3/95

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieure et de la Recherche Scientifique

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT D'ELECTRONIQUE

Projet de fin d'étude pour
l'obtention de diplôme d'ingénieur d'état

ETUDE D'UN SYSTEME DE RECONNAISSANCE DE MOTS ISOLES EN MODE MULTILOCUTEUR PAR LA METHODE GLOBALE

Proposé et dirigé par :

Mr. N.BENIDDIR

Etudié par :

Mr. BOUBAKOUR NOUR-EDDINE

Promotion JUIN 95

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

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique
ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT D'ELECTRONIQUE

Projet de fin d'étude pour
l'obtention de diplôme d'ingénieur d'état

ETUDE D'UN SYSTEME DE RECONNAISSANCE DE MOTS ISOLES EN MODE MULTILOCUTEUR PAR LA METHODE GLOBALE

Proposé et dirigé par:

Mr. N.BENIDDIR

Etudié par :

Mr. BOUBAKOUR NOUR-EDDINE

Promotion JUIN 95

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

Dédicaces

Je dédie ce modeste travail à :

- A la mémoire de D. BOUGUERNE ;
- A mes parents ;
- A mes frères, soeurs et ma belle-soeur ;
- A mes neveux M^{ed} REDHA et HOUSSEM-EDDINE ;
- A mon oncle M^{ed} Mammeri et sa femme ;
- A FOUZI et sa femme ainsi qu'à la petite AMEL ;
- Je pense aussi à ceux qui sont mes amis.

NOUR-EDDINE

Remerciements

Je remercie Mr N.BENIDDIR pour son aide. Qu'il trouve ici le témoignage de ma profonde gratitude.

Je tiens aussi à exprimer ma reconnaissance envers Mr BOUSSEKSOU et Dr.M.GUERTI pour leur soutien.

Je remercie par ailleurs et tout particulièrement Mr M. BENYOUCEF, enseignant à l'université de BATNA ; ainsi que Mr S.BELKACEMI et M. TOUNSI..

Enfin, je tient également à exprimer mes remerciements à tous ceux qui ont contribué, de prés ou de loin, à l'élaboration de ce travail.

SOMMAIRE

Chapitre 3 D.T.W

INTRODUCTION

INTRODUCTION

La parole est le moyen naturel d'échange d'informations entre les personnes. Cependant ce moyen d'échange n'est efficace que si la compréhension entre ces personnes est totale.

Pour nous comprendre, il nous faut parler un même langage avec un niveau sonore suffisant. On laisse par la suite à notre cerveau le soin d'analyser les informations reçues et de réagir en conséquence.

Depuis quelques décennies, on s'intéresse à introduire cette faculté dans une machine. L'intérêt pour la reconnaissance de la parole a commencé vers les années 50. De nombreux projets ont vu le jour depuis cette date. Par exemple le projet ARPA en 1971 avait comme objectif une reconnaissance de la parole continue avec des locuteurs multiples, sur un vocabulaire de 1000 mots avec syntaxe et un taux d'erreur inférieur à 10 %.

On peut citer aussi, les projets ESPRIT du côté européen et DARPA du côté américain. Ces projets ont permis aux chercheurs de mieux cerner la difficulté du problème et les ont ramenés à plus de réalisme et à restreindre leurs objectif.

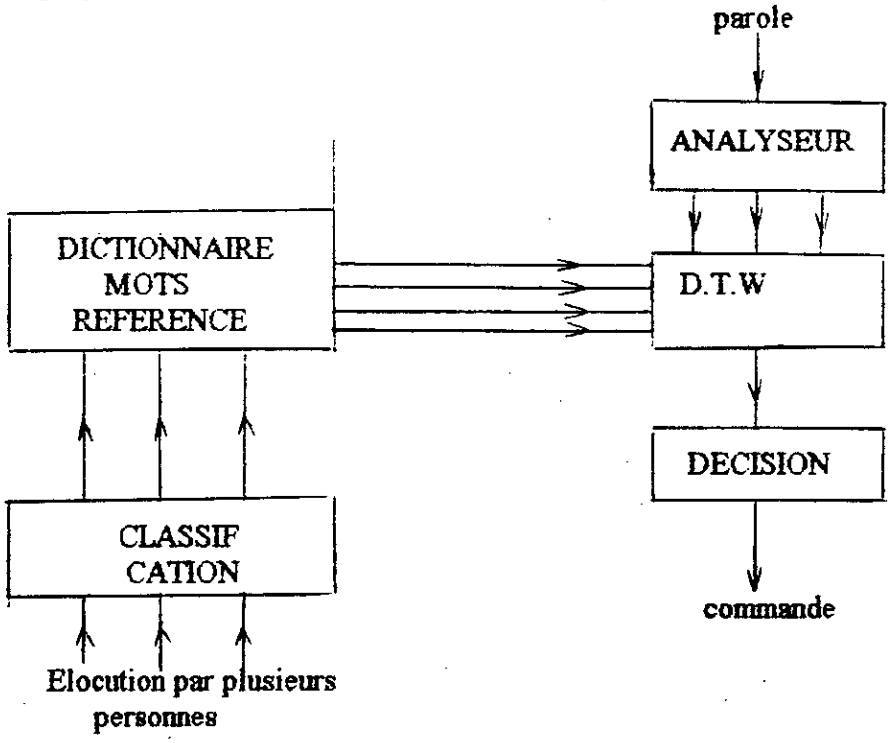
1 - DIFFICULTE DU PROBLEME

La difficulté dans le traitement de la parole est dû essentiellement à la variabilité de ce signal. On distingue quarte classes de variabilité : **intra-locuteur** (vitesse d'élocution, état de santé, hésitation, etc.), **inter-locuteurs**(physiologie de l'appareil de production, accent, etc.), **canal de transmission** (prise de son, bruit ambiant, ligne téléphonique, etc.) et **le vocabulaire de l'application** (langue, coarticulation entre mots, etc.) [8]

2 - Description de l'étude

L'objectif de notre travail est l'étude d'un système de reconnaissance de la parole par les méthodes globales en mode multilocuteurs pour une éventuelle implantation sur le microprocesseur TMS 320.

3 - Synoptique du système de reconnaissance de la parole



nous avons réparti notre travail comme suit:

Le premier chapitre expose des généralités sur le signal vocal, mécanisme de phonation et la définition de la parole ainsi que les différentes approches en RAP, caractéristiques des systèmes de reconnaissance et les différents techniques de reconnaissance.

Le deuxième chapitre décrit les méthodes d'analyse que nous appliquons aux voyelles orales de la langue française.

Le troisième chapitre explique le formalisme de la programmation dynamique et son application à la reconnaissance.

Dans le quatrième chapitre nous étalons l'apprentissage et les méthodes de reconnaissance.

Le cinquième chapitre est consacré à l'étape de décision .

Dans le sixième chapitre nous faisons la simulation des voyelles pour nous servir de fichiers pour les tests.

Le septième chapitre regroupe les résultats des tests que nous avons traités en mode monolocuteur et d'autres en mode multilocuteur.

Le huitième chapitre présente des perspectives d'avenir et possibilités actuelles.

Enfin, dans le dernier chapitre on présente nos conclusions générales.

CHAPITRE 1

ETUDE DU SIGNAL VOCAL

ETUDE DU SIGNAL VOCAL

1- RAPPELS SUR LE SIGNAL DE PAROLE

1.1- DESCRIPTION ANATOMIQUE DES ORGANES PHONATOIRES

Les principaux organes phonatoires sont:

les poumons

- La trachée artère et son extrémité supérieure la larynx qui supportent deux muscles appelés les cordes vocales . L'ouverture séparant ces muscles est la glotte .

- Le conduit vocal est l'ensemble des cavités pharyngo-buccale(pharyngale et buccale) et nasale . Le muscle mobile qui commande le couplage entre ces deux cavités s'appelle le voile du palais ou vélum.
La cavité nasale a une forme fixe et ne peut être obstruée qu'en un seul point; son extrémité postérieure. La cavité buccale, elle est susceptible de prendre des formes variées et de présenter des rétrécissements en divers points .

1.2- CARACTERISTIQUES ARTICULATOIRES ET ACOUSTIQUES DE LA PAROLE

Le signal de parole est un phénomène vibratoire résultant de deux composantes :

- Le passage de l'air expiré à travers les cordes vocales (source d'excitation) produit un signal périodique, dont la fréquence caractérise la hauteur de la voix.

- Un système résonnant, composé de quatre cavités: pharyngale, buccale, nasale et labiale.
Ce système joue un rôle important dans la production de sons de la parole.
Avant d'être rayonné au niveau des lèvres , le signal acoustique se propage à travers le conduit vocal dont lequel il est filtré.

En effet, les cavités suppragiotiques possèdent des fréquences de résonances qui renforcent certaines régions du spectre des sources excitatrices (source sonore ou source bruitée).

Appareil phonatoire

CN : conduit nasal
CB : conduit buccal
PH : pharynx
CV : cordes vocales
LX : larynx
TA : trachée

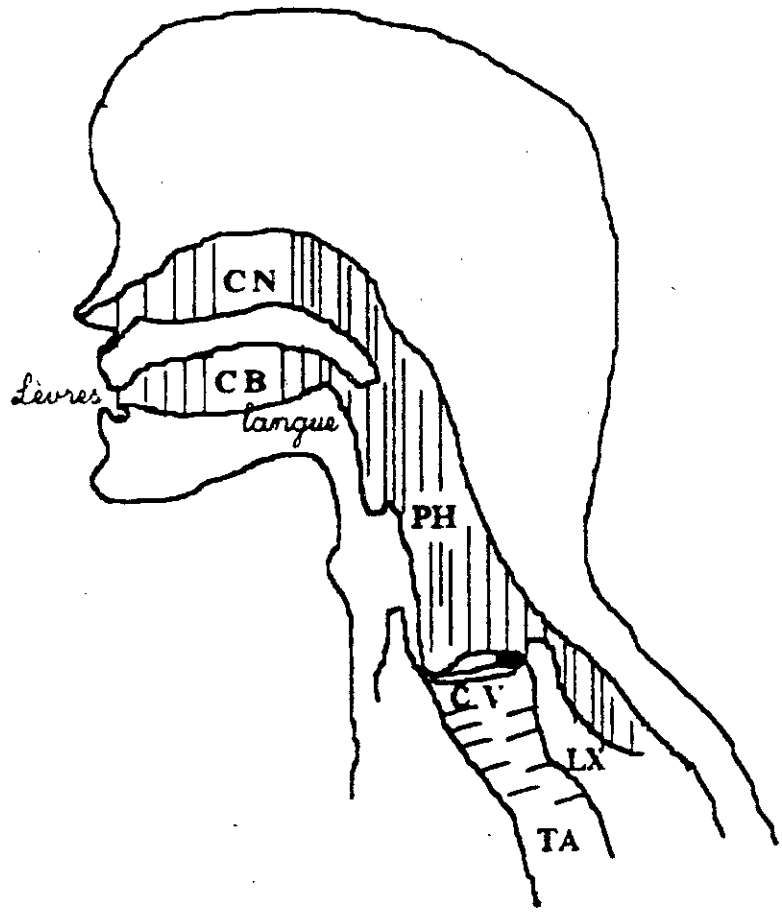


Fig.1.1 Schématisation de l'Appareil Phonatoire

1.3- LES DIFFERENTS TYPES DE SON DE LA PAROLE

On peut établir deux catégories de sons voisés et non voisés: [2]

- Les sons voisés mettent les cordes vocales en vibrations quasi-périodique, sous l'action de la pression de l'air et des muscles du larynx. Le spectre d'un son voisé présente des raies correspondants aux harmoniques. L'enveloppe de ces raies représente les formants. Les trois premiers formants suffiront pour caractériser un spectre vocal.

- Pour les sons non voisés, les cordes vocales n'entrent pas en vibration, et le passage de l'air ne se fait pas librement, ce qui donne naissance à un bruit qui se propage sur les parois du conduit vocal et leurs spectres ne présentent pas de structure de fondamental.

1.4- CLASSIFICATION DES SONS DU LANGAGE

Deux classes sont retenues au mode de production de sons:

1.4.1- Les voyelles

Elles sont caractérisées par un passage libre de l'air. La source d'excitation du conduit vocal est la vibration laryngienne. La fréquence de cette vibration (appelée fondamental ou pitch) varie entre 70 hz et 160 hz pour les voix masculines et 130 hz et 290 pour les voix féminines. Cette fréquence peut dépasser 400 hz pour les voix enfantines.
Certaines voyelles sont dites orales(ex./a/,/u/..)d'autres nasales (ex:/a~/o~/...)

1.4.2- Les consonnes

Elles sont caractérisées par une constriction fermeture, soit momentanée, soit complète du passage de l'air.
Les constrictions peuvent se produire en divers points du conduit vocal. On distinguera plusieurs sortes de consonnes: (voir tableau 1-1)

1.5- LES PARAMETRES ACOUSTIQUES DE LA PAROLE

La parole est constituée d'une succession de phonèmes qui ont été défini par (J.S LIENARD) comme étant "la plus petite unité phonétique susceptible de changer un mot en un autre ". Il a donc été associé au signal de la parole des paramètres phonétiques, liées aux phonèmes et paramètres prosodiques qui se superposent aux précédants.

1.5.1- Les paramètres phonétiques

Ce sont :
- l'intervalle de silence : correspond à une zone de silence dans le signal.
- paramètres formantiques : les premiers paramètres formantiques sont les fréquences des formants, soient les fréquences de résonance du conduit vocal.
- l'anti- résonance d'un phonème: la fréquence à la quelle l'enveloppe du spectre à court terme a une amplitude minimale.

1.5.2- Les paramètres prosodiques

Ce sont :

- la fréquence fondamentale "pitch": liée à la période de vibration des cordes vocales.
- la durée : ce peut être la durée d'un phonème d'un groupe de phonèmes.
- la puissance moyenne du spectre : c'est l'énergie sur un intervalle de temps donné de 10 à 25 ms.

En général, les informations phonétiques sont liées au spectre du signal alors que ces prosodiques renseignent sur l'état du locuteur .

1.6- PROPRIETES STATISTIQUES DE LA PAROLE

La formation des ondes de pressions qui transmettent les sons est un processus aléatoire. l'analyse du signal vocal procède à la statistique court-terme ; son traitement doit se faire sur un laps de temps assez court (10 à 30 ms) afin d'assurer sa stationnarité.

1.7- LES DIFFERENTES APPROCHES EN RECONNAISSANCE [6]

Le système de reconnaissance peut être divisé en deux parties:

- partie analyse acoustique (traitée dans le chap II)
- partie décodage des informations .

Cette dernière partie traitera les informations dérivant de la partie analyse selon deux approches suivant l'existence ou non d'une segmentation du signal par le système de reconnaissance. La première approche est une approche analytique, la seconde est globale.

1.7.1- Approche analytique

Cette approche utilise tout d'abord une segmentation à priori du signal en unités de tailles phonétique, puis chacun des segments est identifié en comparant les mesures acoustiques à des formes de reconnaissance.

Cette identification se fait par émission de plusieurs hypothèses par l'analyseur acoustique pour chacun des segments afin de permettre à l'analyseur lexicale d'émettre des hypothèses de mots qui permettent à l'analyseur syntaxique de déterminer la phrase prononcée en cherchant parmi toutes les phrases syntaxiquement correctes, construites à partir des mots détectés, celle qui est la plus vraisemblable.

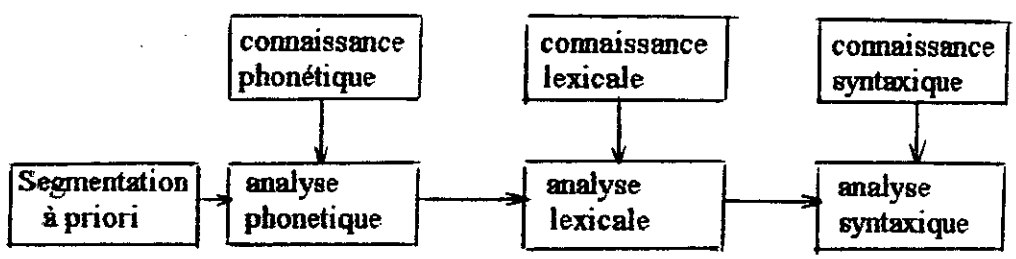


fig: **Exemple d'approche analytique**

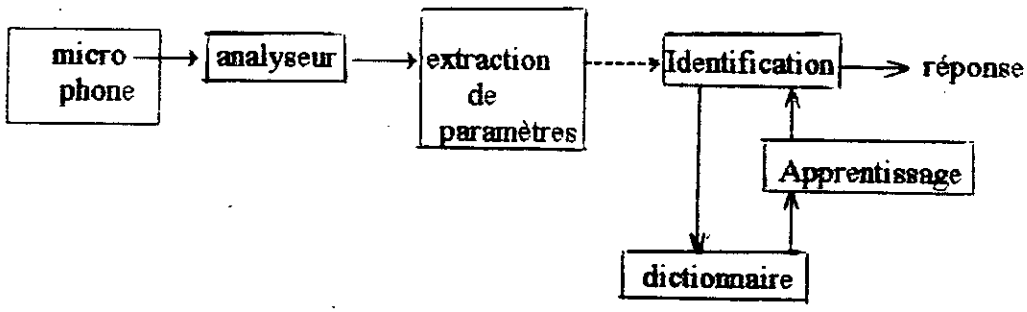
Un des principaux avantages de cette approche réside sans doute dans la facilité théorique d'ajouter des mots dans le vocabulaire puisqu'il suffit de donner la description de ce mot en terme de phonème; comme les distinctions à effectuer sont liées aux phonèmes d'une langue, l'adaptation de cet analyseur à une autre langue sera très difficile, ce qui présente un des principaux inconvénients de cette approche.

Il faut noter aussi qu'il est impossible d'obtenir une segmentation phonétique parfaitement fiable.

1.7.2- Approche globale

C'est une approche qui consiste à faire totalement abstraction des phénomènes linguistiques pour ne retenir que l'aspect acoustique de la parole.

La figure ci dessous représente un système de reconnaissance en "approche globale"



1.8- CARACTERISTIQUES DES SYSTEMES DE RECONNAISSANCE [8]

Les systèmes de reconnaissance de la parole peuvent être distingués selon les trois critères suivants :

Mode de fonctionnement

Indépendant du locuteur (n'importe quel utilisateur peut utiliser le système), monolocuteur (un utilisateur à la fois et cela après apprentissage) ou plurilocuteur (un groupe restreint de personnes).

Mode d'élocution

Mots isolés (un mot à la fois), mots connectés (des séquences de mots avec ou sans pause) ou parole continue (des phrases au sens habituel du terme).

Mode de décodage d'informations

Approche analytique ou approche globale.

1.9- TECHNIQUE DE RECONNAISSANCE

Dans le cas global, trois techniques sont utilisées en phase de reconnaissance :

l'alignement temporel, la méthode statistique et la méthode connexionniste.

1.9.1- Alignement temporel

L'alignement temporel permet de calculer la distance minimale entre une forme acoustique inconnue correspondant au mot à reconnaître et une forme de référence disponible pour chacun des mots du vocabulaire.

Ensuite, la forme de référence fournissant la distance minimale détermine le mot reconnu. Cette méthode repose sur un algorithme de programme dynamique [1] permettant l'ajustement temporel des mots à comparer : l'algorithme DTW (Dynamique Time Warping) [12]. Des contraintes sont fixées préalablement sur la coïncidence des vecteurs de début et de fin de mots et sur les règles d'évolution entre 2 trames consécutives [Itakura,75], [Sakoe,78].

Cette méthode est généralement utilisée en mode dépendant du locuteur. En mode indépendant du locuteur, de nombreuses références sont nécessaires pour chacun des mots du vocabulaire. Ce nombre important de références par mot est dû à la grande variabilité du signal de parole.

Il faut alors une taille mémoire importante pour stocker ces références, et le coût de traitement est également très élevé.

1.9.2- Méthode statistique

Pour palier aux problèmes rencontrés dans la méthode d'alignement temporel (en mode indépendant du locuteur), une méthode qui consiste à remplacer l'ensemble des références acoustiques résultant des différentes prononciations d'un mot, par un modèle de ces prononciations, a été proposée .

Ainsi chaque mot du vocabulaire est modélisé statistiquement par un modèle de MARKOV caché . Ce modèle permet une modélisation des variantes de prononciation des mots et ceci avec un coût d'occupation mémoire moindre .

1.9.3- Méthode connexionniste [11]

Basée sur les réseaux de neurones ; un " neurone formel " (ou simplement" neurone") est un processeur très simple, qui représente, de manière extrêmement simplifiée, voire caricaturale, un neurone réel. La version la plus simple est celle qui a été proposée dès 1943 par McCulloch et Pitts(fig1.1)

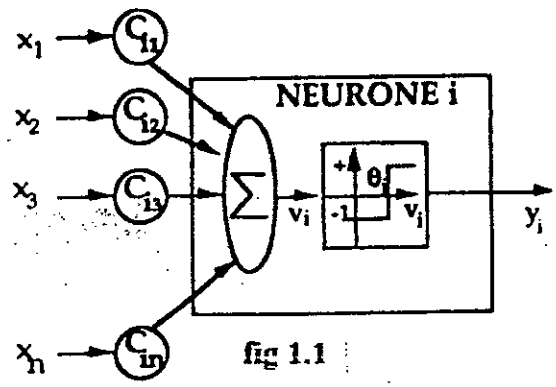


fig 1.1

C'est un automate binaire dont l'état est actif (+1) ou inactif (-1). Il actualise son état de la manière suivante : il calcule son potentiel en faisant la somme pondérée de ses entrées (qui sont les sorties d'autres neurones, ou des informations provenant d'unités d'entée) et il prend une décision en comparant cette somme à un seuil ; si le potentiel est supérieur au seuil, le neurone se met dans l'état actif (+1) ; dans le cas contraire, il se met dans l'état inactif (-1). Il est souvent utile d'introduire une finesse supplémentaire dans le modèle, la décision du neurone n'est pas tranchée, mais elle évolue graduellement entre +1 et -1 en suivant une courbe sigmoïde (fig1.2).

Nous venons de voir qu'un neurone formel ne réalise rien d'autre qu'une somme pondérée suivie d'une non-linéarité. C'est l'association de tels processeurs simples sous forme de réseaux qui permet de réaliser des fonctions utiles pour des applications industrielles (fig.4)

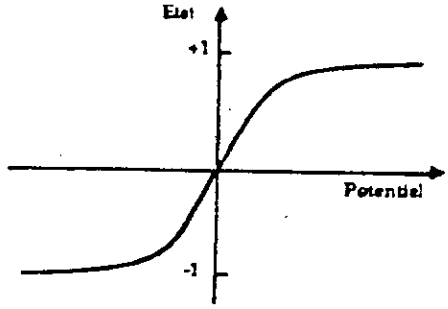


Figure 4.2
Fonction sigmoïde

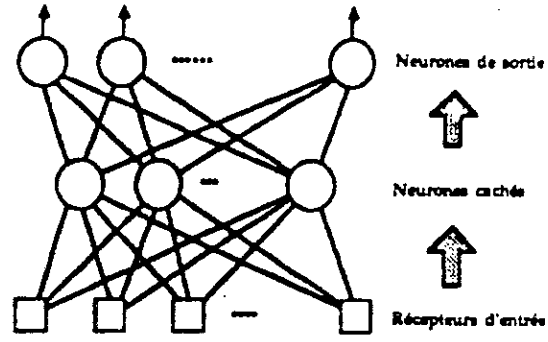


Figure 4.3

Un réseau de neurones est caractérisé par sa topologie (nombre de cellules, connexions entre les cellules, etc) et les caractéristiques des cellules (fonctions de base) .

La phase d'apprentissage consiste à déterminer les poids des connections minimisant l'erreur mesurée entre la sortie réelle du réseau et la sortie désirée.

En phase de reconnaissance, le vecteur d'observation inconnu (vecteur d'entrée) génère un vecteur en sortie. Ce vecteur est comparé aux vecteurs cibles associés aux classes . La classe associée au vecteur cible le plus proche définit la classe d'appartenance du vecteur d'entrée.

	ALPHABET PHONETIQUE	TERMINOLOGIE PRATIQUE	Exemple
VOYELLES ORALES	/i/	i	lit
	/e/	é fermé	été
	/ɛ/	é ouvert	forêt
	/a/	a antérieur	papa
	/ɑ/	a postérieur	âme
	/ɔ/	o ouvert	port
	/o/	o fermé	peau
	/u/	u ou	loup
	/y/	u	tu
	/ø/	eu fermé	creuse
	/œ/	eu ouvert	peur
	/Ə/	e sourd	petit
VOYELLES NASALES	/ɛ̃/	in	brin
	/œ̃/	un	brun
	/ɑ̃/	an	blanc
	/ɔ̃/	on	blond
SEMI-VOYELLES	/j/		yeux
	/ɥ/		lui
	/w/		louis
CONSONNES OCCLUSIVES OU PLOSIVES	/p/ nv		pan
	/b/ v		blanc
	/t/ nv		rente
	/d/ v		rude
	/k/ nv		car
	/g/ v		blague
CONSONNES FRICATIVES	/f/ nv		faux
	/v/ v		veau
	/s/ nv		coussin
	/z/ v		cousin
	/ʃ/		chou
	/ʒ/ v		jour
CONSONNES NASALES	/m/		mer
	/n/		banal
	/ɲ/		agneau
	/ŋ/		camping
CONSONNES LIQUIDES	/l/		lire
	/r/		rire

Tableau I.1 : Classification des phonèmes

CHAPITRE 2

ANALYSE ACOUSTIQUE

ANALYSE ACOUSTIQUE

2.1- INTRODUCTION

Le signal de parole a la particularité d'être naturellement très redondant. Il véhicule des informations ne concernant pas uniquement la signification objective du message; il contient des données sur l'accent, le rythme et l'intonation du locuteur.

Le but de toute analyse est de réduire la redondance du signal de parole, en ne conservant parmi toutes les données disponibles qu'un ensemble de paramètres pertinents pour le caractériser.

Les techniques d'analyse de la parole utilisées en traitement automatique de la parole sont nombreuses, on s'intéresse particulièrement à :

- L'analyse par prédiction linéaire (LPC)
- L'analyse cepstrale.

2.2- Analyses classiques [9]

2.2.1- Prétraitement du signal vocal

Avant d'entamer l'analyse du signal vocal par les techniques LPC et cepstrales, on lui effectue un prétraitement pour rendre plus exploitable son contenu.

2.2.2- Echantillonnage

Echantillonner un signal revient à trouver une représentation discrète, qui permettrait à partir de cette seule représentation de retrouver le signal d'origine avec un minimum d'erreur. SHANNON a proposé un théorème, portant son nom, tant connu et tant utilisé d'ailleurs dans lequel il impose qu'un signal S(t) dont la largeur de la bande est limitée à la fréquence maximale fmax , doit être entièrement déterminé par une suite d'échantillons distants d'une période Te inférieure ou égale à l'inverse du double de la fréquence fmax soit fe $\geq$ 2 fmax .

Dans notre cas, le signal de parole est limité à f_{max} = 6 Khz tout en conservant ses caractéristiques par suite, le théorème de SHANNON nous permis de choisir :

$f_e = 1/T_e = 12.8$ khz $\geq 2\, f_{max}$

où
f_e : fréquence d'échantillonnage ;
f_{max} : fréquence maximale du signal de parole .

2.2.3- Préaccentuation

Cette opération permet de compenser les influences de la source d'excitation du conduit vocal et du rayonnement des lèvres . Il y a donc désadaptation des impédances mécaniques au niveau des lèvres. Par suite le rayonnement du son à l'extérieur s'accompagne d'une baisse d'énergie et d'une distorsion assimilée à une désaccentuation de 6 db / octave sur tout le spectre .

Soient S(n) les échantillons du signal et $S_a(n)$ ceux du signal préaccentué , la préaccentuation de 6 db/octave que nous faisons, n'est autre qu'une dérivation numérique [9];

$$S_a(n+1) = S(n+1) - S(n).$$

2.2.4- Fenêtrage

Le fenêtrage de HAMMING est le plus utilisé en traitement de la parole. En effet, si nous considérons une fenêtre temporelle sur [0,T], l'analyse d'un signal s(t) sur cette dernière est faite en échantillonnant s(t) et en prenant en compte les échantillons s(n) situés à l'intérieur de la fenêtre, ce qui revient a multiplier s(t) par une fonction rectangulaire.

Le spectre du signal ainsi tronqué est obtenu par convolution du spectre initial par une fonction sinus cardinal. Cette opération introduit des irrégularités dues à la nature du sinus cardinal, dont les lobes secondaires contiennent une énergie non négligeable.

Pour palier à cet inconvénient, on multiplie s(t) par une fenêtre plus douce appelée fenêtre de HAMMING définit par son expression:
pour une fenêtre comportant N éléments ,
on a:

$$w(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi (n-1)/N) & 0 \le n \le N-1 \\ 0 & \text{ailleurs} \end{cases}$$

Le domaine spectrale de la fenêtre de HAMMING montre que la majeure partie de la l'énergie de w(f) (plus de 99%) est concentrée dans le lobe principale, les lobes secondaires pouvant être négligés . [7]

Nous voyons ainsi que le fenêtrage de HAMMING a pour effet de réduire la distorsion spectrale qui résulterait d'un simple fenêtrage rectangulaire .

2.3- ANALYSE CEPSTRALE [7]

Nous savons déjà que le signal vocal est très complexe vu sa richesse considérable en informations, il contient des informations phonetiques essentiellement contenues dans le spectre du signal et informations prosodiques essentiellement contenues dans le pitch (la fréquence du fondamental).

Le conduit vocal module le signal pulsé qui est fournit par la vibration des cordes vocales. Il y a une combinaison par convolution des deux filtres qui sont les cordes vocales et le conduit vocal.

Soit h(n) le signal issu de la source d'excitation et x(n) la fonction de transfert du conduit vocal indépendante de h(n). L'expression du signal sonore y(n), obtenu par convolution de h(n) et x(n) échantillonnés, s'écrit:

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

ou * représente le produit de convolution

Le principe de la méthode est donc, de séparer les deux composantes, superposées par un produit de convolution naturel, en une somme des deux composantes. Ceci est obtenu par des traitements homomorphiques. [9]

2.3.1- Déconvolution homomorphiques

Pour trouver l'opérateur de déconvolution D capable de satisfaire à nos besoins c'est à dire qui vérifie les relations suivantes:

$$\begin{aligned} D(y(n)) &= D(\, h(n) * x(n)\,) \\ &= D(\, h(n)\,) + D(x(n)) \end{aligned}$$

Nous portons de la remarque que la fonction logarithme permet de transformer un produit en une somme

$$\ln(\, x.y) = \ln(x) + \ln(y)$$

et la transformée en Z de deux signaux combinés par convolution est justement le produit des deux transformées en Z individuelles

$$TZ \{ h(n) * x(n) \} = TZ \{ h(n) \} . TZ \{ x(n) \}$$

$$\mathbf{Y(Z) = H(Z) . X(Z)}$$

donc, en calculant successivement la transformée en Z du signal de la parole et en calculant le logarithme, on aura réalisé l'opérateur D cherché

$$\mathbf{y(n) \xrightarrow{TZ} Y(z) \xrightarrow{log} Y'(z) = \ln (Y(z)) \xrightarrow{Z^{-1}} y'(n)}$$

* Déconvolution homomorphique *

2.3.2- Définition du cepstre

Nous avons trouvé:

$$Y'(z) = \log (Y(z)) = \log [H(z) . X(z)]$$
$$= \log [H(z)] + \log [X(z)]$$

$$\mathbf{Y'(Z) = H'(z) + X'(z)}$$

La transformation ci-dessus n'est pas applicable en règle générale car le logarithme d'un nombre complexe a une partie imaginaire indéfinie:

$$\mathbf{Y'(z) = \log [y(z)] = \log | y(z)| + j \arg[y(z)]}$$

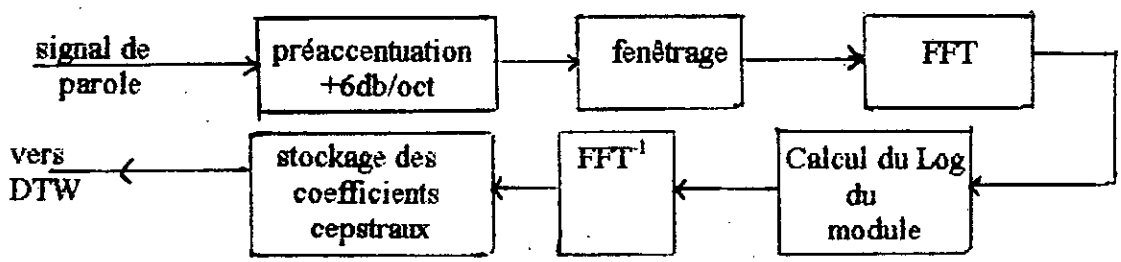
La transformée de FOURIER inverse du logarithme du spectre de notre signal nous donne son cepstre complexe :

$$\mathbf{y(n) = 1/2\pi \int_{-\pi}^{\pi} y'(e^{jw}) e^{jwn} dw}$$

dans notre cas, seule la partie réelle est utilisée. Le cepstre réel d'un signal est défini comme étant la transformée de FOURIER inverse du logarithme du module du spectre du signal .

$$\mathbf{c(n) = 1/2\pi \int_{-\pi}^{\pi} Log \ | y(e^{jw}) | e^{jwn} dw}$$

2.3.3- Etapes d'analyse



_Organigramme (voir Annexe)

2.4- ANALYSE PAR PREDICTION LINEAIRE

2.4.1- Formalisme LPC

L'analyse LPC suppose que le signal vocal traité est approximé par un polynôme d'ordre P (choisi entre 10 et 30). Un échantillon du signal peut être prédit comme une combinaison linéaire des N échantillons précédents d'ou le nom de la prédiction linéaire

$$Sp\ (n) = \sum_{k=1}^{p} a(k)\ s(n-k)$$

ou n = 1..N (indice d'échantillons par fenêtre)
et p l'ordre de prédiction

2.4.2- Problème de la non stationnarité

Le signal vocal étant très chaotique, ne peut être considéré quasi stationnaire que sur des intervalles de temps très limités.

On considère des tranches successives et a estimer un modèle pour chacune d'elles . La procédure usuelle consiste donc a effectuer notre analyse successivement sur des fenêtres de 20 ms avec extraction des paramètres désirés au cours de chacunes d'elles.

Les coefficients LPC a(k) sont calculés en minimisant la somme des carrés des différences entre les échantillons réels de la parole s(m) et les valeurs estimées par combinaison linéaire $S_p(m)$

$$E = \sum_{n=1}^{N+p} [S(n) - \sum_{k=1}^{p} a(k).S(n-k)]^2$$

Il existe plusieurs méthodes de minimisation de l'erreur quadratique citons entre autre : [4]

- La méthode de covariance;
- la méthode d'autocorrélation;
- la méthode en treillis.

La méthode de covariance quand à elle donne des résultats précis mais la stabilité n'est pas assuré.
Bien que la deuxième méthode appelée aussi méthode stationnaire est souvent utilisée pour le calcul des coefficients de prédiction vu la nature symétrique de la matrice qui décrit le système et qui permet d'utiliser des algorithmes de résolution plus aisés.

Méthode d'autocorrélation détermine les coefficients a(k) pour lesquels E est minimale. Cela se fait en annulant la dérivé de E:

$$\delta E / \delta a(i) = \sum_{n=1}^{N+p} 2[S(n) - \sum_{k=1}^{p} a(k).S(n-k)].S(n-i) = 0 \quad (*)$$

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

On pose :

$$C(i,k) = \sum_{n=1}^{N+p} S(n-k).S(n-i)$$

$$\sum_{n=1}^{N+p} S(n).S(n-i) = \sum_{k=1}^{p} a(k).C(i,k)$$

avec $i = 1,p$ et $k = 1,p$

d'ou : $C(i,k) = \sum_{n=1}^{N-i+k} S(n).S(n+i-k)$

avec $i = 1,p$ et $k = 1,p$

C(i,k) est la matrice d'autocorrélation, c'est une matrice carrée d'ordre « p » dite aussi " matrice de toeplitz " car les éléments situés symétriquement de part et d'autre de la diagonale sont égaux.

On note R(k) la fonction d'auto-corrélation définie par:

$$R(k) = \sum_{n=1}^{N+k} S(n).S(n+k)$$

avec $k = 1,p$

$R(k) = R(-k)$
d'ou :

$$C(i,k) = R(|i-k|) \quad (**)$$

L'équation (*) peut s'écrire sous forme:

$$\sum_{k=1}^{p} a(k).\ R(|i-k|) = R(i)$$

avec $i = 1,p$

L'équation (**) peut s'écrire sous forme matricielle :

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

De nombreuses méthodes permettent de résoudre le système d'équation linéaire telles que :

- Méthode de Gauss-Seidel
- Méthode de Jacobi
- Méthode de Gauss-Jordan
- Méthode de durbin

Mais en tenant compte de la rapidité d'exécution et de l'encombrement mémoire réduit, nous avons choisi pour résoudre le système d'équation d'auto-corrélation la méthode de DURBIN.

Méthode de DURBIN

$$D(i) = [\, R(i) - \sum_{j=1}^{i-1} a(j,i-1).R(i-j) \,] / E(i-1)$$

$$E(i) = (1-D^2(i)) .E(i-1)$$
$$A(j,i) = A(j,i-1) - D(i) . A(i-j , i-1)$$
$$A1(j) = A(j,12)$$

ou

D(i) : coefficient de réflexion
E(i) : erreur quadratique
A(j,i) : coefficient de prédiction
A1(j) : coefficient de l'auto-corrélation

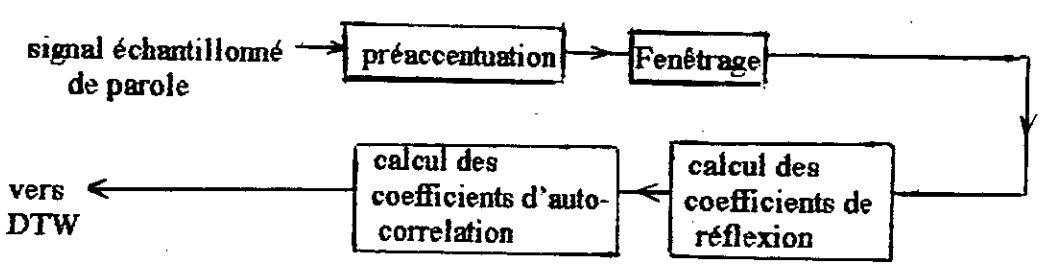
avec $i = 1,12$ et $j = 1, i-1$

Conditions initiales :

$E(1) = R(1)$
$D(2) = (R(2)/R(1))$
$A(2,2) = D(2)$
$E(2) = (1-D^2(2)).R(1)$

Etape d'analyse

Le synoptique suivant décrit les étapes d'analyse de la prédiction linéaire:



Organigramme

(voir ANNEXE)

CHAPITRE 3

D . T . W

D. T. W

3.1- INTRODUCTION

Les méthodes de reconnaissance (alignement temporel, méthode statistique, ou méthode connexioniste) utilisées sont à base d'algorithmes de recherche dans un graphe.

Une méthode de recherche dans un graphe a pour rôle de déterminer l'ordre dans le quel les états doivent être développés afin de retrouver l'état objectif à partir de l'état de départ. Une méthode de recherche sera d'autant "meilleur" qu'elle produira un "petit" ou "peu coûteux " graphe de recherche (par rapport au graphe complet).

Une solution est obtenue dés l'instant où l'état apparaîtra dans le graphe de recherche. La programmation dynamique est basée sur le principe d'optimalité introduit vers les années cinquante par R.BELLMAN et énoncé dans son livre [1] comme suit:

" Une politique est optimale si à une période donnée quelque soient les décisions précédentes les décisions qui restent à prendre constituent une politique optimale au regard du résultat des décisions précédentes".

Aussi tout chemin optimal est constitué de sous chemins optimaux.

3.2- FORMALISME DE LA PROGRAMMATION DYNAMIQUE (D.P)

Soit un système dont l'état est repéré par une variable scalaire ou vectorielle qui évolue sous l'effet de la décision.

Supposons que nous voulions trouver dans l'espace des états possibles du système , une trajectoire (ou politique) optimale pour certaine fonction de valeur, en choisissant, convenablement les variables de contrôle.

Pour comprendre le formalisme de D.P, il est utile d'introduire les notions suivantes:

- Le vecteur d'état: **Ei** (i= 1,2,,,,,,N) caractérise l'état du système à une étape donnée. E_0 état initial, E_N état final.

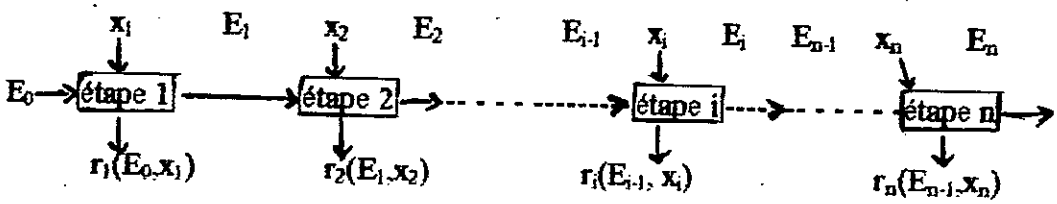
- Le vecteur de commande : **Xi** (i= 1,2 ,,,,,,N).
La transition d'un état vers un autre s'effectue sous l'action d'un vecteur de commande (ou de décision) Xi .

- La fonction de retour: La contrainte d'optimisation est décrite pour une "fonction de retour" (ou fonction coût) r_i, qui à chaque étape, précise la commande optimale Xi.
- La fonction de transfert : Le système est t.q la situation à l'étape (i) ne dépend que de la situation à l'étape (i-1) et de la décision Xi.

On écrit : $\mathbf{E_i = t_i\,(E_{i-1}\,,\,Xi)}$; i= 1,2,,,,N

t_i = fonction de transfert à l'étape (i)

Soit le système séquentiel suivant :



à l'étape (i) on a : $E_i = t_i\,(E_{n-1}, x_n)$
$= t_n\,(t_{n-1}(E_{n-1}, x_{n-1}), x_n)$

finalement $\mathbf{E_n = t_n(t_{n-1}.....t_1(E_0, x_1), x_1)}$

On a donc trouvé une relation qui lie E_0 (état initial) à E_n (état final) par les fonctions de transfert partielles t_i et les décisions x_i.

Le processus d'évolution de E_0 vers E_n se fait par une fonction coût R qui dépend des coûts partiels r_i. $R = (\,r_1(E_0, x_1\,), r_2(E_1,x_2).....,\, r_n(E_{n-1},x_n)\,)$.
Optimiser le système revient à chercher selon le problème le maximum ou le minimum de R.

Dans le cas ou R est décomposable et d'après le principe d'optimalité [1] , on peut écrire pour l'étape 1:

$$\mathbf{F_1(E_0) = \underset{x_1}{opt}\ r_1\,(E_0,x_1)}$$

Remarque

$F_1(E_0)$ est la fonction optimale à l'étape E_0 : elle nous permet à chaque étape de connaître la situation à l'étape suivante du système.
Ainsi à l'étape 1, elle permet de prendre la décision x_1 sous la contrainte r_1 et connaissant l'état initial E_0, puis pour l'étape suivante:

$$F_2(E_1) = \underset{x_2}{\text{opt}} \left(r_2(E_1,x_2), F_1(t_2(E_1,x_2)) \right).$$

On se trouve maintenant à l'étape 2 du processus, sous la contrainte r_2 et connaissant l'état antérieur du système F_1 (résultat du passage de E_0 à E_1 sachant que la décision x_1 a été prise) il s'agit de prendre la décision x_2 qui permet au système d'évoluer vers l'étape suivante.

Dans le cas général :

$$F_i(E_{i-1}) = \underset{x_i}{\text{opt}} \left(R_i(r_i(E_{i-1},x_i), F_{i-1}(t_i(E_{i-1},x_i))\right)$$

Cette formule de récurrence conduit à une résolution séquentielle, à chaque étape, l'optimum porte sur une seule variable de décision mais en fonction du paramètre qui traduit la situation précédente du processus.

En fin de chaîne on a :

$$F_N\ (E_{n-1}) = \underset{x_n}{\text{OPT}} \left(R_n(r_n (E_{n-1},x_n), F_{n-1}(t_n(E_{n-1},x_N))\right)$$

$$= F_n(E_n^*)$$

D'ou E_n^* est l'état final.

Remarque

Les variables conduisent à un optimum sont désignées par une étoile. Les états finaux ne peuvent être qu'optimaux.

A la dernière étape de calcul, E_n^* et x_n^* sont déterminées, on obtient E_{n-1}^* en résolvant :

$E_n^* = t_n(E_{n-1}^*, x_n)$; où $E_{n-1}^* = t_n^{-1}(E_n^*, x_n^*)$;

t_n^{-1} est la fonction de transfert réciproque, puis de proche en proche:
E_i^* et x_i^* déterminent E^*_{i-1} par :
$E_{i-1}^* = t_{i-1}(E_i^*, x_i^*)$
jusqu'à :
$E_0 = t_{i-1}(E_1^*, x_1^*)$.

Schématiquement, le processus peut se traduire par un graphe où les états successifs (les sommets) sont atteintes par divers décisions (arcs).

Soit par exemple, le graphe de la fig (4) où est tracée la dernière partie des chemins allant de la gauche vers la droite. En a,b,c et d les fonctions coût optimales F(a), F(b), F(c) et F(d). On avance étape par étape en considérant tous les états et décisions possibles. On arrive ainsi à l'état I atteint à partir de K par la décision 17.

Pour trouver l'ensemble des décisions optimales (trajectoires optimales), on remonte la chaîne .

Dans cet exemple, une décision est optimale si la taille du segment lui correspond est la plus petite, donc l'ensemble des décisions optimales sera représenté par les décisions 17, 11, 3,... qui conduiront aux états I,K,F,...

Le même problème de recherche de chemin optimal se rencontre en reconnaissance de la parole.

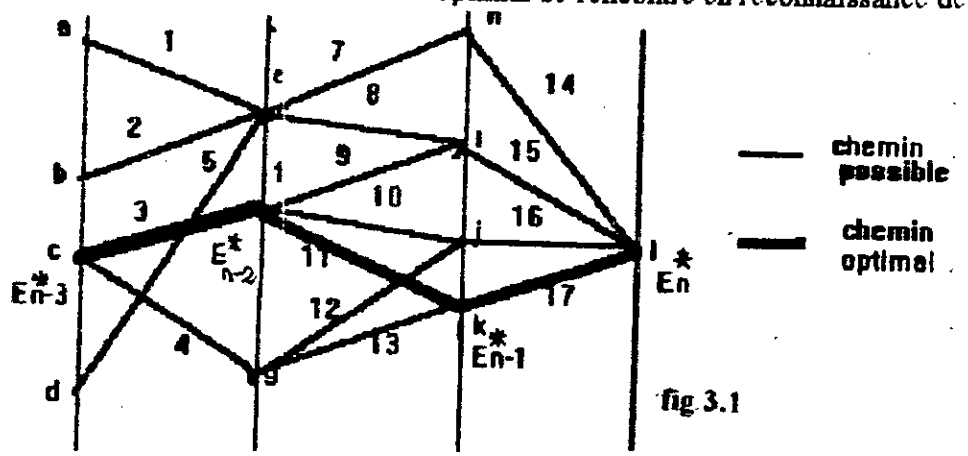


fig.3.1

3.3- NOTION DE DISTANCE OU MESURE DE SIMILITUDE

3.3.1- Introduction

La reconnaissance automatique de la parole (R.A.P) plus particulièrement celle des mots isolés, est basée sur l'évolution de la distance entre le mot à reconnaître et les mots de référence stockées en mémoire.

3.3.2- Notions mathematiques sur les distances

3.3.2.1- Definition

On appelle distance entre deux éléments x et y d'un ensemble E, l'application d définie de E x E dans R+ :

$(x,y) \rightarrow d(x,y)$ appartenant à R+ et vérifiant les propriétés suivantes:

1. $d(x,y) = 0$;
2. $d(x,y) = 0 \Rightarrow x = y$ (réflexivité)
3. quelque soient x, y appartenant à E
   $d(x,y) = d(y,x)$ (symétrie);

4. quelque soient x,y et z appartenant à E
   $d(x,y) \leq d(x,z) + d(z,y)$ (inégalité triangulaire)

L'ensemble E muni de l'application d est dit espace métrique.

3.3.2.2- Differentes formes de distance

La R.A.P est basée essentiellement sur un calcul de distance entre mots test et mot réference d'ou un choix judicieux de cette distance est nécessaire; cependant il existe plusieurs formes de distance applicable à la reconnaissance dont citera quelques unes:

- distance de TCHEBICHEF:

$$d_{TCH}(a_i, b_j) = \sum_{k=1}^{p} \| c_i(k) - c_j(k) \|$$

- distance d'ITAKURA

$$d_{ITA}(a,b) = \text{Log}\ [\ (a_i v_i a_i')/\ (bi\ Vj\ bj)]$$

où V est la matrice d'autocorrélation.

- Distance ceptrale pondérée:

$$d_{cp}(a_i,b_j) = \sum_{i=j}^{p} k^2\ [C_i(k) - C_j(k)]^2$$

3.4- APPLICATION DE LA PROGRAMMATION DYNAMIQUE A LA RECONNAISSANCE

Algorithme DTW (data time warping)

dynamic

3.4.1- Principe

Soient deux mots A (mot du vocabulaire) et B (mot de test) représentés respectivement par :

$A = a_1, a_2 \ldots a_i \ldots\ a_I$. où a_i (i allant de 1 à I) représente la $i^{ème}$ fenêtre (trame) du mot A
I est le nombre total de fenêtres du mot A.
$B = b_1, b_2, \ldots, b_j, \ldots\ b_J$. Ou b_j (j allant de 1 à J) représente la $j^{ème}$ fenêtre (trame) du mot B.
J est le nombre total de fenêtre du mot B.

Chaque fenêtre est représentée par ses p coefficients (de prédiction linéaire, ceptrale), Lors de la phase de reconnaissance le problème consiste à éliminer les différences temporelles entre les deux mots A et B. Les différences temporelles peuvent être traduites par la séquence de pointe : C(1), C(2) ,....,C(k),....,C(k_{max}) où :

$$C(k) = (ai(k),bj(k)) = C(i,j);$$

C(k) représente $k^{ième}$ point de la séquence de comparaison (fig)

Cette séquence de pointe représente une fonction F dite de déformation, qui réalise une représentation graphique de l'axe temporel du mot A sur celui de B.

$$F = C(1), C(2) ,\ldots,C(k),\ldots,C(k_{max}) \qquad \text{(fig)}$$

Cette fonction de déformation F s'écarte plus ou moins de la diagonale et tend à s'en éloigner si les différences temporelles entre les deux mots augmentent.
Pour mesurer les différences entre les vecteurs a et b en définit une distance d tel :

$$d(c) = d(a_i,b_j) = d(i,j) = \|a_i - b_j\| .$$

Les distances d(i,j) pour un i et un j donnés sont appelées distances locales de comparaison entre deux mots comme étant la somme pondérée des distances locales soit :

$$E(F) = \sum_{k=1}^{kmax} d(C(k)).w(k);$$

d(C(k)) : distance locale de comparaison;
w(k) : coefficient de pondération.

Les coefficients de pondération sont des coefficients positifs ou nuls introduits dans le souci de favoriser certains chemins et d'éviter tout écart excessif par rapport à la diagonale.
Les distances globales ne sont pas des mesures exploitables directement.

En effet, chaque distance globale est la somme d'un certain nombre de distances locales variables suivant le mot en référence. Pour pouvoir comparer correctement ce dernier au mot test, la distance globale doit être normalisée à la longueur du chemin de déformation.

De plus

$E(F) = \sum_{k=1}^{k_{max}} d(C(k)).w(k)$ atteint sa valeur minimale lorsque la fonction de déformation est terminée de façon optimale (ie: aucune comparaison n'est faite inutilement et par conséquent introduit une distance additive supplémentaire).

La distance globale temporelle est déduite des remarques précédentes :

$$D(A,B) = \underset{F}{MIN} \left[\sum_{k=1}^{k_{max}} d(C(k)).\, W(k) \right] / \sum_{k=1}^{k_{max}} w(k)$$

3.4.2- Restriction sur la fonction de déformation

La fonction de déformation est un modèle des fluctuations de l'axe temporel, par conséquent ce modèle doit préserver les propriétés essentielles d'un axe temporel. Ces propriétés se traduisent par les conditions suivantes: [12]

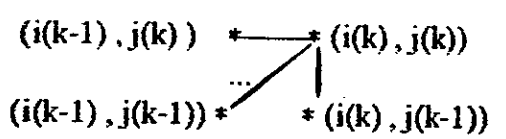
3.4.2.1- Contrainte de monotonie :

$$i(k-1) \leq i(k) \quad \text{et } j(k-1) \leq j(k)$$

3.4.2.2- Condition de continuité :

$$i(k) - i(k-1) \leq 1 \quad \text{et} \quad j(k) - j(k-1) \leq 1$$

Ces conditions imposent aux points C(k) et C(k-1) d'être reliées de la façon suivante :



C(k-1) = (i(k),j(k-1))
(i(k-1), j(k-1))
(i(k-1), j(k))

3.4.2.3- Conditions aux limites

$i(1) = 1$ et $j(1) = 1$

$i(k_{max}) = I$ et $j(k_{max}) = J$

I : nombre total de fenêtre du mot A
J : nombre total de fenêtre du mot B.

3.4.2.4- Fenêtre d'ajustement

$| i(k) - j(k) | \leq r$

ou

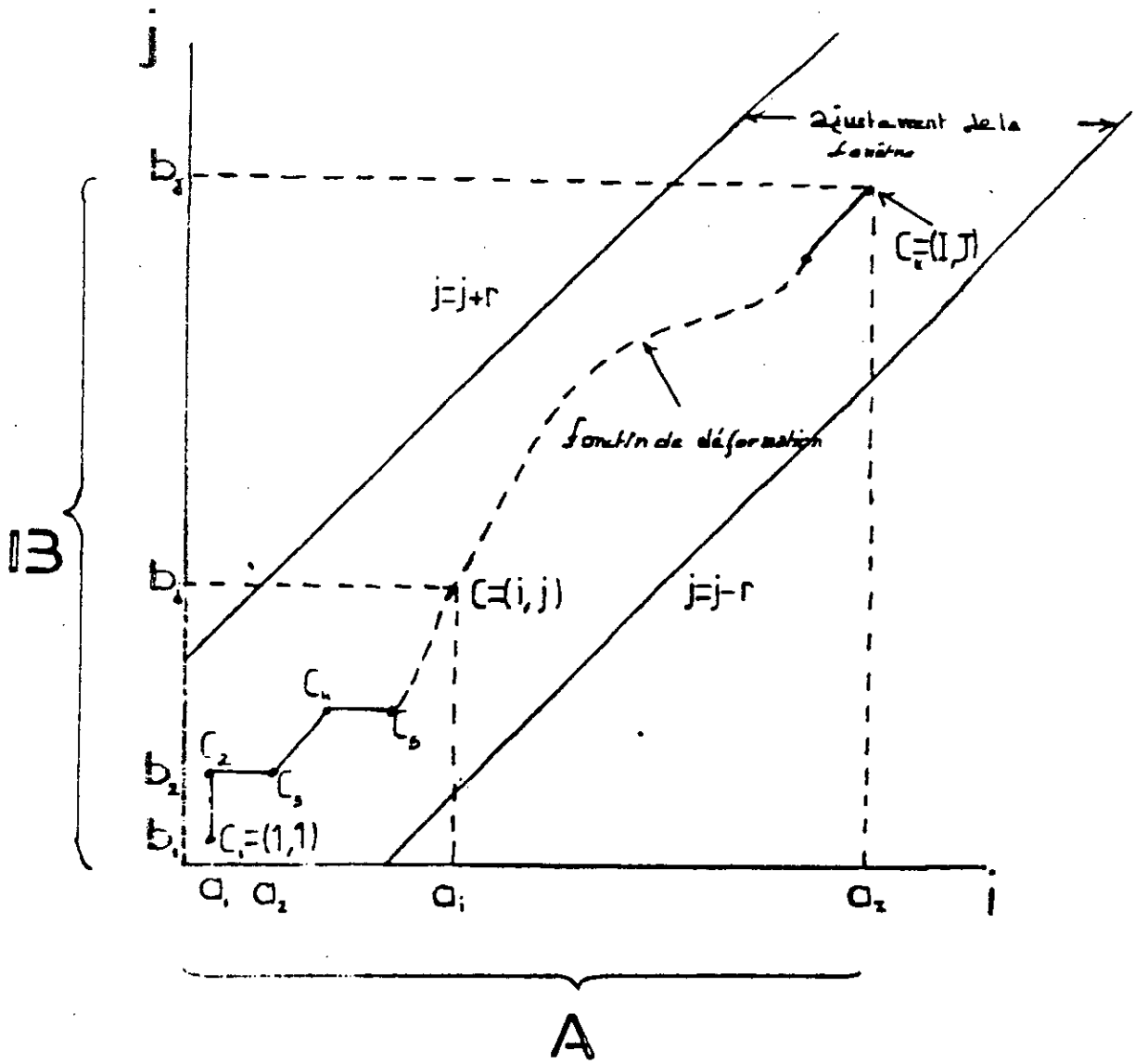


Fig 3.2 Fonction de déformation et ajustement de la fenêtre

r est un entier convenablement choisi et désigne la largeur de la fenêtre d'ajustement. Cette contrainte est introduite dans le cas général les fluctuations de l'axe temporel ne causent jamais d'excessives différences temporelles.

3.4.2.5- Contraintes locales

Les contraintes locales sont l'ensemble des conditions qui limitent les chemins admissibles et donc les points voisins qui peuvent atteindre un point donné. Ces contraintes se basent sur la déformation naturelle qui affecte la parole. Elles permettent des omissions ou des dédoublements de trames pour adapter localement des sons éventuellement identiques mais de longueurs différentes. Ces contraintes peuvent être de la façon suivante :

Si le point C(k) se déplace (m) fois parallèlement à l'axe (I) ou (J) , il ne peut continuer dans cette direction sans au moins (n) pas dans la direction de la diagonale. On évolue l'intensité de la pente p par la mesure $p = n/m$.

Ainsi quand $p = 0$
alors

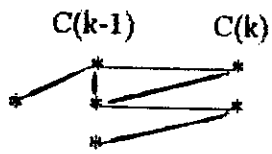
$n = 0$, il n'y a pas donc aucune restriction de pente.

Et quand p est très grand;

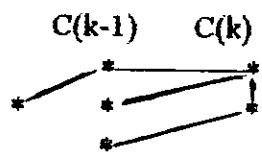
soit $m = 0$, alors la fonction F est restreinte à la diagonale et la normalisation se fait dans ce cas comme s'il n'y avait pas de normalisation temporelle.

Remarque

Afin de diminuer le nombre de chemins à explorer on ajoute une autre contrainte : le chemin de déformation ne doit pas changer orthogonalement de direction



Contrainte de pente
initiale (p = 1)



Contrainte de pente
simplifiée (p = 1)

3.5- LES COEFFICIENTS DE PONDERATION

La distance globale normalisée s'écrit :

$$D(A,B) = MIN \left[\sum_{k=1}^{kmax} d(C(k)).\, W(k) \right] / \sum_{k=1}^{kmax} w(k)$$

posons $M = \sum_{k=1}^{kmax} w(k)$; or $\sum_{k=1} w(k)$ est indépendant de la fonction de déformation F, d'où :

$$D(A,B) = 1/M \,.\, MIN \left[\sum_{k=1}^{kmax} d(C(k).w(k)) \right]$$

Le problème ainsi posé est divisé en étapes élémentaires à optimiser suivant la fonction de

déformation (est justiciable de la technique de programmation dynamique).
SAKOE set CHIBA ont donné deux définitions des coefficients de pondération : l'une pour une forme symétrique et l'autre pour la forme asymétrique.

3.5.1- Formes symétriques

Appelées aussi contraintes locales symétriques, elles traduisent le fait que les insertions les omissions peuvent se faire indifféremment sur les mots de références comme sur les mots test. Dans ce cas, nous pouvons écrire :

$$w(k) = (i(k) - i(k-1)) + (j(k) - j(k-1))$$

avec:

$$M = \sum_{i=i}^{k_{max}} w(k) = I + J$$; où I et J sont respectivement le nombre total de fenêtre des mots A et B.

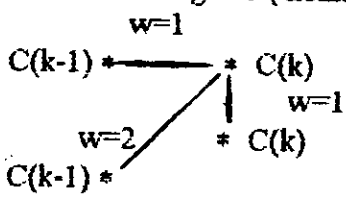
3.5.2- Formes asymétriques

Le caractère asymétrique des contraintes locales fait que les omissions et les insertions, réalisées par programmation dynamique, ne sont permises que sur les mots de références et aucunement sur le mot test, puisque sa longueur sert de norme et doit demeurer identique pour toutes comparaisons de mots.

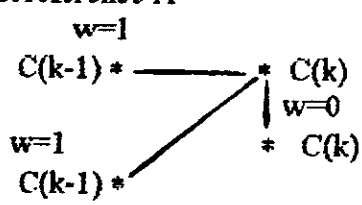
Nous pouvons alors écrire :

$$w(k) = i(k) - i(k-1) \qquad \text{d'où } M = I$$

I : étant la longueur (nombre de fenêtre total) du mot référence A



Forme symétrique



Forme asymétrique

SAKOE et CHIBA ont défini quatre pentes (0, 1/2, 1, 2) pour chaque forme (symétrique ou asymétrique) (voir tableau).

Un autre type de contrainte appelé contrainte globale est également à prendre en compte. Ces contraintes représentent un moyen de rejection des mots du vocabulaire qui sont soient trop longs soient trop courts par rapport au mot inconnu. Cette rejection diminue le risque d'erreur et amélioré le temps de réponse du système de reconnaissance.

3.6- ALGORITHME DE COMPARAISON PAR D.T.W

3.6.1- Cas général

Mots à comparer:

- mot référence : A de longueur I (nombre total de fenêtre)

- mot test : B de longueur J (nombre total de fenêtre)

Définition des contraintes

- Fenêtre d'ajustement : choisir la largeur de la fenêtre d'ajustement .
- Forme symétrique ou asymétrique :
  * forme symétrique : M = I + J;
  * forme asymétrique : M = I;
- Contrainte de pente : choisir P parmi (0,1/2, 1, 2) l'algorithme de base pour le calcul de D(A,B) s'écrit comme suit :

Conditions initiales

(i,j) = (1,1);
k = 1;
$g_1(C(1)) = g_1(1,1) = d(C1)).w(1)$

(*) pour $j-r \leq i \leq j+r$;

$$g_k(C(k)) = g_k(i,j) = \underset{C(k-1)}{MIN} (C(k-1) + d(C(k))).w(k)$$

Si (i,j) différent de (I , J)
alors incrémenter i,j et k puis aller en (*)

sinon
$D(A,B) = 1/M.\ g_{kmax}(I,J)$; (normalisation temporelle de la distance).
(Kmax : nombre de points de la fonction de déformation)

Arrêter.

Pour une pente, forme (symétrique ou asymétrique) préalablement fixées, on explore chaque point de la grille contenu dans la fenêtre d'ajustement choisie; c'est à dire on examine les chemins précédents qui peuvent l'atteindre et on leur affecter un coût, fonction de celui du chemin précédent et de la distance entre trames en court de comparaison.

Cette procédure est répétée pour chaque point de la grille jusqu'au point qui correspond à la fin de comparaison

3.6.2- Exemple d'application

- **Contraintes**

1. Fenêtre d'ajustement de largeur r ;
2. Forme symétrique : M = I + J ;
3. Pas de contrainte de pente : (p = 0)

- **Conditions initiales**

(i,j) = (1,1):
k = 1;
$g_1(1,1) = 2.d(1,1)$

pour

(*) $j-r \leq i \leq j+r$;

$$g_k(i,j) = \text{MIN} \begin{cases} g(i\ ,\ j-1) + d(i\ ,\ j); \\ g(i-1\ ,\ j-1) + 2.d(i\ ,\ j); \\ g(i-1\ ,\ j\ ) + d(i\ ,\ j\ ); \end{cases}$$

Si (i , j) différent de (I , J)
alors incrémenter i, j et k
puis
aller en (*)
Sinon
$D(A,B) = 1/M \ . \ g_{kmax}(I,J)$;

Arrêter..

[12]

pente	chemins	forme	équation de D.P g (i,j) = MIN ...
p = 0		symétrique	g(i,j-1) + d(i,j) g(i-1,j-1) + 2d(i,j) g(i-1,j) + d(i,j)
		asymétrique	g(i,j-1) g(i-1,j-1) + d(i,j) g(i-1,j) + d(i,j)
P = 1/2		symétrique	g(i-1,j-3) + 2d(i,j-2) + d(i,j-1) + d(i,j) g(i-1,j-2) + 2d(i,j-1) + d(i,j) g(i-1,j-1) + 2d(i,j) g(i-2,j-1) + 2d(i-1,j) + d(i,j) g(i-3,j-1) + 2d(i-2,j) + d(i-1,j) + d(i,j)
		asymétrique	g(i-1,j-3) + (d(i,j-2) + d(i,j-1) + d(i,j))/ 3 g(i-1,j-2) + (d(i,j-1) + d(i,j))/ 2 g(i-1,j-1) + d(i,j) g(i-2,j-1) + d(i-1,j) + d(i,j) g(i-3,j-1) + d(i-2,j) + d(i-1,j) + d(i,j)
P = 1		symétrique	g(i-1,j-2) + 2d(i,j-1) + d(i,j) g(i-1,j-1) + 2d(i,j) g(i-2,j-1) + 2d(i-1,j) + d(i,j)
		asymétrique	g(i-1,j-2) + (d(i,j-1) + d(i,j)) / 2 g(i-1,j-1) + d(i,j) g(i-2,j-1) + d(i-1,j) + d(i,j)
P = 2		symétrique	g(i-2,j-3) + 2d(i-1,j-2) +2d(i,j-1) +d(i,j) g(i-1,j-1) + 2d(i,j) g(i-3,j-2) + d(i-2,j-1) +d(i-1,j) +d(i,j)
		asymétrique	g(i-2,j-3) + 2(d(i-1,j-2) +d(i,j-1) d(i,j))/3 g(i-1,j-1) + d(i,j) g(i-3,j-2) + d(i-2,j-1) +d(i-1,j) +d(i,j)

CHAPITRE 4

SIMULATION DES VOYELLES

Simulation des voyelles

4.1- MODELISATION [2]

Le modèle est une abstraction d'une chose réelle dans laquelle les relations entres les éléments réels sont remplacées par des relations convenables entre objets que nous appellerons relations de fonctionnements du modèle .

Tout modèle est une représentation simplifiée de la réalité, et la modélisation consiste à simplifier son étude d'après le modèle choisi.

4.1.1- Modèle de connaissance

Dans ce modèle l'approche vise la compréhension précise des phénomènes, ce qui se traduit par des relations complexes et peuvent être multivariables, aléatoires, non stationnaire etc..., le modèle de connaissance est un cas parfait, difficile à réaliser.

4.1.2- Modèle de représentation

Dans ce modèle, la démarche pour établir un tel modèle consiste à représenter le signal mesuré par un système qu'il s'agit d'identifier.

Dans le cas du traitement du signal, la modélisation d'un processus donné consiste à déterminer un modèle mathématique basé sur les données d'observation pour représenter le système considéré.

4.2- METHODES DE SIMULATIONS

Les sons voisés tels que les voyelles orales, que nous étudions, sont dus à une excitation pseudo-périodique ou les cordes vocales vibrent sous l'action de la pression du flux d'air envoyé par les poumons.

Pour simuler ces voyelles , on propose deux méthodes : on se basant sur les valeurs moyennes de F1 et F2 [5] suivantes :

voyelles	F1(hz)	F2(hz)
A	750	1350
E	400	2200
I	280	2500
O	375	750
U	250	600
Y	250	1800

4.2.1- Méthode 1

Cette méthode est basée sur une somme de sinusoïdes . L'équation de génération des échantillons s'écrit:

$$S(k) = Av \sin(coef\ n\ f1) + Av/\sqrt{2} \sin(coef\ n\ (f1 + 10)) + Av\sqrt{2} \sin(coef\ n(f1-10)) + Av \sin(coef\ n\ f2) + Av\sqrt{2}\sin(coef\ n\ (f2 + 10)) + Av\sqrt{2}\sin(coef\ n\ (f2-10))$$

S(k) : $k^{ème}$ échantillons de la locution considérée ($1 \leq k \leq$ Nech)
Av : amplitude associée
Coef = 2π / Nech
Nech nombre d'échantillons
f1,f2 sont les deux premiers formants caractérisant la voyelle.

4.2.2- Méthode 2

Elle est basée sur le modèle autorégréssif du signal de la parole . Pour les voyelles non nasales, ce modèle ne présente que des pôles. L'excitation U(z) de ce système est une suite périodique d'impulsions de période P égale à la période du pitch

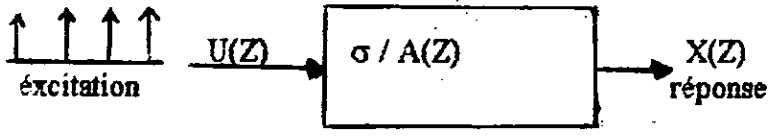


Fig: modèle autorégréssif

La fonction de transfert de ce modèle est :

$$X(z)/U(z) = \sigma/A(z)$$

Avec : $A(z) = \sum_{i=0}^{p} a(i)\, z^{-i}$ et $a(0) = 1$

d'ou :

$$\sigma\, U(z) = X(z)\, A(z) = \sum_{i=0}^{p} a(i)\, X(z)\, Z^{-i}$$

Soient U(n) et X(n) les transformées en Z inverses respectives de U(z) et X(z) tq:

$$U(n) = \sum_{i=0}^{p} \sigma(n-k\ p)$$

par suite : $\sigma U(n) = \sum_{i=0}^{p} a(l)\, X(n-l) = X(n) + \sum_{i=0}^{p} a(l)\, X(n-l)$

alors

$$X(n) = \sigma U(n) - \sum_{i=0}^{p} a(l)\, X(n-l)$$

Donc chaque échantillon peut être évalué par la différence de l'excitation et des p échantillons qui le précèdent les seules inconnues dans cette équation sont le gain σ et les pôles a (i) du modèle .

La fonction de transfert G(z) du conduit vocal pour un modèle tout pôles est donnée par :

$$G(z) = \prod_{k=1}^{*} A_k / (1 + b_{k,1} Z^{-1} + b_{k,2} Z^{-2})$$

ou

k est le nombre de cellules de résonances du conduit vocal.

On prend une seule cellule de résonance pour chaque formant soit:

$$g_k(z) = A_k / (1 + b_{k,1} Z^{-1} + b_{k,2} Z^{-2})$$

avec :

$b_{k,2} = 1 - 2\pi B_k / Fe$

$b_{k,1} = -2\, b_{k,2} \cos(2\pi F_k / Fe)$

$A_k = 1 + b_{k,1} + b_{k,2}$

F_k: fréquence du kième formant
B_k: bande de fréquence à 3 db du kième formant
$b_{k,1}$ et $b_{k,2}$: pôles de G(z).
A_k: amplitude (gain du modèle) du kième formant

Remarquons que dans cette méthode ,c'est le modèle qui fixe l'amplitude des formants selon leurs fréquences et leurs bandes .

Puisque les deux premiers formants f1 et f2 suffisent pour caractériser nos voyelles ,on aurait besoin que de deux cellules de résonance pour la simulation, chacune pour un formant, soit:

$$G(z) = g_1(z) + g_2(z)$$

$$= A_1 A_2 / (1 + b_{1,1} Z^{-1} + b_{1,2} Z^{-2})(1 + b_{2,1} Z^{-1} + b_{2,2} Z^{-2})$$

$b_{1,1}$, $b_{1,2}$, $b_{2,1}$, $b_{2,2}$ ainsi que A_1 et A_2 sont donnés par les formules précédentes .

Par identification au modèle autorégressif , on obtient :

$\sigma = A_1 A_2$

$$A(z) = (1 + b_{1,1} Z^{-1} + b_{1,2} Z^{-2}) . (1 + b_{2,1} Z^{-1} + b_{2,2} Z^{-2})$$
$$= 1 + (b_{1,1} + b_{2,1}) Z^{-1} + (b_{1,2} + b_{2,2} + b_{1,1} b_{2,1}) Z^{-2}$$
$$+ (b_{1,1} b_{2,2} + b_{1,2} b_{2,1}) Z^{-3} + b_{1,2} b_{2,2} Z^{-4}$$

Il en résulte que notre modèle est du quatrième ordre tq:

$a(0) = 1$
$a(1) = b_{1,1} + b_{2,1}$
$a(2) = b_{1,2} + b_{2,2} + b_{1,1} b_{2,1}$
$a(3) = b_{1,1} b_{2,2} + b_{1,2} b_{2,1}$
$a(4) = b_{1,2} b_{2,2}$
d'ou l'équation de génération des échantillons

$$X(n) = \sigma U(n) - \sum_{i=1}^{4} a(i)\, X(n-i)\ ,\quad 0 \le n \le Nech$$

ou encore

$$X(n) = \begin{cases} \sigma & ,\ n=0 \\ \sigma U(n) - \sum_{i=1}^{n} a(i)\, X(n-i) & ,\ n \le 4 \\ \sigma U(n) - \sum_{i=1}^{4} a(i)\, X(n-i) & ,\ 4 \le n \le Nech \end{cases}$$

l'excitation U(n) est donnée par :

$$U(n) = \begin{cases} 1 & \text{si n est un multiple de P} \\ 0 & \text{si non} \end{cases}$$

avec

p = période du pitch / période d'échantillonnage = F échantillonnage (Fe)/F du pitch(Fo)

CHAPITRE 6

CLASSIFICATION

CLASSIFICATION

5.1- INTRODUCTION

Notre système de reconnaissance de la parole a pour but d'identifier une locution prononcée par n'importe quel locuteur . Le dictionnaire aura pour données les résultats des méthodes d'analyse pour chaque locution issue de la population de référence.
Il existe essentiellement deux techniques de classification :

- Dans les méthodes de partitionnement on cherche à regrouper les I éléments (dans notre cas, il s'agit des prononciations d'un mot par divers locuteurs) en K classes aussi homogènes que possible, et aussi différenciées les unes des autres que possible;

- Dans les méthodes hiérarchiques on cherche à regrouper les I éléments en une hiérarchie de classes de plus en plus grandes.

5.2- FORMULATION DE L'APPRENTISSAGE MULTILOCUTEUR

Si dans le cas d'un système monolocuteur une seul référence par mot est en général suffisante, dans le cas d'un système multilocuteur plusieurs références par mot sont nécessaires.

5.2.1- Situation du problème

5.2.1.1- Algorithme de classification

Selon l'approche utilisée , nous pouvons diviser les algorithmes de classification par partitionnement en deux catégories:

- les algorithmes parallèles (ou par échange) ou les classes sont recherchés et produites simultanément;

- les algorithmes séquentiels ou les classes sont recherchés et produites l'une après l'autre.

5.2.1.2- Partitionnement par l'algorithme d'échange basé sur une fonction-critère(AEC)

a - Description de l'algorithme AEC

L'algorithme d'échange basé sur une fonction-critère (AEC) produit de façon itérative et parallèle un nombre k de classes; k est fixé à priori.

On suppose donnée une partition de départ. Lors de chaque itération, un élément X_i est enlevé à sa classe C_k et transféré dans une autre classe C_l .

b - ALGORITHME

1. choisir une partition initiale $C_1, C_2, \ldots, C_k$

2. Trouver X_i appartenant à C pour lequel il existe une classe C_l telle que le transfert de X_i de sa classe C_k vers la classe C_l fasse diminuer la fonction-critère F:
3. Transférer X_i dans la classe C_l, qui donne la plus grande décroissance:
4. Aller en 2.

5.2.1.3- PARTITIONNNEMENT PAR L'ALGORITHME SEQUENTIEL SUR UNE FONCTION-CRITERE (ASC)

a - PRINCIPE

Dans l'algorithme ASC on procède de façon itérative et séquentielle : Les classes sont créées l'une après l'autre; à chaque itération on cherche la meilleure classe parmi les classes-condidates. Les éléments de la meilleure classe sont alors soustraits et l'on continue la procédure jusqu'à ce qu'il n'y ait plus d'éléments à classer.

A Chaque itération intervient un seuil de distance, pour la création des classses-candidates. Dans l'algorithme séquentiel, le nombre de classes n'est en principe pas fixé à priori; par contre, le seuil de distance doit l'être .

b - L'algorithme ASC

Soit $A(X_i)$ la classe-candidate associée à X_i pour un seuil de distance donné T:

$$A(X_i) = \{ X_j \in C' / \ D(X_i,X_j) < T \}$$

et soit $F(A(X_j))$ une fonction-critère.

1. Initialisation

   $k = 1$ (k-ième classe)
   $C' = C$ (le reste est égal à l'ensemble de tous les échantillons)

2. Pour tout $X_i \in C'$ trouver la classe-candidate $A(X_i)$:

$$A(X_i) = \{ X_j \in C' / \ D(X_i,X_j) < T \}$$

3. Déterminer la classe-candidate qui minimise la fonction-critère F :

$$C_k = A(X_i^*) / \ F(A(X_i^*)) \leq F(A(X_i)) \ \forall X_i \in C'$$

4. $C' = C - C_k$

5. Si
   $C' \neq \Phi$: $k = k+1$ et aller en 2.

   Sinon : fin .

Avant d'aborder l'algorithme nous aurons besoin de quelques outils mathématiques que nous définirons ci-après.

5.3- Définitions

Appelons L l'ensemble $\{ L_i \}$ $i = 1,2,....,I$
ensemble qui contient toutes les prononciations d'un même mot.

L'ensemble des classes C_k $k = 1,.....,K$ forment une partition de L c'est à dire :

$C_k \neq \phi \qquad \forall k = 1,....,K$
$C_k \cap C_l = \phi \quad \forall k = 1,.....,k$

$$\bigcup_{k=1}^{k} C_k = L$$

Appelons également $d(L_i, L_k)$ la distance globale normalisée entre les prononciations L_i, L_k et n_k le nombre d'éléments de la classe C_k.

5.3.1- Définition de la métrique

Soit une classe C_k.

Pour avoir une idée qualitative sur la situation d'un élément par rapport aux autres éléments d'une classe donnée C_k, on définit une métrique comme suit :

$$m(L_i, L_k) = [\ (1/n_k - 1) \sum_{\substack{k=1 \\ k \neq i}}^{n_k} d^2 (L_i, L_k)\ ]^{1/2}$$

cette métrique représente la distance quadratique moyenne de L_i à tout les autres $L_k \in C_k$, et $L_k \neq L_i$.

5.3.2- Définition de la fonction d'homogénéité d'une classe

Cette fonction est destiné à nous donner une mesure de la qualité d'une classe, c'est à dire son degré d'homogénéité ou en termes concrets, le degré d'éparpillement des éléments entre eux dans la classe.

Nous définirons simplement par :

$$H(C_k) = (1/n_k) \sum_{i=1}^{nk} m(L_i, L_k)$$

Nous pouvons remarquer que la fonction d'homogénéité h(c) est d'autant plus petite que les éléments de classe sont plus proches les uns des autres

5.3.3- Fonction critère

Cette fonction est destinée à mesurer la qualité relative d'une partition $\{ C_k \}$ $k = 1,\ldots,K$

Pour notre part nous choisissons une fonction critère $F[\{ C_k \}]$ définie comme suit :

$$F = \sum_{k=1}^{k} H(C_k)\,(n_k - 1)$$

F étant par définition une somme de quantité positives , sa diminution , donc la diminution des $H(C_k)$ entraine la formation de meilleures partitions et par voix de conséquence la formation de meilleures classes.

5.4- Algorithme

1. Choix d'une partition initiale (choix arbitraire).

2. Trouver un élément L_i tel que son transfert d'une classe à une autre provoque la diminution de la fonction critère.

$F[\{ C'_k \}] < F[\{ C_k \}]$

sinon aller en 5

3. Transférer cet élément L_i à une classe tel que la diminution de F est maximale .

4. Aller en 2

5. Fin.

Algorithme (voir annexe)

CHAPITRE 6

DECISION

DECISION

6.1- INTRODUCTION

La décision est la dernière étape dans le processus de la R.A.P. Elle consiste à prendre la décision du mot reconnu en exploitant les résultats issus de la comparaison dynamique des mots (distances globales).

A la fin de la phase de comparaison, nous avons un ensemble de distances globales ordonnées suivant un ordre croissant. Toute fois il peut arriver que cet ordre ne soit pas conforme à la réalité , c'est à dire que le mot qui devrait normalement être choisi ne se trouve pas à la première position et ceci peut arriver pour différentes raisons :
- Mauvaise extraction des mots ;
- Ressemblance phonétique entre deux mots du dictionnaire ; ect....

6.2- TECHNIQUE DES K.N.N

La technique des K.N.N (en anglais k Nearst Neighbours) est utilisée pour remédier au problème cité ci-dessus. L'idée est d'utiliser des « K « premières références d'un mot pour établir le classement des mots et non plus un classement référence par référence .

Rappelons que les critères temps de réponse et les contraintes d'espace mémoire sont très importants pour les systèmes de R.A.P. Aussi toute conception doit veiller à optimiser ces deux paramètres .
Dans notre système qui rappelons est multilocuteurs, chaque mot référence est représenté par plusieures élocutions issues de la phase d'apprentissage.

Dans notre système, la reconnaissance fait en deux passages. Pendant le premier passage le mot Test est comparé avec un nombre réduit mais suffisant de références par mots. après ce passage, on dégagera un certain nombre de candidats les plus probables.

Au deuxième passage, on comparera le mot Test avec les références restantes des mots issus du premier passage.

6.3- REJET

Certains mots font l'objet d'une décision de rejet de la part du système c'est à dire que le système ne répond pas à l'ordre lorsque la plus petite distance est plus grande qu'une distance seuil qu'on fixe à l'avance.

Cette procédure est utile et est même nécessaire . Elle a pour but d'écarter une éventuelle détection de bruit ou la prononciation d'un mot étranger c'est à dire n'appartenant pas au dictionnaire de référence.

Il n'y a pas un moyen de prédire à l'avance la distance seuil acceptable, sa détermination est une opération empirique qui se fait sur un grand nombre de données issues de comparaisons entre divers mots.
- **Organigramme (voir annexe)**

CHAPITRE 7

TESTS ET RESULTATS

TEST ET RESULTATS

L'ensemble des fichiers test sont regroupés en Annexe

7.1- INTRODUCTION

Le but de ce chapitre est d'établir un certain nombre de tests pour voir le comportement de nos différents programmes de reconnaissance.

7.2- TESTS MONOLOCUTEUR

7. 2.1- INFLUENCE DE L'ENERGIE

Le but de ce test est d'avoir une idée sur le module analyse acoustique; en effet théoriquement les coéfficients cepstraux sont normalisés en énergie.

Pour ceci nous avons maintenu les formants F1, F2 ainsi que Nech constants et nous avons fait varier l'amplitude Av.

Modèle du Sinus

Ceptrale :

" E1test_ch4 " **Av = 25**

R T	A	O	I	E	U	Y
A	**10.6195**	16.4663	14.7166	18.6521	18.0907	16.7484
O	12.4759	**12.0565**	11.4958	16.3249	14.2093	13.6952
I	13.9513	15.4289	**10.3019**	18.3204	17.1683	14.0391
E	12.4022	13.8726	13.6051	**9.8310**	15.0230	138665
U	12.5560	12.8212	12.3914	14.7598	**10.7255**	12.5888
Y	12.5849	14.6730	11.7827	16.4577	14.2270	**12.0099**

" E2test_ch4 " **Av = 50**

R T	A	O	I	E	U	Y
A	**2.6918**	9.2903	9.1159	11.1844	11.4992	10.7367
O	8.1278	**2.6918**	7.0983	10.3354	8.6756	9.8346
I	8.1278	9.0486	**2.6918**	10.7885	10.6116	9.0316
E	9.2045	9.4706	9.5668	**3.4465**	10.1116	10.9386
U	8.5797	8.1245	7.1556	9.9517	**2.6918**	8.6288
Y	8.3475	9.0877	8.0437	11.1940	8.6706	**3.1847**

Taux de reconnaissance TR = 100 %

- Modèle de sinus

LPC

LE1test_ch4

$A_v = 25$

R T	A	O	I	E	U	Y
A	**0.0000**	10.9461	11.1346	10.7433	9.6276	10.9518
O	10.9461	**0.0000**	7.1139	4.9775	4.2170	7.9464
I	11.1346	7.1139	**0.0000**	6.9009	7.1304	2.9236
E	10.7433	4.9774	6.9009	**0.0000**	3.7944	7.0267
U	9.6276	4.2170	7.1304	3.7944	**0.0000**	7.3214
Y	10.9518	7.9464	2.9236	7.0267	7.3214	**0.0000**

LE2test_ch4

$A_v = 50$

Ref Test	A	O	I	E	U	Y
A	**0.0000**	10.9462	11.1347	10.7434	9.6276	10.9518
O	10.9461	**0.0000**	7.1139	4.9775	4.2170	7.9464
I	11.1346	7.1139	**0.0000**	6.9009	7.1304	2.9236
E	10.7433	4.9774	6.9009	**0.0000**	3.7944	7.0267
U	9.6276	4.2170	7.1304	3.7944	**0.0000**	7.3214
Y	10.9518	7.9464	2.9236	7.0267	7.3214	**0.0000**

Taux de reconnaissance TR = 100 %

7.2.2- INFLUENCE DE LA LONGUEUR DU MOT

Ce test a été élaboré pour voir comment le module normalisation temporelle se comporte vis-à-vis des variabilités temporelles. Dans ce but nous avons crée des fichiers tests en gardant les formants F1 et F2 constants ainsi que l'amplitude Av.

SNE1testX

- **Modèle de sinus**
- **Cepstrale**

Ref \ Test	A	O	I	E	U	Y
A	8.4814	10.2267	8.9764	14.5186	10.0170	12.0345
O	9.1076	**7.6807**	8.4717	12.7713	8.4967	10.4007
I	**7.8593**	9.2471	**7.7944**	12.5455	9.5460	**9.8149**
E	10.2817	10.5746	11.1256	14.7220	10.6144	10.3327
U	9.3849	8.1256	8.3417	**10.5332**	**7.8319**	11.2926
Y	9.1778	10.2918	9.0573	10.4711	9.4470	9.9071

SNE2TSTX . DAT

Ref \ Test	A	O	I	E	U	Y
A	**13.0450**	10.3793	10.4040	13.6430	9.9975	7.9981
O	14.8259	9.6563	9.5757	13.1997	8.8033	8.3203
I	18.7746	9.3630	**9.2809**	14.1282	8.8451	**7.1688**
E	17.4307	9.7801	11.0006	14.3965	10.5022	9.3611
U	16.7045	**9.1908**	9.7465	**10.9896**	**8.4941**	7.4186
Y	15.5711	11.7113	9.7777	12.4238	9.2720	8.6678

Taux de reconnaissance TR = 50 %

- **Modèle de sinus**

LPC

SLE1tstX . DAT

Ref \ Test	A	O	I	E	U	Y
A	8.6616	10.3991	12.5630	10.9937	12.2062	10.6684
O	8.5350	**3.8765**	8.9757	**6.5856**	5.6038	9.1882
I	7.3544	7.4299	**1.8474**	9.9534	10.3671	**3.1152**
E	**6.9432**	4.7286	8.4843	7.8475	5.7559	8.8084
U	7.7913	4.3476	8.8498	7.3698	**5.1594**	9.3942
Y	7.2513	7.7378	3.5870	10.4162	11.1274	3.3784

Taux de reconnaissance TR = 50 %

SLE2TsTX

Ref \ Test	A	O	I	E	U	Y
A	7.5064	8.9617	11.7840	12.8850	12.7455	12.2970
O	7.9175	**7.4584**	9.1158	**4.3366**	4.4718	9.2277
I	7.5391	8.5011	**1.9037**	9.0220	9.2712	3.1001
E	**7.1680**	7.9583	8.7711	5.3317	4.9629	8.6574
U	7.8106	7.7514	8.9194	4.7308	**3.0011**	9.1578
Y	7.3246	8.5469	2.8757	9.6831	9.6573	**1.9480**

Taux de reconnaissance TR = 67 %

- **Modèle AR**
- **Cepstrale**

ARE1TsTX

Ref \ Test	A	O	I	E	U	Y
A	**0.4603**	10.9667	17.5314	14.1449	16.9690	16.2404
O	10.9890	**0.3194**	11.0215	7.1226	10.2699	9.3694
I	17.0005	11.2793	**0.0000**	9.2291	8.6512	10.0048
E	14.2946	7.2751	9.5270	**1.4194**	8.7903	7.7348
U	16.7933	10.5107	8.5016	9.0616	**0.1567**	8.0148
Y	15.2862	9.3923	9.9583	7.6000	7.8968	**1.3648**

ARE2TsT X

Ref \ Test	A	O	I	E	U	Y
A	**6.6172**	11.6288	17.5314	14.7436	16.9690	16.0614
O	23.2463	**0.9271**	11.0215	7.2062	10.2699	9.0744
I	28.1936	10.6386	**0.0000**	9.1574	8.6512	9.6238
E	21.6343	6.8703	9.5270	**0.8393**	8.7903	8.0383
U	23.0535	10.5437	8.5016	9.3155	**0.1567**	8.1489
Y	21.5531	8.6916	9.9583	7.7624	7.7624	**1.4490**

Taux de reconnaissance TR = 100 %

- **Modèle AR**
- LPC

ALE1TsTX . DAT

Ref	Test	A	O	I	E	U	Y
A		**0.0238**	1.9334	1.8545	1.4248	1.7599	1.2605
O		1.9202	**0.0496**	2.7186	2.4424	1.8121	1.9884
I		1.7796	2.7514	**0.1032**	0.6675	2.5727	01.3339
E		1.4374	2.4529	0.7969	**0.0478**	2.4104	1.7685
U		1.6763	1.8986	2.5433	2.3875	**0.0069**	1.8881
Y		1.0501	2.0210	1.3911	0.7365	1.9918	**0.1437**

ALE2TsT

Ref	Test	A	O	I	E	U	Y
A		**0.1766**	1.8006	1.7753	1.6274	1.8714	1.4955
O		1.9034	**0.1835**	2.7470	2.3668	1.8819	1.9932
I		2.0294	2.7162	**0.0200**	0.7720	2.5878	1.6247
E		1.5561	2.4070	0.6884	**0.1147**	2.4524	0.9248
U		1.7516	2.0706	2.5460	2.3283	**0.0767**	2.0136
Y		1.1967	1.9024	1.3120	0.8844	2.0360	**0.1992**

Taux de reconnaissance TR = 100 %

7.2.3- DEPLACEMENT DES FORMANTS

Nous avons réalisés 2 tests en décalant à chaque test les formants F1 et F2 et en gardant les autres paramètres constants c'est à dire Nech et Av.

'F1test_' ch

- **Modèle de sinus**
- **Cepstrale**

Ref \ Test	A	O	I	E	U	Y
A	**6.7773**	8.5837	8.0674	12.4612	9.7274	8.5843
O	8.7610	**6.7200**	8.7175	**10.9823**	8.3270	8.6938
I	7.6375	8.3572	**7.5833**	12.5665	8.6288	8.9495
E	9.4113	8.4270	10.3983	11.1242	9.6647	10.6933
U	9.7886	7.8303	9.3091	11.1454	**7.1350**	7.8515
Y	9.2888	8.6854	9.3762	13.5485	8.5454	**3.2145**

'F2test_'ch

Ref \ Test	A	O	I	E	U	Y
A	**6.0931**	8.6099	8.9550	9.5980	9.3344	10.8119
O	8.2254	**6.9511**	8.1519	9.2203	7.2634	10.3455
I	7.1435	7.3812	**5.0036**	10.2486	8.6776	9.6230
E	9.7187	9.0156	10.6185	**2.5447**	8.6063	11.7487
U	10.4748	8.2415	8.4039	9.7014	**7.2324**	**9.5005**
Y	8.5498	8.3898	7.7561	9.9196	8.2485	10.5022

Taux de reconnaissance TR = 83 %

- Modèle de sinus

LPC

SLF1 tst.ch4 . DAT

Ref \ Test	A	O	I	E	U	Y
A	**7.6420**	12.3414	11.6406	12.0621	11.4840	11.0528
O	14.1767	**1.9906**	8.8102	5.1019	4.1516	9.2646
I	14.2324	8.2943	**1.6911**	8.3992	8.8321	3.4693
E	15.4546	4.5916	8.5746	**1.5025**	3.9514	8.6688
U	15.2447	3.8524	8.8256	4.7731	**2.8631**	9.1967
Y	13.7235	9.2616	3.4466	8.9690	9.1554	**0.9009**

SLF2tst.ch4 . DAT

Ref \ Test	A	O	I	E	U	Y
A	**6.7907**	10.2683	11.6357	11.4826	12.2247	11.7922
O	8.4795	**3.4733**	8.7584	4.8430	4.0619	9.6203
I	7.5709	8.2303	**1.2340**	8.4067	8.7959	3.6394
E	7.2287	4.9836	8.4532	**0.3053**	3.9987	9.1748
U	8.0774	4.4624	8.7904	4.0598	**0.2763**	9.5945
Y	7.4106	8.9909	3.1694	8.8393	9.1307	**1.4612**

Taux de reconnaissance TR = 100 %

- Modèle AR
- Cepstrale

Effet décalage des formants

ARF1TsTX . DAT

Ref \ Test	A	O	I	E	U	Y
A	**1.4127**	11.7847	16.8498	11.6942	17.4784	15.0811
O	10.9664	**2.5897**	10.3539	7.2636	10.0991	8.6035
I	17.4195	12.1327	**3.0576**	10.3649	8.9524	9.6723
E	14.1357	8.7570	9.0604	**5.4174**	9.1892	8.0286
U	16.4191	11.9666	8.2298	9.9223	**1.2898**	7.7200
Y	15.3082	10.4054	9.3416	9.1659	8.8335	**0.9216**

ARF2TsTX . DAT

Ref \ Test	A	O	I	E	U	Y
A	**0.4883**	12.3558	17.8933	14.3962	16.5591	15.8307
O	10.8313	**1.1845**	11.3680	7.1540	9.8884	8.9407
I	17.3782	11.0438	**1.9234**	9.4391	8.3645	9.4808
E	14.3822	7.2651	9.1978	**0.2980**	7.7232	8.0176
U	16.7410	10.3410	7.9433	8.7443	**0.9938**	7.9565
Y	15.5389	9.1551	9.0253	8.1081	8.3268	**2.1383**

Taux de reconnaissance TR = 100 %

- **Modèle AR**
- LPC

ALF1TsRX

Ref \ Test	A	O	I	E	U	Y
A	**0.2758**	2.3239	1.7875	1.4699	1.7246	1.0476
O	2.0852	**0.5634**	2.7411	2.5312	1.7243	2.0376
I	1.7474	2.7180	**0.0769**	0.6092	2.5918	1.3058
E	1.3593	2.5239	0.6926	**0.2717**	2.4073	0.7314
U	1.7850	1.6332	2.5371	2.4344	**0.2498**	1.9468
Y	0.9698	2.3184	1.3125	0.8822	1.9597	**0.0255**

ALF2TsTX

Ref \ Test	A	O	I	E	U	Y
A	**0.0426**	1.9600	1.7778	1.4399	1.6916	1.0716
O	1.9504	**0.4103**	2.7436	2.4722	1.9392	2.0510
I	1.7688	2.7250	**0.0367**	0.6816	2.5805	1.2906
E	1.4164	2.4559	0.6810	**0.0089**	2.3495	0.7132
U	1.7111	1.9530	2.5451	2.3750	**0.1949**	1.9782
Y	1.0302	1.9587	1.3022	0.7415	1.9207	**0.0690**

Taux de reconnaissance TR = 100 %

7.2.4- INFLUENCE DES PARAMETRES DECALAGE DES FORMANTS ET LONGUEUR DES MOTS COMBINES

Dans la réalité les paramètres cités précédement sont toujours combinés simultanément, c'est pour cette raison qu'on a mis au point les deux tests qui suivent:

- **Modèle de sinus**
- **Cepstrale**

' FN1TsT_' X

Ref \ Test	A	O	I	E	U	Y
A	9.4257	12.9068	10.3649	11.6896	10.8535	9.9321
O	8.6808	**9.8615**	10.2906	10.7971	**8.2051**	10.2038
I	8.6972	11.3780	**8.5442**	10.1788	9.5910	9.5588
E	10.8090	15.2087	——	10.2689	11.0766	9.3166
U	9.3029	11.8307	11.0135	**9.2721**	8.3307	10.0481
Y	**8.5433**	12.0954	10.5334	9.8153	10.0101	**8.9750**

Taux de reconnaissance TR = 50 %

' FN2TsT_' X

Ref \ Test	A	O	I	E	U	Y
A	**7.6411**	9.3374	8.2573	11.5219	10.4615	10.7866
O	8.7872	**7.1679**	8.3587	8.3587	9.1208	9.7132
I	8.1897	8.3815	**7.5661**	9.6302	9.5445	**9.4810**
E	10.1738	10.3093	10.6538	10.1459	11.2725	10.2502
U	8.2409	7.4072	9.9294	**8.2635**	**8.0682**	10.9833
Y	9.2453	8.4779	9.5975	10.0435	9.0056	11.2686

Taux de reconnaissance TR = 67 %

- **Modèle de sinus**
- LPC

' SLFE1TsX'

Ref \ Test	A	O	I	E	U	Y
A	8.3139	12.4548	12.5143	9.7220	12.9651	11.2863
O	7.4089	**3.7210**	8.6659	**8.0038**	4.1223	8.3323
I	**7.0965**	9.1251	3.6548	9.3635	8.9212	3.5258
E	7.4419	6.0511	____	8.0128	4.1850	8.5258
U	7.6317	5.5002	8.6375	8.9864	**3.0814**	9.2543
Y	7.2530	10.1722	**2.8313**	9.9731	9.4368	**3.1105**

Taux de reconnaissance TR = 50 %

'LFN2TsTX . DAT '

Ref \ Test	A	O	I	E	U	Y
A	**7.3781**	11.7911	12.4460	10.0842	12.8126	11.8271
O	9.0479	**3.1622**	10.0870	4.4963	5.4103	9.0086
I	8.2325	8.1781	**2.8250**	7.5796	9.9660	3.9235
E	7.7874	4.3302	10.1927	**3.9587**	5.5901	7.9062
U	8.6850	3.4982	10.3245	4.2636	**4.4757**	8.3068
Y	7.8701	8.6419	4.0611	7.5427	10.7360	**3.2356**

Taux de reconnaissance TR = 100 %

- **Modèle AR**
- **Cepstrale**

AFE1TsT

Ref \ Test	A	O	I	E	U	Y
A	**1.4241**	19.7793	19.0199	13.5284	17.3889	16.6231
O	10.8952	**9.4174**	11.6322	7.0550	9.5354	9.5025
I	16.8172	19.8446	**5.7295**	9.3565	8.6218	9.1016
E	13.4532	11.6995	____	**3.8539**	8.6729	7.6631
U	16.5584	14.1311	8.4084	8.9568	**1.4764**	8.1786
Y	15.1965	12.3579	9.6245	8.2901	8.9222	**3.6763**

AFE2TsT

Ref \ Test	A	O	I	E	U	Y
A	**2.0482**	12.9984	17.1777	14.1689	16.9222	16.0022
O	10.9941	**2.6507**	10.7001	7.1256	10.2172	8.6949
I	16.5866	10.7001	**4.1548**	9.1970	8.6193	9.2327
E	13.6439	7.5687	9.1395	**1.6525**	8.7217	8.5132
U	16.0979	10.8307	8.1308	8.9968	**0.2803**	7.6612
Y	15.1662	8.9803	9.3753	7.5650	7.8403	**2.1330**

Taux de reconnaissance TR = 100 %

- **Modèle AR**
- LPC

ALFE1TsX

Ref \ Test	A	O	I	E	U	Y
A	**0.1601**	1.7925	1.7613	1.5983	1.8713	1.2685
O	2.1155	**0.8920**	2.6791	2.3778	1.6101	2.0055
I	1.7639	2.8061	**0.0744**	0.7451	2.6333	1.3153
E	1.3872	2.4564	———	**0.1864**	2.4896	0.7440
U	1.8260	2.1929	2.5274	2.3157	**0.3426**	1.9017
Y	1.0349	1.9863	1.2692	0.9045	2.1249	**0.2310**

ALFE2TsX

Ref \ Test	A	O	I	E	U	Y
A	**0.5646**	1.7829	1.7810	1.6077	1.8690	1.4366
O	1.9832	**0.4836**	2.7349	2.5068	1.8865	1.9648
I	1.7987	2.7062	**0.0793**	0.7796	2.5874	1.3826
E	1.4311	2.3958	0.6875	**0.1379**	2.4496	0.8882
U	1.7989	2.0624	2.5260	2.4819	**0.0967**	1.9344
Y	1.0828	1.9089	1.3014	0.9425	2.0307	**0.1962**

Taux de reconnaissance TR = 100 %

7.2.5- CONCLUSIONS

Les tests effectués sur les différents paramètres avec le modèle des sinus sont nettement plus sensibles que celles avec le modèle AR.

Le taux de reconnaissance dans certain test qu'on juge faible (50 %) au vu de la dimension du dictionnaire avec le modèle des sinusoïdes pour la génération des échantillons ne peut pas à notre sens consister un indice de faiblesse des algorithmes puissants d'analyse, et normalisation; par contre plusieurs paramètres peuvent être retenu comme participant à cette anomalie, dont notamment le modèle réellement insuffisant que nous avons retenu pour simuler les voyelles, c'est un modèle très simple, qui plus est, ne tient pas compte de l'aspect voisé des voyelles.

En effet pour les même tests et avec le modèle AR on a pu arriver à un taux de reconnaissance meilleurs.

Le taux de reconnaissance obtenu à travers les différents tests que nous avons mis au point peuvent être jugés satisfaisant si l'on considère le modèle AR comme modèle de génération des échantillons.

De toute façon pour tirer des conclusions objectives, il aurait fallu travailler avec des échantillons de la parole réelle et des tests adéquats basés sur des faits expérimentaux.

Vu les résultats obtenus en utilise la méthode Lpc et Modèle AR et pour fichier Test " AIFE2TSX " et "ALFE1TSX " respectivement

7.3- TESTS EN MODE MULTILOCUTEUR

LES TESTS EN MODE MULTILOCUTEUR ONT PORTE SEULEMENT SUR L'INFLUENCE DU NOMBRE DE REPRESENTANTS PAR VOYELLE SUR LE TAUX DE RECONNAISSANCE

7.3.1- NEUF REPRESENTANTS (9)

R	T / A	R	T / O	R	T / I	R	T / E	R	T / U	R	T / Y
A5	0.5580	O5	0.6319	I5	**0.0812**	E5	0.1716	U5	0.1424	Y5	0.2189
A6	0.5563	O6	0.6873	I6	0.1116	E6	0.1770	U6	0.1844	Y6	0.2092
A7	0.5609	O7	0.5881	I7	0.0967	E7	0.1728	U7	0.0307	Y7	0.2087
A8	**0.3036**	O8	**0.2928**	I8	0.0852	E8	**0.1423**	U8	**0.0168**	Y8	**0.2001**
A9	0.5409	O9	0.3226	I9	0.1357	E9	0.1818	U9	0.0505	Y9	0.2355
Y5	1.1045	U5	1.9757	E5	0.6830	I5	0.7745	A5	1.7960	E5	0.9022
Y6	1.0808	U6	2.0407	E6	0.7001	I6	0.8152	A6	1.7915	E6	0.9023
Y7	1.0839	U7	1.8663	E7	0.6999	I7	0.8124	A7	1.7847	E7	0.9020
Y8	1.0813	U8	1.8607	E8	0.6804	I8	0.7750	A8	1.8216	E8	0.9033
Y9	1.0826	U9	1.8697	E9	0.6953	I9	0.8048	A9	1.8045	E9	0.9012
E5	1.4390	A5	1.8970	Y5	1.3048	Y5	0.9294	O5	1.9194	A5	1.2325
E6	1.4409	A6	1.8962	Y6	1.2937	Y6	0.9328	O6	1.9213	A6	1.2280
E7	1.4397	A7	1.9448	Y7	1.2950	Y7	0.9352	O7	1.9431	A7	1.2285
E8	1.4366	A8	1.9241	Y8	1.3022	Y8	0.9401	O8	1.7335	A8	1.2175
E9	1.4400	A9	1.8981	Y9	1.2887	Y9	0.9340	O9	1.8404	A9	0.9392
U5	1.8036	Y5	1.8735	A5	1.8491	A5	1.6145	Y5	1.9981	I5	1.3756
U6	2.0589	Y6	1.9284	A6	1.8600	A6	1.6190	Y6	2.0560	I6	1.3453
U7	1.7405	Y7	1.9192	A7	1.9200	A7	1.6415	Y7	2.0460	I7	1.4187
U8	1.7414	Y8	1.9213	A8	1.8052	A8	1.5717	Y8	2.0446	I8	1.3774
U9	1.7396	Y9	1.9319	A9	1.8208	A9	1.5720	Y9	2.0481	I9	1.3415

7.3.2- SIX REPRESENTANTS

R	T/A	R	T/O	R	T/I	R	T/E	R	T/U	R	T/Y
A4	**0.5551**	O4	**0.5403**	I4	0.1206	E4	0.1843	U4	**0.0402**	Y4	0.2192
A5	0.5580	O5	0.6319	I5	**0.0812**	E5	**0.1716**	U5	0.1424	Y5	0.2189
A6	0.5563	O6	0.6873	I6	0.1116	E6	0.1770	U6	0.1844	Y6	**0.2092**
Y4	1.1011	U4	1.8789	E4	0.6967	I4	0.8071	A4	1.8158	E4	0.9092
Y5	1.1045	U5	1.9757	E5	0.6830	I5	0.7745	A5	1.7960	E5	0.9022
Y6	1.0808	U6	2.0407	E6	0.7001	I6	0.8152	A6	1.7915	E6	0.9023
E4	1.4434	A4	1.8829	Y4	1.3178	Y4	0.9394	O4	1.7483	A4	0.9557
E5	1.4390	A5	1.8970	Y5	1.3048	Y5	0.9294	O5	1.9194	A5	1.2325
E6	1.4409	A6	1.8962	Y6	1.2937	Y6	0.9328	O6	1.9213	A6	1.2280
U4	1.7523	Y4	1.9058	A4	1.7943	A4	1.5559	Y4	2.0286	I4	1.3441
U5	1.8036	Y5	1.8735	A5	1.8491	A5	1.6145	Y5	1.9981	I5	1.3756
U6	2.0589	Y6	1.9284	A6	1.8600	A6	1.6190	Y6	2.0560	I6	1.3453

7.3.3- QUATRE REPRESENTANTS (4)

R	T/A	R	T/O	R	T/I	R	T/E	R	T/U	R	T/Y
A3	**0.5545**	O3	0.5938	I3	**0.1099**	E3	**0.1654**	O3	0.0667	Y3	**0.1912**
A4	0.5551	O4	**0.5403**	I4	0.1206	E4	0.1843	O4	**0.0402**	Y4	0.2192
Y3	1.0793	U3	1.8970	E3	0.6941	I3	0.7865	A3	1.8164	E3	0.8993
Y4	1.1011	U4	1.8789	E4	0.6967	I4	0.8071	A4	1.8158	E4	0.9092
E3	1.4388	A3	1.8865	Y3	1.3018	Y3	0.9394	O3	1.9203	A3	1.2328
E4	1.4434	A4	1.8829	Y4	1.3178	Y4	0.9394	O4	1.7483	A4	0.9557
U3	1.7668	Y3	1.9179	A3	1.8001	A3	1.5649	Y3	2.0410	I3	1.3690
U4	1.7523	Y4	1.9058	A4	1.7943	A4	1.5559	Y4	2.0286	I4	1.3441

LE TAUX DE RECONNAISSANCE = 100%

NEUF REPRESENTANTS

R	T / A	R	T / O	R	T / I	R	T / E	R	T / U	R	T / Y
A5	0.2771	O5	0.6883	I5	0.1511	E5	0.1816	U5	0.3009	Y5	0.1918
A6	0.2845	O6	0.7235	I6	0.2695	E6	0.2158	U6	0.4996	Y6	0.2198
A7	0.3221	O7	0.6454	I7	0.1558	E7	0.2194	U7	0.2635	Y7	0.2094
A8	0.2700	O8	0.3659	I8	0.1517	E8	0.1691	U8	0.2438	Y8	0.2325
A9	0.2553	O9	0.4304	I9	0.2522	E9	0.2080	U9	0.2807	Y9	0.1848
Y5	1.0556	A5	1.7828	Y5	1.2729	I5	0.7117	O5	1.6519	E5	0.7603
Y6	1.0346	A6	1.7762	Y6	1.2557	I6	0.7214	O6	1.6853	E6	0.7615
Y7	1.0352	A7	1.7398	Y7	1.2582	I7	0.7759	O7	1.7658	E7	0.7602
Y8	1.0347	A8	1.7127	Y8	1.2651	I8	0.7123	O8	1.7821	E8	0.7393
Y9	1.0387	A9	1.7007	Y9	1.2520	I9	0.7128	O9	1.7002	E9	0.7603
E5	1.3913	Y5	1.9586	A5	1.3048	Y5	0.8825	A5	1.7248	A5	1.2307
E6	1.3901	Y6	2.0053	A6	1.2937	Y6	0.9028	A6	1.7246	A6	1.2348
E7	1.3896	Y7	1.9959	A7	1.2950	Y7	0.8994	A7	1.7782	A7	1.2618
E8	1.3860	Y8	1.9986	A8	1.3022	Y8	0.9081	A8	1.8065	A8	1.2454
E9	1.3897	Y9	2.0091	A9	1.2887	Y9	0.8122	A9	1.8594	A9	1.2080
I5	1.7673	U5	2.0920	U5	2.6141	A5	1.5063	Y5	2.0940	I5	1.2996
I6	1 7975	U6	2.1588	U6	2.6891	A6	1.5095	Y6	2.1484	I6	1.3181
I7	1.7753	U7	2.0318	U7	2.5667	A7	1.5352	Y7	2.1318	I7	1.3412
I8	1.7684	U8	1.9899	U8	2.5687	A8	1.5201	Y8	2.1380	I8	1.3017
I9	1.7915	U9	2.0239	U9	2.5661	A9	1.5070	Y9	2.1373	I9	1.3111

SIX REPRESENTANTS

R	T / A	R	T / O	R	T / I	R	T / E	R	T / U	R	T / Y
A4	0.2295	O4	0 3804	I4	0.2493	E4	0.2095	U4	0.2678	Y4	0.2230
A6	0.2771	O5	0.6883	I5	0.1511	E5	0.1816	U5	0.3009	Y5	0.1618
A7	0.2845	O6	0.7235	I6	0.2695	E6	0.2158	U6	0.4996	Y6	0.2198
Y4	1.0499	A4	1.7241	Y4	1.2851	I4	0.7145	O4	1.7864	E4	0.7674
Y5	1.0346	A5	1.7828	Y5	1.2729	I5	0.7111	O5	1.6519	E5	0.7603
Y6	1.0346	A6	0.7235	Y6	1.2557	I6	0.7214	O6	1.6863	E6	0.7615
E4	1.3939	Y4	1.9824	A4	1.8107	Y4	0.8996	A4	1.8712	A4	1.1915
E5	1.3913	Y5	1.9586	A5	1.8300	Y5	0.8825	A5	1.7248	A5	1.2996
E6	1.7975	Y6	2.0053	A6	1.8400	Y6	0.9028	A6	1.7246	A6	1.2348
I4	1.7912	U4	2.0343	U4	2.5742	A4	1.4915	Y4	2.1218	I4	1.3111
I5	1.7673	U5	2.0920	U5	2.6141	A5	1.5053	Y5	2.0940	I5	1.2996
I6	1.7912	U6	2.1588	U6	2.6891	A6	1.5095	Y6	2.1484	I6	1.3181

QUATRE REPRESENTANTS

R	T / A	R	T / O	R	T / I	R	T / E	R	T / U	R	T / Y
A3	0.2307	O3	0.6318	I3	0.2022	E3	0.2092	U3	0.2554	Y3	0.2413
A4	0.2295	O4	0.3804	I4	0.2493	E4	0.2095	U6	0.2678	Y4	0.2230
Y3	1.0323	A3	1.7148	Y3	1.2657	I3	0.7042	O3	1.7227	E3	0.7576
Y4	1.0499	A4	1.7241	Y6	1.2557	I4	0.7145	O4	1.7864	E4	0.7674
E3	1.3899	Y3	2.0078	A3	1.8129	Y3	0.9071	A3	1.8696	A3	1.2212
E4	1.3939	Y4	1.9824	A4	1.8107	Y4	0.8996	A4	1.8712	A4	1.1915
I3	1.7746	U3	2.0512	U3	2.5776	A3	1.4986	Y3	2.1348	I3	1.3028
I4	1.7912	U4	2.0343	U4	2.5742	A4	1.4915	Y4	2.1218	I4	1.3111

CHAPITRE 8

PERSPECTIVE D'AVENIR ET POSSIBILITES ACTUELLES

PERSPECTIVE D'AVENIR ET POSSIBILITES ACTUELLES

La plupart des système de reconnaissance utilisés aujourd'hui, emploient pour la phase de décodage une méthode statistique. Cette dernière est à base de modèles de MARCOV caches. Ces modèles ont eu beaucoup de succès grâce aux taux de reconnaissance élevés obtenus aujourd'hui.

Cependant ce taux de reconnaissance chute lorsque le système est mis en exploitation (locuteur peu coopératif, bruit ambiant, etc.). Que faut-il faire dans ce cas ? Faut-il imposer à l'utilisateur d'un tel système un cahier de recommandations à suivre du style: éloignez votre chien, ne toussez pas, ne criez pas, ne vous énerver pas, etc.
A coup sûr le système de reconnaissance réalisé sera mis aux oubliettes.

Une autre approche est d'améliorer les performances du système de reconnaissance en évaluant l'apport de la détection bruit/parole, de la reconnaissance dans en environnement bruité, de la recherche des N meilleures solutions, etc.

8.1- LES METHODES DE RECHERCHE DES N MEILLEURES SOLUTIONS [8]

Sont essentiellement apparues, en reconnaissance automatique de la parole, vers la fin des années 80 et début des années 90.
Le critère de validité de chaque méthode est l'optimalité.

Une méthode est optimale si les N solutions fournies sont les N meilleures (i.e. s'il n'y a pas d'oubli).

Les N solutions obtenues sont des solutions syntaxiquement différentes et non pas des alignements différents d'une même solution.

8.1.1-Description de la méthode

La fig.8.1 présente le post-traitement segmental appliqué aux N meilleures solutions proposées par le module markovien. A la sortie de ce module markovien, on récupère le score, l'alignement et la séquence de mots de chaque solution. Pour chaque solution, l'alignement fournit un découpage en segments de parole.

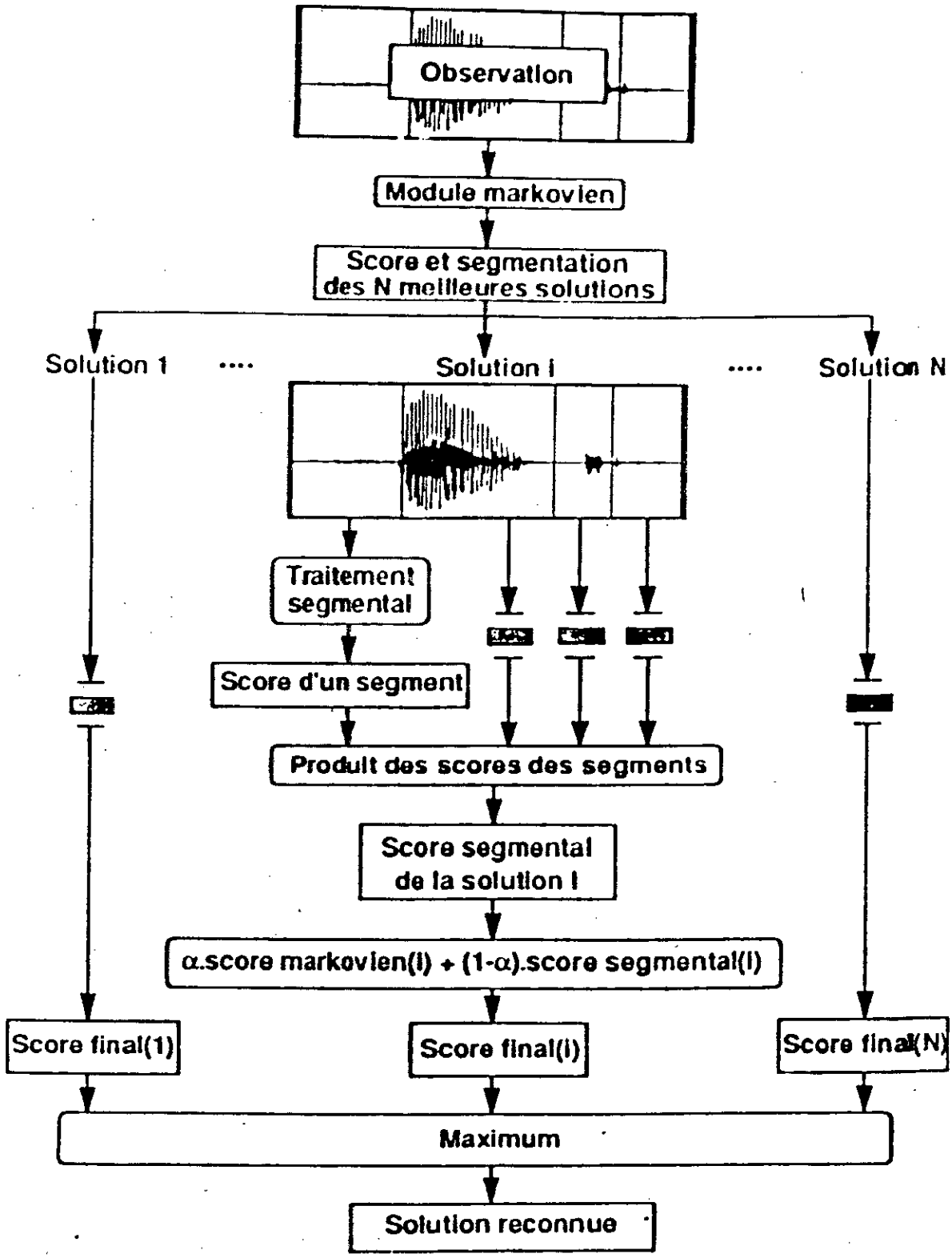


fig 8.1 Recherche des *N* meilleures solutions et post-traitement segmental statistique

Un segment de parole est défini comme un ensemble de trames acoustiques et identifié par une étiquette. Afin de calculer des scores segmentaux pour chaque solution, un modèle statistique est associé à chaque étiquette. Ainsi on parle de post-traitement segmental statistique.

Le traitement segmental consiste à calculer un score de vraisemblance pour chaque segment de chacune des solutions. Le score global d'une solution est égal au produit des scores des segments la constituant. Le score final d'une solution est obtenu par une combinaison linéaire (α,1-α), de son score segmental et de son score markovien.
La solution reconnue parmi les N développés est celle ayant le score le plus élevé.

8.1.2- Description de l'observation

Pour une observation inconnue X, on développe les N meilleures solutions. L'alignement obtenu pour chaque solution par le module markovien définit les différents segments utilisés pour décrire la solution.

La fig.8.2 donne un exemple des 3 meilleures solutions pour le mot prononcé « 5 » et illustre les différents segments mis en jeu pour chaque solution. L'observation correspondant à la $k^{ème}$ meilleure solution est notée X_{ik}

Ainsi, on remarque qu'on n'a plus une observation unique mais N observations correspondant aux meilleures solutions fournies par le premier module. Afin de ne pas privilégier une segmentation plus qu'une autre, on définit l'observation globale X comme étant la concaténation des observation associées à chacune des solutions proposées:

$$X = \{X_{i1}, X_{i2}, X_{i3},....,X_{iN}\}$$

où N est le nombre de solutions développées.

8.2- QUELQUES SYSTEMES DE RECONNAISSANCE

Ce paragraphe présente quelques systèmes de reconnaissance de la parole. Ceux utilisés aujourd'hui sont pour la plupart à base de HMM.

BYLOS est un système de reconnaissance de parole continue dépendant du locuteur. Il est conçu pour un très large vocabulaire. Ce système est à base de modèles de MARCOV cachés.

MOZART a été développé pour la reconnaissance de petits vocabulaires (quelques centaines de mots prononcés de manière isolée ou connectée).
Depuis et afin d'aborder la reconnaissance de plus grand vocabulaires (de l'ordre de plusieurs milliers de mots) un nouveau système de reconnaissance a vu le jour ,

AMADEUS ce système a été conçu pour la reconnaissance de 5000 mots en mode isolé et environ 300 mots en mode connecté.

Depuis un système de reconnaissance a vu le jour. Ce système fonctionne en mode indépendant du locuteur utilise une approche globale statistique et comme unité de reconnaissance le phonème en contexte.

SERAFINE (CNET) est un système de reconnaissance fonctionnant en mode monolocuteur et permettant la reconnaissance d'un petit vocabulaire, 100 mots en mode isolé et des courtes phrases (maximum 10 mots).

L'algorithme utilisé pour la phase de reconnaissance est la DTW. Par la suite un nouveau logiciel, appelé PHIL86 , a été élaboré pour permettre la reconnaissance en mode indépendant du locuteur en utilisant l'approche par modèles de Markov cachés.

Des améliorations ont été apportées depuis sur les différents modules du système de reconnaissance PHIL86 pour donner naissance à PHIL90 .

SPICOS (Philips) est un système de reconnaissance de parole continue dépendant du locuteur. La taille du vocabulaire utilisé a voisine les 1000 mots.

Les modèles de MARCOV cachés sont utilisés dans la description acoustique des phonèmes. L'originalité de ce système réside dans l'utilisation des fonctions de probabilités de Laplace au lieu des fonctions de probabilités gaussiennes habituellement employées.

SPHINX II (CMU) est un système de reconnaissance de la parole continue indépendant du locuteur pour un vocabulaire de 1000 mots environ.

TANGORA (IBM) est un système de reconnaissance monolocuteur conçu pour des vocabulaires de très grande taille (20000 mots). Il est réalisé à partir d'une approche globale à base de modules de MARCOV cachés.

Le développement de cartes spécialisées a permis l'intégration d'un tel système sur un PC-AT pour un vocabulaire de 5000 à 20000 mots. La reconnaissance est réalisée en temps réel et chaque prononciation est effectué en mots isolés.

Des applications ont été développées avec le système TANGORA pour la reconnaissance de la parole automatique dictée.
Une adaptation du système de reconnaissance TANGORA à la langue française a été réalisée par l'équipe IBM-France .

PARSYFAL est un système de reconnaissance développé pour la langue française, par l'équipe du centre IBM-France. La parole est prononcée en syllabes isolées . Le système utilise des dictionnaires contenant de 10000 à 200000 mots afin de couvrir la totalité du vocabulaire.

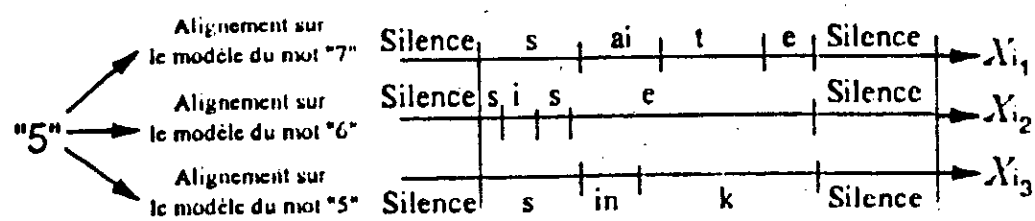


fig 8.2 Exemple d'alignements acoustiques obtenus en prononçant le mot "5".

CONCLUSION

CONCLUSION

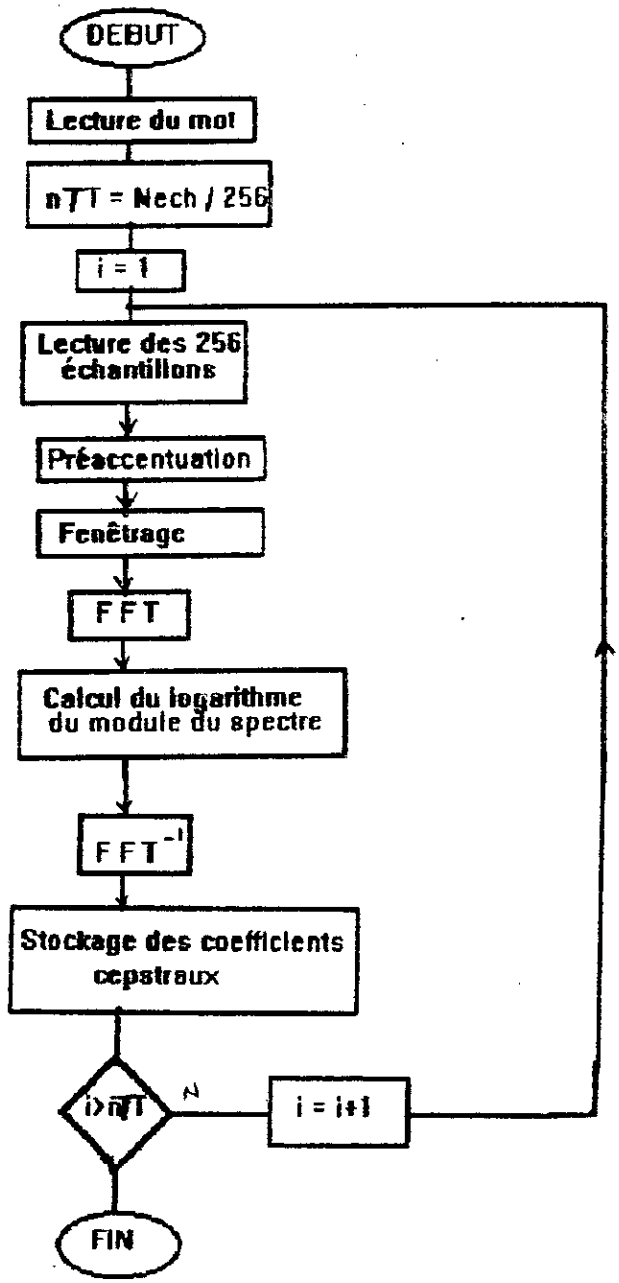
L'objectif initial de notre travail était au départ, l'étude d'un système de reconnaissance de la parole par les méthodes globales en mode multilocuteurs en vue d'une éventuelle implantation sur le microprocesseur TMS320. Cependant, nous étions donc forcés de nous limiter uniquement à la partie étude où nous avons mis au point des programmes opérationnels.

Avec l'analyse par prédiction linéaire et le modèle AR on a pu arriver à de bon résultats. Cette méthode est très intéressante vu qu'elle est programmable, aussi très rapide; d'ou l'augmentation de vitesse de fonctionnement.
Sa supériorité par rapport aux autres techniques d'analyse provient du fait qu'elle est fondée sur un modèle simple de production du signal vocal.
De plus c'est une méthode très appropriée à l'algorithme utilisé dans notre étude à savoir l'algorithme DTW (data time warping)

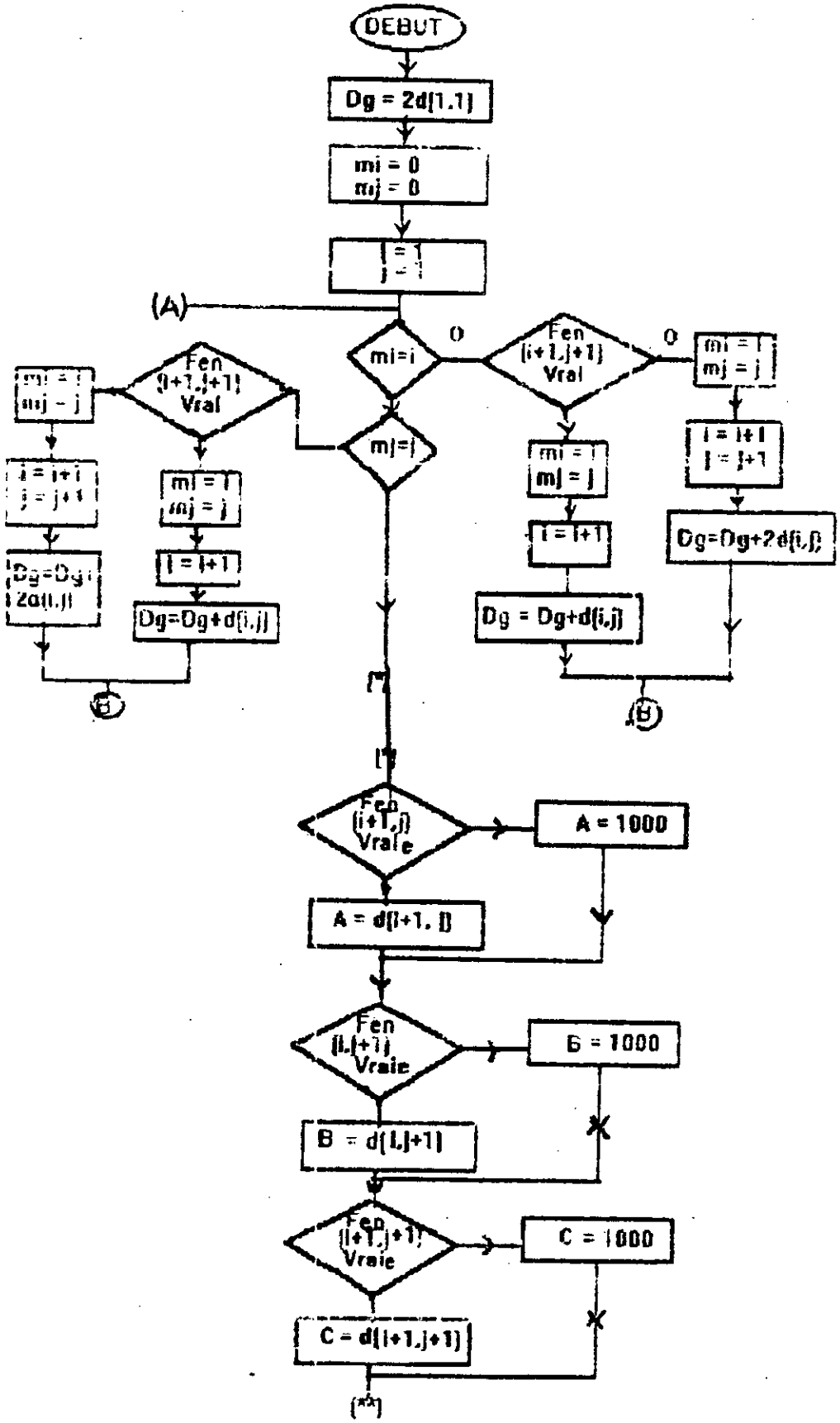
Par ailleurs, nous insistons sur le fait que la recherche sur reconnaissance de la parole est devenue, de nos jours, presque totalement expérimentale et exige beaucoup de moyen pour pouvoir arriver à un quelconque résultat. Des données théoriques relatives à la parole ne peuvent plus servir de base dés lors que les procédures actuelles des chercheurs, commencent par émettre des hypothèses puis procéder immédiatement à l'expérience pour confirmation ou infirmation.

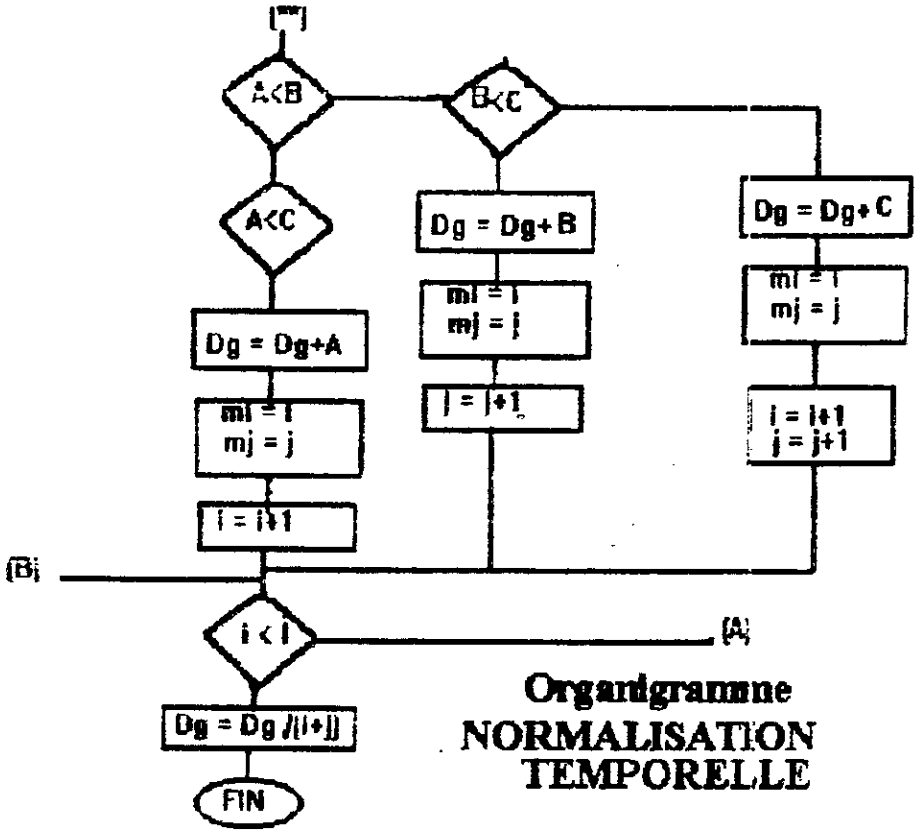
En effet une simulation du signal de la parole ne peut prendre en compte toutes les subtilités et toute la dynamique naturel qui caractérisent le signal acoustique.

ANNEXE

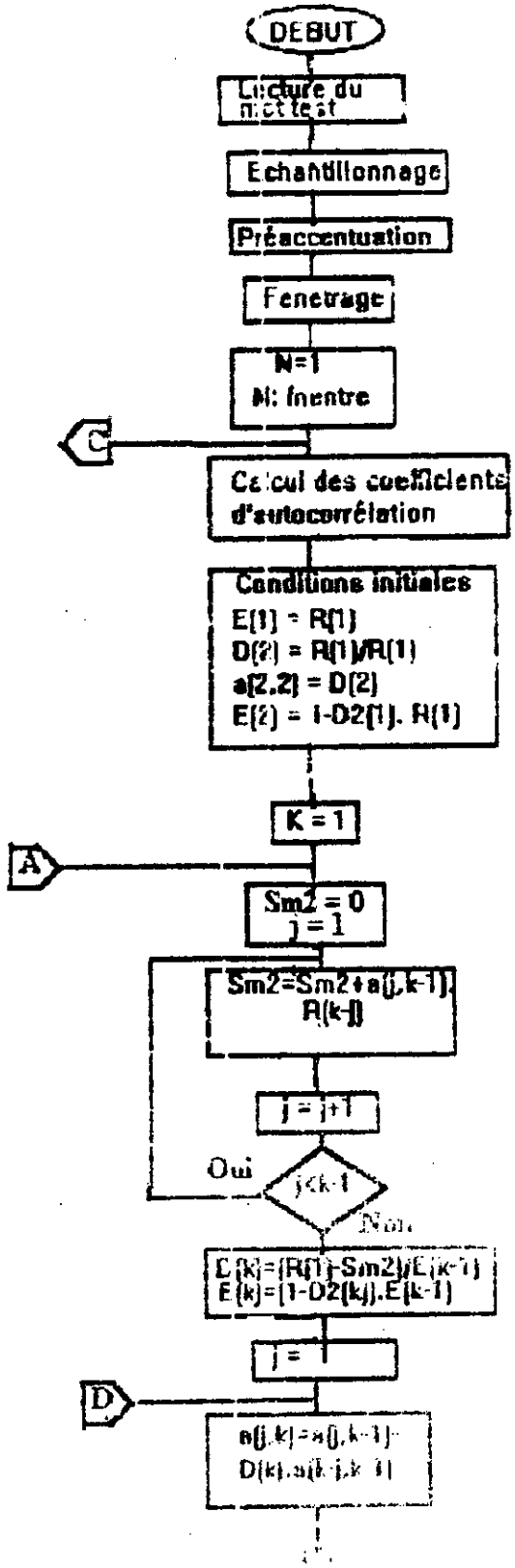


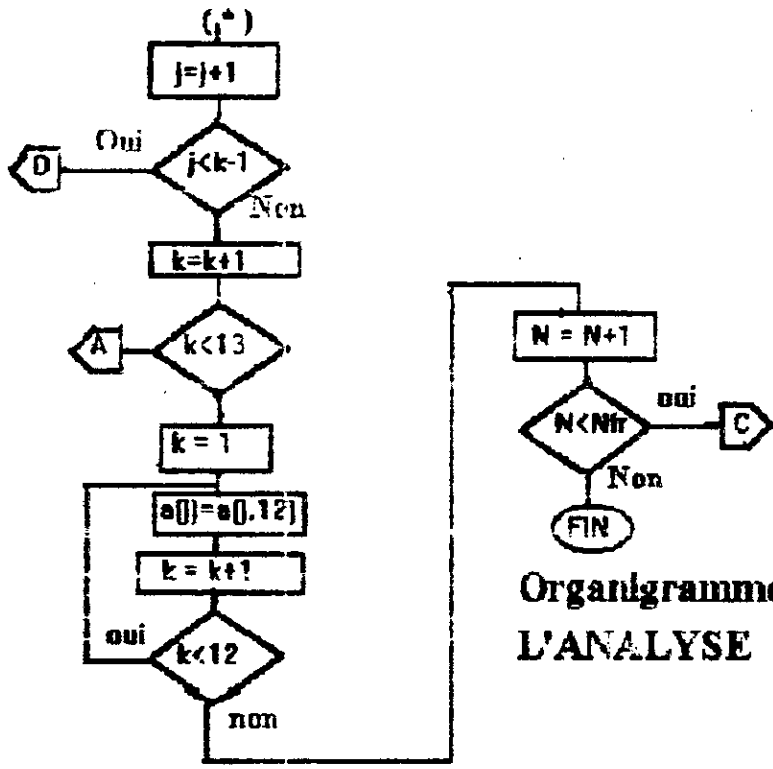
organigramme de
L'ANALYSE CEPSTRALE


DEBUT
Dg = 2d[1,1]
mi = 0
mj = 0
i = 1
j = 1
(A)
mi=i
O
Fen
[i+1,j+1]
Vrai
O
mi = i
mj = j
i = i+1
j = j+1
Dg=Dg+2d[i,j]
mj=j
Fen
[i+1,j+1]
Vrai
mi = i
mj = j
i = i+1
j = j+1
Dg=Dg+2d[i,j]
mi = i
mj = j
i = i+1
Dg = Dg+d[i,j]
mi = i
mj = j
i = i+1
Dg=Dg+d[i,j]
(B)
(B)
Fen
[i+1,j]
Vrai
A = 1000
A = d[i+1, j]
Fen
[i,j+1]
Vrai
B = 1000
B = d[i,j+1]
Fen
[i+1,j+1]
Vrai
C = 1000
C = d[i+1,j+1]
(**)

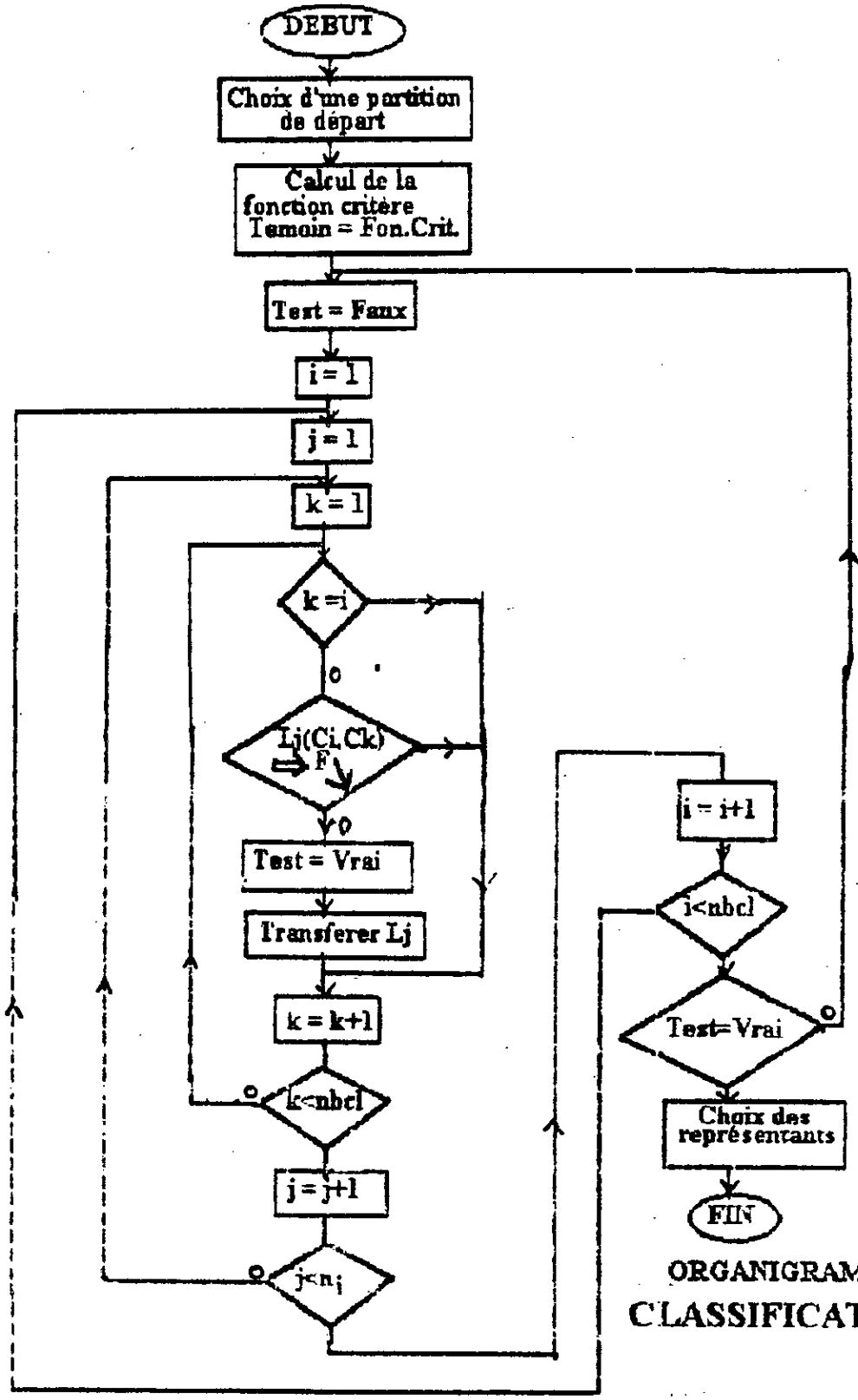


Organigramme
NORMALISATION TEMPORELLE


DEBUT
Lecture du mot test
Echantillonnage
Préaccentuation
Fenetrage
N=1
N: fnentre
C
Calcul des coefficients d'autocorrélation
Conditions initiales
E(1) = R(1)
D(2) = R(1)/R(1)
a[2,2] = D(2)
E(2) = 1-D2(1). R(1)
K = 1
A
Sm2 = 0
j = 1
Sm2=Sm2+a(j,k-1).R(k-j)
j = j+1
Oui
j<k-1
Non
D(k)=(R(1)-Sm2)/E(k-1)
E(k)=(1-D2(k)).E(k-1)
j = 1
D
a(j,k)=a(j,k-1)-D(k).a(k-j,k-1)

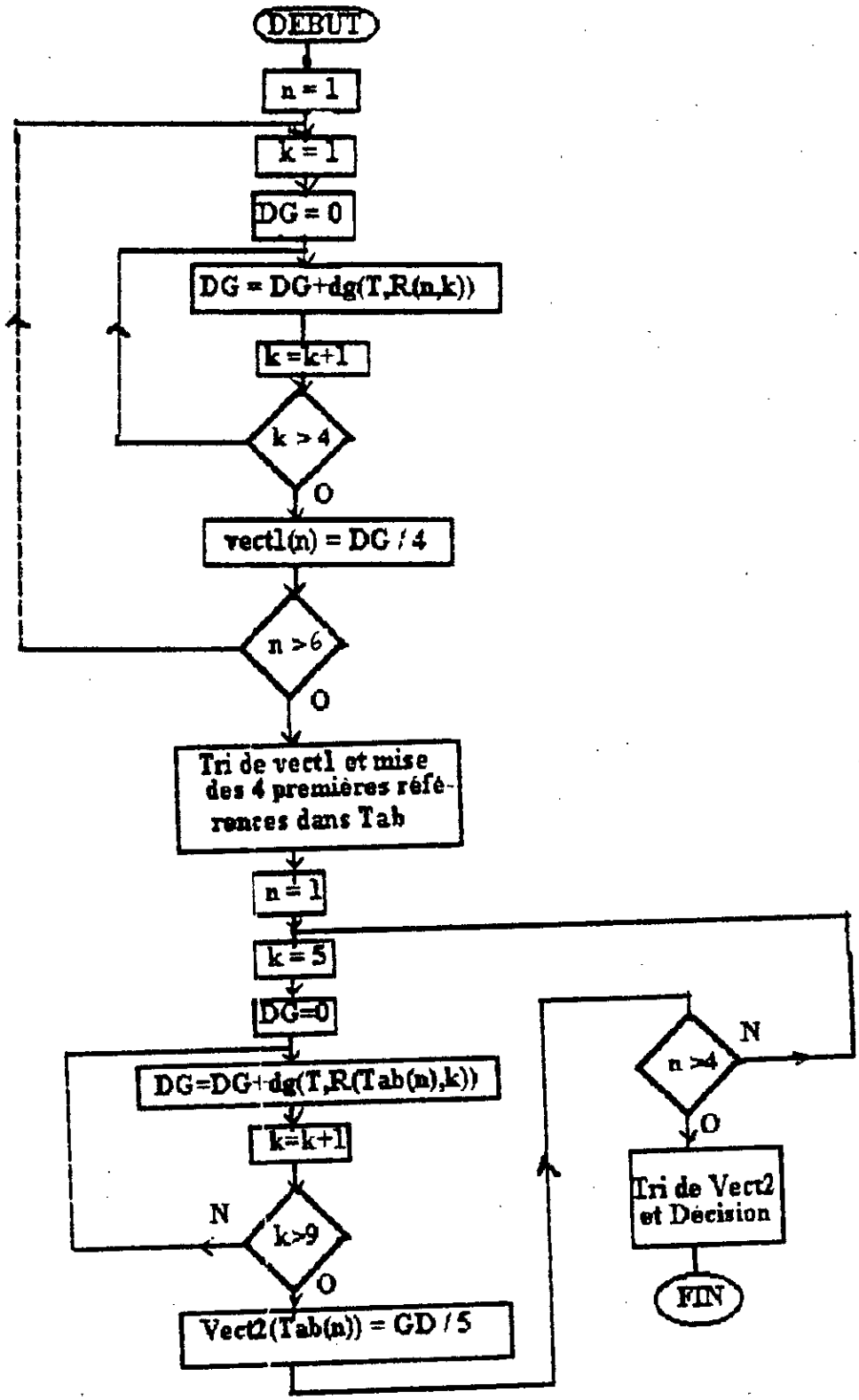


Organigramme de
L'ANALYSE LPC



ORGANIGRAMME

CLASSIFICATION



Organigramme DECISION

test Longueur de mots

	Test E1		Test E2	
Nech	A	4090	A	4097
	O	3600	O	3130
	I	4209	I	4280
	E	2700	U	2455
	U	3800	E	7647
	Y	3000	Y	3400

- Déplacement des formants

Test F1

	F1	F2
A	760	1360
O	365	740
I	255	2509
E	450	2230
U	240	590
Y	260	1800

Test F2

	F1	F2
A	754	1352
O	382	757
I	258	2500
E	400	2202
U	255	609
Y	264	1806

Décalage des formants et longueurs des mots combinés

Test FE 1

	F1	F2	Nech
A	755	1354	4600
O	380	770	4100
I	241	2489	4800
E	408	2220	3000
U	240	590	4000
Y	255	1809	3000

Test FE 2

	F1	F2	Nech
A	755	1362	4235
O	387	756	4225
I	251	2506	4291
E	401	2201	2643
U	250	601	3681
Y	255	1800	3653

BIBLIOGRAPHIE

[1] R. E BELLMAN : ' **Dynamic programming** ' ; princeton, N. J.,princeton University Press, 1957.

[2] M. BENYOUCEF : ' **Reconnaissance automatique de la parole pour la commande des systèmes** ' thèse de Magister, 1994.

[3] R. BOITE, M. KUNT: ' **Traitement de la parole** ' ; Presses Polytechniques Romandes, 1987.

[4] CALLIOPE : ' **Traitement automatique de la parole** ' ; Edition Masson, 1989.

[5] E. Emerit : ' **Cours de phonétique** ' ; SNED, 1977.

[6] J-P. HATON : ' **Reconnaissance automatique de la parole** ' ; Edition Dunod, 1991.

[7] M. KUNT : ' **Traitement Numerique des signaux** '; Edition Dunod, 1987.

[8] M.N. Lokbani : ' **Recherche des N Meilleures Solutions et Post-Traitements en Reconnaissance de la parole** ' ; Thèse de Doctorat, université de Paris XI Orsay, 1993.

[9] A. Menacer : ' **Reconnaissance de la parole en mode multilocuteur** ' ; Thèse de Docteur-ingenieur, université de Neutchatel, 1984.

[10] A. Mokkedem : ' **Reconnaissance de la parole en mode multilocuteur de mots isolés par les systèmes miniaturisés** ' ; Thèse de doctorat, université de Neutchatel, 1985.

[11] Revue « Les ingenieurs Supélec » : ' **Les réseaux neuronaux** ' ; Revue de la société des ingenieurs de l'école supérieure d'électricité.

[12] H. Sakoe ,S. Chiba : ' **Dynamic programming Algorithm optimization for word recognition** ' ; IEEE, ASSP, vol.26, N° 10 1978.