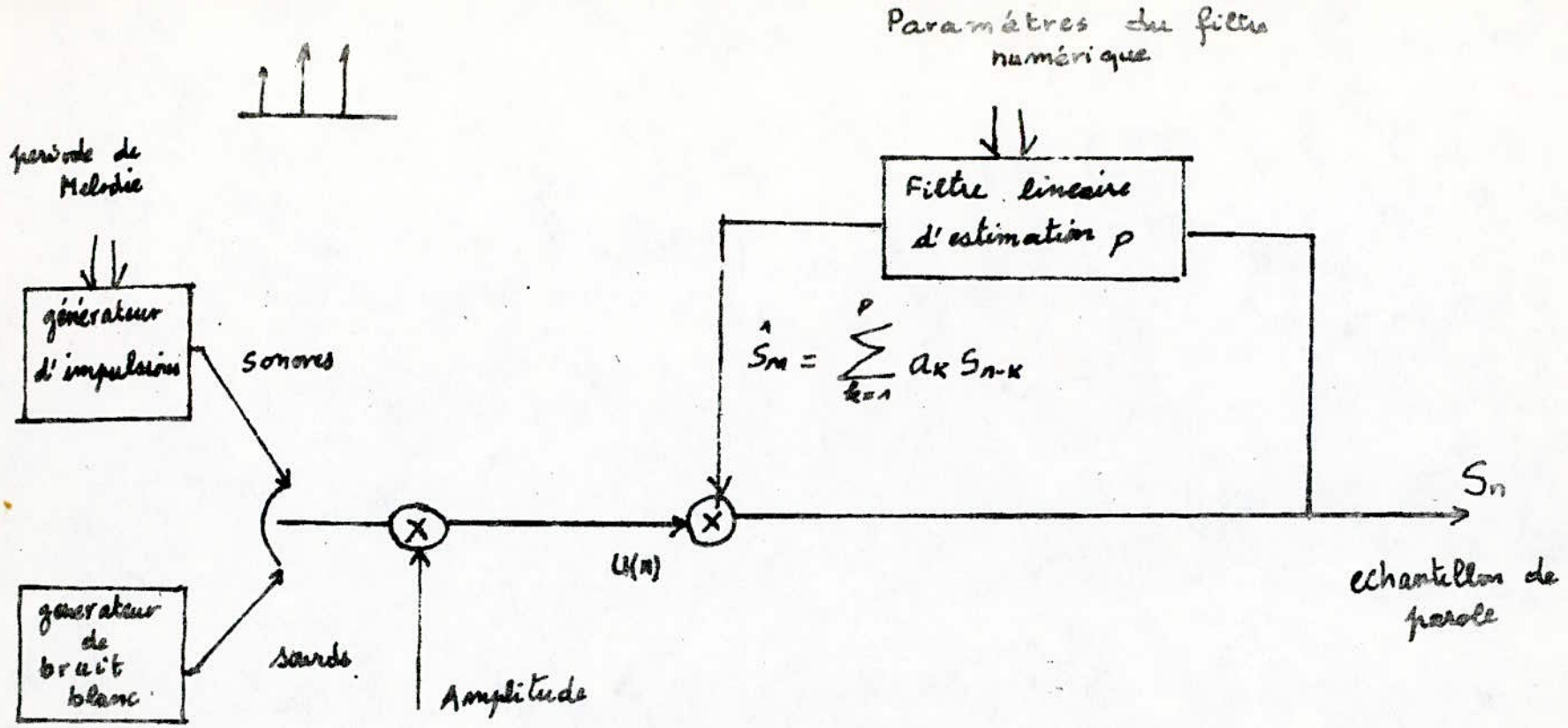
où est une constante d'adaptation d'energie

la transforfée en z de (2) est:

$$S(z) = S(z)\left[\sum_{k=1}^{p} a_k z^{-k}\right] + \sigma e(z) \qquad (3)$$

le modele du filtrage est le suivant:

$$S(z) = \frac{\sigma}{1 + \sum_{k=1}^{p} a_k z^{-k}} E(z) = H(z) \cdot E(z) \qquad (4)$$



MODELE NUMERIQUE DE LA PRODUCTION DE LA PAROLE

Fig 15

donc le modèle de la production de la parole, H (Z) représente la fonction de transfert du conduit vocal.

II.33 METHODES DE CALCUL DES COEFFICIENTS DE LA L.P.C

Dans la théorie il existe une quantité illimitée de façons à calculer les coefficients de la LPC dépendant du critère d'approximation désiré entre le modèle et le signal.

dans notre cas on veut minimiser la quantité définie par

$$E = \sum_{n} \left(S(n) - \hat{S}(n) \right)^2 = \sum_{n} \left(S(n) - \sum_{k=1}^{p} a_k S(n-k) \right)^2 \qquad (5)$$

cette quantité est minimale si

$$\frac{\partial E}{\partial a_k} = 0 \iff \sum_{k=1}^{p} a_k \sum_{n} S(n-k)\, S(n-i) = \sum_{n} S(n-i)\, S(n) \qquad (6)$$

le système linéaire d'équation se resoud soit par la méthode de covariance, soit par l'auto-corrélation. Le choix de la méthode dépend de l'intervalle d'analyse du signal.

II-32-1 METHODE DE CORARIANCE.

Cette méthode fait les suppositions suivantes

a)- le signal est de fini par P+N échantillons pù p est l'ordre du prédicteur et N la taille de l'échantillon estimé.

b)-Un échantillon est prédit par les p échantillons qui le précéde. Ceci est valable exclusivement pour les N échantillons successifs.

c)-l'érreur est minimisée exclusivement sur les N échantillons .

Le système (6) se traduit par

$$\sum_{k=1}^{p} a_k \phi_{ik} = \phi_{io} \qquad i = 1, 2, \ldots . p$$

avec $\phi_{ik} = S(n-i) \cdot S(n-k)$ $\quad i = 1, \ldots . p$ $\quad k = 1, \ldots . p$

la forme matricielle est:

$$\begin{pmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1p} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2p} \\ \vdots & \vdots & & \\ \phi_{p1} & \phi_{p2} & \cdots & \phi_{pp} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \phi_{10} \\ \phi_{20} \\ \vdots \\ \phi_{p0} \end{pmatrix}$$

La matrice de gauche estla matrice de covariance du signal $S(n)$, la solution des coéfficients de la L.P.C est réalisé par la méthode de CHOLESKY, cet algorithme nécéssite $\frac{p^2}{2}$ mots mémoires et efféctue $p^3/6$ produits.

II_322 METHODE D'AUTO_CORRELLATION

Cette méthode fait les suppositions suivantes:

a) le signal est nul à l'éxterieur de l'intervalle considéré. Ceci est réalisé en multipliant le signal par une fenêtre temporelle de largeur N.

b) chaque échantillon est prédit par ses p échantillons qui le pré-cédent et ceci pour tout le temps. ($n \in]-\infty, +\infty[$)

c) l'erreur quadratique totale entre le signal fenêtré et le modéle est minimisé de $-\infty$ à $+\infty$

ces considerations donnent les équations normales d'auto-corréllation à partir de (6).

$$\sum_{k=1}^{p} a_k R_{|i-k|} = R_i \qquad i = 1, 2, \ldots p$$

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

Les équations d'auto-coréllation ont la forme matricielle suivante

$$\begin{pmatrix} R_0 & R_1 & \cdots & R_{p-1} \\ R_1 & R_0 & & R_{p-2} \\ \vdots & \vdots & & \vdots \\ R_{p-1} & R_{p-2} & \cdots & R_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_p \end{pmatrix}$$

la matrice de gauche est carrée , en plus d'être symetrique, a les mêmes valeurs dans les lignes paralleles à la diagonale principale.
Une matrice de ce type est une forme particulière de matrice de TOEPLITZ. La solution des equations est donnée par l'algorithme de Lévinon. Ce dernier necessite 2 P mots mémoires et effectue P (p+1) produits

CONCLUSION

En synthese l'analyse par la L.P.C donne de bons resultats pour le calcul des paramètres de base de la parole.
En reconnaissance ces resultats sont discutés à cause des hypothéses prises au départ sur la stationnarité du signal.

La méthode de covariance assume que le signal de la parole est encore non stationnaire pour l'intervalle considéré dans le calcul des coefficient ak . Cette méthode n'assure pas la stabilité du modèle.

La méthode d'auto-correllation simplifie ces difficultés, en effet elle considére le signal stationnaire, le fenetrage du signal avant l'estimation assure la stabilité du modéle. Son avantage c'est d'utiliser un minimum d'information pour le calcul des coefficients de la L.P.C

II-4/ ANALYSE CEPSTRALE

41/ INTRODUCTION

Les systémes linéaires occuppent une place important dans le traitement des signaux car ils sont décrits par des relations mathématiques simples et d'utilisation souples, donc ils sont faciles à analyser. Malleurheureusments, ce n'est paslecas pour les systémes non linéaires, il n'est pas aisé et meme dificile à représenter mathématiquement.

Le traitement homomorphique est le plus adapté pour l'analyse de ces signaux non linéaires, basé sur le principe de supérposition généralisé.

42/ SUPERPOSITION GENERALISEE

on dit qu'un systéme est linéaire si et seulement si

$$L\{x_1(k) + x_2(k)\} = L(x_1(k)) + L(x_2(k)) \qquad (1)$$

$$L\{a(x(k))\} = a\,L(x(k)) \qquad (2)$$

le filtrage linéaire peut étre utilisé pour séparer deux signaux occuppants deux bandes de fréquence différentes, il ne peut etre utiliséeque si le signal rechérché est combiné par addition.

Ex: $x(k) = x_1(k) + x_2(k)$ et que le signal utile est $x_1(k)$. Il suffit d'utiliser un filtre passe-bande qui laisse passer uniquement la bande de fréquence de $x_1(k)$

La quéstion qui se pose maintenant est la suivante:
comment extraire le signal x(k) s'il etait combiné avec un autre signal par multiplication ou par convolution?

ex: $x(k) = x_1(k) \cdot x_2(k)$ ou $x(k) = \sum_{-\infty} x(l) \cdot x(k-l)$

pour généraliser les relations (1) et (2) on écrira:

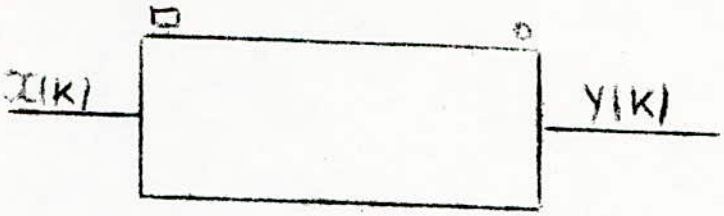
$H(x(k)) \circ (x(k)) = H(x(k) \circ x(k))$

$H(c \circ x(k)) = H(x(k)) \circ c$

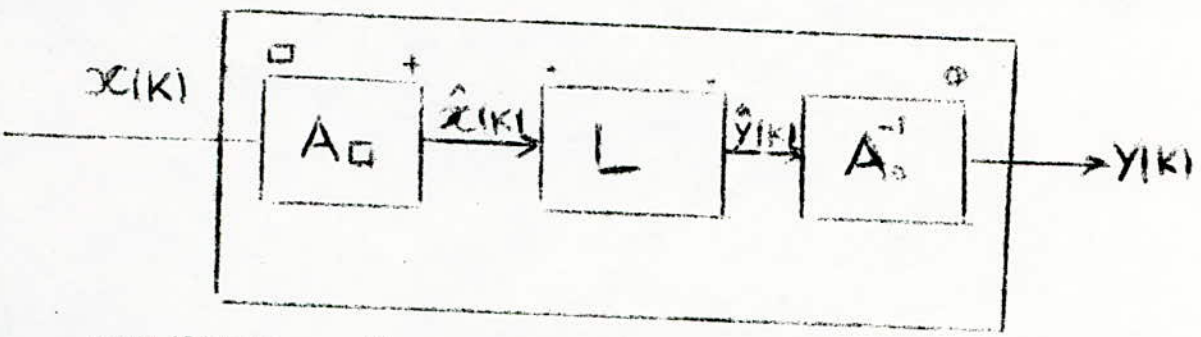
H représente l'opérateur du systéme

ainsi H est une transformée linéaire (structure algébrique d'un espace vectoriél). CES systémes sont représenés entre éspaces vectoriéls sont appelles homomorphiques.

un systéme homomorphique H avec la loi d'entrée et de sortie respectivement $\square$ et o est representé schématiquement par:



OPPENHEIM a demontré qu'il est possible de representer un système homomorphique par 3 systèmes en serie, dont celui du milieu est un système linéaire conventionnel, c'est la forme dite canonique des systèmes homomorphiques (fig 1)



pour illustrer cette representation, considerons les systèmes homomorphiques multiplicatifs et convolutifs.

SYSTEME HOMOMORPHIQUE MULTIPLICATIF.

les signaux d'entrée possible sont de la forme

$$x(K) = [x_1(K)]^{a_1} \times [x_2(K)]^{a_2}$$

a_1, a_2 sont desconstantes arbitraires

le probléme qui se pose est de trouver un espace où les signaux sont séparés, c à trouver un système qui satifsfait la relation suivante

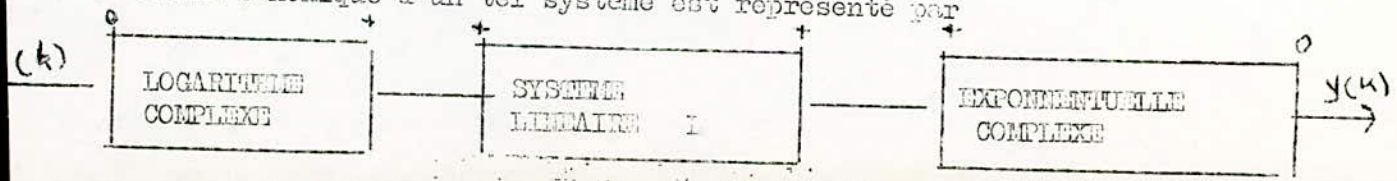
$$A\left[x_1^{a_1} \times x_2^{a_2}(K)\right] = a_1 A\left[x_1(K)\right] + a_2 A\left[x_2(K)\right]$$

$$Log\left[x_1^{a_1} \times x_2^{a_2}(K)\right] = a_1 Log\left[x_1(K)\right] + a_2 Log\left[x_2(K)\right]$$

dans le cas des signaux de la parole $x_1(K)$ et $x_2(K)$ peuvent être négatifs, on considére alors la fonction ligarithme complexe

$$x(K) = |x(K)| \exp(j \arg x(K))$$

la forme canonique d'un tel système est representé par



LES SYSTEMES HOMOMORPHIQUES CONVOLUTIFS

Les signaux sont tels que

$y(n) = \sum_{k=-\infty}^{+\infty} h(n-k) \cdot x(k) = h(n) * x(n)$

$*$ représente le produit de convolution sur des signaux réguliérement échantillonnés. La déconvolution vérifie (D) l'équation:

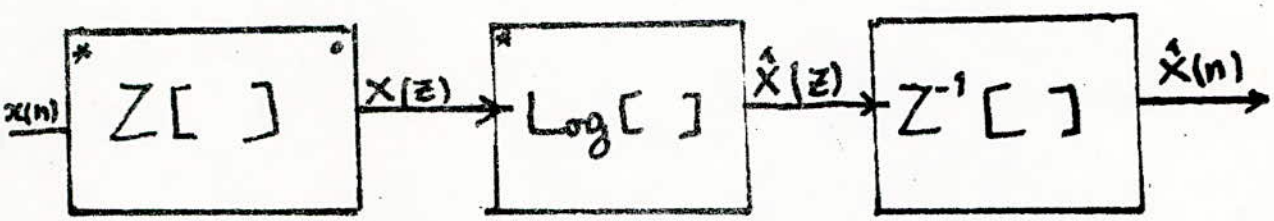
$$D(y(n)) = D(h(n) * x(n)) = D(h(n)) + D(x(n))$$

nous savons que le produit de convolution de deux signaux se transforme en une multiplication avec les transformées en z de chaque signal

$$Z(h(n) * x(n)) = H(z) \cdot X(z) = Y(z)$$

il suffit de prendre le logarithme compléxe de $Y(z)$ pour que le produit devienne une somme. Ainsi on peut représenter le système de déconvolution par la forme suivante:

$$\text{Log } Y(z) = \text{Log}h(z) + \text{Log}x(z)$$



déconvolution homomorphique

LE CEPSTRE

LE CEPSTRE EST LA TRANSFORMÉE de fourier inverse du logarithme du spe d'un signal

Le cepstre $\hat{x}(k)$ d'un signal $x(k)$ est défini ppar:

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

$$\hat{x}(0) = \int_{-1/2}^{1/2} \text{Log}|X(f)|\, df$$

PROPRITE DU CEPSTRE

nous allnns voir une propriete du cepstre que nous démontrerons pas soit une fonction de transfert ayant la forme suivante:

$$X(f) = A. \frac{\prod_{n=1}^{z_i} (1 - \alpha_n e^{-j2\pi f}) \left(\prod_{n=1}^{z_e} (1 - \beta_n e^{j2\pi f_e}) \right.}{\prod_{n=1}^{p_i} (1 - \gamma_n e^{-j2\pi f}) \prod_{n=1}^{p_e} (1 - \delta_n e^{j2\pi f_e})}$$

où A >0 et $|\alpha_n|, |\beta_n|, |\gamma_n|$ et $|\delta_n|$ est inferieur à 1

les coéfficients cepstraux sont donnés par

$$\hat{x}(k) = \begin{cases} -\sum_{n=1}^{z_i} \frac{\alpha_n^k}{k} + \sum_{n=1}^{z_e} \frac{\gamma_n^k}{k} & k > 0 \\ \sum_{n=1}^{p_i} \frac{\beta_n^k}{k} + \sum_{n=1}^{p_e} \frac{\delta_n^k}{k} & k < 0 \end{cases}$$

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

si le signal x(k) est causal c'est àdire que:

$x(k)=o$ pour $k<o \Rightarrow \hat{x}(k)=o$ pour $k<o$

ceci se traduit par le fait que la fonction de transfert n'a ni zéro, ni pôles en dehors des cercle unite

$$X(f)= A.\frac{\prod_{n=1}^{Zi}(1-\alpha_n e^{-j2\pi f})}{\prod_{n=1}^{R}(1-\gamma_n e^{-j2\pi f})}$$

ECHELLE MEL DES COEFFICIENTS CEPSTRAUX.

On remarque les coefficients cepstraux decroissent en $\frac{1}{K}$, il suffit donc d'un nombre reduit pour caracteriser le signal

l'echelle spectrale dans le calcul du cepstre est liniáire les études physicologiques et perceptives de l'oreille semble indiquer que cette dernière est sensible à une échelle logarithmique ~~que~~ de la fréquence. c'est pour cette raison qu'on utilise une echelle quasi- logarithmique (echelle MEL) linéaire sur le premier KHZ et logarithmique au dela de 1 KHZ.

Les coefficients cepstraux sont obtenus à l'aide d'un vocodeur à canaux composés de 25 filtres passe bande de forme triangulaires (fig 7)Les 10 premiers filtres ont une largeur uniforme et les autres ont une largeur qui suivant une progression géométrique . Les coefficients cepstraux dans l'echelle MEL se calcule à partir de l'equation suivantes:

$$MFCC(n) = \frac{1}{NF}\sum_{i=1}^{NF} Log\ E(k)\ \cos(n(k-i))\frac{\pi}{NF}$$

où NF: nombre de filtres triangulaire (en générale 25)

M le nombre de coefficients cepstraux

E (k)sont les reponses d'énergie des filtres triangulaire (voir fig 8)

C O N C L U S I O N

Le cepstre a ces avantages suivants

a)le spectre logarithmique ren d visible les faibles densités spectrales dans les bandes de fréquence sensible à la perception

b)la large gamme dynamique de la densité spectrale est reduite par la conversion logarithmique quantitativement. La consequence de cette opération se represente dans le domaine temporel par une concentration de l'énergie autour de l'origine c'est à dire la plupart de l'énergie du cepstre se trouve dans la première dizaine de coefficients.

.../...

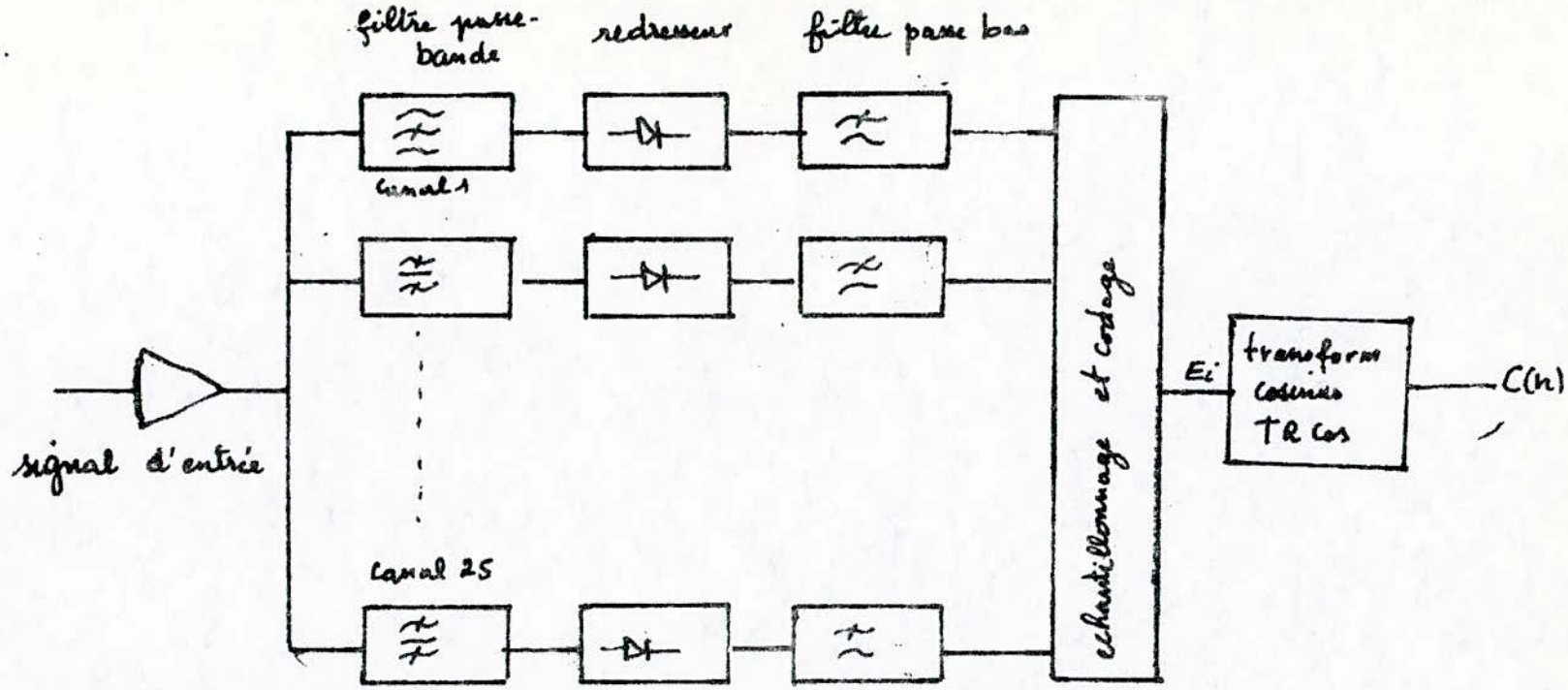


fig (7)

OBTENTION DES COEFFICIENTS CEPSTRAUX MEL PAR UN VOCODEUR A CANAUX

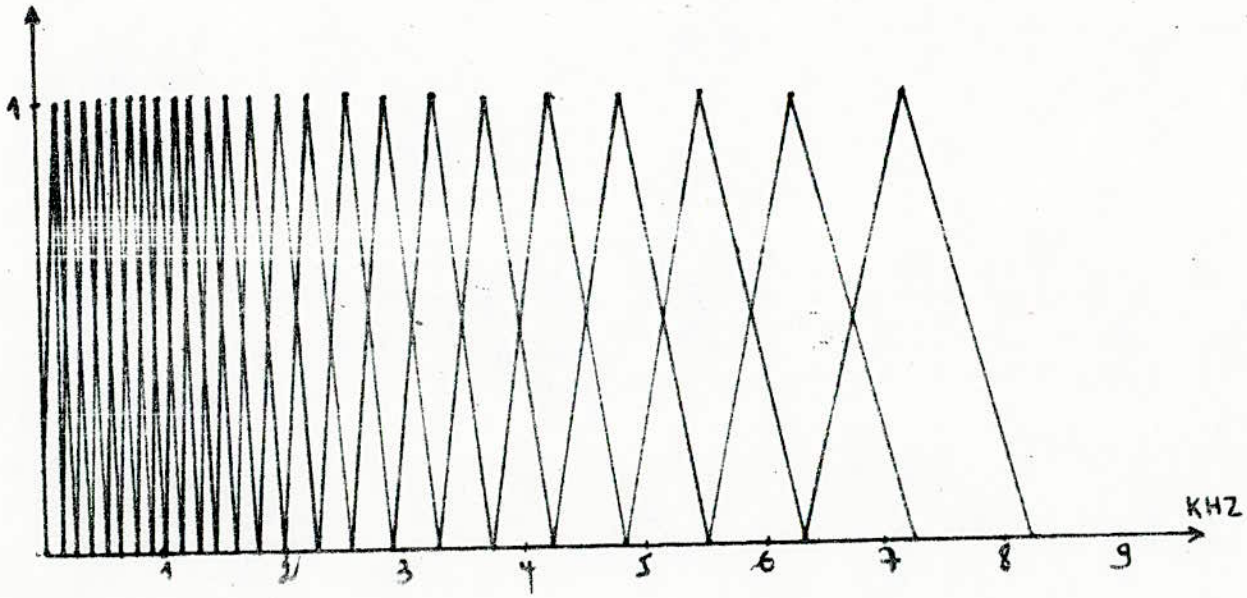


Fig (8) Bancs de Filtres triangulaires

c) la conversion logarithmique souleve les formants de tous faibles énergie localisés fréquemment dans la bande fréquentielle perceptible (voir fig 6) le cepstre a méanmoinsun inconvenient l'utilisation de la distance cepstrale comme indice de similitude dans les systèmes de reconnaissance, nécéssite un temps de calcul considerable.

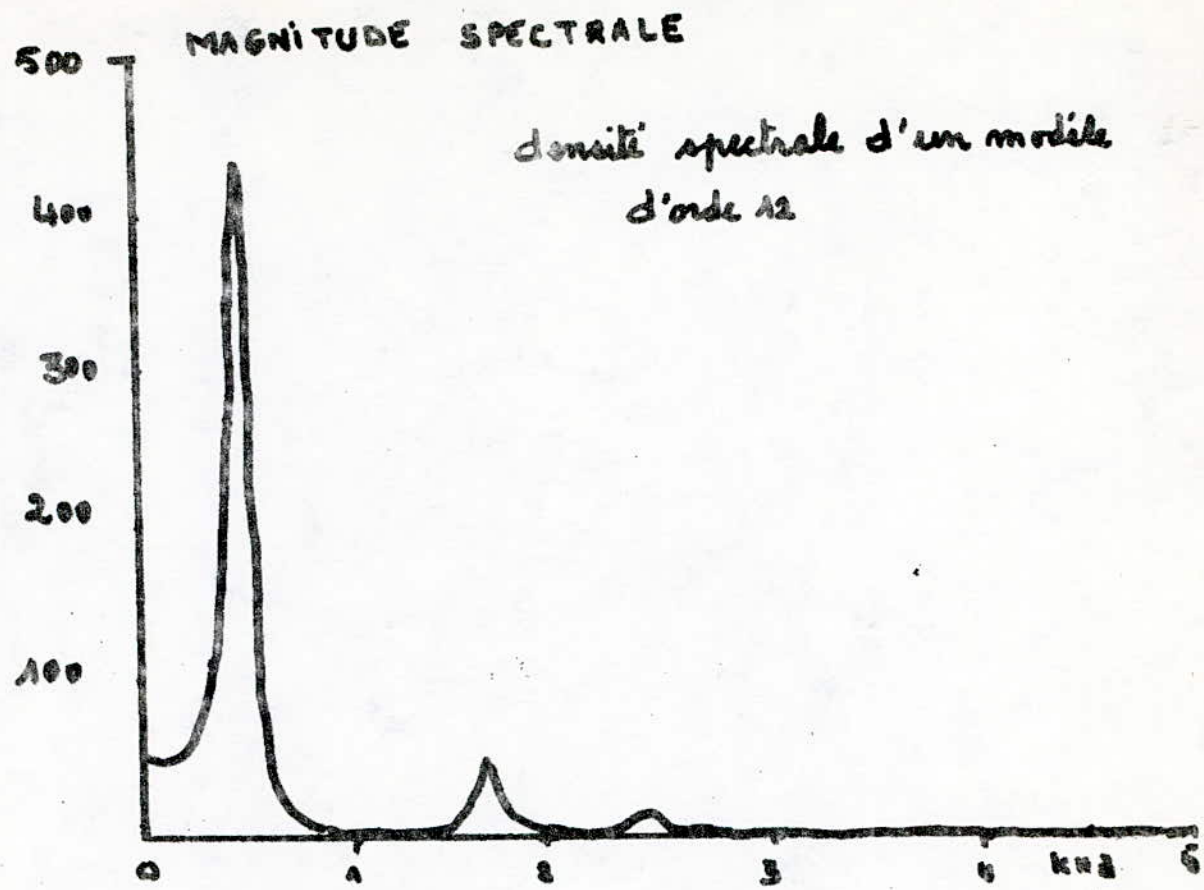

MAGNITUDE SPECTRALE
densité spectrale d'un modèle d'ordre 12
500
400
300
200
100
0
1
2
3
4
KHz
5

fig (5a)

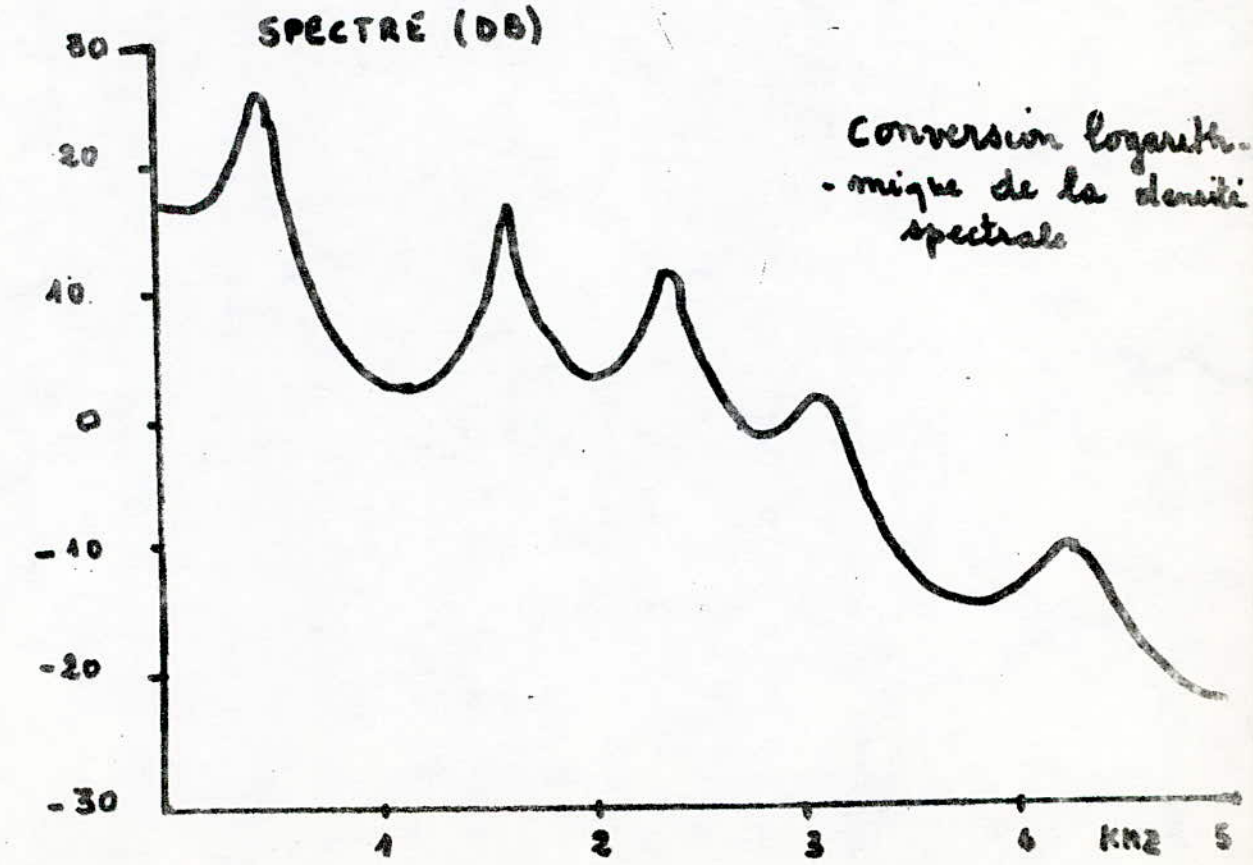

SPECTRE (DB)
Conversion logarithmique de la densité spectrale
30
20
10
0
-10
-20
-30
1
2
3
4
KHz
5

fig (5b)

II-5 ANALYSE TEMPORELLE

Ces méthodes d'analyse prévilégent l'évolution temporelle de certains paramétres dont les principaux sont la fréquence du fondamentale fo, les formants $F_1, F_2 \ldots$ et certains coefficients de la prédiction linéaire.

II-51 LA MESURE DE L'ENERGIE

L'énergie E d'un signal x (n) échantillonné est donnée par la relation suivante

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

on considére généralement des intérralles de temps trés courts (environ 20 ms) pour lequel le signal de la parole est stationnaire l'énergie dans cet intervalle est :

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

où h(m) est une fenêtre temporelle dont le but est de donner le moins d'importance aux échantillons éloignés dans le temps cette énergie E est un paramétre simple et utile pour détecter les silences ou séparer les syllabes.

II-52 LE NOMBRE DE PASSAGES PAR ZERO DU SIGNAL

La méthode pour calculer le nombre de passage par zéro est facile à mettre en oeuvre, en effet, pour un signal discret, il suffit de vérifier la relation suivante : $x(n) \cdot x(n-1) < 0$

ce parametre est utilisé pour la mesure de la fréquence du fondamentale et pour estimer la fréquence des formants. Cette méthode présente certains avantages,

-elle est indépendante de l'amplitude du signal

-elle est trés rapide et pourrait fonctionner en temps réel

-elle permet de séparer les fricatives non voisées des phonémes voisés

II-53 LA FONCTION D'AUTO CORRELATION A COURT-TERME

Cette fonction est définie pour un signal x(t) échantillonné par

$$\phi(m) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} x(n)\,x(n+m)$$

le signal de la parole étant considéré stationnaire sur un intervalle court (15 à 20 ms), on définit alors la fonction d'auto-corréllation à court terme par

$$\begin{cases} \phi_N(m) = \frac{1}{N} \sum_{i=0}^{N-1-m-1} x(i)\, x(i-m) \\ \phi_N(m) = 0 \quad \text{pour } m \geq N \end{cases} \qquad m = 0, \pm 1, \ldots, \pm(N-1)$$

si le signal x (t) est péricdique alors ϕ_n (m) est aussi periodique ceci laisse entrevoir un moyen de détécter les périodicités qui peuvent être invisibles au premier examen. Cette fonction permet aussi de calculer les coefficients de la prédicition linéaire.

CHAPITRE III

CHAPITRE III

CLASSIFICATION

1 INTRODUCTION

Nous avons vu au chapitre précédent les différents types d'analyse. Chaque méthode nous donne un certain nombre de paramétres, ces derniers serviront de données pour la classification afin de constituer un système de reconnaissance automatique de la parole.

La classification est le rangement d'entités en groupes prédéfinis. Donc l'espace dans lequel évolue ou sont contenus les éléments à classer est un espace dont les caractéristiques sont entièrement connues. La classification quant à elle est une famille d'algorithmes dont le but est, partant d'un espace totalement inconnu, de définir des classes à partir de l'ensemble des éléments qui le composent.

L'ensemble des données (les distances) concernant les léléments de cet espace inconnu suffit à le définir. La classification s'attache à recouvrir la structure profonde des données. ces données initiales sont complexes et en quantité importante, donc difficile à cerner.

La définition des classes par la classification permet dans une étape suivante et en supposant que ces classes recouvrent tout l'espace d'étude, de procéder à un simple classement, la première étape de la classification dite d'apprentissage permet de choisir des représentants pour chaque classe. Ces derniers constitueront le dictionnaire des références, ensuite vient la deuxiéme étape dite de reconnaissance qui permet d'identifier un élément inconnu c'est à dire retrouver dans le dictionnaire l'élément qui lui ressemble le plus.

III-2 Les principes de la classification numerique.

Les prinicpes de la taxonomie (étude de la classification, ses bases, ses opérations et ses régles) ont été définis lors de l'introduction desprocédures numériques dans les méthodes de groupement d'entités en classes. Ces procédures basant leur décision sur des caractéres codés ou numériques lors des observations experimentales, obligent à formuler des hypothesés concernant la nature des entités groupées.

.../...

0 LASSIFICATION

La classification comme nous l'avons definie précédemment est une procédure mais le resultat de cette opération est souvent aussi appellé classification. à priori on peut classer les éléments d'un espace inconnu de plusieurs manières, par conséquent on doit avoir une idée claire quant au resultat qu'on veut obtenir il s'agit là de selectionner efficacement les propriètés utilisées comme base de notre classification.

TAXONOMIE NUMERIQUE.

L'ouvrage de SNCATH P et SOKAL R " THE principales of numerical taxonomy" a marqué l'introduction des méthodes de classification automatique toutes ces méthodes démarrent sur le choix d'un indice de similitude entre les éléments à classer et d'un algorithme qui établit une classification ou une hierarchie de classification susceptibles de respecter les relations de similitude la base fondamentale de ces méthodes est la conversion de l'information portée sur les objets en quantités numériques et ceci afin de pouvoir estimer le coefficient de similitude.

Les principes fondamentaux de la taxonomie numérique peuvent se résumer ainsi.

1.Une classification est d'autant meilleur que la quantité des caractéres sur laquelle elle est fondée est grande et que l'information portée pour chacun des caractères considérés est aussi grande que possible .

2.Pour créer des classes naturelles, on doit affecter un même poids pour chaque caractère considéré.

3.La classification est basée sur un indice global de similitude

4.La similitude globale entre deux éléments est une fonction des similitudes entre chacun des caractères simultanément comparés.

II.2. ETUDE DE LA DISTANCE

Nous avons vu que toutes les méthodes de classification automatique démarre par le choix d'un indice de similitude, celui-ci guide les décisions de l'algorithme de classification l'indice de dissemblance est en général une distance calculée à partir d'une des representations paramétriques.

En général la fonction distance $d(x,y)$ entre deux éléments d'un espace de demension n est definie par l'application

$$\mathbb{R}^n \times \mathbb{R}^n \xrightarrow{d} \mathbb{R}^+$$

$$(x,y) \qquad d(x,y)$$

Cette distance doit avoir les propriété suivantes:

$d(x,y)=o \iff x=y$

$d(x,y) \geq 0$

$d(x,y) = d(y,x)$ symetrie

$d(x,z) \leq d(x,y) + d(y,z)$ inégalité triangulaire

Dans le cas de la parole, cettedistance doit avoir une interprétation spectrale et doit etre aussi simple à calculer.

soient $x = (\alpha_1, \alpha_2 \ldots \alpha_n)$

$y = (\alpha'_1, \alpha'_2 \ldots \alpha_n)$

$dx,y) = \sum_{i=1}^{n} | \alpha_i - \alpha'_i |$ distance de Minkowski

$d(x,y) = \left[\sum (\alpha_i - \alpha'_i)\right]$ " " " " Euclidiénne

$d(x,y) = \text{Max} | \alpha_i - \alpha'_i |$ " " " " " Chebytchev

CE sont les principales distances

III 4 L'APPRENTISSAGE

4 I INTRODUCTION

Le but de l'apprentissage comme nous l'avons signalé au début du chapitre, est de constituer un dictionnaire d'éléments prototypes capable de recouvrir l'éspace étudié.

A Partir de la séquence de vecteurs formes obsérvés lors de l'analyse du processus, on détérmine des classes à l'aide d'un algorithme de classification.
On distingue trois(3) types :

- L'apprentissage supérvisé: il y'aune suite d'échantillons classés dont on connait la répartition de tous les éléments (échantillons d'apprentissage classés).

- L'apprentissage non supérvisé :les estimations des paramétres se forment d'aprés les échantillons non classés.

- S'il se produit durant l'apprentissageune modification des paramétres(ou de la structure de l'algorithme de classification) optimisant ces algorithmes au point de vue du critére de qualité choisi.Ces dispositifs sont dits adaptatifs .

4-2 Les méthodes d'apprentissage non supérvisées

Dans ces cas là on ignore le nombre et la nature des classes , il s'agit donc de grouper les vecteurs formes en sous ensembles. Nous citons les principales méthodes

Méthode des centroides:

On sait par exemple qu'il existe 3 classes à déterminer,on choisit 3 vécteurs formes quelconques de la séquence.Ces dérniers sérviront de premiéres éstimations des centroides de ces trois classes.On associe alors à chaque centroide les éléments qui lui sont proches,on ainsi coupé notre espace en trois domaines.On détérmine alors les centres de gravités de ces domaines,puis on classe à nouveau les formes,

ce qui donne trois nouveaux domaines.
On démontre qu'au bout d'un fini nombre fini de cycles, les trois domaines et leurs centres cessent d'évoluer. On a ainsi isolé les trois classes.

Méthode de groupement en chaine

Elle ne suppose à priori aucune connaissance sur le nombre de classes. Lorsque les amas sont assez denses et séparés on peut prendre les formes une à une dans un ordre quelconque, la première etant prise comme spécimen de la première classe, on calcule la distance qui la sépare de la deuxieme forme. Si cette distance excéde un certain seuil choisi à l'avance, on commenceune deuxiéme classe, dans le cas contraire la deuxiéme forme est ajustée à la premiére classe. On procéde ainsi jusqu'à l'épuisement de toutes les formes.

Méthode utilisant une mesure de similitude fondée sur le nombre de voisins communs :

soient n points $X_1, X_2 \ldots\ldots X_n$

on dresse pour chacun des points la liste des k plus proches voisins chaque point est à lui méme son plus proche voisin d'ordre zéro;. Les k autres voisins sont rangés par ordre de distance croissante à X_i. On considére toutes les paires (X_i, X_j), on regarde d'abord si X_i figure parmi les k plus proches voisins de X_j et réciproquement. Si cette double condition est vérifiée et si en outre X_i ou X_j ont en commun un nombre ks de voisins ($ks < k$). On considére que X_i ET X_j appartiennent à la méme classe, le nombre minimale ks de voisins communs corréspend à un seuil de similitude. Les classes ainsi obtenu-es, pouvant etre à leur tour regroupées en super classes jusqu'à ce que le nombre et la densité des groupements obtenus convient à l'application envisagée.

METHODE D'APPRENTISSAGE SUPERVISEE

On connait la répartition de tout les éléments c'est à dire le nombre et la nature des classes.On doit simplement éstimer certains paramétres tels que la moyenne et la matrice de corréllation de chaque classe.

soient m+I classes

$X_1^0 , X_2^0 \ldots\ldots X_{n_0}^0$

$X_1^1 , X_2^1 \ldots\ldots X_{n_1}^1$

$X_1^m , X_2^m \ldots\ldots X_{n_m}^m$

à partir de ces échantillons classés,on estime les paramétres de la densité de probalité(moyenne et matrice de corréllation)
La moyenne de chaque classe est definie par:

$$A_K = \frac{1}{n_K} \sum_{i=1}^{n_K} X_i^K$$

et la matrice de corréllation par:

$$|M_K| = \frac{1}{n_K} \sum_{i=1}^{n_K} (A_K - X_i^K)(A_K - X_i^K)^t$$

APPRENTISSAGE ADAPTATIFS

Les formes à reconnaitre peuvent subir des modifi-cations au cours du temps, ainsi la voix du locuteur sous l'effet de la fatigue change.
Un certain nombre d'algorithmes permettent une adaptation de la machine à de telles dérives des formes ,par conséquent ,il faut

reajuster les centres de gravités des classes.
Par exemple: soit H_0 l'hyperplan séparateur de 2 centres de gravités G_0^I ET G_0^2 respectivement de 2 sous-ensembles E_0^1 et E_0^1 dont les formes sont connues , au moment de l'apprentissage comme appartenant aux classes C_1 et C_2. Lorsque une fournée de formes est présentée elle se trouve divisée par H_0 en 2 sous ensembles E_1^1 ET E_1^2 dont la machine déterminera les nouveaux centres de gravités G_1^1 et G_1^2, d'où un nouvel hyperplan H_0^1 qui sera le plan médiateur de ces derniers.
Le processus est repris dès qu'une nouvelle fournée se présente, on voit qu'il permet de suivre l'évolution du phénoméne. L'intervention de l'être humain est limité au tout premier stade.

III-5 LA RECONNAISSANCE.

Il s'agit dans l'étape de reconnaissance de ranger un élément inconnu dans une classe se trouvant déjà dans le dictionnaire. On procède de plusieurs façons suivant la structure de ce dernier .

-Chaque classe est representée par un seul élément prototype, on calcule alors toutes les distances separant l'élément à classéràtous les representants des classes . Le candidat retenu est celui dont la distance est minimale: c'est la méthode du plus proche voisin.

-Le dictionnaire est formé de plusieurs références par mot (classe) . La phase de reconnaissance fournit un ensemble de candidats ordonnés par leur distance au mot à reconnaitre.

Les techniques Knn (les " k nearst neighbours", les k plus proches voisins) prend l'ordre des mots non plus référence par référence, mais en prenant les 2,3 ou K premières reférences de chaque mot Knn.

les principales variantes des Knn sont :

a)-en partant de la première référence vers les références de distance croissante, dés que le K^{ieme} representant d'un mot est atteint, ce mot est retenu comme candidats dans le rang où il apparait. C'est une méthode suffisante mais elle ignore l'ordre d'apparition et les distances des références des mots procedants la K^{ieme}.

b)-l'ordre des mots candidats est obtenu par ordonnancement de la somme des distances des K premiers représentants de chaque mots est ainsi pris en compte par une contribution à la somme des distances.

CONCLUSION

Le choix de la méthode de classification depend essentiellement de la nature des éléments à classer et du resultat qu'on veut obtenir, l'opération éssentielle dans les problèmes de reconnaissance de forme est l'attribution d'une classe (appartenant à un ensemble fini de classes d'assignation) à un élément dit d'entrée ou inconnu.

Il y'a plusieurs méthodes d'assignation dépendant du type d'apprentissage.

III-6.ANALYSE EN COMPOSANTES PRINCIPALES

61 INTRODUCTION.

On a vu au chapitre II qu'il est possible de représenter le signal de la parole dans plusieurs espaces. La question qui se pose maintenant est de savoir combien de composantes faut il pour avoir une représentation suffisante pour une classification efficace.

Le resultat de la paramètrisation du signal de la parole est un tableau de dimension (N,P) car à chaque intervale d'analyse (10 à 20 ms) on a un jeu de paramètres (énergies dans les bandes de fréquence, coefficients cepstraux...) et N représente la durée du segment analysé.

Souvent la quantité N X P de données est trés grande, l'analyse en composante principale inspiré des méthodes d'analyse de données permet de reduire la dimension du tableau (N,P). Ce dernier est considéré comme étant le representant des coordonnées de N points dans un espace à p dimensions affectés chacun de la masse unité.

L'idée de base de l'analyse en composantes principale est de projeter ce nuage de points sur des axes previllegiés (axes factoriels). Dans cette opération de projection il ne s'agit pas de conserver la particularité de chacun des points mais d'extraire de ces points les caracteristiques moyennes .

62 PRINCIPE DE L'ANALYSE EN COMPOSANTES PRINCIPALES.

soient Xi i=1,2 .....N avec Xi $\in$ Rp

Xi représente les données à classer, il y a donc N X p données

soit Dkl la distance entre deux points Mk et Ml de l'espace Rp

on remplace Dkl par une projection dkl de telle manière à avoir les quantités suivantes:

$$\frac{1}{N}\sum_{K=1}^{N} D_{kl}^2 \quad \text{et} \quad \frac{1}{N}\sum_{K=1}^{N} d_{kl}^2$$

les plus proches possibles c'est à dire que la projection apportera le minimum de déformation du nuage des points considérés initialement.

63 DETERMINATION DES AXES FACTORIELS

On considère l'occurence étudiée, c'est à dire le tableau

$$X = \left(Xi \quad / X \in R^p \ , \ i = 1,2 \ldots\ldots N \right)$$

Comme étant un tirage particulier réalisé sur une variable aléatoire de R^p dont la distribution statistique est à déterminer , l'estimation statistique de lamoyenne de la distribution cherchée et de sa covariance conduit aux valeurs suivantes.

$$\mu = \frac{1}{N} \sum \left\{ X_i / X_i \in R^p \right\}$$

$$\sum = \frac{1}{N} \sum \left\{ (X_i - \mu)(X_i - \mu)^t \right\}$$

les axes principaux d'inertics du nuage sont aussi les axes factoriels. Ceux ci sont determinés d'après la matrice de covariance

soit u le vecteur initaire de R^p, les points Xi peuvent être considérés comme des vecteurs de R^p. Les projections sur la direction de u des points du nuage X sont les points de la droite portée par u et dont les coordonnées sont $yi = \langle u, Xi \rangle$ (produit scolaire de u par Xi .

En notation matricielle $yi = u^t Xi$

le premier axe factoriel est l'axe sur lequel les projections des points du nuages ont une dispersion maximale. La variance de projection V(u) est

$$V(u) = E \left((yi - \bar{y}) \ (yi - \bar{y}) \ ^t \right)$$

où $\bar{y} = u^t u$ est la moyenne des projections

$$V(u) = \frac{1}{N-1} \sum_{i=1}^{N} (U^t (Xi - u))^t u^t (Xi - u)$$

$$V(u) = u^t Tu$$

On remarque que T n'est autre que la matrice de corariance de l'échantillon X.

Le premier axe factoriel est le vecteur propre de T.correspondant à la plus grande valeur propre il s'agit de déterminer le vecteur unitaire u minimisant V(u), en utilisant le multiplicateur de lagrange , il faut donc maximiser le critère J défini par :

$$J = u^t Tu - \lambda (u^t u^{-1})$$

derivons J par rapport à u (derivation matricielle).

$$\frac{\partial J}{\partial u} = 2Tu - 2\lambda u = o \iff Tu = \lambda u$$

On voit bien que la valeur du critère V(u) est égale à la valeur propre λ , le maximum de V (u) est obtenu pour la plus grande valeur de λ de même les axtes factoriels successifs du nuage sont les vecteurs propres de la matrice T de covariance de X rangés dans l'ordre des valeurs propres décroissantes

6-4 PROPRIETES DES AXES FACTORIELS

La valeur propre λ_k rapportée à la somme des valeurs propres represente la part d'inertie ou de variance totale du nuage porté sur l'axe factoriel défini par le vecteur u_k. Les axes factoriels ont de ce fait une propriété interessante pour la réduction des données. En effet, on peut reduire la dimension de l'espace initial R^p sans pour autant déformer le nuage de points.
une autre propriété des axes factoriels est la decorrellation réalisée sur les coordonnées des points.

65 CONCLUSION.

Dans le cas de la reconnaissance automatique de la parole, trois axes factoriels suffisent à décrire tout le nuage des points. Ceci dit l'application de cette méthode diminue sensiblement le taux de reconnaissance.

CHAPITRE IV

CHAPITRE APPLICATION DE LA CLASSIFICATION A LA RECONNAISSANCE DES PHONEMES

(méthode des K-moyennes)

1- Introduction

L'etude et l'analyse du signal de la parole nous ont pérmis d'éxtraire des paramétres déja cités au chapitre II .Ces paramétres caractérisent les éléments d'un éspace.L'algorithme des K- moyennes, diréctement inspiré des méthodes des nuées dynamiques,a pour but de définir des classes à partir des cet éspace.

Cette étape de la classification précéde celle de la reconnaissance. Le but de la premiére étape dite d'apprentissage est de munir l'éspace déscriptif du processus d'une structure de représentation.Cette dér--niére appellée dictionnaire,est capable de recouvrir l'éspace défini par une structure de classe.L'étape suivante ,celle de la reconnaissance consiste à assigner un élément quelconque à une des classes.Cette assignation est réalisée par une comparaison d'un élément inconnu à la totalité des représentants des classes possibles et une décision indicatrice de la classe la plus proche de l'élément.

2- Nuées dynamiques

21-Introduction

La méthode des nuées dynamiques est une méthode de classification automatique sans professeur due à DIDAY .Elle détérmine sur un éspace donné une structure de classes.Cette méthode est itérative, l'itération du processus se poursuit jusqu'à la stationnarité.

22-principe de la méthode

La méthode consiste à détérminer sur une population donnée une partition en un nombre k de classes,k fixé à priori.A chaque classe est associée un noyau.

On agrége la partition à partir des noyaux,puis à partir des partitions obtenues,trouver de nouveaux noyaux capable de générer une partition meilleur que la précédente .On itére le processus jusqu'à ce qu'il devienne stationnaire.

23-fondement de la méthode:

Soient I l'ensemble à classer

P_k l'ensemble des partitions de I en k classes disjointes.

L_k l'ensemble des noyaux, un élément de L_k est la réunion des k noyaux.

On se donne deux applications f et g telles que:

$f(L)=P$ est la fonction qui à un noyau associe la partition P, cette application réalise la reconnaissance.

$g(P)=L$ est la fonction associantà une partition P l'ensemble des noyaux L elle réalise une opération d'apprentissage

D'une partition P_i on passe à la partition $P_{(i+1)}$ par:

$P_{(i+1)}=(fog)(i)= (fog)^{i+1}(P_o)$

Soit D la distance d'un élément de L à un noyau α_i: $D(X, \alpha_i)$

dans notre cas, I est formé de vécteurs, et la distance:

$D(X_n, \alpha_i)= \alpha_i^t X_n$

La fonction f associe tout point au noyau dont il est le plus proche.

$f(L)=P$ avec $P_i=\{X / X \in I,\ D(X,di) \leq D(X, \alpha_j)\ \ j \neq i$

La fonction associe à une classe de I le vecteur optimal. En choisissant comme critére J:

$$J_i = \sum_{j \in P_i} \alpha_i^t X_j$$

, le noyau optimal est celui qui associe à la classe P_i le vecteur (calculé par l'algorithme de DURBIN) sur le centre de gravité de la classe.

DIDAY montre que l'algorithme convérge si la distance R par laquelle les noyaux sont choisis vérifie la proprité d'être carrée. Sans préciser plus cette propriété, signalons que le citére J que nous adoptons définit précisemment une distance carrée et ésquissons la démonstration de la convérgence de l'algorithme .

Le critére J_i est positif ou nul. Nous associerons à la partition P_i le critére

$$J = \sum_{i=1}^{k} J_i$$

ce critére vaut donc $$J = \sum_{i=1}^{k} \alpha_i^t \sum_{j \in P_i} X_j = J(\alpha, P)$$

montrons que:

$$J(\alpha^n, P^n) \geq J(\alpha^{n+1}, P^n) \geq J(\alpha^{n+1}, P^{n+1})$$

où α^n et P^n sont les noyaux et les partitions obtenues à la Niéme itération. La première inégalité est liée à l'optimalité des noyaux α^{n+1} calculés sur P^n. Pour chaque classe P_i^n, le noyau α_i^{n+1} calculé sur P^n rend le critére J_i minimum:

$$J_i(\alpha^n, P^n) \geq J_i(\alpha^{n+1}, P_n)$$

La seconde inègalité vient de ce que les points de I sont ou bien dans la même classe entre la partition $P(n+1)$ et la partition $P(n)$ (et alors leur contribution au critére reste le même), ou bien s'ils sont changés de classe, leur contribution $\alpha_j^{n+1} x$ est devenue $\alpha_j^{n+1} x$ qui a diminué (le point est classé avec la classe j qui minimise $\alpha_j^{n+1} x$), et le critére global J ne peut que diminuer.

La suite des critéres $J = J(\alpha^n, P^n)$ est décroissante, bornée par zéro, elle convérge donc. En outre l'ensemble P_k est fini car l'ensemble I est fini et le critére J^n ne peut prendre qu'un nombre fini de valeurs.

La suite J^n est donc stationnaire, à partir d'un rang N les centres ne se modifient plus.

24- Conclusion:

Cette méthode de classification se justifie par sa simplicité, la convér-gence y est réalisée. Cela veut dire que la stationnarité est obtenue à partir d'uncertain rang puisque l'ensemble à classer est fini.

Cet algorithme de classification requiert assez de temps pour réaliser la convérgence, néanmoins il est trés pérformant sur le signal de la parole.

III_3_ALGORITME DES K_MOYENNE

1_SEQUENCE D'APPRENTISSAGE

1_1 regles phonologique et coarticulation

Dans la conception de tout systéme de reconnaissance analytique, il est indispensable de tenir compte des influences qu'exercent les phonemes sur leur voisins .cela est dû à l'effet de coarticulation par exemple, le sonagramme revele le K est à plus haute fréquence dans (ki) que dans (kou) ; bien que l'oreille ne fasse pas la différence.

On envisage donc d'enrigistrer plusieurs locuteurs ,chacun des pronémes plusieurs fois une série de mots ; Ainsi les réglas decri--vant les variations de prononciation affectant les phonémes dans tel ou tel contexte sont prisés en compte .

Impliciteme nt les diverses prononciations figurent dans le dictio--nnaire.

1_2 déscription de la forme et prétraitement

A chaque période d'analyse on a un jeu de paramétres (vecteur), donc pour un segment donné,n a un emsemble de vecteurs(matrice).

Le pretraitement consiste à faire une segmentation bruit-parole, puis une segmentation en phonéme dans le cas de la reconnaissance des phonémes.

Le pretraitement effectue une normalisation en temps,cela veut dire que pour une matrice temps-fréquense,on se raméne à un vecteur forme car chaque phonéme à une durée propre qui dépend de sa nature et du locuteur qui le prononce.

Supposons par exemple numéeisées les sorties d'un vocodeur à 15 canaux, soit 'k' le nombre de prélevements correspondant à la durée du mot analysé et soit ki le nombre de prelevements correspondant aux divers phonemes i qui composent ce mot. On obtient dons un ensemble de matrices au nombre i et de dimension (ki,15) qui decrivent tout le mot.

La normalisation permet de transformer la matrice i en un vecteur, ainsi chaque vecteur decrit un phonéme.C'est sur ces données que va opérer l'algorithme des k-moyennes.

1-3 La classification

- principe de la methode

La méthode consiste à calculer le centroide de toute la séquence d'apprentissage. Ce centre est pérturbé ,c'est à dire remplacé par un couple d' éléments voisins. On associe alors à chacun de ces deux éléments une classe dont on déterminera le centroide. Les deux représentants obtenus sont à nouveau pérturbés pour obtenir quatre éléments d'ou quatre nouvelles classes et quatre centroides.

A chaque étape le nombre de représentants est multiplié par deux, ainsi si k est une puissance de deux ,on obtient au bout de log(k) itérations un dictionnaire de taille k .

La langue française se compose de 36 phonémes mais il suffirait d'une trentaine pour qu'elle soit intelligible

Dans notre cas on prendrait 32 qui est une puissance de deux.

- Mise en oeuvre de l'algorithme

Supposant que la séquence d'apprentissage soit formée de N vecteurs

$$X^1, X^2, \ldots\ldots\ldots X^N$$

Soit $y(k) = \{ y_i(k) \quad i=1, \quad k \}$ le dictionnaire à la $k^{iéme}$ itération.

La premiére étape de l'itération consiste à partitionner la séquence d'apprentissage en k classes $S_i(k)$ définies par:

$$X^h \in S_i(k) \Leftrightarrow d(X^h, y_i(k)) \leqslant d(X^h, y_j(k)) \quad j=1,\ldots k$$

L'erreur de quantificatinn est calculée pour chaque classe $S_i(k)$ par:

$$D_i(k) = \frac{1}{N} \sum_{X \in S_i(k)} \left(d(X, y_i(k)) \right)$$

L'espérance mathématique est remplacée par une sommation, chaque vécteur de la séquence

$X^1, X^2, \ldots\ldots X^N$ se voyant attribuer la probabilité $1/N$

Dans une deuxiéme étape, on calcule le centroide ou représentant de chaque classe c'est à dire le vécteur C_i qui minimise $D_i(k)$.
on calculera C_i en annullant les dérivées partiélles de :

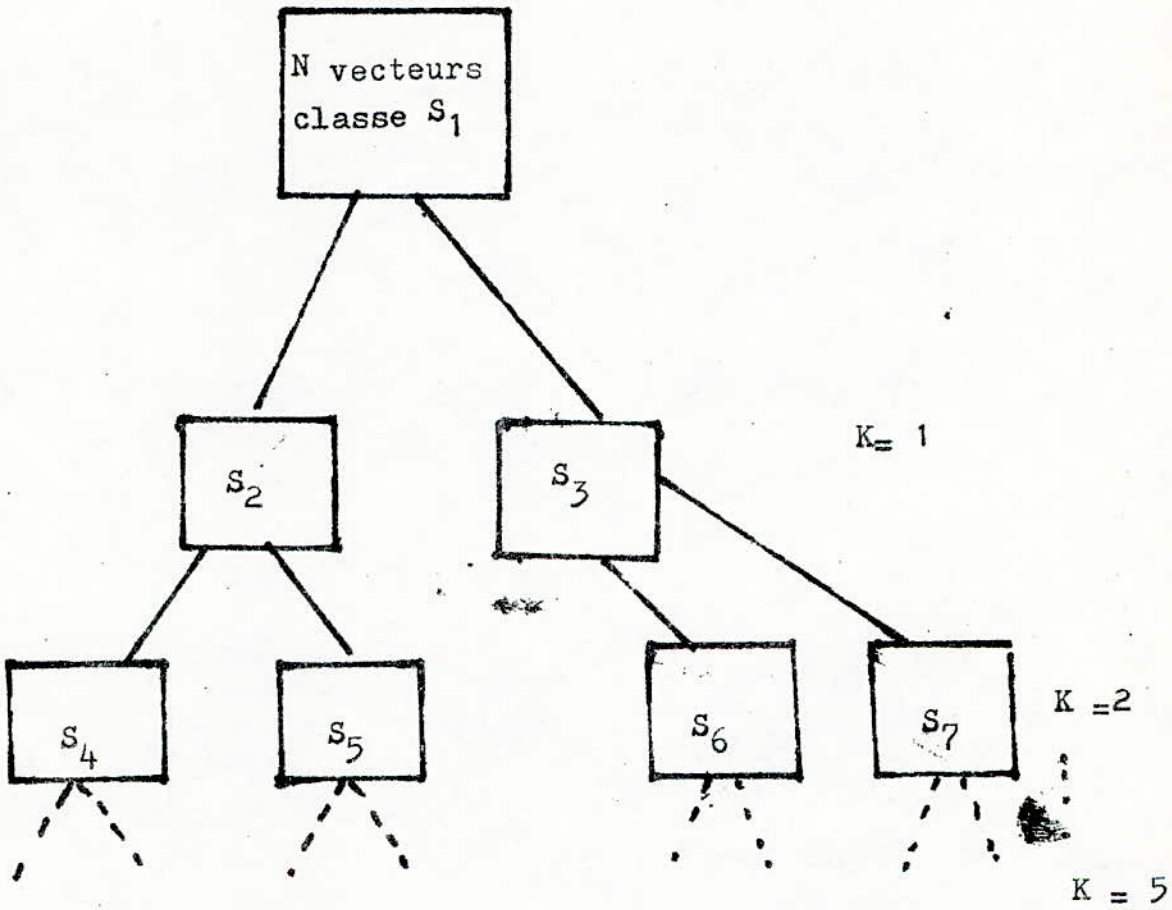
$$D_i(k) = 1/N \sum_{X \in S_i(k)} \left(d(X, y_i(k)) \right)$$

$$\frac{\partial D_i}{\partial y_1} = 0 \quad , \quad \ldots\ldots \quad \frac{\partial D_i}{\partial y_m} = 0 \qquad y = \{ y_1, y_2 \cdots y_m \}$$

ce qui nous donne $y_j = 1/N \sum X_{ij}$

Le nouveau dictionnaire est formé de k représentants ainsi définis.
Le processus se poursuit jusqu'à ce qu'un état stationnaire soit atteint
La somme des erreurs associées à chaque classe a atteint un minimum local. L'initialisation peut s'éffectuer de différentes maniéres, la plus simple consiste à choisir aléatoirement k vécteurs de la séquence d'apprentissage.
Une fois que l'algorithme des k moyennes est élaboré, on sauvegarde par progammation toutes les classes struturées durant toutes les itérations.
Ainsi on confére au dictionnaire une structure en arbre comportant 63 classes(voir figure) .

SEQUENCE D'APPRENTISSAGE



STRUCTURE HIERARCHIQUE BINAIRE

ALGORITHME PRINCIPALE DES K_ MOYENNES

LECTURE DU TABLEAU X

S(1)= 1 ; S(2) =N+1

Pour I = 1 à 33

    New S(I) ⟸ N + 1

FFaire

Appel centroide (1, N, C)

    y(1) = C

Pour K = 1 à 5

    Faire pour I=1 à 2^(K-1)

    Appel voisins (I, y(2**(K-1) + I-1), y1, y2)

    Appel classes (I, y1, y2, S1, S2)

    Appel centroide (S1, S2-1, C1)

    Appel centroide (S2+1, S(I+1) - 1 , C2)

    y(2^K + 2I -2) = C1

    y(2^K + 2I -1) = C2

    New S(2I - 1) = S1

    New S(2I) = S2

    Fin faire

    S = NEW S

FFaire

ALGORITHME DE CALCUL DU CENTROIDE DE LA SEQUENCE D'APPRENTISSAGE

Centroide (ind^1 , ind^2 , C)

```
Pour  j= 1      à m
|     C1 ← o
|     Pour a = ind1  à  ind2
|     |   Cj = Cj + X(a,j)
|     FFaire
|       C(j) ← C1/(ind2 - ind1 + 1 )
FFaire
```

COMMENTAIRES

ind^1 , ind^2 indices du début et de la fin de la séquence

C : centroide

ALGORITHME DE RECHERCHE DES PLUS PROCHES VOISINS DU CENTROIDE

Voisins(I, C, y_1, y_2)

$y_1 \leftarrow X(S(I))$; $D_{min1} = d(X(S(I)), C)$

$y_2 \leftarrow X(S(I)+1)$; $D_{min2} = d(X(S(I)+1), C)$

Pour $A = S(I)$ à $S(I+1) - 1$

Faire
$D = d(X(A), C)$

Si $D < D_{min1}$
alors $D_{min1} \leftarrow D$
$y_1 \leftarrow X(A)$

Si $D > D_{min1}$ et $D < D_{min2}$
alors $D_{min2} \leftarrow D$
$y_2 \leftarrow X(A)$

Commentaires

S(I) : indice du début de la séquence

C : centroide de la séquence

y_1 ; y_2 les olus proches voisins de C

ALGORITHME DE CLASSEMENT

Classes(I, y_1, y_2, S_1, S_2)

$Q \leftarrow 1$, $P = 0$

Pour $j = S(I)$ à $S(I+1)-1$

Faire

si $d(X(j), y_1) < d(X(j), y_2)$

/ X appartient à la classe S_1 /

alors

$TAB(Q) \leftarrow X(j)$

$Q \leftarrow Q+1$

sinon / X appartient à la classe S_2 /

$TAB(S(I+1)-1-P) \leftarrow X(j)$

$P \leftarrow P+1$

Fsi

$S_1 \leftarrow S(I)$

$S_2 \leftarrow S(I)+Q-1$

$Q \leftarrow 1$

Pour $j = S(I)$ à $S(I+1)-1$

Faire $X(j) = TAB(Q)$

$Q \leftarrow Q+1$

FFaire

Commentaires

y_1 , y_2 voisins du centroide

I : indice du début de la séquence

S_1 et S_2 classes associées à y_1 et y_2

III-4 RECONNAISSANCE DES PHONEMES

L'étape de reconnaissance consiste à identifier un phonéme quelconque, ce phonéme inconnu est comparé à la totalité des centroides(représen--tant de chaque phonéme du dictionnaire).

La donnée de base de l'algorithme de reconnaissance est une distance. L'algorithme des k- moyennes a conféré au dictionnaire une structure hiérarchique binaire. Au lieu de calculer 32 distances et prendre l'élément le plus proche, on calcule simplement $2Log_2(N)$ distances, d'où un gain en temps-calcul considérable.

L'algorithme de reconnaissance aura comme entrées un vécteur repré--sentant le phonéme à identifier et la distance D

```
PHONE ( V ,étiq , D )
   Ind 1 = 2
   Ind 2 = 3
   Pour K = 1 à 5
        Faire
        D1 = distance (y( ind 1 ), V )
        D2 = distance (y( ind 2 ), V )
        si  D1 < D2
            alors  D ← D1
                   ind ← ind 1
                   ind1 ← 2 x ind 1
                   ind2 ← 2 x ind 1 + 1
            Sinon
                   D ← D2
                   ind1 ← 2 x ind1
                   ind2 ← 2 x ind1 + 1
        Fin si

étiq = TRANS (ind - 31 )

FIN
```

5- Conclusion

Pour diminuer le volume des caclùls nécéssaires à la recherche du meilleur représentant d'un vécteur, il est nécéssaire de hiérarchiser les données en conférant au dictionnaire une structure en arbre.
L'algorithme que nous avons élaboré pérmet,à partir de la racine l'arbre,d'emprunter à chaque noeud la branche qui minimise la distortion.

LE PROGRAMME D'APPRENTISSAGE:

Le programme est écrit en P.L.1,nous avons choisi ce langage car il est puissant et universelle.Il est le plus adapté à notre probléme,de plus il pérmet la sauvegarde de la structure en arbre du tableau des centroides de toutes les classes.
Les données de ce programme sont des vecteurs de dimension quelconque. Nous avons choisi arbitrairement cette dimension, mais elle peut être ajustée facilement en fonction du type de représentation obtenue lors de l'analyse des phonémes.
Le type de distance utilisée est l'érreur quadratique pondérée.
La pondération consiste à donner plus d'importance à certaines composantes du vecteur lorsque ces derniers portent plus d'informatims que les autres.

PROGRAMME D'APPRENTISSAGE (NUEE)

```
1   0   NUEES  : PROC ;
2   1     DCL SYSIN FILE STEAM INPUT ;
3   1     dcl SYSPRINT FILE SREAM OUTPUT ;
4   1     DCL  X(N,4) BIN FIXED(31) CTL ,
5   1          C(4) DEC FIXED (31,5) ,
6   1          Y(63,4) BIN FIXED(31) ,
7   1          (Y1(31),Y2(31)) BIN FIXED (31) ,
8   1          (S1,S2) BIN FIXED (31) ,
9   1          (C1,S2) BIN FIXED (31) ,
10  1          (C1(4),C2(4)) DEC FIXED(31,5) ,
11  1          (I,LK,J,LG,N) BIN FIXED(31) ,
12  1   /PROCEDURE DE RECHERCHE DES VOISINS
13  1          D'UN CENTROIDE
14  1   VOISINS ; PROC(IND,CENT,YP1,YP2) ;
15  2     DCL IND BIN FIXED(31) ,
16  2     (YP1(4) ,YP2(4)) BIN FIXED(31) ,
17  2        CENT(4) DEC FIXED(31,5)
18  2     (D,IND1,IND2, DMIN1,DMIN2) BIN FIXED,
19  2     /FONCTION: DISTANCE ENTRE DEUX VECTEURS/
20  2     DIST : PROC(A,B) RETURN(BIN FIXED(31))
21  3     DCL(A(4),(4)) BIN FIXED(31);
22  3     DCL(I,SOMM) BIN FIXED (31);
23  3     SOMM=0 ;
24  3     DO I= 1  TO  4 ;
25  4     SOMM = SOMM+(5-I)*((A(I)-B(I))**2
26  4     END ;
27  3     RETURN(SOMM) ;
28  3     END DIST ;
29  3     DMIN1=DIST(X(S(LG,IND) ),CENT) ;
30  2     YP1=X(S(LG,IND) , )
```

SUITE

```
31    2     DMIN2=DIST(X(S(LG,IND)+1,*),CENT) ;
32    2     YP2=X(S(LG,IND)+1,*);
33    2     IND1=S(LG,IND);
34    2     IND2 S(LG,IND+1)-1 ;
35    2     DO J=IND1   TO  IND2 ;
36    3     D=DIST(X(J, *),CENT) ;
37    3     IF  D<DMIN1  THEN DO ;
38    4     YP1=X(J,*);
39    4     DMIN1=D
40    4               END ;
41    3     IF (D<DMIN1 ET D<DMIN2)  THEN  DO ;
42    4     DMIN2=D ;
43    4     YP2=X(J,*) ;
44    4               END  ;
45    3     END  ;
46    2     END VOISINS ;

47    1     / PROCEDURE DE CLASSEMENT /

48    1     CLASSES  : PROC(IND,YP1,YP2,BL1,BL2) ;
49    2     DCL (Q,IND1,IND2,J,IND,P,BL1,BL2,SIZE ) BIN FIXED (31)
50    2         (YP1(4),YP2(4))  BIN FIXED (31)
51    2         INTERX(SIZE,4)  BIN FIXED(31) CTL ;
52    2         /FONCTIN DISTANCE ENTRE DEUX VECTEURS/
53    2        DIST : PROC(A,B) RETURNS(BIN FIXED(31))  ;
54    3        DCL (A(4),B(4))  BIN FIXED(31) ;
55    3        DCL (I,SOMM)  BIN FIXED(31) ;
56    3        SOMM=0  ;
57    3        DO I=1  TO  4 ;
58    4        SOMM=SOMM+(5-I)*((A(I)-B(I))**2)  ;
59    4        END ;
```

SUITE

```
60      3       RETURN(SOMM) ;
61      3       END DIST  ;
62      2       SIZE S(LG,IND+1)-S(LG,IND) ;
63      2       ALLOCATE INTERX ;
64      2       Q=1 ;
65      2       P=0 ;
66      2      IND1=S(LG,IND) ;
67      2      IND2=S(LG,IND+1)-1  ;
68      2      DO J=IND1  TO IND2 ;
69      3        IF  DIST(X(J,*),XP1)  =DIST(X(J,*),XP2)) THEN DO ;
70      4        INTERX(Q,*)=X(J,*) ;
71      4        Q=Q+1 ;
72      4                                        END  ;
73      3                                    ELSE  DO  ;
74      4        INTERX(S(LG,IND+1)-S(LG,IND)-P,*)=X(J,*)  ;
75      4        P=P+1 ;
76      4                                        END  ;
77      3        END  ;
78      2
79      2        BL2=Q-1+S(LG,IND)  ;
80      2        BL1=S(LG,IND)  ;
81      2        Q=1 ;
82      2        DO  J=IND1  TO  IND2 ;
83      3        X(J, )=INTERX(Q,*)  ;
84      3        Q=Q+1 ;
85      3        END  ;
86      2        FREE INTERX  ;
87      2        END CLADES ;
```

```
88    1   CENTROIDE : PROC(LR,IND,CENT) ;
89    2     DCL (I,LR,IND1,IND2,J,IND,SOM) BIN FIXED(31) ;
90    2        CENT(4) DEC FIXED(31,5) ;
91    2        DO J=1 TO 4 ;
92    3        SOM= 0 ;
93    3        IND1=S(LG,IND) ;
94    3        IND2=S(LG,IND+1)-1 ;
95    3        DO I=IND1 TO IND2 ;
96    4        SOM=SOM+X(I,J) ;
97    4        END ;
98    3        CENT(J)=SOM/(S(LG,IND+1)-S(LG,IND)) ;
99    3     END CENTROIDE ;
100   1   / PROGRAMME PRINCIPALE /
101   1
102   1        GET LIST(N);
103   1        ALLOCATE X
104   1        DO I=1 TO N ;
105   2        DO J=1 TO 4 ;
106   3        GET LIST(X(I,J)) ;
107   3        END ;
108   2        PUT EDIT X(I,J)) (SKIP,4(X(10),F(8))) ;
109   2        END ;
110   1        DO I=1 TO 6
111   2        DO J=1 TO 33 ;
112   3        S(I,J)=N+1
113   3        END ;
114   3        END ;
115   1        S (1,1)=1
116   1        LG=1 ;
117   1        I=1 ;
118   1        CALL CENTROIDE (LG,I,C) ;
119   1        PUT EDIT (' ')(SKIP X(10) A(10))
120   1        Y(1,1)=1 ;
121   1        DO LG=1 TO 5 ;
```

SUITE

```
122    2        DO I=1  TO   5  ;
123    3        C=Y((2  (LG-1))+I-1,* )  ;
124    3        CALL VOISINS(I,C,Y1,Y2) ;
125    3        CALL CLASSES(I,Y1,Y2,S1,S2)  ;
126    3             S(LG+1,2 I-1)=S1  ;
127    3             S(LG+1,2 I)=S2  ;
128    3        PUT EDIT(S1,S2)(SKIP,X(10),F(5),F(5)  ;
129    3             S1=2 I-1  ;
130    3             S2=2 I
131    3        PUT EDIT('          ')(SKIP X(10),A(11)  ;
132    3            CALL CENTROIDE(LK,S1,C1)  ;
133    3            CALL CENTROIDE LK,S2,C2)  ;
134    3            Y((2  LG)+(2 I)-2,*)=C1  ;
135    3            Y((2  LG)+(2 I)-1,*)=C2  ;
136    3            END  ;
137    2        END  ;
138    2        DO  I=1  TO  33
139    1        DO  J=S(6,I)  TO   S(6,I+1)-1)  ;
140    2        PUT EDIT X(J,*))(SKIP,4(X(10),F(8)))  ;
141    3        END  ;
142    2        END  ;
143    1        DO  I=1  TO  63
144    2        PUT EDIT  (Y(I,*))(SKIP,4(X(10),F(8,4)))  ;
145    2        END  ;
146    1        END  NUEES
```

F I N DU P R O G R A M M E

SUITE

```
122     2       DO I=1  TO   5  ;
123     3       C=Y((2  (LG-1))+I-1,* )  ;
124     3       CALL VOISINS(I,C,Y1,Y2) ;
125     3       CALL CLASSES(I,Y1,Y2,S1,S2)  ;
126     3            S(LG+1,2 I-1)=S1  ;
127     3            S(LG+1,2 I)=S2  ;
128     3       PUT EDIT(S1,S2)(SKIP,X(10),F(5),F(5)  ;
129     3            S1=2 I-1  ;
130     3            S2=2 I
131     3       PUT EDIT('          ')(SKIP X(10),A(11)  ;
132     3           CALL CENTROIDE(LK,S1,C1)  ;
133     3           CALL CENTROIDE LK,S2,C2)  ;
134     3           Y((2  LG)+(2 I)-2,*)=C1  ;
135     3           Y((2  LG)+(2 I)-1,*)=C2  ;
136     3           END  ;
137     2       END  ;
138     2       DO  I=1  TO  33
139     1       DO  J=S(6,I)  TO   S(6,I+1)-1)  ;
140     2       PUT EDIT X(J,*))(SKIP,4(X(10),F(8)))  ;
141     3       END  ;
142     2       END  ;
143     1       DO  I=1  TO  63
144     2       PUT EDIT  (Y(I,*))(SKIP,4(X(10),F(8,4)))  ;
145     2       END  ;
146     1       END  NUEES
```

F I N DU P R O G R A M M E

PROGRAMME DE RECONNAISSANCE (PHONE)

```
146     3       PHONE:  PROC(V,ETIQ,D) ,
147                     DCL V(4) BIN FIXED(31)
148     2       (B,D1,D2)  BIN  FIXED(31)
149     2       ETIQ CHAR(2) VARYING ,
150     2       TRANSL(32) CHAR(2) VARYING INIT(A1,A2,A3,A4,
151     2       A5,A7,A8,A9,B1,B2,B3,B4,B5,B6,B7,B8,B9,C1,
152     2       C2,C3,C4,C5,C6,C7,C8,C9,D1,D2,D3,D4,D5)
153     2       DCL(IND,IND1,IND2(K) BIN FIXED(31)
154     2       DIST: PROC(A, B )RETURNS(BIN FIXED(31))
155     3       DCL(A(4),(B(4)  BIN FIXED(31)) ,
156     3       (I,SOM) BIN FIXED(31),H
157     3       SOM=0 ,
158     3       DO  I=1  TO  4 ,
159     4       SOM SOM+(5-I) *(A(I) B I))**2
160     4       END  ;
161     3       RETURN(SOM)  ;
162     3       END DIST ;
163     2       IND1=2 ,IND2=3  ;
164     2       DO  K=1  TO  4 ;
165     3       D1=DIST(Y(IND1,  ),V) ;
166     3       D2 DIST(Y(IND2,  ),V) ;
167     3       IF D1  D2  THEN  DO ;
168     4       IND=IND1 ;
169     4       D=D1 ;
170     4       IND1=2  IND1 ;
17      4       IND2=2  IND1+1  ;
172     4                  END ;
173     4       END ;
```

```
I74        2        ETIQ=TRANSL(IND-3I)  ;
I75        2        END  PHONE ;
I75        I      / PROGRAMME PRINCIPAL /
I76        I        CALL NUEE  ;
I77        I        GET LIST(PRON) ;
I78        I        CALL  PHONE(PRON,PHNNEM,DIST) ;
I79        I        PUT EDIT (PRON)(SKIP,4(X(IO),F(5)))  ;
I80        I        PUT EDIT('PHONEME RECONNU')(SKIP X(IO),A(35)) ;
I8I        I        PUT EDIT  (PHONEME)(SKIP,X(IO),A(2))  ;
I82        I        PUT EDIT ('!!!!!!!!!!!!!!!!!!!!')(SKIP,X(IO),A(36))  ;
I83        I        END TEST  ;
```

Resultats obtenus

Nous avons élaboré un programme constitué essensiellement de deux parties:

- un programme d'apprentissage (NUEE)
- " " " " de reconnaissance (PHONE)

Nous avons simulé 64 données en entrée, le programme élabore un dictionnaire de 32 classes. Parmi ces 32 classes obtenues, 16 sont vides. Cela s'explique par le fait que la taille de l'échantillon simulé à l'entrée est reduite.

Ensuite nous avons simulé une donnée (un vecteur) en vue de sa reconnaissance. Le programme de reconnaissance ne fournit pas de résultat car il dépend du programme d'apprentissage.

CONCLUSION:

La reconnaissance automatique de la parole est un sujet difficile et complexe, qui est toujours au stade de la recherche fondamentale.

La simple analyse acoustique du message parlé ne peut suffire à le resoudre. LA parole n'est pas une simple juxtaposition de phonemes pour sa reconnaissanceil fautfaire intervenir, en plus des notions de critère et de phonéme, la notion de syllabe et de la redondanceliée au vocabulaire et aux idees.

On a vu qu'à travers ce travail les différentes étapes qui interviennent dans un probléme de reconnaissance de la parole peuvent être abordés par diverses techniques mathematiques, leur mise en oeuvre demande beaucoup de simplification et d'approximation mais elles permettent de fonder la reconnaissance sur des bases solides. LE choix de la methode d'analyse dépend du type de méthode de reconnaissance utilisée et du résultat qu'on veut obtenir.

On distingue principalement deux méthodes de reconnaissance l'une dite globale basée sur une comparaison directe du signal à des descriptions de mots ou de sequence de mots en termes de modeles acoustiques. Cette methode semble devoir aboutir dans un avenir proche à des realisations operationnelles. Ses limites tiennent surtout au volume de calcul et de memoire, croissante directement avec la taille du vocabulaire utilisable.

La deuxiéme methode est basée sur la reconnaissance des phonemes, elle comporte des etapes: le pretraitement, la reconnaissance des phonémes et le décodage lexical. Les performances des procédures de reconnaissance des phonemes sont médiocres et depassent rarement 60%; ce qui complique davantage latâche du décodage lexical. L'evaluation des résultats de la segmentation et d'identification n'est pas aisée car plusieurs facteurs entrent en jeu. En ce qui concerne la segmentation en phonemes , on peut dire que les résultats sont superieurs à 90% avec en général peu de phonemes omis (2 à 10 %) et parfois plus de phonemes inseres (5 à 15 %). Les resultats de l'identification des phonemes est difficile à évaluer, on peut seulement dire pour une trentaine de phonemes prononcés par un ou quelques locuteurs , les résultats varient de 50 à 70% de reconnaissance.

B I B L I O G R A P H I E

GUIBERT,J. " La parole: compréhension et synthèse par les ordinateurs" P.U.F(I979)

LEINARD,J.S (I977) "Les processus de lacommunication parlée. Introduction à l'analyse et la synthèse "

MAX,J. (I98I) "Méthodes et techniques de traitement du signal et application aux mésures phonétiques" Tome I , 3ième édition Masson .

BENZEKRI,A "Analyse de données" Tome I taxonomie

MERCIER,G. (I977) "Analyse acoustique et transcription phonétique du signal de la parole" Information générale IG/DAS/SST/I

BOUSSEKSOU,B.(I985) " Reconnaissance automatique dela parole par les méthodes globales. APPlication aux particularitée linguistiques de l'arabe standard" Thése de magister

GUERTI, .(I984) "Contribution de lasynthése de la parole en arabe standard" Thése de magister

GRMATER,A.(977) "Identification du locuteur et adaptation au locuteur d'un systéme de RECONNAISSANCE phonémique" These de docteur-ingénieur.