République Algérienne Démocratique et Populaire

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DE BLIDA
INSTITUT D'ELECTRONIQUE

MEMOIRE

présenté par :
DJAHID RAGGAI

EN VUE DE L'OBTENTION DU DIPLOME
DE MAGISTERE EN ELECTRONIQUE
OPTION : COMMUNICATION

THEME

RECONNAISSANCE AUTOMATIQUE DE LA PAROLE PAR LES MODELES DE MARKOV CACHES

devant le jury :

Mr H.SALHI	*maître de conférence (Univ. Blida)*	Président
Mr D. BERKANI	*professeur (ENP Alger)*	Examinateur
Mr M.BENSEBTI	*maître de conférence (Univ. Blida)*	Examinateur
Mr M.AIT AKKACHE	*maître assistant (Univ. Blida)*	Invité
Mr H.MELIANI	*maître de conférence (Univ. Blida)*	Rapporteur
Mr A.GUESSOUM	*professeur (univ.Blida)*	Rapporteur

BLIDA, ALGERIE SEPTEMBRE 2000

بسم الله الرحمن الرحيم

Remerciements

Toute ma gratitude et mes vifs remerciements vont à Mr H.MELIANI qui a dirigé ce travail et qui n'a jamais cessé de m'encourager et de me porter soutien .

Je remercie vivement Mr GUESSOUM pour avoir accepter de prendre la responsabilité d'encadrement et de finition de ce travail.

Mes remerciements vont également à Mr A.BENALLEL pour la mise en disponibilité du logiciel de traitement de signal « ICHARA ».

Mes sincères remerciements à : Mr D.BERKANI , Mr H.SALHI , Mr M.AIT AKKACHE et Mr M.BENSEBTI qui m'ont fait l'honneur d'être membres du jury d'évaluation de ce travail.

Que Mr M.AIT AKKACHE trouve ici L'expression de mes sentiments les plus respectueux pour l'aide qu'il m'a apportée pour la documentation.

Qu'il me soit permis d'exprimer ici toute ma reconnaissance à mes collègues pour leur soutien et leur encouragements.

D.RAGGAI

Dédicaces

Ce travail est dédié à :

Mon père et ma mère (que dieu ait pitié de son ame)

Mes frères et ma sœur,

Toute ma famille,

Tous mes amis,

D. Raggai

RESUME

Ce travail porte sur la reconnaissance automatique de mots isolés en mode mono locuteur en utilisant les modèles de Markov cachés(HMM). Les paramètres des HMMs sont obtenus après présentation d'un corpus d'apprentissage. La reconnaissance a été testée sur les chiffres de un à neuf prononcés en langue Française.

ABSTRACT

In this work, Hidden Markov Models(HMM) are used to perform the recognition of isolated words. After training, the HMMs are used to recognise the french digits from one to nine .

SOMMAIRE

Chapitre IV: RESULTATS DE RECONNAISSANCE

'Il découvrit bientôt plusieurs livres d'images , des lectures enfantines et un gros dictionnaire . Il les examina avec la plus grande attention . Les images surtout, frappaient son imagination, mais les petits insectes bizarres qui couraient sur les pages l'intriguaient au plus haut point.

Accroupi sur la table, Tarzan, l'enfant-singe, le petit sauvage au corps nu et bruni, sa longue chevelure noire tombant autour d'un visage bien fait qu'éclairaient deux yeux vifs et intelligents se penchait avec un interet évident sur le livre que tenaient ces mains fines et musclées. C'était un touchant tableau , une image allégorique de l'homme tatonnant vers la lumière de la science à travers la nuit de l'ignorance. Son petit visage est tendu par l'effort, car il venait de découvrir une piste, encore obscure et incertaine, qui devait lui permettre de percer le mystère des petits insectes .(. ..). Peu à peu, il progressait ; c'était une tache ardue et de longue halène, dans laquelle il venait de se jeter à l'aveuglette, une tache qui pourrait nous sembler impossible : apprendre à lire sans avoir la moindre notion de langage écrit et de lettre, ni meme avoir la moindre idée que de telles choses puissent exister !

Bien sur il ne réussit pas en un jour, ni une semaine ni meme une année ; c'est seulement lentement, très lentement qu'il comprit ce qu'étaient ces petits insectes, après avoir étudié patiemment toutes leurs possibilités. Vers l'age de quinze ans, il connaissait toutes les combinaisons de lettres qui figuraient au bas des dessins du petit manuel de lecture et de deux ou trois livres d'images.(. . .)

Vers l'age de dix-sept ans, il était parfaitement capable de lire le livre de lectures enfantines et les merveilleux petits insectes avaient perdu leur mystère à ses yeux.(. . .) .

Il commencait à savoir lire'

E. R. Burroughs
Tarzan le seigneur de la jungle
Extrait du chapitre 7 : 'la lumière de la science'

INTRODUCTION GENERALE

Les recherches intensives dans le domaine de la parole durent depuis plusieurs décennies. Qu'est ce qui motive ces recherches ?
Les linguistes sont les premiers à s'être intéressés à la génération de la parole, aux modes articulatoires combinés de tous les organes du conduit vocal, aux sons appelés phonèmes pouvant suffire à la production d'une langue, aux signaux sonores perçus comme signaux pertinents et renfermant de l'information, à la structure des langues à partir du vocabulaire les composant, langues tant écrites que parlées, l'historique des langues, de leurs influences mutuelles dans leurs évolutions.

Dans les années cinquante, l'avènement de nouveaux moyens de calculs incite les chercheurs à les utiliser pour le traitement numérique de la parole. Cet intérêt ne s'est guerre démenti ni infléchi.Bien au contraire, des méthodes nouvelles de traitement du signal en général et du signal de parole en particulier ne cessent de voir le jour.

Depuis l'arrivée des ordinateurs donc, la parole est sortie du domaine de la linguistique pour être explorée par tous les scientifiques, acousticiens, psycho-acousticiens, physiciens(mécanique des fluides), informaticiens .....et devenir ainsi du ressort(non exclusif bien sur) de la science dite exacte.

Le champ de recherche sur la parole est vaste. Il exige des connaissances très variées, multidisciplinaires, condamnant les chercheurs de divers horizons à conjuguer leurs efforts. Le succès est à ces conditions.

A l'intérêt purement scientifique manifesté depuis le début par les universitaires, s'est ajouté l'intérêt économique de dizaines d'entreprises qui commercialisent actuellement des circuits intégrés spécialisés, des cartes de reconnaissance et de synthèse à usage général....et enfin l'intérêt stratégique puisque les militaires voient tout le parti qui peut en être tiré, et cela a mené à l'un des projets les ambitieux qui soient : le projet **A.R.P.A / S.U.R (Advanced Research Project Agency / Speech Understanding Research)** de 15 millions de dollars de financement [1] , lancé par le ministère de la défense Américain en 1971 avec pour objectif la réalisation d'un système ayant les caractéristiques suivantes :

- Compréhension de la parole continue.
- Multilocuteur.
- Vocabulaire de 1000 mots avec syntaxe.
- Moins de 10% d'erreurs de compréhension(de sémantique).
- Réponse en temps réel avec un calculateur de 300 millions d'instruction par seconde.

Les applications premières du traitement de la parole couvre les trois domaines suivants :

1. **Le codage**

- Réduction de débit pour de longues transmissions par un traitement adapté du signal.
- Paramétrisation de la parole avant soit une synthèse soit une reconnaissance.
- Traitement du signal. Par exemple, rendre intelligible la parole distordue par des atmosphères inhabituelles en composition et en pression comme pour les plongeurs en mer profonde et pour les astronautes .

2. **La reconnaissance**

- Adaptation de la machine à l'homme : maillon initial de la chaîne parlée dans le dialogue homme/machine. La parole est un moyen de communication naturel pour l'homme.
- Transmission orale d'ordres à la machine : dicter à une machine à écrire le courrier à taper.
- Communication multimodale : la commande orale s'ajoute à la panoplie des moyens de communication physiques comme les boutons va et vient, poussoirs, rotatifs, les touches sensibles ,les claviers .....

- Réduction de débit par transmission du code des mots reconnus, plutôt que le signal lui même.

3. **La synthèse**

- Module de sortie comme maillon final de la chaîne parlée dans le dialogue homme/machine.
- Synthèse à partir de texte (text to speech).
- Fournir la 'parole' aux muets.

L'ordre de présentation de ce mémoire a pour but de mettre en valeur ce qui nous intéresse, à cet effet nous avons opté pour la chronologie suivante :
Après un aperçu général sur la recherche dans le domaine du traitement de la parole on expose dans le premier chapitre les notions générales sur le signal de parole. Sont aussi exposées dans ce chapitre les approches les plus connues dans la reconnaissance automatique de la parole.
Le deuxième chapitre est consacré au traitement du signal de parole, les principales méthodes d'analyse du signal vocal (pour des fins de reconnaissance) y sont exposées .
L'objet de ce mémoire étant la reconnaissance de la parole par HMM, le troisième chapitre est entièrement consacré aux modèles de markov cachés et à leur utilisation dans la reconnaissance de la parole.
Le quatrième chapitre expose les résultats de reconnaissance obtenus par utilisation des HMMs. On terminera enfin par une conclusion générale.

Chapitre I

GENERALITES SUR LA PAROLE ET SA RECONNAISSANCE

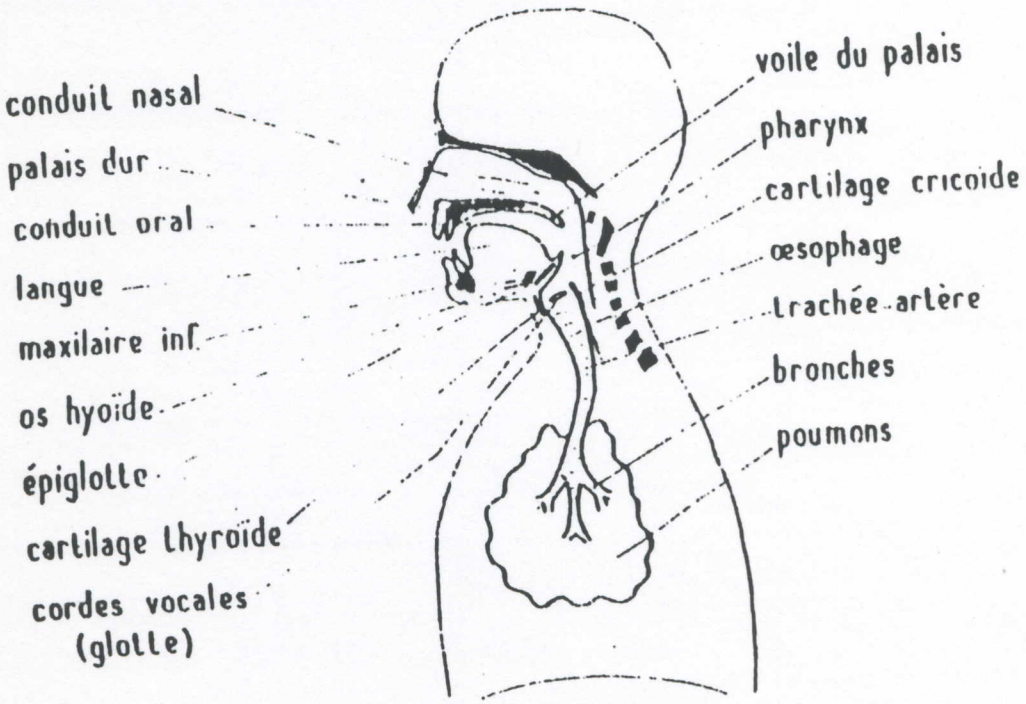
I-1-INTRODUCTION :

Dans ce chapitre on décrira brièvement les caractéristiques du signal vocal ainsi que les principales approches de reconnaissance de la parole.

I-2-L'APPAREIL PHONATOIRE :

La parole est un phénomène sonore produit par ce qu'on appelle l'appareil phonatoire - figure(1-1)-. L'énergie provient de l'air envoyé par les poumons ; les cordes vocales impriment des pulsations à l'air envoyé par les poumons si elles se mettent au travers de son chemin. On parle dans ce cas de sons voisés. Si les cordes vocales s'écartent , les sons sont non voisés.

Le conduit vocal est une suite de cavités qui servent de résonateurs et suivant leurs formes , il apparaît des résonances à des fréquences variables appelées formants. On distingue la cavité pharyngale , la cavité buccale et enfin la cavité nasale. Cette dernière est soit inutilisée soit elle se met en dérivation sur la cavité buccale par l'abaissement du vélum. La forme des cavités est affectée par les articulateurs tels les lèvres , le vélum , la mâchoire... Ces divers éléments combinés permettent plusieurs modes de production sonore et l'émission d'un ensemble de sons voisés.



Figure(1-1) : l'appareil phonatoire

I-3-PHYSIQUE DU SIGNAL DE PAROLE :[2]

Le signal de parole est d'abord un signal sonore, c'est donc une onde matérielle de faible puissance se déplaçant dans l'air.

Avant même que nous puissions le capter, le signal est sujet à des perturbations importantes : bruit ambiant, résonances, phénomènes d'écho, etc . Le signal de parole couvre quasiment toute l'étendue du spectre audible. En pratique on peut se limiter à la bande 50-5000 Hz.

Les mode de propagation du signal sonore dépendent aussi de la fréquence, les plus hautes (>2000 Hz pour un locuteur ordinaire) se déplacent de façon plus isotrope. On parle d'ondes planes et d'ondes sphériques bien que ces modèles représentent des cas limites. Ces problèmes sont bien connues des preneurs de son.

I-4-CONTENU DU SIGNAL DE PAROLE: [1]

Le signal de parole est un signal très complexe . Il contient une quantité importante d'informations imbriquées entre elles, ce qui rend difficile leurs extraction. La parole transmet l'information phonétique, une information sur le locuteur (homme, femme ou enfant), sur son état psychologique (joyeux , en colère ……), sur son état physique (respire la santé, fatigué, malade….).

Le spectre du signal de parole s'étend comme mentionné précédemment de quelques dizaines de Hz à plusieurs KHz. Les voyelles(A, I , O, …..), les liquides (L…..), les nasales(M , N…) et d'autres encore ont des spectres largement compris dans les 4 premiers KHz ; au delà leurs spectres sont négligeables. Par contre les fricatives (S , Z …….)et certaines occlusives ont des spectres étendus qui décroissent peu même au delà de 8 ou 10 KHz . Le débit phonétique de la parole est de 10 à 20 phonèmes maximum par seconde.

1-5- RECONNAISSANCE AUTOMATIQUE DE LA PAROLE(RAP) :

Le but de la (RAP) est de transcrire une suite de vecteurs acoustiques représentant le signal en une suite de symboles (lettres, mots ...), c'est donc le passage du signal au message . Les chercheurs espèrent par là, pouvoir un jour communiquer en langage naturel avec une machine(un ordinateur).

I-5-1- FACTEURS DE COMPLEXITE :

Les facteurs déterminants sont les suivants :

- Reconnaissance monolocuteur ou non.
- Reconnaissance de mots isolés ou de parole continue.
- Taille du vocabulaire ; à noter aussi que le choix du vocabulaire est un paramètre important, car on peut disposer de deux vocabulaires de même taille alors que la

reconnaissance sera plus difficile dans l'un que dans l'autre à cause du fait que les mots choisis sont phonétiquement voisins.

- Langage.
- Environnement :environnement acoustique, conditions de prise de son ...

On distingue en fait deux voix de recherche principales, l'une visant à résoudre le problème de reconnaissance de parole continue, multilocuteur, à grand vocabulaire; l'autre visant d'abord un sous problème moins difficile mais susceptible de déboucher rapidement sur des réalisations industrielles, la reconnaissance de mots isolés en nombre limité et pour un seul locuteur(c'est dans ce cadre que se place notre travail).

I-5-2- STRUCTURE D'UN SYSTEME DE (RAP) :

Un système de reconnaissance de la parole est habituellement constitué de trois composants indépendants :

- Un dispositif d'acquisition du signal, usuellement un microphone ou une bande magnétique, associés à un système d'amplification et de filtrage. Le signal est ensuite numérisé afin d'être traité.
- On pratique ensuite des prétraitements dont le but est d'une part de réduire le flot de données dans le système à un niveau acceptable, et d'autre part de présenter les données sous un format compatible avec l'étape de reconnaissance.
- La troisième étape consiste à extraire de ces données transformées une séquence de symboles(des mots ou des phonèmes par exemple). C'est le point le plus délicat des systèmes de reconnaissance.

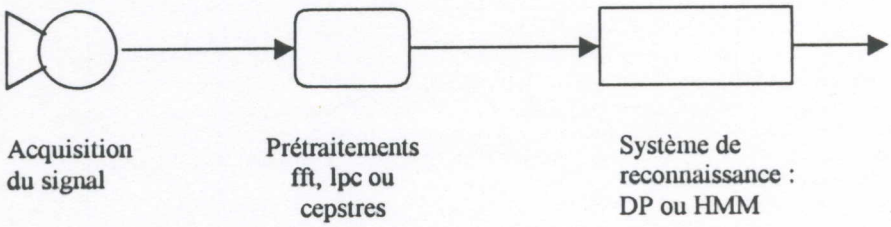


Fig. (1-2) : Une chaîne de reconnaissance de la parole

I-5-3-METHODES DE RECONNAISSANCE :

Il n y a qu'à voir le volume des travaux effectués dans ce domaine(voir par exemple [3] [4] [5] [6] [7] [8] .....) pour ce rendre compte de la difficulté de cerner un tel sujet, nous allons toutefois essayer d'exposer les principales approches connues dans la reconnaissance automatique de la parole.

Les système de reconnaissance peuvent être classés en deux catégories :systèmes globaux et systèmes analytiques, la reconnaissance d'une courte phrase peut être effectuée suivant l'une ou l'autre des deux méthodes :

Dans la méthode globale, la phrase à reconnaître est comparée dans son ensemble avec des phrases constitués par des mots du vocabulaire.

La méthode analytique au contraire , procède à une segmentation en unités de base (phonèmes , syllabes,...)et à leur identification; toutefois, la co-articulation a pour conséquence un taux d'erreurs important.

Parmi les techniques utilisées dans la RAP citons trois approches très connues :

- **RECONNAISSANCE PAR ALIGNEMENT TEMPOREL**:[2][9] (algorithme DTW)

La programmation dynamique (DTW : Dynamic Time Warping) permet de faire l'alignement temporel optimal entre le mot(signal) à reconnaître et les mots(signaux) de référence. Chaque signal après prétraitement se présente sous forme d'une séquence de vecteurs spectraux ; la comparaison entre deux mots revient donc à comparer deux séquences de vecteurs spectraux. On définit une métrique permettant de calculer la distance entre deux vecteurs spectraux(distance locale) ,un algorithme permet ensuite de calculer, à partir de distances locales, une distance globale(un coût) entre deux mots. On dispose d'une bibliothèque contenants les mots de référence . Lorsqu'un mot inconnu se présente, on calcule la distance entre ce mot et tous les mots du la bibliothèque. La décision se fera au profit du mot de la bibliothèque correspondant à la distance minimale. Le calcul de la distance se fait comme suit :

Soient deux mots $R(r_1 ..... r_J)$ et $T(t_1 .....t_I)$, avec r_j et t_i des vecteurs de R^p (espace des vecteurs de dimension p) .

La figure(1-3-a) montre la matrice constituée par les distances entre chaque vecteur de R et de T (distances locales), donc à chaque point (i,j) de cette matrice est associée une distance locale. L'ajustement optimal entre les deux mots est représenté par un chemin croissant dans la grille de la figure(1-3-b); le coût de ce chemin est la somme de toutes les distances locales

sur tout les points du chemin. On cherche donc un chemin (c) de coût optimal : $c_1(1,1)$ à $c_l(I,J)$ (L longueur du chemin) ne comportant que des transitions bien définies, par exemple :

$g(i,j) = d(t_i , r_j) + \min \{g(i,j-1) , g(i-1,j) , g(i-1 , j-1)\}$ avec g(i , j) coût optimal du chemin aboutissant à (i , j).

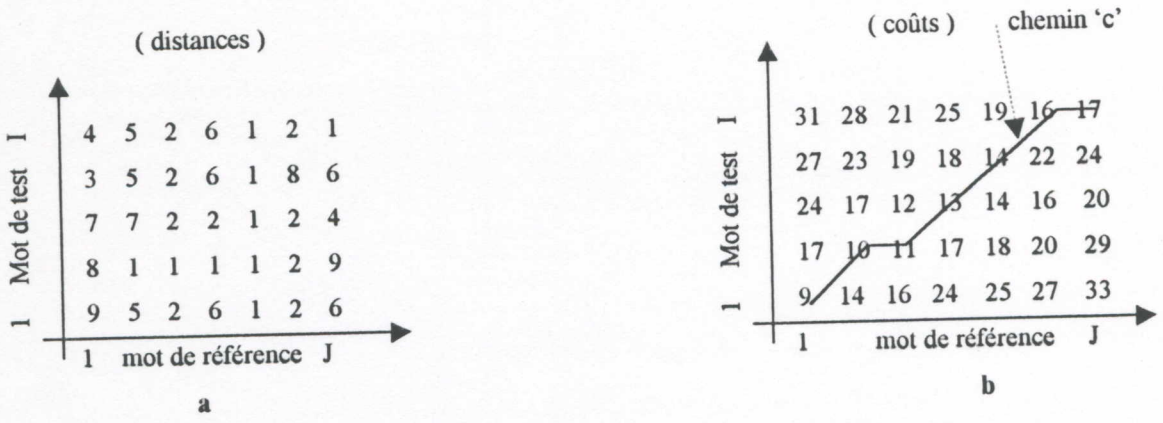
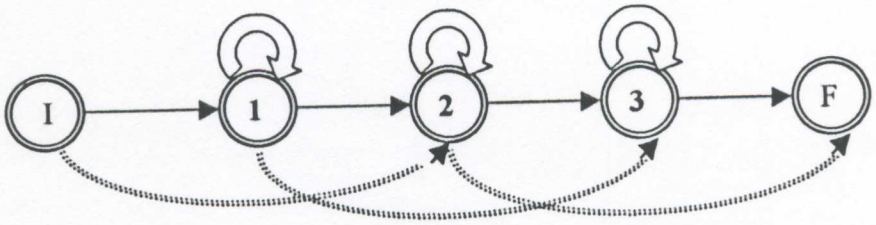


Fig.(1-3) : les coûts g(i,j) à droite sont calculés à partir du tableau des distances , 'c' représente le chemin d'appariement

- **APPROCHE STATISTIQUE :**(modèles de Markov cachés HMM)

Dans cette approche on essaie de modéliser la production d'une unité linguistique : phonème ,syllabe, mot ...(on utilisera dans ce qui suit le terme générique de **mot**).
Ainsi donc chaque mot du vocabulaire sera représenté par un modèle censé générer ce mot avec une certaine probabilité(cette probabilité est appelée vraisemblance du mot). Pour que ce modèle représente assez fidèlement le mot lui correspondant , il faut que la vraisemblance de ce mot (calculée dans ce modèle) soit plus élevée que celle des autres mots du vocabulaire, on parle ainsi du principe de maximum de vraisemblance(ML maximum likelihood). Chaque modèle est entraîné par plusieurs versions du même mot pour capter les statistiques de chaque mot du vocabulaire.
La figure (1-3) montre un exemple de modèle de Markov caché , les cercles représentent les états du modèle, et les flèches les transitions entre ces états. I et F représentent les état initial et final respectivement .Les états 1 , 2, et 3 sont des états émetteurs.
Les modèles de Markov cachés ont été utilisés avec succès dans la RAP, parmi les travaux effectués dans ce domaine citons [4] [5][6][7][8][10].



Figure(1-4) exemple de modèle HMM

- **INTELLIGENCE ARTIFICIELLE :**(les réseaux de neurones artificiels)

Les réseaux de neurones artificiels ont prouvé leur utilité dans la RAP ([3] [4] [12][13][14][15]) .
Il n'est absolument pas question de cerner ce sujet dans ces quelques lignes. Disons seulement que ces systèmes basés sur cette approche essaient d'imiter le comportement biologique de l'être humain capable **d'apprendre** à partir de son environnement. **L'apprentissage** est donc Un aspect capital dans les réseaux de neurones. A partir d'un ensemble d'exemples présentés à son entrée, le système apprendra à classer les mots à reconnaître en autant de classes qu'il y a de mots dans le vocabulaire. Enfin, il est intéressant de souligner le lien entre les réseaux de neurones et les modèles de Markov cachés. En effet les modèles de Markov cachés sont utilisés en conjonction avec les réseaux de neurones pour former ce qu'on appelle les HMMs discriminants [2][4][13][14][16].

CONCLUSION :

Ont été exposées dans ce chapitres les principales approches dans la reconnaissance de la parole. L'intérêt dans ce travail étant porté à l'approche statistique, nous allons exposer dans le troisième chapitre le formalisme mathématique nécessaire à l'application des HMMs dans la reconnaissance de la parole .Toutefois comme mentionné dans la section (I-V-2), avant de se présenter au niveau du système de reconnaissance, le signal doit subir des prétraitements. On se doit donc de parler de ces prétraitements et c'est la raison d'être du chapitre suivant.

Chapitre II

ANALYSE ACOUSTIQUE

II-1-INTRODUCTION :

L'analyse acoustique est une partie importante dans le traitement que subit le signal sonore pour pouvoir réaliser un système de reconnaissance de la parole. Il s'agit de tirer du signal de parole les paramètres pertinents susceptibles de représenter correctement toutes ses caractéristiques. L'analyse doit être suffisamment fine pour permettre de séparer les divers sons élémentaires dans l'espace de décision. L'analyse dépendra de l'usage auquel elle est destinée . En reconnaissance multilocuteur il est indispensable de normaliser tout ce qui pourrait nuire à l'identification des mots et ainsi pouvoir comparer ce qui est comparable, alors qu'en reconnaissance du locuteur toute caractéristique (du locuteur) reproductible peut servir d'information de reconnaissance qu'il faut éviter d'éliminer.

Les paramètres d'une analyse acoustique peuvent découler d'un traitement quelconque et simple qui ignore la nature du signal comme par exemple le calcul spectral.....La prise en compte de la nature du signal peut apporter un plus et permettre la découverte de paramètres optimaux qui renferment plus d'informations pertinentes avec un nombre réduit de coefficients. Il faut faire un tri dans toutes les informations disponibles et donner à chacune un poids suivant son importance.

La démarche historique a été initialement de considérer le signal de parole un signal comme un autre, puis des techniques ont été développées spécialement pour le signal de parole, comme la prédiction linéaire (coefficient LPC), l'analyse homomorphique (coefficients cepstraux) .....

De nouvelles techniques ne cessent de naître. Certaines cherchent à imiter le traitement que réalise par réflexe le corps humain car nous pouvons supposer dans un premier temps que la solution appliquée par la nature est idéale; comme par exemple l'échelle MEL en fréquence qui est plus performante que l'échelle linéaire usuellement utilisée.

II-2-ANALYSE ACOUSTIQUE :

Avant d'aborder les principales méthodes d'analyse du signal de parole, commençons par les prétraitements que constituent le fenêtrage et la préaccentuation.

II-2-1-Fenetrage :

La théorie du signal montre que les variables temps et fréquence sont conjuguées, il y a entre eux une relation d'incertitude du type :

$\Delta T \cdot \Delta F = 1/\pi$.....(2-1).

Où ΔT est la largeur de la fenêtre temporelle d'analyse d'un signal et ΔF le plus petit détail que nous puissions espérer observer dans son spectre. Autrement dit, l'analyse temporelle d'un signal sinusoïdal pur sur une fenêtre de largeur ΔT fournit un spectre de largeur ΔF au lieu d'une raie unique. Le fenêtrage que nous somme obligés d'effectuer sur le signal provoque deux inconvénients. Les raies s'élargissent et bavent d'une part, et d'autre part apparaissent des rebonds de part et d'autre des raies centrales dont les niveaux maximaux peuvent être importants ce qui fait qu'en pratique une résonance (un formant) peut porter son influence très loin du lieu où elle se trouve et perturber certaines zones peu énergétiques du spectre. Ceci explique la difficulté d'observer les antiformants dans les spectres des nasales. La forme de la fenêtre influence l'allure générale de la réponse spectrale; la figure (2-1) donne la forme des spectres pour un signal sinusoïdal pur analysé à travers deux types de fenêtres (rectangulaire, Hamming)

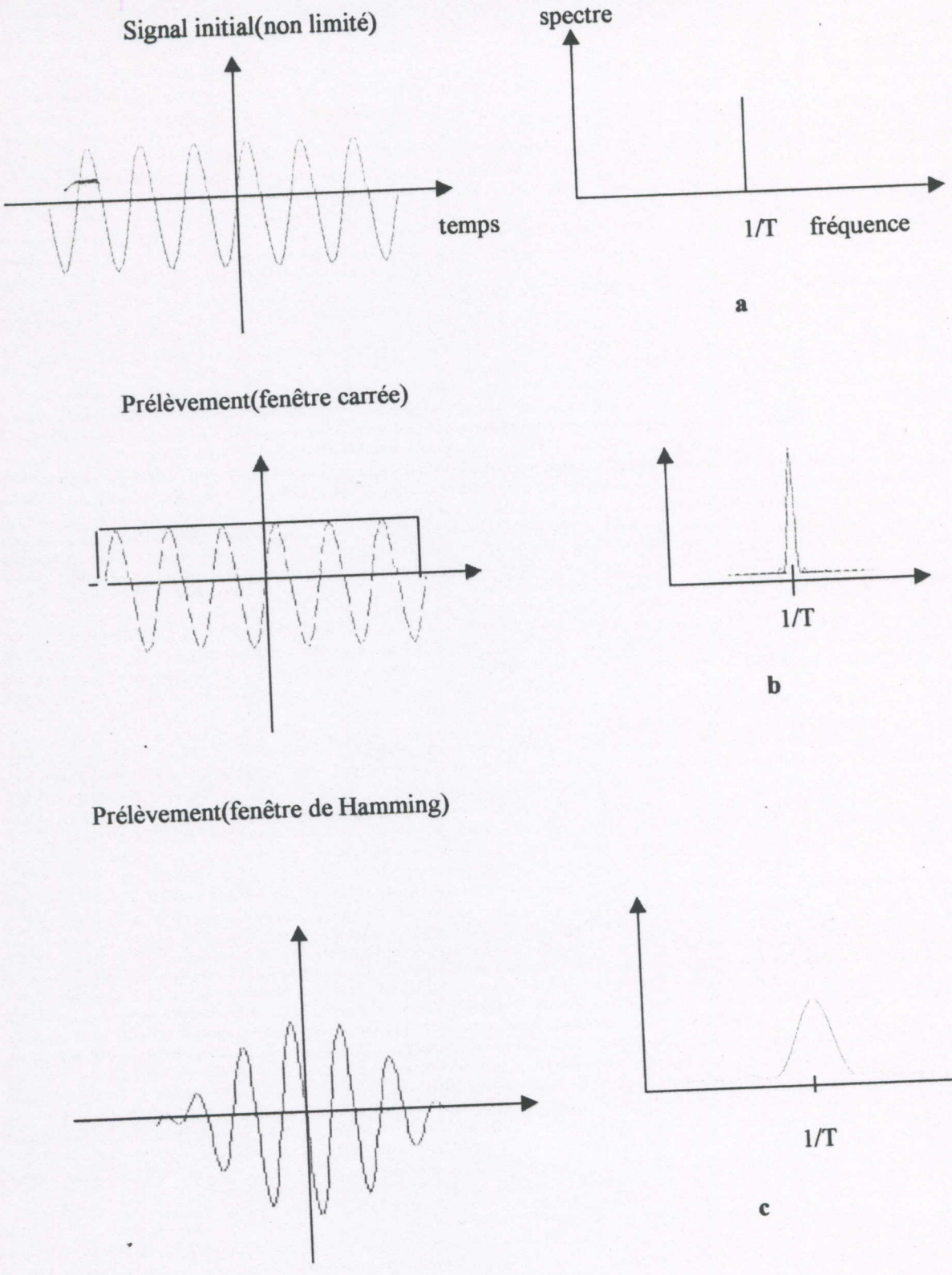


Figure (2-1) : fenêtrage d'un signal sinusoïdal par la fenêtre carrée et celle de Hamming

Une fenêtre rectangulaire donne la largeur du pic principal la plus faible possible mais provoque l'apparition de pics secondaires très énergétiques dont les maximas décroissent très lentement. La fenêtre de HAMMING (cas particulier de la fenêtre de HANNING), de forme sinusoïdale est préférée. Elle ne provoque l'apparition que de raies secondaires peu énergétiques dont les maximas décroissent vite. En contre partie, puisque nous ne pouvons gagner sur tous les tableaux, la largeur fréquentielle du pic principal augmente par rapport à celle de la fenêtre rectangulaire. Bien d'autres fenêtres sont envisageables mais nous portons notre choix sur la fenêtre de HAMMING dont l'équation est :

$$FEN(n) = \begin{cases} 0.54 - 0.46\cos(2.\pi(n+1/2)/N) \\ 0 \sin on \end{cases} \quad \text{avec } 0 \leq n \leq N-1 \ldots\ldots (2\text{-}1)$$

II-2-2-Préaccentuation du signal:

Le signal sonore diffuse de la bouche vers l'extérieur. Le son est la transmission d'une onde dans un milieu mécanique : l'air ou tout autre gaz.
Le milieu mécanique est définit par son impédance mécanique (inertie). En débouchant des lèvres l'onde doit attaquer un milieu nettement plus important que celui contenu dans le conduit vocal. Il y a désadaptation des impédances mécaniques au niveau des lèvres. Le rayonnement du son à l'extérieur s'accompagne d'une baisse d'énergie par unité de surface véhiculée par l'onde et fait plus important il s'accompagne aussi d'une distorsion qui peut être assimilée à une désaccentuation de 6dB par octave sur tout le spectre.

Nous le voyons la préaccentuation numérique du signal de 6dB par octave a pour but de rétablir le signal tel qu'il était avant de déboucher des lèvres. Soient S(n) les échantillons du signal et $S_a(n)$ ceux du signal préaccentué. Cette préaccentuation de 6dB par octave est très facilement réalisée sur les échantillons du signal S(n) par l'utilisation d'un filtre non récursif :

$S_a(n) = S(n) - S(n-1)$, $a = 0.9$. .....(2-2)

Pourquoi désirons nous rétablir le signal tel qu'il était avant de sortir de la bouche ? La question est très importante. Les chercheurs désirent par transformation inverse estimer l'évolution de la forme du conduit vocal et les lieux d'articulation et en déduire ainsi la parole prononcée. La transformation inverse demande la connaissance de tous les phénomènes secondaires qui peuvent intervenir et qu'il faut éliminer pour remonter à l'information principale : la forme du conduit vocal.

L'examen des spectres du signal sonore montre en général une décroissance du maximum des formants de rangs de plus en plus élevés. Le but de l'analyse acoustique du signal est de permettre de discriminer autant que possible les divers sons élémentaires. Il est possible pour cela d'appliquer une préaccentuation (supplémentaire) plus importante pour que les quelques premiers formants soient d'énergies comparables. Sans cela le premier formant se taille la part de lion bien qu'il soit peu discriminant en reconnaissance. Il n'est nul besoin de connaître tous les phénomènes et de remonter à leurs sources pour faire de bonnes reconnaissances. Il suffit de discriminer les divers classes de sons et tous les moyens sont bons pour y arriver même si les méthodes retenues sont purement empiriques.

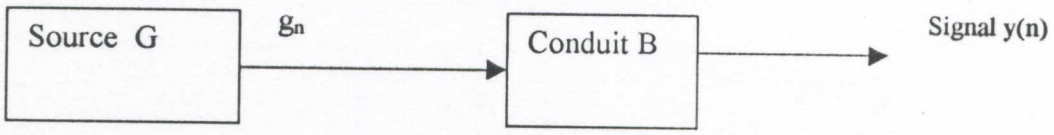
2-2-3-Méthodes d'analyse :

L'analyse a pour but la paramétrisation du signal vocal, plusieurs techniques d'analyse sont utilisées parmi lesquelles on cite :

-l'analyse par prédiction linéaire.
-l'analyse par banc de filtres.
-l'analyse homomorphique.
La technique utilisée dans ce travail est l'analyse homomorphique dont la description est donnée ci-après.

2-2-3-3-L'analyse homomorphique : [18]

Le défaut majeur de la FFT pour le calcul du spectre réside dans l'intermodulation source/conduit qui rend difficile la mesure des formants (figure (2-3)) et la mesure du fondamental caractéristiques précisément du conduit et de la source. Le lissage cepstral ou cepstre est une méthode qui vise à séparer leur contribution respective par déconvolution.
Pour cela on fait l'hypothèse que le signal vocal y(n) est produit par un signal excitateur g_n (source glottique) traversant un système linéaire passif de réponse impultionnelle b_n (conduit oral et nasal).



Figure(2-2) digramme bloc simplifié d'un système linéaire de production de la parole

On peut écrire :
$\forall n \rangle 0 : y(n) = g_n * b_n$ .....(2-3)
soit S(Z) = G(Z) . B(Z)..... (2-4)
Pour déconvoluer y(n) c'est à dire pour retrouver les composantes g_n et b_n il suffit de transposer le problème par homomorphisme dans un espace où l'opérateur de convolution (*) devient un opérateur additif (+) . Soit D_*^+ cet homomorphisme

$$y(n) = g_n * b_n \xrightarrow{D_*^+} \hat{y}(n) = \hat{g}_n + \hat{b}_n$$

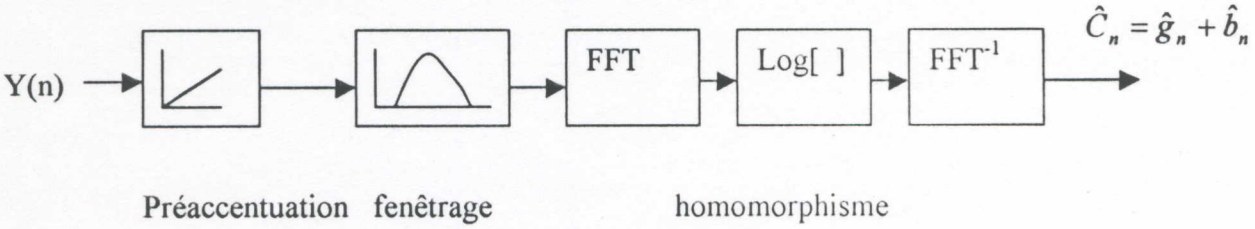
après séparation de $\hat{g}_n$ et $\hat{b}_n$, si la transformation inverse existe(D_+^*) on aura :

$$\hat{g}_n \xrightarrow{D_+^*} g_n$$
$$\hat{b}_n \xrightarrow{D_+^*} b_n$$

Les homomorphismes D_*^+ et D_+^* sont inverses l'un de l'autre et peuvent se définir par :

$$D_*^+ = Z(.) \circ \log(.) \circ Z^{-1}(.)$$
$$D_+^* = Z(.) \circ \exp(.) \circ Z^{-1}(.)$$

où Z et Z^{-1} désignent respectivement les transformées en Z directe et inverse, log et exp le logarithme et exponentielle complexes respectivement. Un problème surgit ici dans la mesure où $\log(re^{j\varphi}) = \log r + j\varphi$ et que φ n'est défini qu'à 2π près. Cette difficulté d'inversion du logarithme est levée dans le cas de la parole où l'on s'intéresse rarement à la phase. On considère donc dans les calculs le logarithme du module (LM) du spectre; le calcul des coefficients cepstraux revient donc à calculer la transformée en Z inverse du logarithme du module de spectre du signal. En pratique la TZ est avantageusement remplacée par la FFT qui possède les mêmes propriétés de linéarité. Le schéma suivant illustre les différentes étapes du traitement homomorphique.



Figure(2-5) traitement homomorphique

Les coefficients $\hat{C}_n$ ainsi obtenus appartiennent à un domaine pseudo-temporel réel appelé domaine quéfrenciel. Pour séparer ensuite la contribution de la source de celle du conduit vocal (celle ci représentant la nature du son émis) on s'appuie sur les deux hypothèses suivantes :

- $\hat{g}_n$ se réduit théoriquement à une séquence d'impulsions séparées par n_0 échantillons(n_0 période du pitch).
- $\hat{b}_n$ décroît rapidement (en 1/n) et devient rapidement négligeable du moins pour $n > n_0$ (ceci est surtout vrai pour les hommes pour lesquels $F_0 < 150$Hz.

Dans ces conditions on peut admettre que les premiers coefficients cepstraux contiennent essentiellement la contribution du conduit; sur ce point ils sont vraiment pratiques[1].
Les propriétés des coefficients cepstraux sont les suivantes :

1. Le cepstre est la transformée de Fourrier inverse du logarithme du module du spectre :

$$\hat{C}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|y(e^{j\omega})\right| e^{j\omega n} d\omega \quad \ldots\ldots\ldots\ldots\ldots\ldots (2-5)$$

cette relation montre que les coefficients cepstraux sont normalisés en énergie. La multiplication d'un signal par un nombre ne modifie en rien la valeur des cepstres.

2. La relation (2-11) montre aussi que Le premier coefficient $\hat{C}_0$ est une infirmation sur l'énergie du spectre.
3. Les coefficients cepstraux décroissent en 1/n, c'est pour cela que peu de coefficients suffisent en pratique.

D'autres types de paramètres acoustiques sont utilisés en reconnaissance automatique de la parole et qui sont en quelque sorte des dérivés des paramètres vus précédemment, des exemples de tels types de paramètres peuvent être trouvés dans [5].
Tous les types de paramètres vus précédemment sont extraits du signal vocal après que celui ci soit sorti de la bouche et qu'il ait traversé l'air avant d'être acquisitionné au niveau du microphone. Toutes ces considérations montrent que la pertinence des paramètres extraits dépend en grande partie des conditions de propagation et d'acquisition du signal. On comprend alors qu'un type de paramètres qui pourrait s'affranchir de ces conditions soit beaucoup plus robuste et plus représentatif du signal vocal avant sa sortie de la bouche. Dans [19]on trouve un exemple de tels types de paramètres, où la forme de la bouche lors de la prononciation d'un son (surface de l'ouverture par exemple) est prise comme paramètre acoustique. On disposera ainsi de deux types de paramètres (audio et visuels) rendant ainsi le système de reconnaissance plus robuste.

II-2-3-2-INDICES ET TRAITS PHONETIQUES :

L'expérience des chercheurs sur le signal de parole a permis de prendre en compte des paramètres globaux tels que le nombre de passages par zéro (ou un autre seuil) , l'énergie totale, l'énergie haute fréquence (au dessus d'un seuil donné), le rapport entre les énergies en basse et en haute fréquence .....L'ensemble de ces paramètres permet de construire des indices. Comme indice citons : l'indice de nasalité, le voisement, indice d'ouverture (la voyelle « A » est plus ouverte que la voyelle « I » par exemple).....
Des traits phonétiques sont estimés existants (avec une certaine probabilité) ou non par la confrontation de l'ensemble de ces indices. Parmi les traits phonétiques citons : les fricatives, les occlusives, les nasales, les liquides.....elles représentent une classification des sons de la parole en un nombre restreint de classes.

II-3-LA QUANTIFICATION VECTORIELLE :

Etant donné que ce travail porte sur les HMMs discrets, le signal à l'entrée du système de reconnaissance doit se présenter sous forme de vecteurs quantifiés. La technique permettant de réaliser cette quantification est appelée : quantification vectorielle.
La quantification vectorielle est une opération qui généralise la quantification scalaire. Elle concerne la représentation d'un vecteur x dont les composantes sont à valeurs réelles continues ($x \in R^N$) par un vecteur appartenant à un ensemble fini $\{ y_i \in R^N , i = 1 , 2, ...,L \}$.
Les vecteurs dits quantifiés sont en nombre fini; ils constituent un dictionnaire(code-book) de points dans R^N .
La quantification vectorielle doit être organisée pour minimiser la distorsion moyenne(moyenne des erreurs de quantification) pour un dictionnaire de taille L donnée.

En traitement de la parole, on peut effectuer la quantification vectorielle des formes d'ondes et la quantification vectorielle des vecteurs de paramètres. Dans le premier cas les composantes de chaque vecteur sont N échantillons consécutifs du signal, dans le second cas il s'agit d'un modèle autoregressif (vecteurs de paramètres ou vecteurs spectraux). C'est ce dernier cas qui nous intéresse dans ce travail.

Les applications de la quantification vectorielle sont essentiellement le codage à faible débit (moins de 8kbits/seconde) et la reconnaissance.
Par exemple en traitement de la parole, lorsqu'un ensemble de paramètres (représentant un vecteur) est utilisé pour représenter l'enveloppe spectrale d'un son; la quantification vectorielle peut être considérée comme une technique de reconnaissance de formes où la forme d'entrée est approximée par une des formes d'un ensemble prédéterminé de formes standards.

Dans quelques applications la quantification vectorielle est utilisée comme une étape préliminaire visant à réduire la complexité des calculs d'étapes ultérieurs dans un algorithme; c'est précisément le cas dans ce travail. En effet on verra ultérieurement que la quantification vectorielle permet de simplifier l'algorithme d'apprentissage des modèles de Markov cachés (HMM). La section qui suit présente le formalisme mathématique de la quantification vectorielle.

II-3-1-FORMULATION DU PROBLEME :

Les sections qui suivent s'appuient essentiellement sur les travaux de recherche [20][21][22][23].

Soit $x = [\, x_1 x_2 \ldots\ldots x_N \,]^T$ un vecteur de dimension N dont les composantes $\{ x_k \,,\ 1 \le k \le N \}$ sont des variables aléatoires réelles continues (l'exposant T dénote le transposé). La quantification vectorielle transforme le vecteur x en un vecteur y appartenant à un ensemble fini $Y = \{ y_i \in R^N \,,\ i = 1,\ldots\ldots,L \}$. On dit que x est quantifié en y, et que y est la valeur quantifiée de x. On écrit :

$y = q(x)$ où q (.) est l'opérateur de quantification. y est aussi appelé le vecteur de reconstruction ou bien le vecteur de sortie correspondant à x ou encore centroide. L'ensemble Y constitue le dictionnaire du quantificateur, dans la littérature de reconnaissance de formes les y_i sont aussi appelés formes de référence.

L est la taille du dictionnaire, appelé aussi nombre de niveaux, un terme emprunté de la terminologie de la quantification scalaire; on parle alors de dictionnaire à L niveaux ou bien de quantificateur à L niveaux.

Pour concevoir le dictionnaire on découpe l'espace de la variable aléatoire x en L régions ou cellules $\{ C_i \,,\ 1 \le i \le L \}$ et on associe à chaque cellule C_i un vecteur y_i. Le quantificateur assigne alors y_i à x si x se trouve dans C_i : $q(x) = y_i$, si $x \in C_i$.

La figure (2-7) montre une partition de l'espace bidimensionnel (N = 2) pour des fins de quantification vectorielle. La région entourée par des lignes en gras est la cellule C_i. Tout vecteur à l'intérieur de C_i est quantifié en y_i. les positions des autres centroides sont marquées par des points. Une cellule non limitée est dite surchargée et l'ensemble des cellules non limitées constitue la région de surcharge [22] .

Lorsque x est quantifié en y, il en résulte une erreur de quantification et une mesure de distorsion d(x,y) peut être définie entre x et y. d(x,y) est aussi considérée comme une mesure de dissemblance ou de distance .

Pour un ensemble de vecteurs quantifiés on définit la distorsion moyenne :

$$d_w(x,y) = (x-y)^T W (x-y)$$

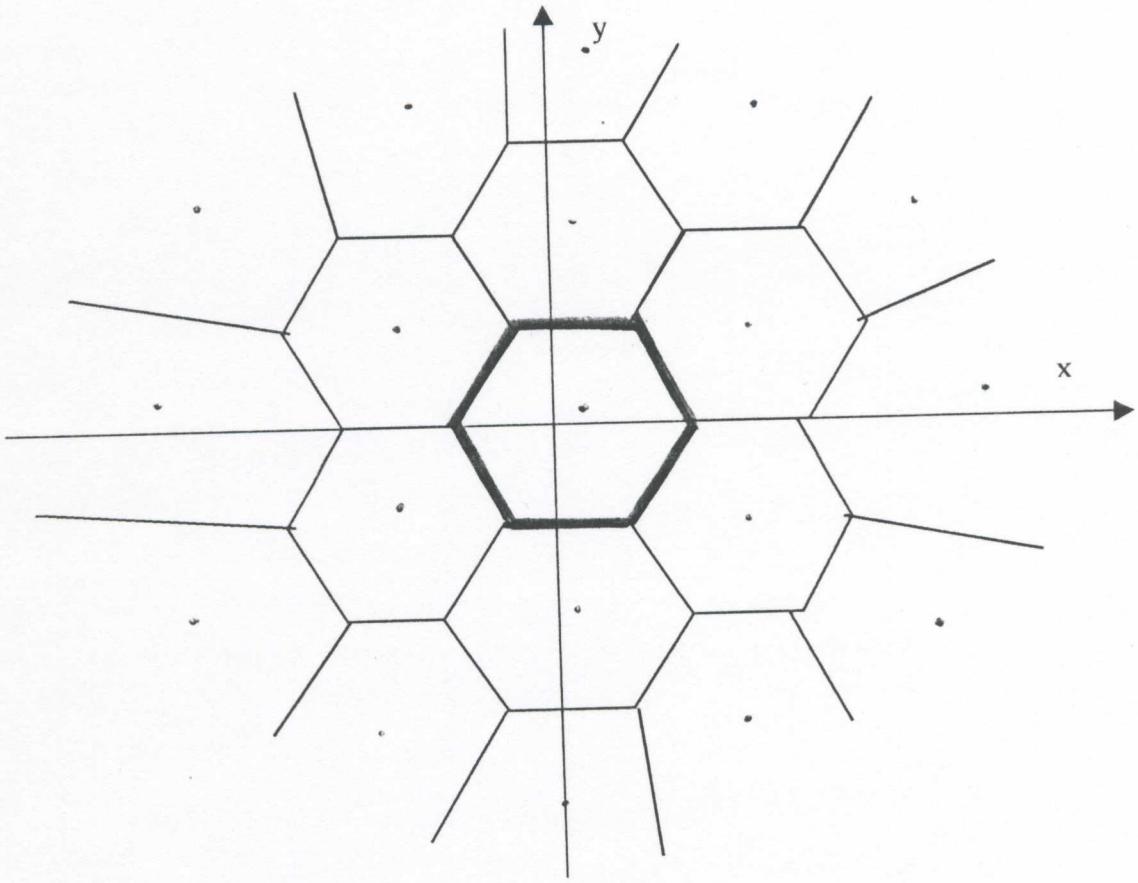
Si le processus de x(n) est stationnaire et ergodique (c-à-d que la moyenne temporelle est égale à la moyenne d'ensemble) alors la moyenne temporelle tend en limite vers l'espérance :

$$D = \zeta[d(x,y)]$$

$$= \sum_{i=1}^{L} P(x \in C_i)\, \zeta[d(x,y_i) / x \in C_i]$$

$$= \sum_{i=1}^{L} P(x \in C_i) \int_{x \in C_i} d(x,y_i) P(x) dx \ldots\ldots\ldots (2-6)$$

où $P(x \in C_i)$ est la probabilité pour que x appartienne à C_i; P(x) est la densité de probabilité multidimensionnelle de x et l'intégrale est prise sur toutes les composantes de x.



Figure(2-4) : exemple de partition d'un espace à deux dimensions

II-3-2-MESURES DE DISTORSION :[9][20][21]

Une définition particulière de la distance entre deux vecteurs spectraux doit être :

- Significative sur le plan acoustique.
- Formalisable d'une manière efficiente sur le plan mathématique.
- Définie dans un espace de paramètres judicieusement choisi.

Parmi les mesures de distorsion les plus utilisées on cite :
- L'erreur quadratique moyenne (EQM)
- L'erreur quadratique moyenne pondérée
- Mesure de distorsion de prédiction linéaire

Dans ce travail nous avons utilisé l'erreur quadratique moyenne dont l'expression ci-après :

$$d_2(x,y) = \frac{1}{N}[(x-y)(x-y)^T]^{1/2} = \frac{1}{N}[\sum_{k=1}^{N}(x_k - y_k)^2]^{1/2} \quad .....(2-7)$$

La popularité de l'EQM provient essentiellement de sa simplicité et sa maniabilité sur le plan mathématique.

II-3-3-CONCEPTION DU DICTIONNAIRE :

Comme mentionné précédemment, pour concevoir un dictionnaire à L niveaux on divise l'espace (de dimension N) en L cellules { C_i , $1 \le i \le L$ } et on associe à chaque cellule C_i un vecteur y_i (centroide). Pour un vecteur x donné, le quantificateur donne en sortie le vecteur y_i si x est dans C_i. Le quantificateur est dit optimal (distorsion minimale) si la distorsion dans (2-13) est minimisée sur tous les L niveaux du quantificateur.
Les deux conditions nécessaires d'optimalité sont :

- **Règle de sélection par plus proche voisin :**

$Q(x) = y_i$ si $d(x,y_i) \le d(x,y_j)$, $j \ne i$, $1 \le j \le L$ .....(2-9)

- **Condition du centroide :**

Les y_i sont choisis de telle sorte qu'ils minimisent :

$$D_i = \zeta[d(x,y_i)/x \in C_i] = \int_{x \in C_i} d(x,y)P(x)dx \quad .....(2\text{-}10)$$

y_i est appelé centroide de la cellule C_i et on écrit : $y_i = cent(C_i)$ ....(2-11).
Le calcul du centroide pour une région particulière dépend de la définition de la mesure de distorsion.
En pratique on dispose d'un ensemble de vecteurs d'apprentissage{ x(n) ,$1 \le n \le M$},un sous ensemble M_i de ces vecteurs appartiendra à la cellule C_i, la distorsion moyenne est alors donnée par :

$$D_i = \frac{1}{M_i} \sum_{x \in C_i} d(x,y_i) \quad .....(2\text{-}12).$$

Pour l'EQM et l'EQM pondérée D_i est minimisée par :

$$y_i = \frac{1}{M_i} \sum_{x \in C_i} x(n) \quad .....(2\text{-}13).$$

Dans ce cas y_i est la moyenne de tous les vecteurs dans C_i.

Une méthode itérative de construction du dictionnaire est connue dans la littérature de reconnaissance de formes sous la forme de l'algorithme K-Means. Le nom K-Means est due à MacQueen. Dans un article non publié, Lloyd a indépendamment développé en 1957 le même algorithme que celui de Forgy mais pour le problème de quantification scalaire, cet article ne fut publié que plus tard (en 1982). Linde, Buzo et Gray ont généralisé l'algorithme au cas vectoriel ce qui lui a value le nom de l'algorithme de Lloyd généralisé nommé aussi algorithme LBG [20].

Dans notre cas K=L (nombre de centroides). L'algorithme divise l'ensemble des vecteurs d'apprentissage { x(n) } en L clusters (cellules) Ci de manière à ce que les deux conditions nécessaires d'optimalité (citées précédemment) soient satisfaites. Dans ce qui suit

m dénotera l'indexe de l'itération et Ci(m) est le cluster 'i' à l'itération 'm' avec Yi comme centroide . L'algorithme est le suivant:

ALGORITHME K-MEANS :

Etape 1 :

initialisation $m = 0$, choisir convenablement un ensemble de centroides $y_i(0)$, $1 \leq i \leq L$.

Etape 2 : classification :

Classer l'ensemble des vecteurs d'apprentissage $\{ x(n), 1 \leq n \leq M \}$ dans les clusters C_i par la règle du plus proche voisin .

$x \in C_i(m)$, si $d[x , y_i(m)] \leq d[x , y_j(m)]$ pour tout $j \neq i$

Etape 3 : Mise à jour des centroides (y_i) :

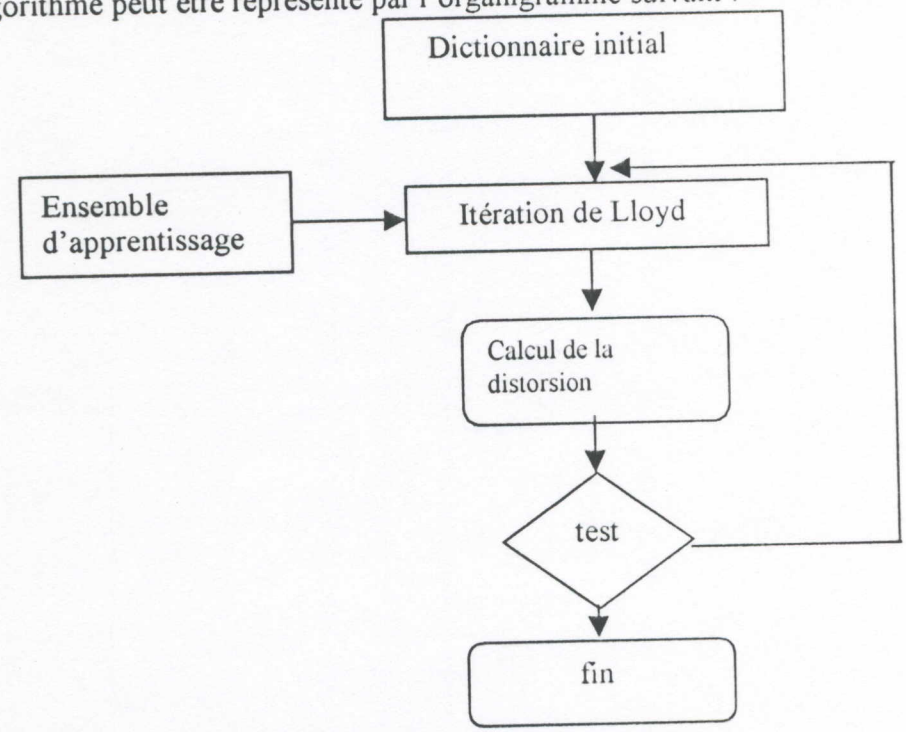
$m \leftarrow m+1$, mettre à jour les y_i de chaque cellule par calcul des centroides des vecteurs d'apprentissage dans chaque cellule

$y_i(m) = cent(C_i(m))$, $1 \leq i \leq L$.

Etape 4 : Test d'arrêt :

Si la diminution de la distorsion totale D(m) à l'itération 'm' relativement à D(m-1)est inférieure à un certain seuil alors STOP, sinon aller à l'étape 2.

L'algorithme peut être représenté par l'organigramme suivant :



Figure(2-8) : organigramme de l'algorithme de Lloyd

L'algorithme converge vers un optimum local qui dépend du choix du dictionnaire initial[19].

Pour avoir une solution acceptable plusieurs méthodes de choix du dictionnaire initial sont utilisées parmi lesquelles on cite:

- **QUANTIFICATION ALEATOIRE:**
  Cette approche très simple consiste à sélectionner les L vecteurs du dictionnaire initial de manière aléatoire, on pourra par exemple choisir les L premiers vecteurs de la séquence

D'apprentissage. si les données on choisira des vecteurs dispersés, disons chaque $K^{ième}$ vecteur d'apprentissage.

- **Elimination :**(Prunning)

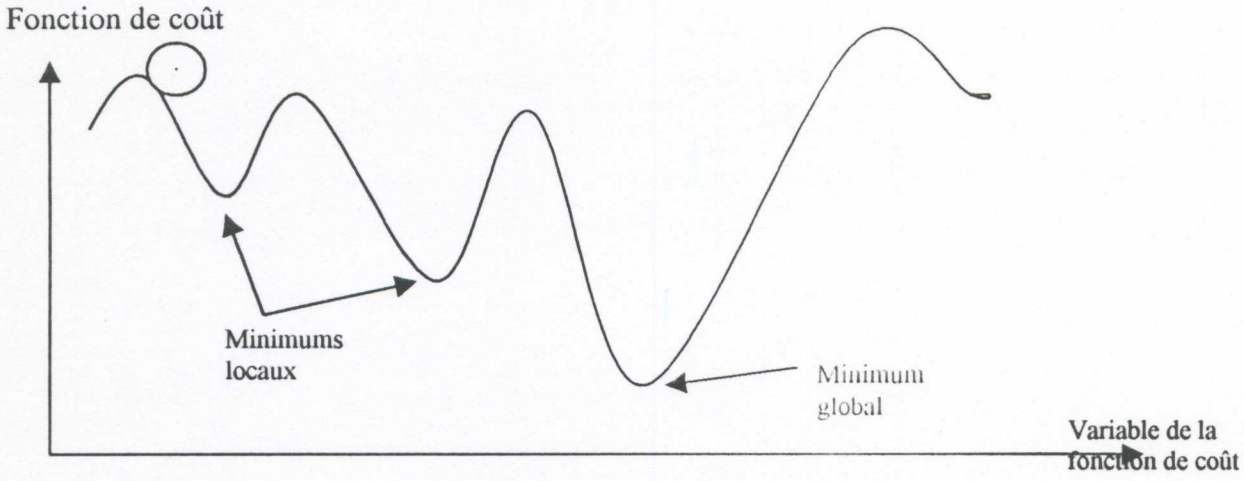
  On commence par l'ensemble entier des vecteurs d'apprentissage et on procède ensuite par élimination de ces vecteurs comme centroides candidats jusqu'à obtenir le dictionnaire; la procédure est comme suit :
  - Mettre le premier vecteur d'apprentissage dans le dictionnaire.
  - Calculer ensuite la distorsion entre ce vecteur et le deuxième vecteur d'apprentissage, si cette distorsion est inférieure à un certain seuil alors continuer, sinon ajouter ce deuxième vecteur au dictionnaire.
  - Pour chaque nouveau vecteur d'apprentissage trouver le plus proche centroide du dictionnaire, si la distorsion résultante est supérieure à un certain seuil, ajouter alors le vecteur d'apprentissage au dictionnaire.
  - Continuer jusqu'à obtenir le dictionnaire entier.

D'autres méthodes encore peuvent être utilisées, les détails peuvent être trouvés dans [19][23][20].

II-2-4-4-LA RELAXATION STOCHASTIQUE :

La relaxation stochastique est une famille de techniques d'optimisation ayant pour caractéristique commune le fait que chaque itération de recherche du minimum de la fonction de coût (la distorsion moyenne dans notre cas) consiste à perturber l'état qui est l'ensemble des variables indépendantes , de la fonction de coût (le dictionnaire dans notre cas) , d'une manière aléatoire. L'amplitude de la perturbation décroît avec le temps. Un exemple important est le recuit simulé (simulated annealing). Le but recherché par ces techniques est d'éviter les optimums locaux de la fonction de coût (ou du moins avoir de bons minimas).

Cette approche peut être illustrée dans la figure (2-9) , dans laquelle une bille glisse sur un parcours, ce parcours contient des vallées (minimas de la fonction de coût). Si la boule 'tombe ' dans un minimum local, alors en lui donnant une perturbation elle a beaucoup de chances de « s'échapper » de ce minimum. Si par contre cette boule tombe dans le minimum global(voir figure(2-9)), alors même en la perturbant, la boule n'a pas beaucoup de chances pour « échapper » à ce minimum.



Figure(2-9) illustration de la technique d'optimisation par recuit simulé

II-3-5-LE K-MEANS FLOU :(fuzzy clustring)

Les techniques de relaxation stochastique pour la conception du dictionnaire d'un quantificateur vectoriel sont coûteuses en temps de calcul [22], le « fuzzy clustring »est une technique qui a été développée pour remédier à cet inconvénient .

Dans la conception du dictionnaire à partir d'un ensemble d'apprentissage les formules pour le calcul du centroide d'un cluster et de la distorsion moyenne respectivement peuvent être écrites sous la forme :

$$y_i = \frac{\sum_{i=1}^{M} x_i S_j(x_i)}{\sum_{i=1}^{M} S_j(x_i)} \qquad \text{.....(2-27)}$$ avec $S_j(x_i) = 1$ si $x_i \in C_j$; sinon $S_j(x_i) = 0$

.

$$D = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{L} d(x_i, y_j) S_j(x_i) \text{ .....(2-28)}$$

avec :
L nombre de centroides (taille du dictionnaire),
M nombre de vecteurs d'apprentissage.
$S_j(x_i)$ est appelée fonction sélecteur , elle peut être considérée comme une fonction d'appartenance puisque $S_j(x_i)$ n'est égale à un que si « x_i »appartient à la classe (cluster) « j » , sinon cette fonction est nulle. Un cluster est dit « ensemble flou » si on peut assigner à chaque vecteur d'apprentissage x_i un degré d'appartenance ou bien une valeur d'appartenance partielle entre zéro et un qui indique à quel degré ce vecteur x_i peut il être considéré comme appartenant à cet ensemble. L'idée donc est de généraliser la fonction sélecteur $S_j(x_i)$ et de la rendre une fonction d'appartenance. On définit alors une distorsion floue comme étant :

$$D_f = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{L} d(x_i, y_j) [S_j(x_i)]^q \qquad \text{.....(2-29).}$$

q est un paramètre permettant de contrôler le caractère flou de la distorsion .
L'algorithme résultant est appelé K-MEANS FLOU (FUZZY K-MEANS) .
Notons enfin que le travail dans le domaine de la quantification vectorielle n'est guère à ses fins. En effet plusieurs techniques ne cessent de voir le jour [25]. Il importe cependant de noter que ces techniques ne sont que des variantes fondées principalement sur l'algorithme K-Means décrit précédemment.

II-4-CONCLUSION:

Ont été exposées dans ce chapitre les prétraitements que le signal vocal doit subir avant de se présenter à l'entrée de l'étage de décision(reconnaissance), dans le chapitre suivant nous allons aborder les modèles de Markov cachés et leur utilisation dans la reconnaissance de la parole.

Chapitre III

LES MODELES DE MARKOV CACHES

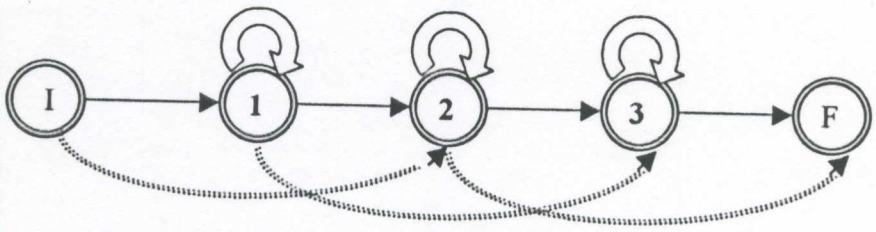
III-1-INTRODUCTION :

Ce chapitre expose le formalisme mathématique nécessaire pour l'utilisation des modèles de Markov cachés dans la reconnaissance automatique de la parole. Après quelques définitions et notations mathématiques, les modèles de Markov sont présentés comme outil de reconnaissance de la parole. L'apprentissage est un aspect capital dans les modèles de Markov cachés, à cet effet une bonne partie de ce chapitre lui a été consacrée. L'adaptation des HMMs à la reconnaissance de la parole sera ensuite abordée. On discutera enfin de la capacité discriminante (et donc des performances des HMMs) dans la reconnaissance de la parole.

La figure(3-1) montre un exemple de modèles de Markov cachés. On se donne donc un HMM, c'est à dire un ensemble d'états et une matrice de probabilités de transitions entre ces états :

A chaque instant, cette 'machine' saute de l'état courant vers un nouvel état, avec une probabilité déduite de la matrice de transitions. On suppose en outre qu'à chaque instant également, un signal élémentaire (on dit une observation) est émis par l'état courant, selon une loi de probabilité paramétrée qui lui est propre.

Une telle machine a pour effet de produire une séquence aléatoire d'observations, et constitue de ce fait un modèle de production de séquences, paramétré par la matrice de probabilités de transitions d'une part, et les paramètres des lois d'émission d'observations d'autre part.



Figure(3-1) exemple de modèle HMM

III-2-DEFINITIONS :

Soient :

Ω l'espace des observations, c'est à dire l'espace mesurable dans lequel les séquences à modéliser prennent leurs valeurs . Dans le cas de la parole, cela peut être par exemple l'espace R^N des vecteurs de paramètres issus d'une analyse LPC, cepstrale ou par banc de filtres.

O une séquence d'observations, c'est à dire un T-uplet $(O_1 \ldots\ldots\ldots O_T) \in \Omega^T$ (dans le cas de la parole ce T-uplet est une suite de T vecteurs spectraux).
On appelle chaîne de Markov la donnée de :

- S un ensemble d'états (cinq dans le cas de la figure(3-1)), désignons un état au temps t par S_t.
- $\mathbf{A_0} = [a_i]$ le vecteur de probabilités initiales des états, que l'on suppose à priori connues.(a_i représentant la probabilité de se trouver initialement à l'état i) leur somme doit être égale à 1. Signalons un cas particulier important :
  définir $a_I = 1$ et $a_i = 0$, $\forall i \neq I$ est équivalent à définir un état initial **I** .
- $\mathbf{A} = [a_{ij}]$ la matrice des probabilités de transition entre états; a_{ij} est la probabilité de transition de l'état $i \in S$ vers l'état $j \in S$. On peut bien sur imposer des contraintes sur cette matrice, en imposant une valeur nulle à certains de ses termes. Cela revient à autoriser ou interdire certaines transitions.

L'égalité suivante est toujours vérifiée :

$$\sum_{j \in S} aij = 1 \ldots\ldots\ldots\ldots (3\text{-}1)$$

- $\mathbf{b_i}$ un ensemble de lois sur Ω, associées aux états $i \in S$.

dans la suite la notation $bi(O_t)$ désignera, soit dans le cas d'un espace Ω discret, la probabilité de O_t, soit dans le cas continu la densité de la loi b_i au point O_t (O_t étant l'observation à l'instant t). Si les O_t sont quantifiées on parle de HMM discrets.

III-2-1-Remarques sur la loi d'émission :

Dans certains cas, la loi b_i est associée à chaque transition, cela conduit à des notations un peu plus lourdes, sans changer significativement le sens des démonstrations.

Un autre point délicat concerne la nature de ces lois. On souhaite un modèle paramétré : dans le cas discret, cette loi peut être donnée exhaustivement. Dans le cas continu, on considère généralement des lois normales ou des sommes de lois normales, comme donné ci dessous, où m_i et Σi sont respectivement le vecteur moyen et la matrice définie positive des covariances de la loi normale.

$$b_i(o_t) = \frac{1}{(2\Pi)^{N/2}\sqrt{\det(\sum_i)}} e^{\frac{-(o_t - m_i)' \sum_i^{-1} (o_t - m_i)}{2}} \ldots\ldots\ldots\ldots\ldots (3\text{-}2).$$

III-2-2-Hypothèses implicites :

La définition des chaînes de Markov implique deux hypothèses sur la nature des séquences à modéliser.

- La probabilité à priori d'un état ne dépend que de l'état précédent; c'est l'hypothèse de Markov : $P(S_t / S_{t-1}) = a_{S_{t-1}, S_t}$.
- Les observations ne dépendent que de l'état courant: toute la dépendance entre deux observations à deux instants donnés est exprimée au travers des états correspondants.

On dispose maintenant de séquences d'observations aléatoires O^k (des mots), et l'on cherche à estimer les meilleurs paramètres $\lambda = \{ \lambda_i \}$ du modèle de Markov caché qui conduisent à la détermination des meilleurs densités de ces séquences, dans l'espace des Ω^T .

III-3-DETERMINATION DES PARAMETRES D'UN HMM : [02][16]

Il nous faut donc maximiser la vraisemblance des séquences d'observations de référence.

Pour fixer les idées, on cherche à créer le modèle acoustique d'un mot (séquence d'observations). A l'aide de quelques répétitions de ce mot, on va déterminer les meilleurs paramètres du modèle de Markov correspondant. Mais avant de maximiser la vraisemblance, il faut d'abord trouver l'expression de celle ci; pour cela on définit :

- L_λ : vraisemblance de la séquence d'observations générée par le modèle de Markov caché paramétré par l'ensemble $\lambda = \{ \lambda_i \}$.
- $\alpha_i(t) = P(O_1\ O_2 \ldots\ldots\ldots O_t, i \text{ à } t \ / \ \lambda)$: c'est la probabilité du chemin partiel aboutissant à l'état i à l'instant t.
- $\beta_i(t) = P(O_{t+1}\ O_{t+2} \ldots\ldots O_T \ / \ i \text{ à } t, \lambda)$: c'est la probabilité de se trouver à l'état i à l'instant t sachant $O_{t+1}\ O_{t+2} \ldots O_T$.

$\alpha_i(t)$ et $\beta_i(t)$ peuvent être calculés de manière récursive :

$$\alpha_i(t) = \sum_{j \in S} \alpha_j(t-1) L_\lambda(O_t, S_t = i / S_{t-1} = j) = \sum_{j \in S} \alpha_j(t-1) b_i(O_t) a_{ji} \ldots\ldots\ldots\ldots (3\text{-}3)$$

$$\beta_i(t) = \sum_{j \in S} \beta_j(t+1) L_\lambda(O_{t+1}, S_{t+1} = j / S_t = i) = \sum_{j \in S} \beta_j(t+1) b_j(O_t) a_{ij} \ldots\ldots\ldots\ldots (3\text{-}4)$$

III-3-1-CALCUL DE LA VRAISEMBLANCE :

- **CALCUL DIRECT :**

Grâce aux hypothèses d'indépendance entre les observations lorsque la séquence d'états est fixée, on peut écrire la vraisemblance dans notre modèle d'une séquence d'observations ($O_1\ O_2 \ldots\ldots O_T$), sachant la séquence d'états ($S_1\ S_2 \ldots\ldots S_T$) :

$$L_\lambda(O_1 O_2 \ldots\ldots O_T / S_1 S_2 \ldots\ldots S_T) = \prod_{t=1}^{T} b_{s(t)}(O_t) \ldots\ldots\ldots\ldots\ldots (3-5).$$

$$L_\lambda(O_1 O_2 \ldots O_T, S_1 S_2 \ldots\ldots S_T) = P(S_1 S_2 \ldots\ldots S_T) L_\lambda(O_1 O_2 \ldots\ldots O_T / S_1 S_2 \ldots S_T) \ldots\ldots (3-6)$$

$$L_\lambda(O_1 O_2 \ldots\ldots O_T, S_1 S_2 \ldots S_T) = \prod_{t=1}^{T} a_{s(t-1)s(t)} b_{s(t)}(O_t) \ldots\ldots\ldots\ldots\ldots (3-7)$$

On peut alors écrire la vraisemblance d'une observation comme étant la somme sur toutes les séquences d'états des $L_\lambda(O_1\ O_2 \ldots\ldots O_t)$:

$$L\ (O_1 O_2 \ldots\ldots O_T) = \sum_{(S_1 S_2 \ldots S_T) \in \Gamma} L_\lambda(O_1 O_2 \ldots\ldots O_T, S_1 S_2 \ldots\ldots S_T) \ldots\ldots (3-8)$$

$$L\ (O_1 O_2 \ldots\ldots O_T) = \sum_{(S_1 S_2 \ldots S_T) \in \Gamma} \prod_{t=1}^{T} a_{S(t-1)S(t)} b_{S(t)}(O_t) \ldots\ldots\ldots\ldots\ldots (3-9).$$

Γ étant l'ensemble de toutes les séquences d'états possibles.

En d'autres termes, notre séquence d'observations peut être produite en suivant tous les parcours issus de l'état initial et aboutissant à l'état final. Le long de chaque parcours cette

séquence d'observations est émise avec une probabilité égale au produit des probabilités des transitions effectuées multiplié par le produit des probabilités d'émissions successives. Le calcul de la vraisemblance par la méthode directe nécessite donc la recherche de tous les parcours possibles allant de **(I)** et aboutissant à **(F)**. Une telle recherche exhaustive n'est évidemment pas envisageable dans la pratique. On utilise donc une méthode itérative décrite dans ce qui suit :

- **CALCUL RECURSIF :**

La vraisemblance peut s'écrire :

$$L_\lambda(O_1O_2.....O_T) = \sum_{j\in S} L_\lambda(O_1O_2...O_t, S_t = j)L_\lambda(O_{t+1}.....O_T / O_1.....O_t, S_t = j) \quad .....(3\text{-}10)$$

Les hypothèses sur les sources de Markov spécifient que (O_{t+1}.....O_t) ne dépend de (O_1.....O_t) qu'au travers de l'état S_t. On peut donc écrire :

$$L_\lambda(O_1O_2.....O_T) = \sum_{j\in S} L_\lambda(O_1O_2....O_t, S_t = j)L_\lambda(O_{t+1}.....O_T / S_t = j) \quad .....(3\text{-}11).$$

d'où :

$$L_\lambda(O_1O_2.....O_T) = \sum_{j\in S} \alpha_j(t)\beta_j(t) \qquad \forall t\in[1 .....T] \qquad .....(3\text{-}12).$$

Les formules (3-3) (3-4) nous permettent donc de calculer la vraisemblance d'une façon algorithmiquement efficace.
Ayant calculé la vraisemblance il nous faut maintenant la maximiser.

III-3-2-REESTIMATION DES PARAMETRES :

L'expression de $L_\lambda(O_1O_2.....O_T)$ est malheureusement bien trop compliquée pour envisager une solution analytique du problème de maximisation de la vraisemblance. Les solutions utilisées ne présentent que des optimisations locales telles que les techniques de gradient et les procédures de reestimation.
Connaissant un jeu de paramètres λ , on cherche un nouveau jeu de paramètres λ' , tel que $L_{\lambda'}(O_1O_2.....O_T) > L_\lambda(O_1O_2.....O_T)$, à l'aide d'une astuce donnée par le théorème suivant .

THEOREME DE BAUM :

Soit la fonction auxiliaire :

$$Q(\lambda,\lambda') = \sum_{(S_1S_2...S_T)\in\Gamma} L_\lambda(O_1O_2...O_T)\log\{L_{\lambda'}(O_1O_2....O_T, S_1S_2.....S_T)\} \quad .....(3\text{-}13).$$

Si $Q(\lambda , \lambda') \geq Q(\lambda , \lambda)$ alors $L_{\lambda'}(O_1O_2.....O_T) \geq L_\lambda(O_1O_2.....O_T)$

On peut donc commencer par choisir λ^0 arbitrairement, puis maximiser $Q(\lambda^0 , \lambda)$ par rapport à λ.'. Appelons le maximum obtenu λ^1. Evidemment $Q(\lambda^0 ,\lambda^1) > Q(\lambda^0 ,\lambda^0)$ d'où $L_{\lambda^1}(O_1O_2.....O_T) \geq L_{\lambda^0}(O_1O_2.....O_T)$. Dans la prochaine étape on maximise $Q(\lambda^1,\lambda)$ par rapport à λ et on appelle le maximum λ^2. De la même façon

$L_\lambda^2(O_1O_2.....O_T) \geq L_\lambda^1(O_1O_2.....O_T)$. En procédant de cette manière, on obtient une séquence $\lambda^0, \lambda^1, ..., \lambda^m$ telle que :

$L_\lambda^0(O_1O_2.....O_T) \leq L_\lambda^1(O_1O_2.....O_T) \leq .....\leq L_\lambda^m(O_1O_2.....O_T)$.

Il s'agit donc d'un algorithme itératif où à chaque itération la valeur de la vraisemblance est améliorée. $Q(\lambda, \lambda')$ est calculée comme suit :

$$Q(\lambda,\lambda') = \sum_{(S_1S_2...S_T)\in\Gamma} L_\lambda(O_1O_2....O_T)\log(L_{\lambda'}(O_1O_2....O_T))$$

$$Q(\lambda,\lambda') = \sum_{(S_1...S_T)\in\Gamma} L_\lambda(O_1O_2...O_T, S_1S_2....S_T)(\sum_{t=1}^{T}\log(a'_{S(t-1)S(t)}) + \log(b'_{S(t)}(O_t))) ....(3\text{-}14).$$

On cherche à maximiser la fonction $Q(\lambda,\lambda')$ par rapport à λ, où :

$\lambda = \{ a_{i0}, a_{ij}, b_j(\xi_n) \quad 1 \leq i,j \leq N, 1 \leq n \leq L \}$

$\lambda' = \{ a_{i0}', a_{ij}', b_j'(\xi_n) \quad 1 \leq i,j \leq N, 1 \leq n \leq L \}$

les paramètres λ' sont des variables réelle dans l'intervalle [0 1], ces variables doivent satisfaire les conditions :

$$\sum_{i=1}^{N} a_{0i}' = 1 \quad , \quad \sum_{i=1}^{N} a_{ij}' = 1 \quad , \sum_{n=1}^{L} b_j'(\xi_n) = 1$$

Le problème s'écrit :

$$MAX_{\lambda'}(Q(\lambda,\lambda'))$$

$$C_0(\lambda') = \sum_{i=1}^{N} a_{0j} = 1 \qquad , \qquad 1 \leq j \leq N$$

$$C_i(\lambda') = \sum_{i=1}^{N} a_{ij} = 1 \qquad , \qquad 1 \leq i, j \leq N \qquad .....(3\text{-}15)$$

$$C_{N+j}(\lambda') = \sum_{n=1}^{L} b_j'(\xi_n) = 1 \quad , \qquad 1 \leq n \leq L \text{ et } 1 \leq j \leq N$$

On peut alors calculer l'optimum sous contraintes de $Q(\lambda,\lambda')$ en écrivant la dérivée du lagrangien :

$$\nabla_{\lambda'\mu}\left\{Q(\lambda,\lambda') - \sum_{i=0}^{2N}(\mu_i C_i(\lambda') - 1)\right\} = 0. .....(3\text{-}16).$$

Dans laquelle les $C_i(\lambda')$ représentent les contraintes sur les paramètres. La résolution de l'équation précédente conduit aux formules de Baum-Welch données ci-après.

III-3-3-FORMULES DE BAUM DANS LE CAS DISCRET :

Dans le cas discret, les paramètres sont les probabilités de transition a_{ij} et les probabilités d'émission $b_j(\xi_n)$, où les ξ_n sont les observations quantifiées. Les formules de réestimation correspondantes sont :

$$a_{ij}' = \frac{\sum_{t=1}^{T} \alpha_i(t-1) a_{ij} b_j(O_t) \beta_j(t)}{\sum_{t=1}^{T} \alpha_i(t-1) \beta_i(t-1)} \quad \ldots\ldots\ldots\ldots (3-17).$$

$$b_i'(\xi_n) = \frac{\sum_{O_t=\xi_n} \alpha_i(t) \beta_i(t)}{\sum_{t=1}^{T} \alpha_i(t) \beta_i(t)} \quad \ldots\ldots\ldots\ldots (3-18).$$

III-3-4-FORMULES DE BAUM DANS LE CAS CONTINU :

Dans le cas gaussien , les probabilités d'émission sont modélisées par des lois normales (3-2). On obtient alors les formules ci dessous ; les a_{ij} étant les probabilités de transition, les M_i' les centres des gaussiennes et les Σ_i' les matrices de covariance des gaussiennes. On note x^t le vecteur transposé de x

$$a_{ij}' = \frac{\sum_{t=1}^{T} \alpha_i(t-1) a_{ij} b_j(O_t) \beta_j(t)}{\sum_{t=1}^{T} \alpha_i(t-1) \beta_i(t-1)} \quad \ldots\ldots\ldots\ldots (3-19)$$

$$M_i' = \frac{\sum_{t=1}^{T} \alpha_i(t) \beta_i(t) O_t}{\sum_{t=1}^{T} \alpha_i(t) \beta_i(t)} \quad \ldots\ldots\ldots\ldots (3-20).$$

$$\sum_i' = \frac{\sum_{t=1}^{T} \alpha_i(t) \beta_j(t) O_t O_t^t}{\sum_{t=1}^{T} \alpha_i(t) \beta_i(t)} - M_i' M_i'^t \quad \ldots\ldots (3-21)$$

III-4-CONVERGENCE :

On ne peut assurer que la convergence vers un maximum local [2] [9][14].
La pratique montre que l'algorithme converge vers une solution acceptable.
Le nombre d'itérations est fixé empiriquement , le test d'arrêt porte en général sur la variation relative des paramètres du HMM. En général, quatre itérations sont suffisantes [5] .

REMARQUE :

En pratique on dispose de plusieurs séquences d'observations(versions) ($O_1^k O_2^k \ldots O_T^k$) pour entraîner notre modèle.

Le principe du maximum de vraisemblance consiste donc à rechercher le maximum de :

$$\prod_k L_\lambda(O_1^k O_2^k \ldots O_T^k) = \prod_k \sum_{(S_1 S_2 \ldots S_T)\in\Gamma} L_\lambda(O_1 O_2 \ldots\ldots O_T) \ldots\ldots\ldots (3\text{-}22)$$

Il faut alors développer ce polynôme ,formuler la fonction auxiliaire et annuler les dérivées de son lagrangien. Les reéstimmations correspondantes sont comparables aux formules de Baum , mais une sommation supplémentaire sur les exemples, pondérée par les inverses de leurs vraisemblances, se glisse au numérateurs et dénominateurs. Les formules de reestimation sont dans ce cas :

$$a_{ij}^{'} = \frac{\sum_K \frac{1}{L_\lambda(O_1^K O_2^K \ldots O_T^K)} \sum_{t=1}^{T} \alpha_i^K(t-1) a_{ij} b_j(O_t) \beta_j^K(t)}{\sum_K \frac{1}{L_\lambda(O_1^K O_2^K \ldots O_T^K)} \sum_{t=1}^{T} \alpha_i^K(t-1) \beta_i^K(t-1)} \ldots\ldots\ldots\ldots\ldots (3-23)$$

$$b_i^{'}(\xi_n) = \frac{\sum_K \frac{1}{L_\lambda(O_1^K O_2^K \ldots O_T^K)} \sum_{O_t=\xi_n} \alpha_i^K(t) \beta_i^K(t)}{\sum_K \frac{1}{L_\lambda(O_1^K O_2^K \ldots O_T^K)} \sum_{t=1}^{T} \alpha_i^K(t) \beta_i^K(t)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3-24)$$

III-5-PROBLEMES NUMERIQUES :(problème de l'underflow)

Cet algorithme pose également d'importants problèmes numériques. Les calculs des $\alpha_i(t)$, $\beta_i(t)$ consiste essentiellement à multiplier entre elles des probabilités inférieures à un.Toutes ces valeurs tendent vers zéro très vite. Les formules de Baum deviennent alors des quotients de la forme 0/0. Une méthode pour s'affranchir de ces problèmes numériques est proposée par Levinson [2].
La suite N_i étant donnée, on pose :
$\alpha_i'(t) = C_t \alpha_i(t)$ et $\beta_i'(t) = D_t \beta_i(t)$ avec :

$$C_t = \prod_{i=1}^{t} N_i \ldots\ldots\ldots\ldots (3\text{-}25).$$

$$D_t = \prod_{i=t+1}^{T} N_i \ldots\ldots\ldots\ldots (3\text{-}26).$$

En reprenant les équations (3-3) (3-4) on obtient :

$$\alpha_i'(t) = \frac{C_t}{C_{t-1}} \sum_{j \in S} \alpha_j'(t-1) b_i(O_t) a_{ji} = N_t \sum_{j \in S} \alpha_j'(t-1) b_i(O_t) a_{ji} \quad \ldots\ldots (3-27).$$

$$\beta_i'(t) = \frac{D_t}{D_{t+1}} \sum_{j \in S} \beta_j'(t+1) b_i(O_t) a_{ji} = N_{t+1} \sum_{j \in S} \beta_j'(t+1) b_j(O_{t+1}) a_{ji} \quad \ldots\ldots (3-28).$$

Ces formules de récurrence permettent de calculer les α_i'(t) et les β_i'(t).On détermine alors les N_t au cours de la passe avant , de telle sorte que $\sum_{j \in S} \alpha_i'(t) = 1$.

De plus , on peut faire apparaître les coefficients C_t et D_t dans les formules de Baum. On obtient alors des formules de réetimation équivalentes aux formules de Baum , dans lesquelles les α_i'(t) et β_i'(t) remplacent les $\alpha_i(t)$ et $\beta_i(t)$. Les formules de reestimation deviennent alors :

$$a_{ij}' = \frac{\sum_K \frac{1}{L'_\lambda (O_1^K O_2^K \ldots O_T^K)} \sum_{t=1}^{T} \alpha_i^{K\,\prime}(t-1) N_t a_{ij} b_j(O_t) \beta_j^{K\,\prime}(t)}{\sum_K \frac{1}{L'_\lambda (O_1^K O_2^K \ldots O_T^K)} \sum_{t=1}^{T} \alpha_i^{K'}(t-1) \beta_i^{K'}(t-1)} \quad \ldots\ldots (3-29)$$

$$b_i'(\xi_n) = \frac{\sum_K \frac{1}{L'_\lambda (O_1^K O_2^K \ldots O_T^K)} \sum_{O_t = \xi_n} \alpha_i^{K\,\prime}(t) \beta_i^{K\,\prime}(t)}{\sum_K \frac{1}{L'_\lambda (O_1^K O_2^K \ldots O_T^K)} \sum_{t=1}^{T} \alpha_i^{K\,\prime}(t) \beta_i^{K\,\prime}(t)} \quad \ldots\ldots (3-30)$$

Une deuxième solution [9] consiste à introduire un facteur d'échelle k > 1 dans (3-3) (3-4).Toutefois le choix du facteur k est très délicat, surtout pour l'entraînement des modèles, pendant lequel les paramètres estimés peuvent varier d'une façon notable.
Dans ce travail, c'est la première solution qui sera adoptée puisqu'elle permet d'éviter les problèmes rencontrés dans la deuxième approche.

III-6-HMM POUR LA RECONNAISSANCE DE LA PAROLE :

III-6-1-REGLE BAYSIENNE :

On se place dans le problème de reconnaissance des mot isolés d'un dictionnaire de taille finie { $\omega_1 \ldots\ldots \omega_N$ }. Soient $W_1 \ldots\ldots W_N$ les modèles de Markov censés représenter chacun des mots ω_i .
On a vu qu'un modèle de Markov est un modèle de production de signal. Les HMMs peuvent cependant facilement être utilisés pour la reconnaissance, en utilisant le théorème de Bayes.
Si l'on dispose d'une séquence d'observations ($O_1O_2 \ldots O_T$), correspondant à un mot inconnu de l'ensemble { $\omega_1 \ldots\ldots \omega_N$ }. On peut approcher la probabilité conditionnelle de chaque mot ω_i par : $P(W_i / O_1O_2 \ldots\ldots O_T)$. D'après la formule de Bayes sur les densités :

$$P(\omega_i / O_1 O_2 \ldots O_T) \approx P(W_i / O_1 O_2 \ldots O_T) = \frac{L(O_1 O_2 \ldots O_T) / W_i) P(W_i)}{L((O_1 O_2 \ldots O_T)} \quad \ldots\ldots (3-31).$$

Pour une séquence d'observations fixée, le dénominateur de cette expression est constant. Le mot reconnu est donc :

$\omega^* = \arg\max_{\{\omega i\}} (L(O_1O_2...O_T / W_i)P(W_i))$.....(3-32).

$L(O_1O_2...O_T / W_i)$ peut être calculé pour notre modèle W_i et en tenant compte des contraintes sur les chemins admissibles , on obtient :

$$L(O_1O_2.....O_T / W_k) = \alpha_{S_F(W_k)}(T) \quad \text{.....(3-33).}$$

$$\alpha_{S_1(W_k)}(0) = 1$$

Cet algorithme est appelé algorithme Forward. En effet, le calcul récursif de $\alpha_i(t)$ dans(3-12) se fait à l'aide de ces même valeurs à l'instant précédent, c'est à dire dans le sens du temps.

Reste à estimer $P(W_i)$. Le modèle W_i représentant le mot ω_i , on évalue cette probabilité en se donnant un modèle de langage, c'est à dire en déterminant empiriquement la probabilité d'occurrence de chaque mot . $P(W_i) \approx P(\omega_i)$ ce qui donne :

$\omega^* = \arg\max_{\{\omega i\}} (L(O_1O_2...O_T / W_i)P(W_i))$.

A noter que dans notre cas (reconnaissance de mots isolés) on adoptera des probabilités d'apparition égales pour tous les mots du vocabulaire.

III-6-2-ALGORITHME DE VITERBI :

L'algorithme de Viterbi est une approximation de l'algorithme de Baum. Il consiste à calculer la vraisemblance de la séquence d'observations le long du chemin le plus probable c'est à dire que :

$$L^{Viterbi} = \max_{(S_1...S_T)} \{(O_1.....O_2, S_1.....S_T / W_k\} \quad \text{.....(3-34).}$$

Comme dans le cas de l'algorithme Forward , on peut effectuer ce calcul récursivement, dans le sens du temps. Il suffit de remplacer tous les opérateurs de sommation (Σ) en opérateurs de maximisation (Max) :

$$L^{Viterbi}(O1.....O_T / W_k) = \alpha^V_{S_F(W_k)}(T) \quad \text{.....(3-35).}$$

$$\alpha^V_{S_1(W_k)}(0) = 1 \quad \text{.....(3-36)}$$

$$\alpha^V_i(t) = \left\{\alpha^V_j(t-1) b_i(O_t) a_{ji}\right\}_{j \in S} \quad \text{.....(3-37).}$$

Comparé avec l'algorithme Forward, l'algorithme de Viterbi possède deux qualités majeures et un défaut important [2] :

- On peut retrouver explicitement la meilleure séquence d'états. Il suffit de se rappeler des arguments de l'opérateur (Max) qui ont été sélectionnés. Ce point est crucial pour la parole continue.

- L'algorithme de Viterbi se prête à plusieurs simplifications. On peut pratiquer l'élimination de séquences d'états déjà fort improbables à un instant donné. On peut également pratiquer un codage logarithmique des probabilités, et changer ainsi les produits en sommes.
- Mais il faut se rappeler que l'algorithme Forward et l'algorithme de Viterbi ne sont pas équivalents. En toute rigueur il faudrait appliquer l'algorithme Forward. Rien ne prouve en effet que la vraisemblance de la séquence d'états la plus probable fournit une bonne indication sur la vraisemblance le long de toutes les séquences d'états possibles pour la production d'un mot.

III-6-3-CAPACITE DISCRIMINANTE :

L'une des particularités de l'apprentissage de ces modèles est son aspect peu discriminant : chaque modèle de mot peut être entraîné séparément. C'est extrêmement utile du point de vu pratique car ceci implique qu'on peut à tout moment ajouter un nouveau mot à notre vocabulaire.
L'algorithme de Baum-Welch n'extrait pas des caractéristiques permettant de discriminer deux mots, mais essaie de faire une approximation de la densité de probabilité du mot dans l'espace de toutes les séquences d'observations.
De nombreux travaux ont été effectués pour améliorer la capacité discriminante des HMM ([5], [7].....). Il s'agit des modèles de Markov discriminants
Contrairement aux HMMs (dits standards) faisant l'objet de ce travail , les modèles de Markov discriminants s'appuient sur le principe de maximum à posteriori (MAP). Il faut cependant noter la complexité des algorithmes d'apprentissage de ce type de HMM [5] [2] , ajouter à cela le fait que : une fois l'apprentissage effectué pour un vocabulaire donné, on ne pourra plus ajouter un nouveau mot au vocabulaire (à moins que l'apprentissage ne soit repris entièrement) .

III-7-CONCLUSION :

D'après ce qu'on vient de voir, la modélisation Markovienne repose sur un formalisme à la foi simple et rigoureux. A ces deux points forts viendra s'ajouter un troisième qui est la robustesse comme on le verra à travers les résultats de reconnaissance exposés dans le chapitre suivant.

Chapitre IV

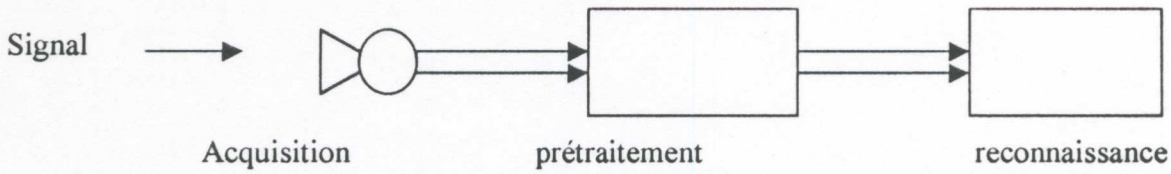
RESULTATS DE RECONNAISSANCE

IV-1-INTRODUCTION :

Ce chapitre est consacré à l'application des HMMs à la reconnaissance de la parole.
La deuxième section décrit la procédure de reconnaissance à partir de l'acquisition des signaux jusqu'à l'étage de décision. Les procédures d'apprentissage et de reconnaissance sont ensuite exposées dans la troisième et quatrième section. On y trouvera les détails de calcul à travers des organigrammes et des algorithmes détaillés.
Enfin, dans la cinquième section, on expose les résultats de reconnaissance obtenus. Ces résultats seront à chaque fois suivis de commentaires.

IV-2-DESCRIPTION FONCTIONNELLE DU SYSTEME :

La figure (4-1) montre l'organisation de la procédure de reconnaissance :



Figure(4-1) schéma bloc des différentes étapes de la procédure de reconnaissance

IV-2-1-AQUISITION :

Les enregistrements ont été effectués au niveau du laboratoire de traitement de la parole de l'institut d'électronique (université de Blida).
L'acquisition a été faite par un microphone dont les spécifications sont les suivantes :

- microphone omnidirectionnel.
- sensibilité 58-68 dB
- réponse fréquentielle 50-13000 Hz.
- rapport signal sur bruit >40 dB.

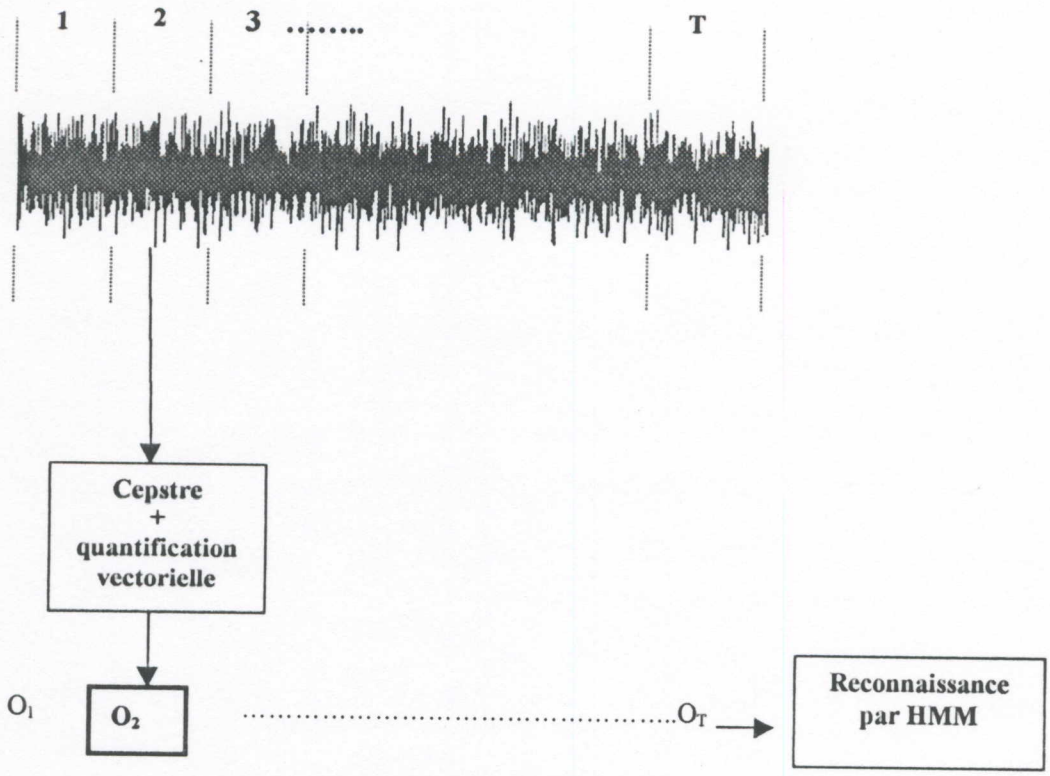
Le signal est échantillonné à 8 kHz(spectre utile jusqu'à 4Khz) et codé en entier 16 bits.
Une fenêtre de Hamming est appliquée tout les 128 échantillons (soit 16 ms de signal) avec un chevauchement de 50% entre fenêtres successives..

IV-2-2-CODAGE :

Sur chaque fenêtre du signal on extrait les coefficients cepstraux, le signal sera donc sous forme d'une suite de vecteurs spectraux représentant chacun une fenêtre de 16 ms de signal.

IV-2-3-QUANTIFICATION VECTORIELLE :

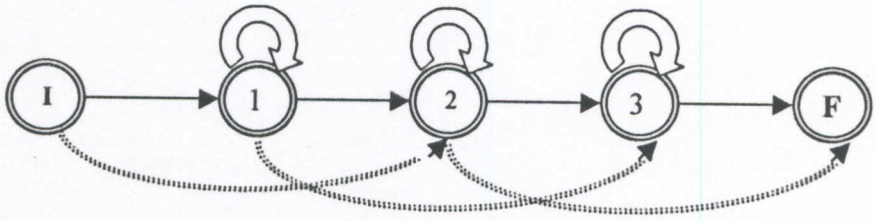
On opère ensuite une quantification vectorielle des vecteurs résultants, en utilisant l'algorithme K-MEANS décrit dans la section(II-2-4-3)du deuxième chapitre. Cet algorithme est disponible sous MATLAB(version 5-2).
La taille du dictionnaire est L = 128 (valeur usuellement utilisée[9]). Le signal se présentera donc à l'entrée de l'étage de reconnaissance sous forme d'une suite de vecteurs spectraux appartenant à un ensemble discret (les 128 centroides).
La figure (4-2) illustre ce qui a été dit précédemment.



Figure(4-2) du signal au message

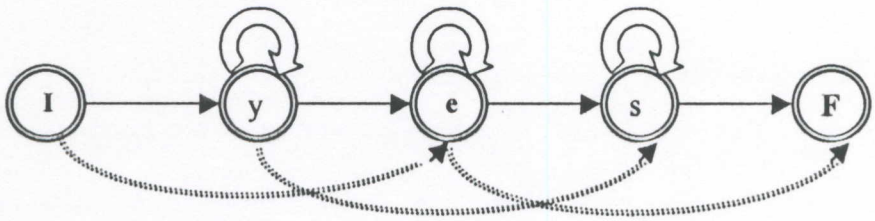
IV-2-4-CONCEPTION DES MODELES :

On dispose d'un certain vocabulaire et on veut modéliser chacun de ses éléments par un HMM .
Les éléments du vocabulaire peuvent être des mots, des phonèmes ou n'importe qu'elle autre unité linguistique. Dans la reconnaissance phonémique le modèle le plus utilisé (figure(4-3)) est un modèle à trois états émetteurs[8][9]; les sauts sont interdits(flèches en pointillés) en ce sens qu'ils ont une probabilité très faible.
Les états 1 et 3 dans la figure (4-3) correspondent aux parties transitoires du phonème, alors que l'état 2 correspond à la partie stable. L'état initial est noté (I) et l'état final(*F*).



Figure(4-3) modèle HMM d'un phonème

Dans la reconnaissance de mots, on peut associer un état à chaque phonème. Le nombre d'états correspondra donc au nombre de phonèmes contenus dans le mot (figure(4-4)).
Si dans ce cas on désire avoir une topologie standard pour tous les modèles des mots du vocabulaire, alors on choisira le modèle correspondant au mot le plus long; ce modèle sera ensuite adopté pour tout le reste des mots du vocabulaire.



Figure(4-4) exemple d'un modèle HMM
Pour le mot anglais yes

IV-2-5-VOCABULAIRE UTILISE :

L'objet de ce travail étant l'application des HMMs dans la reconnaissance de la parole, nous avons choisi dans un premier temps de travailler sur un vocabulaire simple constitué par les cinq alphabets français (A B C I O). A travers ce vocabulaire on va évaluer les performances de la modélisation Markovienne en spéculant sur plusieurs paramètres à savoir : le nombre d'exemples d'apprentissage, le nombre de coefficients cepstraux ainsi que le nombre de centroides utilisés dans le dictionnaire du quantificateur vectoriel. L'effet de la taille du vocabulaire sera mis en évidence à travers un vocabulaire de taille plus importante constitué par les chiffres de un à neuf prononcés en langue française.

IV-3-PROCEDURE D'APPRENTISSAGE :

Le modèle adopté pour chaque mot du vocabulaire est celui de la figure (4-3).La définition des états (**I**)et (**F**) implique que :

$\alpha_I(0) = 1$

$\alpha_i(0) = 0$ pour i≠**I**

$\alpha_I(t) = 0$ pour t≠0

$\beta_F(T+1) = 1$

$\beta_F(t) = 0$ pour t≠T+1

T étant la longueur du mot (nombre de vecteurs spectraux).

Notre mot peut donc être produit en suivant tous les parcours allant de (**I**) et aboutissant en (**F**). La vraisemblance du mot est donc donnée par :

$$L_\lambda(O_1O_2.....O_T) = \alpha_F(T+1)$$
$$= \sum_{i\in S} \alpha_i(T)a_{iF} \quad , \text{ S étant l'ensemble des trois états émetteurs.}$$

Les formules de réestimation sont :

$$a_{ij}^{'} = \frac{\sum_K \frac{1}{L_\lambda(O_1^K O_2^K ... O_T^K)} \sum_{t=1}^{T} \alpha_i^K(t-1)a_{ij}b_j(O_t)\beta_j^K(t)}{\sum_K \frac{1}{L_\lambda(O_1^K O_2^K ... O_T^K)} \sum_{t=1}^{T} \alpha_i^K(t-1)\beta_i^K(t-1)}$$

$$b_i^{'}(\xi_n) = \frac{\sum_K \frac{1}{L_\lambda(O_1^K O_2^K ... O_T^K)} \sum_{O_t=\xi_n} \alpha_i^K(t)\beta_i^K(t)}{\sum_K \frac{1}{L_\lambda(O_1^K O_2^K ... O_T^K)} \sum_{t=1}^{T} \alpha_i^K(t)\beta_i^K(t)}$$

Rappelons que a_{ij} représente la probabilité de transition de l'état i vers l'état j et que $b_i(\xi_n)$ représente la probabilité d'émission par l'état i de l'observation quantifiée ξ_n (qui est un des centroides du dictionnaire du quantificateur). Vue le modèle adopté, la matrice de probabilités de transition $\{a_{ij}\}$ est de dimension 5×5, et la matrice des probabilités d'émission est de dimension 3×L ('3' pour le nombre d'états émetteurs et L pour le nombre de centroides).
La procédure d'apprentissage sera donc comme suit :

1-Initialiser les paramètres

a_{ij} $\quad 1 \le i \le 5$, $1 \le j \le 5$

$b_i(\xi_n)$ $\quad 1 \le i \le 3$, $1 \le n \le L$

2- Poser Z=1

3- Présenter le corpus ; calculer les $\alpha_i^k(t)$ et $\beta_i^k(t)$ avec :

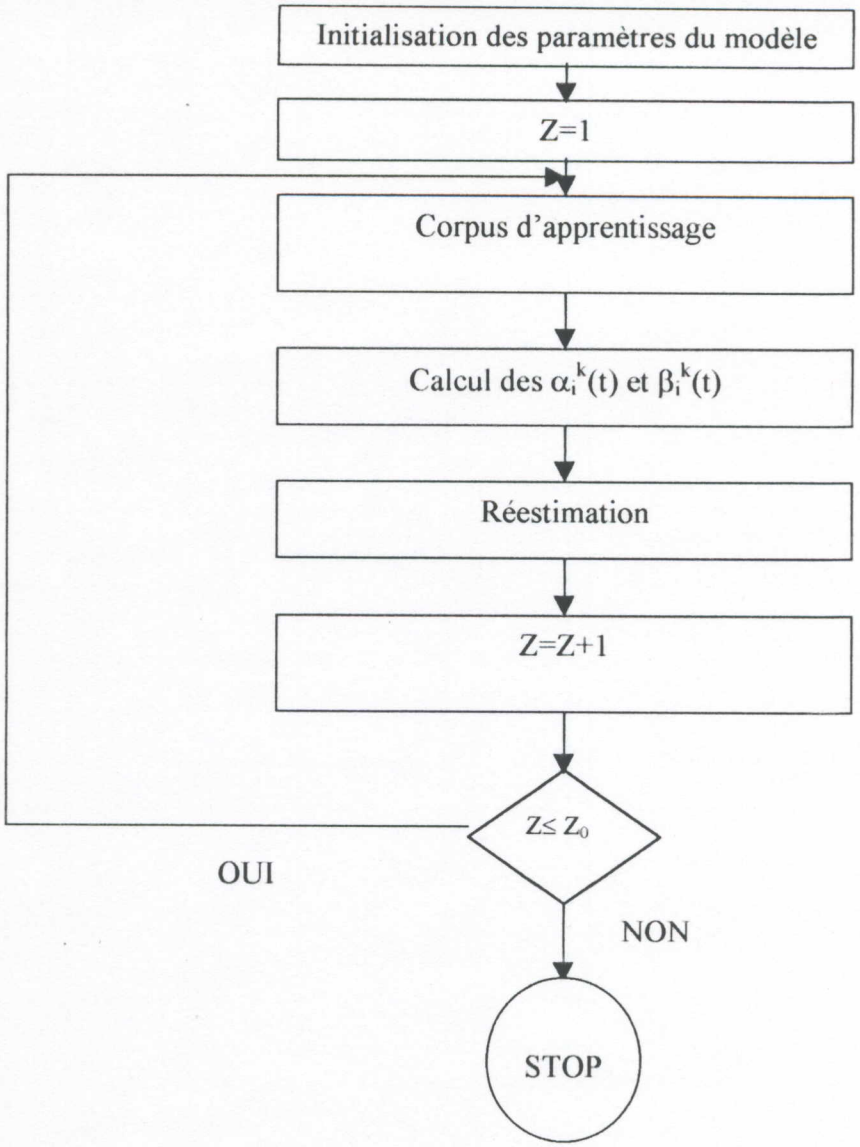
$1 \le i \le 5$, $0 \le t \le T$, $1 \le k \le K$ (K taille du corpus).

4- Calculer les nouvelles valeurs de a_{ij} et $b_i(\xi_n)$ ($a_{ij}^{'}$ et $b_i^{'}(\xi_n)$) à l'aide des formules(3-23) et (3-24).

5- Z=Z+1

6- Si $Z \le Z_0$ (nombre d'itérations) alors aller à l'étape 3,
sinon STOP.

L'organigramme résultant est le suivant :



Figure(4-5) :organigramme de la procédure d'apprentissage

Les variables $\alpha_i(t)$ et $\beta_i(t)$ pour un exemple du corpus d'apprentissage sont calculées à l'aide des formules (3-3) (3-4) respectivement. Les algorithmes résultant (appelés FORWARD et BACKWARD respectivement) sont les suivants :

ALGORITHME FORWARD :

1- Initialisation :

$\alpha_1(0) = 1$,

$\alpha_j(1) = a_{1j} b(O_1)$, $\quad 2 \leq j \leq 4$

2- Récursion :

$$\alpha_j(t+1) = \sum_{i=1}^{5} \alpha_i(t) a_{ij} b_j(O_{t+1}) \quad , \quad 1 \leq t \leq T \text{ et } 2 \leq j \leq 4$$

ALGORITHME BACKWARD :

1-Initialisation :

$\beta_F(T+1) = 1$

$\beta_i(T) = a_{iF}$

2-Récursion :

$$\beta_j(t) = \sum_{i=1}^{5} \beta_i(t+1) b_i(O_{t+1}) a_{ji} \quad , \quad 0 \le t \le T \text{ et } 1 \le i \le 4$$

Nous avons évoqué dans le chapitre précédent le problème de l'underflow ainsi que la solution adoptée pour le vaincre. Rappelons que la solution consiste à remplacer les variables $\alpha(t)$ et $\beta(t)$ par les variables $\alpha'(t)$ et $\beta'(t)$ telles que :

$$\alpha_i'(t) = C_t \alpha_i(t)$$
$$\beta_i'(t) = D_t \beta_i(t)$$

avec : $C_t = \prod_{i=1}^{t} N_i$ et $D_t = \prod_{i=t+1}^{T} N_i$

Les N_i sont determinés de telle sorte que $\sum_{j=1}^{5} \alpha_j(t) = 1$.

Nous avons donc appliqué cette solution pour obtenir les algorithmes FORWARD et BACKWARD modifiés suivants:

ALGORITHME FORWARD MODIFIE :

1-Initialisation :

$\alpha_1(0) = 1$

$\alpha_i(1) = a_{1i} b_i(O_1)$

$$N_1 = \frac{1}{\sum_{i=2}^{4} \alpha_i(1)} \qquad , 2 \le i \le 4$$

$\alpha_i'(1) = N_1 \alpha_i(1)$

$C_1 = N_1$

2-Récursion :

$$\alpha_j(t+1) = \sum_{i=1}^{5} \alpha_i(t) a_{ij} b_j(O_{t+1})$$

$$N_{t+1} = \frac{1}{\sum_{j=1}^{5} \alpha_j(t+1)}$$

$$\alpha_j'(t+1) = N_{t+1}\alpha_j(t+1)$$

$$C_{t+1} = \prod_{i=1}^{t+1} N_i$$

ALGORITHME BACKWARD MODIFIE :

1-Initialisation :

$$\beta_F(T+1) = 1$$

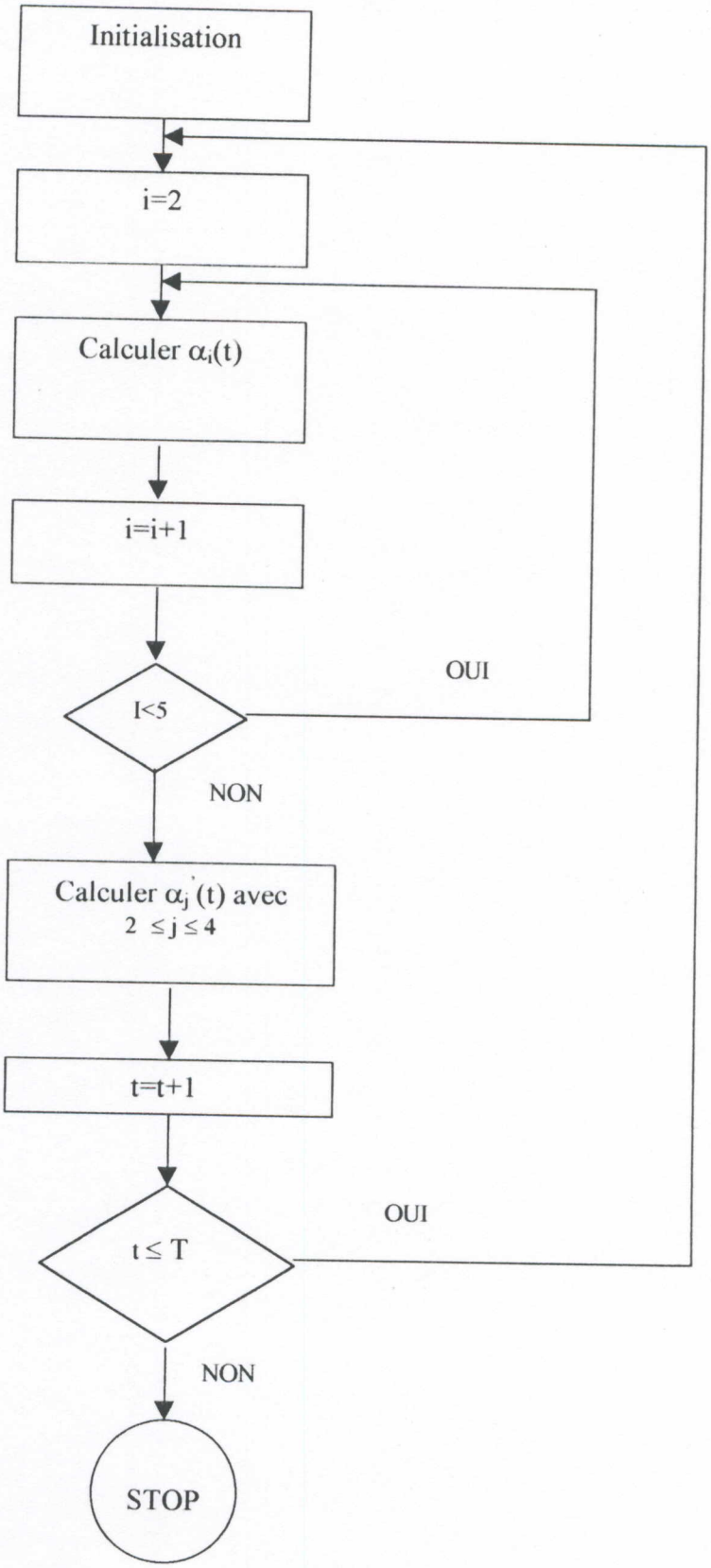
$$\beta_i(T) = a_{iF} \qquad , \ 1 \le i \le 4$$

2-Récursion :

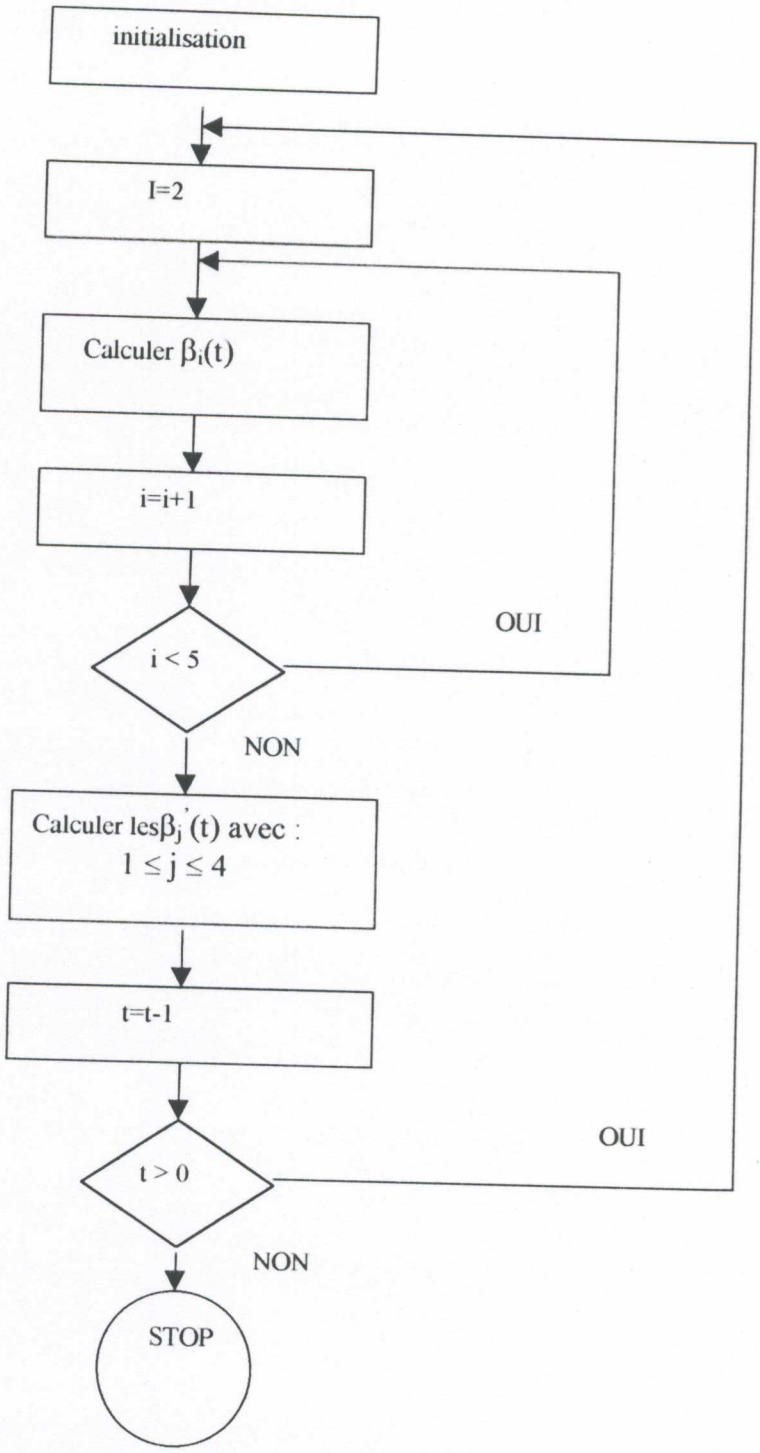
$$\beta_j(t) = \sum_{i=1}^{4} \beta_i(t+1) b_i(O_{t+1}) a_{ji}$$

$$\beta_j'(t) = N_{t+1}\beta_j(t) \qquad , \ 1 \le i \le 4 \ \text{ et } \ 0 \le t \le T\text{-}1$$

L'organigramme de la figure (4-5) reste évidemment inchangé sauf que les $\alpha_i(t)$ et les $\beta_i(t)$ seront remplacées par les $\alpha'_i(t)$ et les $\beta'_i(t)$; et les formules (3-3), (3-4) seront remplacées par (3-27), (3-28).
Les organigrammes suivants illustrent le calcul des $\alpha'_i(t)$ et $\beta'_i(t)$:



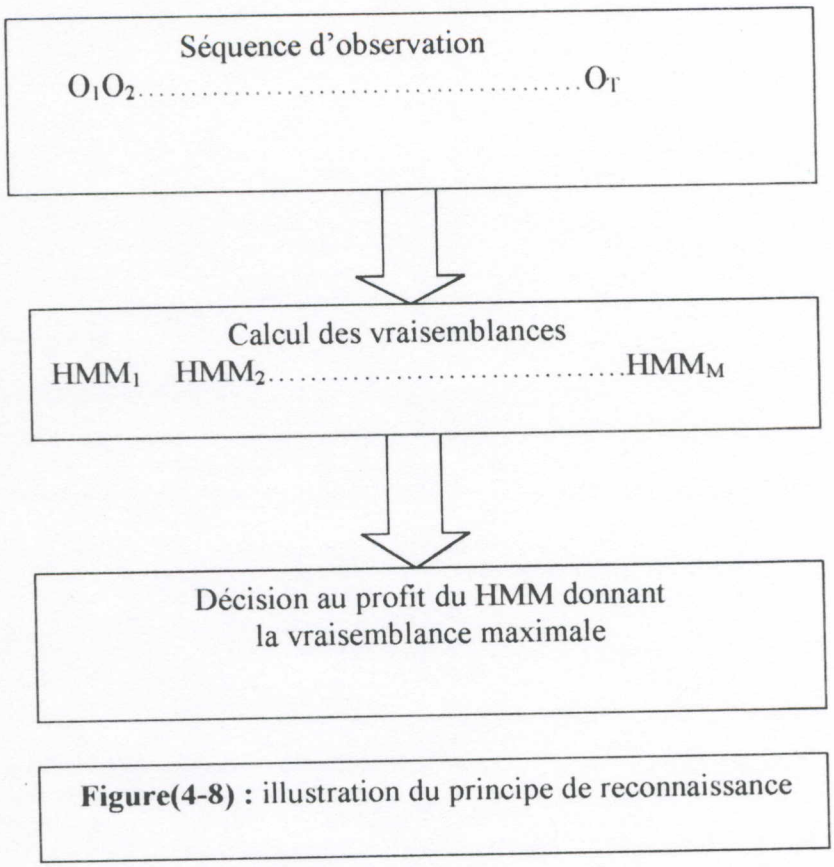
Figure(4-6) :organigramme de calcul des $\alpha_i(t)$



Figure(4-7) :organigramme de calcul des $\beta_i'(t)$

IV-4-PROCEDURE DE RECONNAISSANCE :

La reconnaissance est basée sur le principe illustré dans la figure (4-8).



Figure(4-8) : illustration du principe de reconnaissance

Dans la figure(4-8) : HMM1 , HMM2 , ..., HMM_M représentent les modèles de Markov des mots du vocabulaire (M représentant la taille du vocabulaire). Ces modèles sont obtenus par l'apprentissage décrit dans la section précédente. La reconnaissance consiste donc à calculer la vraisemblance du mot inconnu à partir de tous les modèles HMM des mots du vocabulaire puis de décider au profit du modèle donnant le maximum de vraisemblance.
La vraisemblance d'un mot est donnée par :

$$L\lambda(O_1O_2.....O_T) = \alpha_F(T+1)$$

$$= \beta_1(0)$$

$$\alpha_F(T+1) = \sum_{i=2}^{4} \alpha_i(T) a_{iF}$$

$$\beta_1(0) = \sum_{i=2}^{4} \beta_i(1) a_{1i} b_i(O_1)$$

Le calcul de la vraisemblance revient donc à calculer les $\alpha_i(t)$ ou bien les $\beta_i(t)$. Ce calcul a été déjà décrit précédemment.
Dans ce qui suit , a_{ij} représentera la probabilité de transition de l'état i vers l'état j. Les matrices des probabilités d'émission seront représentées sous forme de tableaux dont la

première colonne représente les numéros(indice j) des centroides, les trois autres colonnes représentent leurs probabilités d'émission par les états (**I1,I2,I3**); seuls les centroides ayant une probabilité d'émission non nulle seront indiqués.

IV-5-RESULTATS D'APPRENTISSAGE ET DE RECONNAISSANCE :

IV-5-1-EFFET DU NOMBRE D'EXEMPLES :

Il est assez intuitif de penser que le fait d'augmenter la taille du corpus d'apprentissage permet d'améliorer le taux de reconnaissance. En effet, ceci permet de capturer les statistiques de chaque mot du vocabulaire, rendant plus meilleures les performances de reconnaissance de nos modèles HMM. Dans un premier temps nous avons utilisé cinq exemples d'apprentissage pour chaque mot du vocabulaire, le nombre d 'exemples d'apprentissage sera ensuite augmenté à dix puis à vingt afin de voir l'effet de la taille du corpus d'apprentissage sur le taux de reconnaissance.

IV-5-1-1-APPRENTISSAGE AVEC CINQ EXEMPLES :

Les paramètres des modèles obtenus après apprentissage sont :

- **Modèle de la voyelle « A » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9085 & 0.0915 & 0 & 0 \\ 0 & 0 & 0.9213 & 0.0787 & 0 \\ 0 & 0 & 0 & 0.9484 & 0.0516 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

j	I=1	I=2	I=3
9	0.1281	1.1717 (10^{-4})	0.3407
26	0.0366	0	0
27	0	0.4250	0
29	0.0366	0	0.2271
32	0.0732	0	0
34	0.3112	0	0
44	0.0055	0.0581	0
50	0	0.0315	0
52	0	0.0629	0
62	0.0015	0.0617	0
63	0.0353	0.0011	0
64	0.2563	0	0
66	0	0.2204	0
85	0	0	0.0103
87	0.0237	0.0581	1.4632 (10^{-4})
93	0.0366	0	0
118	0.0549	0	0.4216
122	0.0785	0	0

- **Modèle de la consonne « B » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9403 & 0.0597 & 0 & 0 \\ 0 & 0 & 0.9389 & 0.0611 & 0 \\ 0 & 0 & 0 & 0.9328 & 0.0672 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I=1	I=2	I=3
1	0	0.0061	0.0201
3	0	0	0
9	0.0937	0.0752	0
10	0	0.0019	0.0402
19	0	0.0122	0.1074
25	0.0478	0	0
26	0	0	0.0269
27	0	0.2444	0
29	0.1113	0.0572	0
32	0.0239	0	0
34	0.4897	0	0
38	0	0	0.0134
40	0	0	0.1746
45	0	0	0.0134
55	0.0358	0	0
64	0.0717	0	0
65	$2.2527\ (10^{-4})$	0.0121	0.0402
66	0	0.1467	0
67	0.0239	0	0
74	0.0119	0	0.0806
85	0	0.1222	0
92	0.0016	0.1450	0
93	0.0358	0	0
95	0.0119	0.0122	0.3492
96	0	0	0.0268
97	0.0167	0.0196	
103	0	0	0.0269
117	0	0.1451	0.0151
118	0.0239	0	0.0134

- **Modèle de la consonne « C » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9320 & 0.0680 & 0 & 0 \\ 0 & 0 & 0.9431 & 0.0569 & 0 \\ 0 & 0 & 0 & 0.9364 & 0.0636 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

j	I=1	I=2	I=3
1	0	0.0090	0.0154
3	0	0.0093	0.0405
6	0	1.2269 (10^{-4})	0.0126
9	0.0952	0	0.1909
10	0	0	0.0381
21	0.0544	0	0
26	0.0404	3.1936 (10^{-4})	0.0255
27	0	0.0341	0
29	0	0	0.0255
31	0.0027	0.0091	0
32	0.0680	0	0
34	0.3946	0	0
40	0.0046	0.0189	0.0891
55	0.0136	0	0
61	0	0.1137	0
63	0	0.3639	0
64	0.1361	0	0
65	0	0	0.0763
66	0	0.0796	0
67	0.0272	0	0
74	0	0	0.0382
77	0.0272	0	0
89	0	0	0.0126
93	0.0544	0	0
95	0	0.0114	0.2291
103	0	0	0.0254
107	0	0	0.0255
117	0	0.2841	2.5277 (10^{-4})
118	0.0562	0.0212	0.1527
120	0	0.0341	0
123	0.0251	0.0018	0
125	0	0.0091	0.0025

- **Modèle de la voyelle « I » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9608 & 0.0392 & 0 & 0 \\ 0 & 0 & 0.9107 & 0.0893 & 0 \\ 0 & 0 & 0 & 0.9116 & 0.0884 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d 'émission :

j	I=1	I=2	I=3
1	0	0.0536	0
3	0	0.0742	0.0149
6	0	0.0179	0
9	0.0392	0	0.2476
24	0	0	0
25	0.0078	0	0
26	0	0	0.0884
29	0.0471	0	0.0354
32	0.0628	0	0
34	0.4313	3.4935 (10^{-4})	0
35	0	2.7984 (10^{-4})	0.0705
38	0	0.0178	0.1238
40	0.0157	0	0
49	0	0.1589	0.0018
63		0.0178	0
64	0.2901	2.3205 (10^{-4})	0.0177
65	0.0078	0	0
66	0	0.0179	0
79	0	0	0.0354
81	0	3.2924 (10^{-4})	0.0177
92	0	0	0
93	0.0117	0.0091	0
95	0.0549	0	0
96	0	0.2272	0.0050
103	0	0.0355	0
104	0	0.0117	0.0238
107	0	0	0.0177
117	0	0.3573	0
118	0.0314	0	0
123	0	0	0.2476

- **Modèle de la voyelle « O » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9573 & 0.0427 & 0 & 0 \\ 0 & 0 & 0.9196 & 0.0804 & 0 \\ 0 & 0 & 0 & 0.9178 & 0.0822 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

j	I=1	I=2	I=3
8	0	0.0161	0
9	0.0513	0	0.1645
13	0	0.0804	0
20	0	0.0161	0
25	0.0171	0	0
27	0	0.2411	0
29	0.0769	$1.2078\ (10^{-4})$	0.3618
32	0.0681	$4.5774\ (10^{-4})$	0
34	0.3932	0	0
41	0	0.0482	0
42	0	$1.2900\ (10^{-4})$	0.0328
44	0	0.0482	0
55	0.0085	0	0
58	0	0.0321	0
61	0.0085	0	0
63	0.0085	0	0
64	0.2821	0	0.1480
66	0	0.0321	0
67	0.0598	$1.4266\ (10^{-4})$	0
69	0	0.1267	0.0019
70	0	0.0964	0
87	0	0.0286	0.0034
90	0	0.0804	0
91	0	0.0161	0
92	0	0	0.0164
93	0.0256	0	0
97	0	0	0.0164
101	0	0.0161	0
102	0	0.0161	0
106	0	0.0263	0.0060
108	0	0.0013	0.0480
110	0	0.0604	0.0038
111	0	$2.8262\ (10^{-4})$	0.0490
112	0	0	0.0164
118	0	0.0161	0.1315

COMMENTAIRES :

- D'après les résultats obtenus, on remarque que les probabilités de transition avec saut – figure (4-3)- (a_{13},a_{24}, a_{35}) sont nulles, ceci confirme ce qui a été dit dans la section(IV-2).

- Pour chaque modèle et pour un état donné de ce modèle, seuls quelques centroides sont émis avec une probabilité non nulle, ceci est du au fait que le mot correspondant à ce modèle se présente sous forme d'une suite de centroides plus ou moins identiques, ce sont ces centroides qui auront des probabilités d'émission non nulles; ce fait est d'ailleurs confirmé dans [9].

TESTS DE RECONNAISSANCE :

Trente versions de chaque mot ont été utilisées pour le test, ce qui donne un corpus total de 150 mots.
Le tableau ci-après présente les résultats de reconnaissance obtenus, le signe « + » indique une réussite de reconnaissance, tandis que le signe « - »indique un échec, dans ce cas le mot reconnu par erreur sera mentionné dans la même case.

Versions de 1 à 14

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	+	+	- C	+	+	+	+	+	+	- O	+	- O	+	+
B	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C	+	+	+	+	+	+	+	+	+	+	+	+	+	+
I	+	+	+	+	+	- C	- B	+	+	+	+	+	+	- C
O	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Versions de 15 à 30

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
A	- O	+	- O	+	+	- O	+	+	+	+	+	+	-B	+	- O	+
B	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C	+	+	+	- I	+	+	+	+	+	+	+	+	+	+	+	+
I	- B	+	+	-O	+	+	+	+	+	+	+	+	+	+	+	+
O	+	+	+	+	+	+	+	+	- C	+	+	+	+	+	+	+

Tableau (4-1) : résultats du test de reconnaissance après apprentissage avec 5 exemples pour chaque HMM

Les taux de reconnaissance sont donnés ci-après dans le tableau (4-2) :

Vocabulaire	Taux de reconnaissance
A	73.33 %
B	100 %
C	96.99 %
I	83.33 %
O	96.66 %

Tableau(4-2) : taux de reconnaissance pour un apprentissage par 5 exemples et un quantificateur à 128 centroides de dimension 10

COMMENTAIRES :

- Le taux de reconnaissance est assez bon surtout pour la consonnes B pour laquelle on enregistre un taux de reconnaissance de 100 %.

- Pour la voyelle A on enregistre un taux de reconnaissance de 73.33 %, soit un taux d'erreur de 26.67 % . On remarque qu'il y a souvent confusion entre le A et le O et parfois (une seule foi dans le tableau (4-1)) avec le B et le C .

- Pour la voyelle I un taux de réussite de 86.66 % a été obtenu soit un taux d'erreur de 13.34 %. On enregistre des confusions avec le C et parfois avec le B et le O.

- Le taux de reconnaissance de la consonne C est de 96.99 % soit un taux d'erreur de3.01 % , une confusion avec le I a été enregistrée.

- Enfin pour la voyelle O on enregistre un taux de réussite de 96.66 % , soit un taux d'erreur de 6.67 %. Une confusion avec le C a été enregistrée.

- Le taux global de réussite est de 90.728 % , soit un taux global d'erreur de 9.272 %.
  Ces résultats montrent la robustesse de la modélisation Markovienne (du moins pour le vocabulaire utilisé) étant donné le faible nombre d'exemples utilisés dans l'apprentissage des modèles HMM.

IV-5-1-2-APPRENTISSAGE AVEC DIX EXEMPLES :

Les paramètres des modèles obtenus après apprentissage sont :

- **Modèle de la voyelle « A » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9163 & 0.0837 & 0 & 0 \\ 0 & 0 & 0.9261 & 0.0739 & 0 \\ 0 & 0 & 0 & 0.9430 & 0.0570 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

j	I1	I2	I3
9	0.1004	0	0.2680
25	0.0083	0	0
26	0.0167	0	0
27	0	0.4732	0
29	0.0418	0	0.1254
32	0.0586	0	0
34	0.4437	0	0
41	0	0.0073	0
44	0.0023	0.0495	$1.1597\ (10^{-4})$
50	0	0.0147	0
52	$2.2004\ (10^{-4})$	0.0441	0
62	0	0.0368	0
63	0	$1.0607\ (10^{-4})$	0
64	0	0	0
66	0	0.1996	0
85	0	0	0.0057
87	0	0.0842	0.0021
91	0	0.0147	0
92	0	0	0.0057
93	0	0	0
118	0	0	0.5927
122	0	0	0.07390

- **Modèle de la consonne « B » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9382 & 0.0618 & 0 & 0 \\ 0 & 0 & 0.9086 & 0.0914 & 0 \\ 0 & 0 & 0 & 0.9370 & 0.0630 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d 'émission :

j	I1	I2	I3
1	0	0	0.0378
3	0	0	0.0252
6	0	0	0.0189
9	0.0611	0.1286	0.0063
10	0	0	0.0441
19	0	0.0274	0
25	0.0375	0	0
26	0	0	0.0504
27	0	0.1610	0.1031
29	0.0786	0,0937	0
32	0.0308	0	0
34	0.5373	0	0
35	0	0	0.0063
38	0	0	0.0126
40	0	0	0.1197
45	0	0	0.0189
49	0	0	0.0063
55	0.0307	1.2272 (10^{-4})	0
63	0.0011	0.0074	0.0063
64	0.1420	0	0
65	0	0.0091	0.0189
66	0	0.0668	0.1177
67	0.0123	0	0
74	0.0032	0.0115	0.0454
77	0.0061	0	0
83	0	0	0.0063
85	0	0.1918	0
92	0	0.2003	0
93	0.0370	0	0
95	0.0061	0.0143	0.1414
96	0	0	0.0189
97	0.0026	0.0508	0
101	0	0.0091	0
103	0	0	0.0126
107	0	0	0.0063
117	0	6.4524 (10^{-4})	0.1634
118	0.0132	0.0169	0
120	0	0.0091	0
125	0	0	0.0063

- **Modèle de la consonne « C » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9293 & 0.0707 & 0 & 0 \\ 0 & 0 & 0.9310 & 0.0690 & 0 \\ 0 & 0 & 0 & 0.9303 & 0.0697 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

j	I1	I2	I3
1	0	0	0.0279
3	0	0	0.0279
6	0	0	0.0209
9	0.1201	0	0.0418
10	0	0	0.0558
25	0.0141	0	0
26	0.0288	0.0064	0.0348
27	0	0.1160	0.0013
29	0.0707	0	0
31	0.0019	0.0119	0
32	0.0424	0	0
34	0.4099	0	0
38	0	0	0.0070
40	0.0015	0.0192	0.1324
55	0.0212	0	0
61	0	0.1931	0
63	0	0.3655	0.0140
64	0.1343	0	0
65	0	0	0.0627
66	0	0.1313	0.0067
67	0.0141	0	0
74	0	0	0.0488
77	0.0141	0	0
86	0	0	0.0070
89	0	0	0.0070
93	0.0565	0	0
95	$4.3014\ (10^{-4})$	0.0134	0.1603
96	0	0	0.0139
103	0	0	0.0209
107	0	0	0.0139
117	0	0.0804	0.2323
118	0.0446	0.0323	0.0558
120	0	0.0276	0
123	0.0252	0.0030	0
125			0.0070
128	0	0	0

- **Modèle de la voyelle «I » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9550 & 0.0450 & 0 & 0 \\ 0 & 0 & 0.9206 & 0.0794 & 0 \\ 0 & 0 & 0 & 0.9239 & 0.0761 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

j	I1	I2	I3
1	0	0.5550	0
3	0	0.0695	0.0018
6	0	0.0235	0
9	0.0314	0	0.3347
15	0	0.0079	0
24	0	0	0.0152
25	0.0044	0	
26	0	0.0078	0.0836
27	0	0.0635	0
29	0.0673	0	0.0532
32	0.0539	0	0
34	0.4626	0	0
35	0	0.0014	0.0594
38	0	0.0122	0.0719
40	0.0089	0	0
49	0	0.0713	$1.3317\ (10^{-4})$
55	0.0089	0	0
63	$1.2423\ (10^{-4})$	0.0077	0
64	0.2784	0	0.0152
65	0.0044	0	0
66	0	0.0873	0
67	0.0134	0	0
74	0	0.0088	0.0295
79	0	0	0.0152
81	0	0.0010	0.0065
83	0	0.0079	0
92	0	0	0.0076
93	0.0154	0.0045	0
95	0.0314	0	0
96	0	0.1503	$4.8317\ (10^{-4})$
103	$5.1834\ (10^{-4})$	0.0229	0
104	0	0.0147	0.0087
107	0	$3.2330\ (10^{-4})$	0.0072
108	0	0	0.0380
117	0	0.3732	0
118	0.0179	0	0
123	0	0.0078	0.2434

- **Modèle de la voyelle « O » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9579 & 0.0395 & 0.0026 & 0 \\ 0 & 0 & 0.8685 & 0.1315 & 0 \\ 0 & 0 & 0 & 0.9174 & 0.0826 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
8	0	0.0127	0.0127
9	0.0505	0	0.1156
13	0	0	0.0188
18	0	0.0140	0.0083
20	0	0.0420	0
25	0.0084	0	0
26	0.0084	0	0.0165
27	0	0.0645	0.2428
29	0.0560	0.0030	0.1362
32	0.0654	0.0065	0
34	0.4286	0.0024	0
40	0.0084	0.0024	0
41	0	0.0560	0
42	0	0	0.0165
43	0	0	0.0083
44	0	0.0660	0.0024
48	0	0.0140	0
51	0	0.0129	$6.4659\ (10^{-4})$
55	0.0084	0	0
58	0	0.0391	0.0017
61	0.0042	0	0
63	0.0042	0	0
64	0.2734	$6.1103\ (10^{-4})$	0.0083
66	0	0.0536	0.0262
67	0.0449	0.0048	0
69	0	0.1506	0.0351
70	0	0.1168	0.0220
74	0	0	0. 0165
87	0	0.0238	0.0025
90	0	0.1081	0.0024
91	0	0.0140	0
92	0	0	0.0083
93	0.0253	0	0
101	0	0	0.0083
102	0	0.0280	0
105	0	0	0.0083
106	0	0	0.0246
108	0	$2.8205\ (10^{-4})$	0.0578
110	0	0.0650	0.0112
111	0	0	0.0578
112	0	0	0.0330
118	0.0054	0.0132	0.0973
123	0.0084	0	0

TESTS DE RECONNAISSANCE :

Versions de 1 à 14

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	+	+	- C	+	+	+	+	+	+	+	+	+	+	- O
B	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C	+	+	+	+	+	+	+	+	+	+	+	+	+	+
I	+	+	+	+	+	- B	+	+	+	+	+	+	+	+
O	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Versions de 15 à 30

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
A	- O	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
B	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
I	+	+	+	- O	+	+	+	+	+	+	+	+	+	+	+	+
O	+	+	+	+	+	+	- B	+	- C	+	+	+	+	+	+	+

Tableau (4-3) : résultats du test de reconnaissance après apprentissage avec 10 exemples pour chaque HMM

Les taux de reconnaissance sont donnés ci-après dans le tableau (4-4) :

Vocabulaire	Taux de reconnaissance
A	90 %
B	100 %
C	100 %
I	93.33 %
O	93.33 %

Tableau(4-4) : taux de reconnaissance pour un apprentissage par 10 exemples et un quantificateur à 128 centroides de dimension 10

COMMENTAIRES :

- Le taux de reconnaissance c'est nettement amélioré pour tous les mots du vocabulaire
  - le taux de reconnaissance de la voyelle A qui était de 73.33 % est monté à 90 %.
  - le taux de reconnaissance de la consonne C qui était de 96.99 % est monté à100 %.
  - Le taux de reconnaissance de la voyelle I qui était de 86.66 % est monté à 93.33 %.
  - Le taux global de reconnaissance est donc monté de 90.728% à 95.332%,

  ceci confirme l'effet positif de l'augmentation du nombre d'exemples d'apprentissage.

Nous avons donc augmenté le nombre d'exemples d'apprentissage à 20 les résultats sont présentés ci-après.

IV-5-1-3-APPRENTISSAGE AVEC VINGT EXEMPLES :

Les paramètres des modèles obtenus après apprentissage sont :

- **Modèle de la voyelle « A » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9383 & 0.0617 & 0 & 0 \\ 0 & 0 & 0.9288 & 0.0712 & 0 \\ 0 & 0 & 0 & 0.9436 & 0.0564 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

matrice des probabilités d'émission :

J	I1	I2	I3
9	0.0678	0	0.1381
25	0.0092	0	0
26	0	0	0
27	0	0.4665	0
29	0.0863	0	0.10150
32	0.0493	0	0
34	0.4038	0	0
41	0	0.0036	0
44	0	0.0314	5.1496 (10^{-4})
50	0	0.0071	0
52	0	0.0748	0
55	0.0154	0	0.0056
62	4.0184 (10^{-4})	0.0423	0
63	0.0370	0	0
64	0.1788	0	0.0085
66	0	0.2030	0
67	0.0247	0	0
74	7.5154 (10^{-4})	0.0027	0
80	0	0.0106	0
87	0.0036	0.0455	0.0312
91	0	0.0264	0.0017
92	0	0	0.0056
93	0.0339	0	0
95	0.0092	0	0
117	0	0.0214	0
118	0.0458	7.2534 (10^{-4})	0.7073
122	0	0.0641	0
123	0.0123	0	0

- **Modèle de la voyelle « B » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9399 & 0.0601 & 0 & 0 \\ 0 & 0 & 0.9404 & 0.0596 & 0 \\ 0 & 0 & 0 & 0.9314 & 0.0686 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0	0.0059	0.0344
3	0	0	0.0240
6	0	0	0.0171
9	0.0485	0.0592	0.0514
10	0	$3.3371\ (10^{-4})$	0.0442
19	0	0.0149	0
25	0.0391	0	0
26	0	0	0.0651
27	0	0.2055	0
29	0.1136	0.0344	0
31	0	0	0.0034
32	0.0331	0	0
34	0.5142	0	0
35	0	0	0.0240
38	0	0	0.0240
40	0	0.0030	0.1920
45	0	0	0.0206
49	0	$8.1189\ (10^{-4})$	0.0059
55	0.0271	0	0
58	0	0.0030	0
63	0.0020	0.0040	0.0137
64	0.1413	0	0
65	0	0.0060	0.0137
66	0	0.1638	0
67	0.0120	0	0
74	0.0028	0.0062	0.0651
77	0.0030	0	0
83	0	0	0.0034
85	$1.1690\ (10^{-4})$	0.1518	0.0034
92	$3.0586\ (10^{-4})$	0.1427	0
93	0.0421	0	0
95	0.0030	0.0089	0.2949
96	0	$3.4575\ (10^{-4})$	0.0202
97	0.0097	0.0113	0
101	0.0030	0.0030	0
103	0	0	0.0274
104	0	0	0.0034
107	0	0	0.0034
108	0	0	0.0130
117	0	0.1505	0.0051
118	0.0081	0.0188	0.0137
120	0	0.0060	0
123	0	0	0.0069
125	0	$9.8802\ (10^{-4})$	0.0023
126	0	0	0.0069

- **Modèle de la consonne « C » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9392 & 0.0608 & 0 & 0 \\ 0 & 0 & 0.9330 & 0.0670 & 0 \\ 0 & 0 & 0 & 0.9399 & 0.0601 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0	0	0.0301
3	0	0	0.0120
6	0	0	0.0090
9	0.1367	$1.0213\ (10^{-4})$	0.0631
10	0	0	0.0721
25	0.0122	0	0
26	0.0090	0.0303	0.0391
27	0	0.0975	0.0358
29	0.0517	0	0.0060
31	0	0.0066	0
32	0.0425	0	0
34	0.4376	0	0
38	0	0	0.0120
40	0	0.0168	0.1473
45	0	0	0.0090
55	0.0152	0	0
61	0	0.1642	0
63	0	0.3720	0.0091
64	0.1732	0	0
65	0	0.0032	0.0423
66	0	0.1533	0.0248
67	0.0122	0	0
74	0	0	0.0391
77	0.0122	0	0
86	0	0	0.0030
89	0	0	0.0030
93	0.0608	0	0
95	0.0067	0.1533	0.1383
96	0	0	0.0150
103	0	0	0.0150
107	0	0	0.0060
108	0	0	0.0080
112	0	0	0.0030
117	0	0.0290	0.1935
118	0.0157	0.0564	0.0601
120	0	0.0268	0
123	0.0145	0.0343	0
125	0	0	0.0030

- **Modèle de la voyelle « I » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9545 & 0.0455 & 0 & 0 \\ 0 & 0 & 0.9227 & 0.0773 & 0 \\ 0 & 0 & 0 & 0.9237 & 0.0763 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0	0.0541	0
3	0	0.0669	0.0027
6	0	0.0114	0
9	0.0569	0	0.3052
10	0	0.0033	$5.4742\ (10^{-4})$
15	0	0.0039	0
24	0	0	0.0191
25	0.0159	0	0
26	0	0.0041	0
27	0	0.0425	0.0951
29	0.0660	0	0.0305
32	0.0569	0	0
34	0.4507	0	0
35	0	0.0358	0.0334
38	0	0.0141	0.0662
40	0.0046	0	0.0076
49	0	0.0424	$1.1300\ (10^{-4})$
55	0.0091	0	0
59	0	0	0.0114
63	0	0.0037	0
64	0.2618	0	0.0076
65	0.0023	0	0
66	0	0.1121	0
67	0.0273	0	0
74	0	0.0095	0.0211
79	0	0	0.0076
81	0	$3.9317\ (10^{-4})$	0.0034
83	0	0.0039	0
86	0	0	0.0038
89	0	0	0.0038
92	0	0	0.0038
93	0.0203	$2.8755\ (10^{-4})$	0
95	0.0159	0	0.0191
96	0	0.1157	$2.6805\ (10^{-4})$
103	$5.7902\ (10^{-4})$	0.0183	0
104	0	0.0125	0.0067
107	0	$1.2495\ (10^{-4})$	0.0037
108	0	0	0.0801
117	0	0.4409	0
118	0.0114	0	0.0038
123	0	0.0038	0.2632

- **Modèle de la voyelle « O » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9595 & 0.0405 & 0 & 0 \\ 0 & 0 & 0.9174 & 0.0826 & 0 \\ 0 & 0 & 0 & 0.9108 & 0.0892 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

matrice des probabilités d'émission :

J	I1	I2	I3
8	0	0.0124	0
9	0.0628	0	0.2944
12	0	0.0041	0
13	0	0.0578	0
14	0	0.0041	0
18	0	0.0150	0.0150
20	0	0.0289	0
25	0.0101	0	0
26	0.0041	0	0.0089
27	0	0.3966	0
29	0.0587	0	0.2810
32	0.0647	1.3956 (10^{-4})	0
34	0.0410	0	0
40	0.0061	0	0
41	0	0.0248	0
42	0	0	0.0089
43	0	0.0022	0.0066
44	0	0.0409	4.4410 (10^{-4})
47	0	0.0041	0
48	0	0.0124	0
51	0	0.0041	0
55	0.0081	0	0
58	0	0.0330	0
61	0.0020	0	0
63	0.0041	0	0
64	0.2977	0	0.0625
66	0	0.0661	0
67	0.0362	5.3044 (10^{-4})	0
69	0	0.0863	0.0184
70	0	0.0577	1.8352 (10^{-4})
74	0	0	0.0089
87	0	0.0066	0.0017
90	0	0.0577	1.4465 (10^{-4})
91	0	0.0164	0
92	0.0020	0	0.0045
93	0.0203	0	0
97	0	0	0.0045
101	0	0.0041	0
102	0	0.0165	0
105	0	0.0048	0.0082
106	0	0.0115	0.0010
108	0	0	0.0491
110	0	0.0183	0.0249
111	0	0	0.0713
112	0	0	0.0223
118	0.0181	0.0045	0.1071
119	0	0.0083	0
123	0.0041	0	0

TESTS DE RECONNAISSANCE :

Versions de 1 à 14

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	+	+	+	+	+	+	+	+	+	+	+	+	+	- O
B	+	+	+	+	+	+	+	+	+	+	+	+	+	- O
C	+	+	+	+	+	+	+	+	+	+	+	+	+	+
I	+	+	+	+	+	+	- B	+	+	+	+	+	+	+
O	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Versions de 15 à 30

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
B	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
I	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
O	+	+	+	+	+	+	+	+	- I	+	+	+	+	+	+	+

Tableau (4-5) : résultats du test de reconnaissance après apprentissage avec 20 exemples pour chaque HMM

Les taux de reconnaissance sont donnés ci-après dans le tableau (4-6) :

Vocabulaire	Taux de reconnaissance (%)
A	96.66
B	96.66
C	100
I	96.66
O	96.66

Tableau(4-6) : taux de reconnaissance pour un apprentissage par 20 exemples et un quantificateur de 128 centroides de dimension 10

COMMENTAIRES :

Les résultats de reconnaissance obtenus montrent une très nette amélioration du taux de reconnaissance :

- Le taux de reconnaissance global a atteint 97.328 % soit un taux d'erreur de 2.672 %.

- Les valeurs des probabilités de transition n'ont pas changé remarquablement lors de l'augmentation du nombre d'exemples d'apprentissage, d'ailleurs entre deux HMMs correspondant à deux mots différents ces paramètres ne sont pas très différents, ceci montre que les a_{ij} ne participent que faiblement dans le processus de reconnaissance

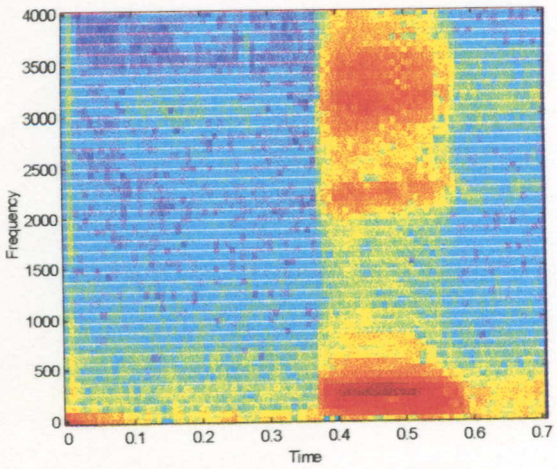
contrairement aux probabilités d'émission qui varient dans une large mesure, ce fait est d'ailleurs confirmé dans [5]

- Remarquons aussi un fait qui semble bizarre à première vue :
  La 14^{ième} version de test des voyelles « A » et « B » ont été reconnues après apprentissage par cinq exemples de chaque HMM, l'on pourrait s'attendre à ce que ces versions le soient (d'avantage) après apprentissage par vingt exemples pour chaque HMM, ce qui n'a pas été le cas. N'ayant changé aucun paramètre à part le nombre d'exemples d'apprentissage l'explication de ce phénomène peut être la suivante :
  Nous avons souligné dans la section (III-6-3) du chapitre – III – le caractère peu discriminant des HMMs standards, chaque HMM est entraîné séparément de manière à approximer la densité de probabilité du mot correspondant .En changeant donc le nombre d'exemples d'apprentissage, ce sont les frontières de décision dans l'espace des mots qui changent; ce qui s'est donc probablement passé c'est que la 14^{ième} version de test de « B » qui appartenait à la frontière de décision de la classe (B) lors de l'apprentissage par cinq exemples, est devenue appartenant à la frontière de décision de la classe(O). La meme explication est valable pour la voyelle « A ». Nul besoin de noter que ces cas 'singuliers' n'affectent en rien l'apport de l'augmentation du nombre d'exemples d'apprentissage prouvé d'ailleurs par l'augmentation du taux global de reconnaissance de 90.728 à 96.662 %. Toutefois on peut se poser la question suivante : pourquoi les frontières de décision de la classe (A) ont elles interféré avec celles de la classe (O) plutôt qu'avec celles d'une autre classe. Pour répondre à cette question nous sommes allé chercher dans les spectrogrammes des différents mots du vocabulaire. Ces spectrogrammes sont présentés dans les figures (4-9) et (4-10). On remarque une très nette ressemblance entre le spectrogramme de la voyelle « A » et celui de la voyelle « O » ce qui explique la confusion souvent remarquée entre le A et le O

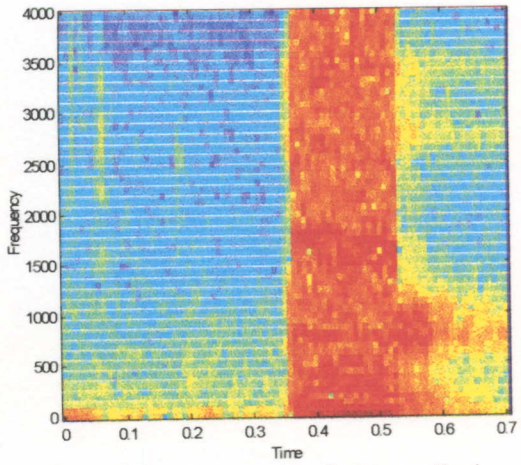
Le tableau suivant récapitule les résultats obtenus précédemment :

	Voyelle A	Consonne B	Consonne C	Voyelle I	Voyelle O	Taux global
Apprentissage 5 exemple	73.33 %	100 %	96.66 %	83.33 %	96.66 %	89.996
Apprentissage 10 exemple	90 %	100 %	100 %	93.33 %	93.33 %	95.332 %
Apprentissage 20 exemple	96.66 %	96.66 %	100 %	96.66 %	96.66 %	97.328 %

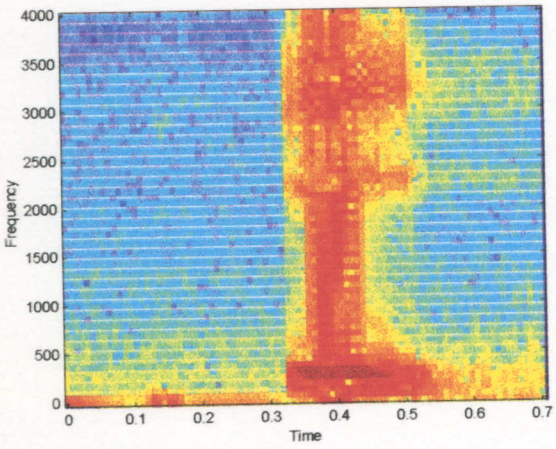
Tableau(4-7) :récapitulation des résultats de reconnaissance



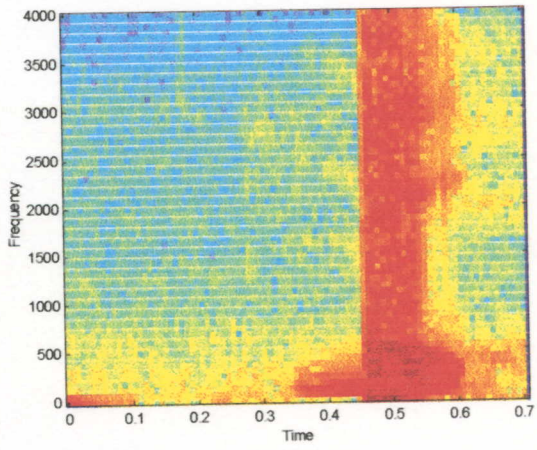
a- spectrogramme de la voyelle I



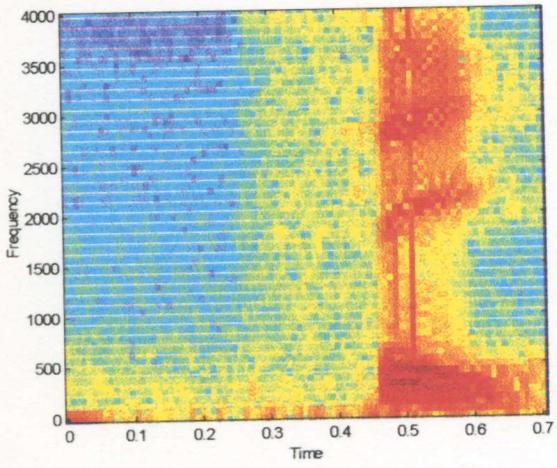
b- spectrogramme de la voyelle A



c- spectrogramme de la voyelle O

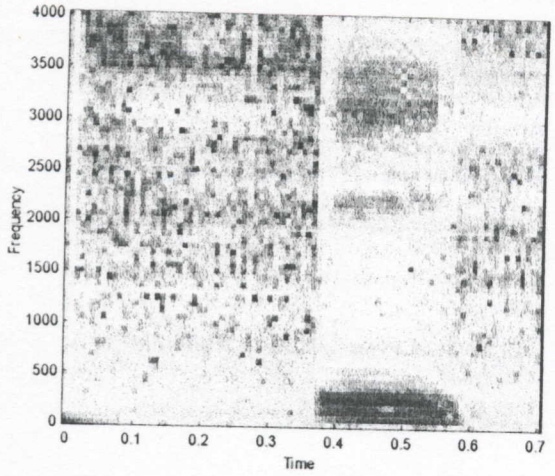


d- spectrogramme de la consonne B

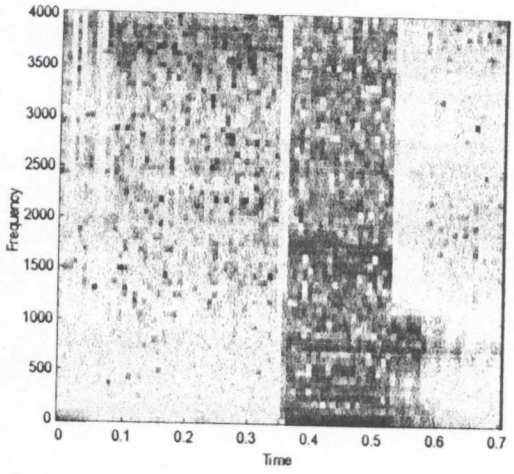


e- spectrogramme de la consonne C

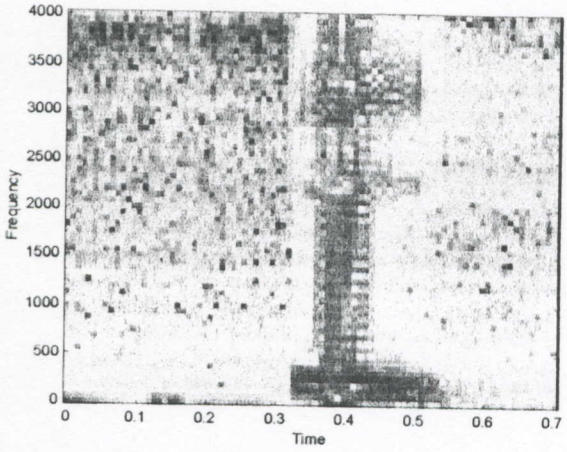
Figure(4-9) : spectrogrammes des alphabets **A, B, C, I, O**



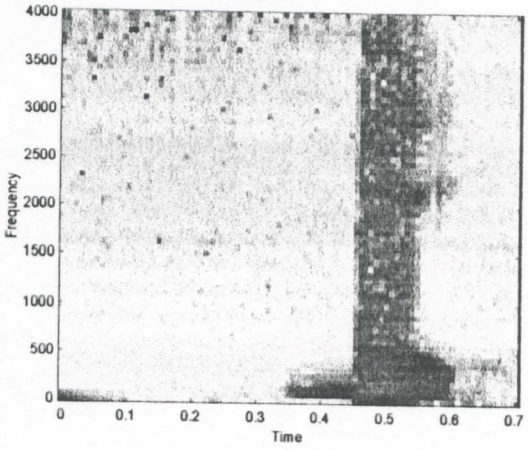
a- spectrogramme de la voyelle I



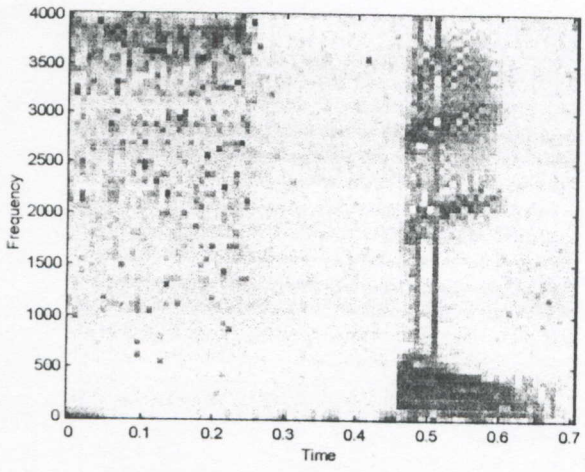
b- spectrogramme de la voyelle A



c- spectrogramme de la voyelle O



d- spectrogramme de la consonne B



e- spectrogramme de la consonne C

Figure(4-10) : spectrogrammes en niveaux de gris des alphabets **A, B, C, I, O**

IV-5-2-EFFET DE LA QUANTIFICATION VECTORIELLE :

Pour mettre en évidence l'effet de la quantification vectorielle, nous avons baissé le nombre de centroides du quantificateur vectoriel à 64 centroides, tout en laissant la taille du corpus d'apprentissage égale à 20. Les paramètres des modèles obtenus après apprentissage sont les suivants :

- **Modèle de la voyelle « A » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9578 & 0.0418 & 0.0004 & 0 \\ 0 & 0 & 0.8529 & 0.1382 & 0.0089 \\ 0 & 0 & 0 & 0.9251 & 0.0749 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0.0084	0	0
3	0	0	0.0119
4	0.0169	0	0
6	0.0337	0	0.0040
7	0	0.0383	0.0790
8	0.0042	0	0
9	0.0169	0	0
10	0	4.0428 (10^{-4})	0.0954
13	0.0021	0	0
15	0.0729	0.0031	0
17	0	0.0210	0.0763
19	0	0	0.2111
20	0.1240	0.0088	0
24	0.0064	0.0072	0
25	0.0127	0	0
27	0.0270	0.0014	0.0119
28	0	0	0.1155
29	0.0063	0	0
30	0	0.0220	0.0559
31	0.0084	0.0066	4.2057 (10^{-4})
33	0.0114	0.0046	0
34	0	0.1698	0.0204
35	0	0	0.0717
37	0.0016	0.0019	0
40	0.0328	0.0035	0
44	0	0.1323	0.0047
45	0.0211	0	0
46	0.0253	0.0071	0.0122
47	0.3343	0.4714	0
48	0.1053	5.0832 (10^{-4})	0
52	0	0.0279	0.0169
53	0	1.0662 (10^{-4})	0.0438
56	0.0356	0.0676	0
57	0	0	0.1673
58	0.0063	0	0
59	0.0232	0	0
61	0.0632	0.0047	0.0016

- **Modèle de la consonne « B » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.9931 & 0.0069 & 0 & 0 \\ 0 & 0.8867 & 0.1132 & 0 & 0 \\ 0 & 0 & 0.9554 & 0.0413 & 0.0033 \\ 0 & 0 & 0 & 0.9218 & 0.0782 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
3	7.7513 (10^{-4})	0.0019	0.3380
6	0.0285	0	0
7	0	0	0.0042
9	0.0057	0	0
10	0	0	0.0085
13	0	0.0045	0
14	0	0	0.0085
15	0.0806	0.0019	0
16	0	0	0.0042
19	0	0	0.0042
20	0.2391	0.1059	0.0154
21	1.6097 (10^{-4})	0	0
25	0	0.0045	0
27	0.0684	0	0
28	0	0	0.0042
29	0.0043	0.0050	0
30	0	0	0.0296
33	0.0029	0.0100	0
35	0	0	0.0042
37	0.0114	0	0
40	0.0341	0.0017	0.0011
45	0.0057	0.0043	3.3385 (10^{-4})
46	0.0114	0	0.0253
47	0.1351	0.6570	0
48	0.1585	0.0320	0.0163
50	0	0	0.4774
55	1.7175 (10^{-4})	0.0022	0
56	0.0659	0.1528	0
57	0	0	0.0423
58	0.0057	0	0
59	0.0390	0.0067	0.0049
61	0.1224	0.0075	0.0113

- **Modèle de la consonne « C » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.9999 & 0.0001 & 0 & 0 \\ 0 & 0.9202 & 0.0797 & 0.0001 & 0 \\ 0 & 0 & 0.9471 & 0.0399 & 0.0131 \\ 0 & 0 & 0 & 0.9350 & 0.0650 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0.0039	0.0027	0
4	0	0	0.0949
5	0.0040	0	0.1338
6	0.0086	0.0312	0.0045
7	0	0	0.0129
8	0	0	0.1208
9	0.0040	0.0071	0.0963
10	0	0	0.0129
13	0	0.0026	0
14	0	0	0.0043
15	0.1355	0.0213	0
20	0.1751	0.0241	0
24	0	0	0.1597
25	0.0107	0.0035	0
27	0.1128	0.0258	0
28	0	0	0.0086
29	0.0079	0	0
30	0	0	0.0043
31	0	0	0.0820
33	0.0190	0.0086	0
34	0	0	0.0043
37	0.0080	0.0026	0
40	0	0.0136	0
45	0	0	0
46	0	0.0256	0.0015
47	0	0.7393	0
48	0	0.0200	0
53	0	0	0.0129
54	0	0	0.0043
56	0	0.0479	0
57	0	0	0.0302
58	0	0	0.1079
59	0	0.0078	$2.2173\ (10^{-4})$
60	0	$1.0697\ (10^{-4})$	0.0948
61	0	0.0162	0
64	0	0	0.0086

- **Modèle de la voyelle « I » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.9413 & 0.0587 & 0 & 0 \\ 0 & 0.9066 & 0.0933 & 0 & 0 \\ 0 & 0 & 0.9658 & 0.0226 & 0.0116 \\ 0 & 0 & 0 & 0.8197 & 0.1803 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
5	0.0050	0	0.0272
6	0.0545	0	0.0136
9	0.0496	0	0.0136
13	0.0097	0	0
14	0	0	0.0272
15	0.1032	3.8206 (10^{-4})	0.0268
16	0	0	0.1769
20	0.1668	0.2053	0.0172
24	0	0	0.0408
25	0.0061	0.0164	0.0024
27	0.0545	0	0.0136
28	0	0	0.0136
29	0.0082	0.0040	0
30	0	0	0.0272
31	0.0050	0	0
33	0.0130	0.0313	0.0011
35	0	0	0.1769
36	0.0050	0	0
37	0.0050	0	0
40	0.0301	0.0084	0
45	0.0274	0.0230	0
46	0.0446	0	0
47	0.0519	0.5346	0.0799
48	0.1777	0.0562	0.0042
50	0.0050	0	0
53	0	0	0.0544
54	0	0	0.1906
55	5.9783 (10^{-4})	0.0032	0
56	0.0138	0.1051	0.0110
57	0	0	0.0136
59	0.0041	0.0035	0
61	0.0992	0.0085	0
64	0	0	0.0681

- **Modèle de la voyelle « O » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9392 & 0.0608 & 0 & 0 \\ 0 & 0 & 0.9330 & 0.0670 & 0 \\ 0 & 0 & 0 & 0.9399 & 0.0601 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0.0037	0	0.0063
3	0	0	0.0032
4	0.0074	0	0
5	0.0037	0	0
6	0.0409	0	0
9	0.0186	0	0
13	0.0036	1.4298 (10^{-4})	0
15	0.0852	0.0040	0
20	0.0959	0.3152	0.1660
24	0.0112	0	0
25	0.0050	0.0085	0.0168
27	0.0484	0.0072	0.0032
29	0	5.8480 (10^{-4})	0.0122
30	0.0037	0	0.0063
31	0.0149	0	0
33	0.0225	0.0426	0.0197
40	0.0856	0.0136	8.4411 (10^{-4})
45	0.0275	0.0412	0.0325
46	0.0074	0	0.0032
47	0.0329	0.4067	0.6148
48	0.2150	0.0786	0.0240
49	0	0	0.0348
55	0	0.0063	0.0040
56	0.0119	0.0532	0.0448
59	0.1039	0.0070	0.0069
60	0.0074	0	0
61	0.1438	0.0152	5.4073 (10^{-4})

TESTS DE RECONNAISSANCE :

Les taux de reconnaissance sont donnés ci-après dans le tableau (4-8) ,les résultats de reconnaissance obtenus dans le cas précédent sont aussi mentionnés pour des fins de comparaison:

	Voyelle A	Consonne B	Consonne C	Voyelle I	Voyelle O	Taux global
Cas de 128 centroides	96.66 %	96.66 %	100 %	96.66 %	96.66 %	97.33 %
Cas de 64 centroiodes	63.33 %	73.33 %	63.33 %	40 %	56.66 %	59.33 %

Tableau(4-8) : tableau comparatif pour un apprentissage par 20 exemples et des centroides de dimension 10

COMMENTAIRES :

- Le taux de reconnaissance a baissé d'une manière très remarquable; toutefois cette baisse était prévisible. En effet la diminution du nombre de centroides réduit la finesse de la quantification vectorielle (l'erreur de quantification augmente).

- Il faut aussi noter que cette diminution du nombre de centroides à 64 à permis de réduire le temps nécessaire à l'apprentissage ainsi qu'à la reconnaissance dans un rapport de 1/2 Une telle diminution du temps de calcul peut être qualifiée de très importante pour une application en temps réel telle que la reconnaissance (le problème se pose de manière moins critique pour l'apprentissage qui se fait une fois pour toute et en temps différé). Le but prioritaire recherché étant l'amélioration du taux de reconnaissance, nous avons tenté de garder le nombre de centroides à 64 mais en augmentant la taille des centroides (nombre de coefficients cepstraux); les résultats obtenus sont présentés dans la section suivante.

IV-5-3-EFFET DU NOMBRE DE COEFFICIENTS CEPSTRAUX :

Il va sans dire que le nombre de coefficients cepstraux joue un rôle important dans la qualité de codage du signal vocal. En effet plus ce nombre est élevé plus le spectre court terme (sur une fenêtre) est fidèlement représenté. Le nombre de coefficients utilisé jusqu'à maintenant étant 10, nous avons augmenté ce nombre à 16, dans une tentative d'amélioration des résultats du cas précédent (c à d un corpus d'apprentissage de taille20et un quantificateur à 64 centroides).
Les paramètres des modèles obtenus après apprentissage sont :

- **Modèle de la voyelle « A » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.9856 & 0.0144 & 0 & 0 \\ 0 & 0.9376 & 0.0605 & 0.0019 & 0 \\ 0 & 0 & 0.9584 & 0.0352 & 0.0064 \\ 0 & 0 & 0 & 0.9042 & 0.0958 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0.0032	0	0
2	0.0036	0.0146	$2.0215\ (10^{-4})$
4	0	$2.3455\ (10^{-4})$	0.0106
5	0.0158	$2.3920\ (10^{-4})$	0.0050
7	0.0063	$1.6574\ (10^{-4})$	0.0052
8	0.0093	0.0065	0
10	0	0	0.0282
11	0.1266	0.0321	0
14	0.0190	0	0
16	0	$4.8028\ (10^{-4})$	0.0325
17	0.0113	0.0031	0
20	0.1285	0.5883	$3.1832\ (10^{-4})$
22	0.0111	0.0353	0
23	0	0	0.0225
24	0	0	0.0167
25	0.1589	0.1217	0
26	0	0	0.0619
28	0	0	0.0113
29	0.2122	0.1114	0
30	0.0758	0.0194	0
31	0.0430	0.0052	0
32	0.0091	0.0024	0
34	0.0032	0	0
35	0	0	0.0732
40	0.0032	0	0
41	0	0.0086	0
43	0	0.0021	0
44	0	0	0.0113
45	0.0029	0.0045	0
46	0	0	0.0337
49	0	0	0.0056
54	0	0	0.0282
58	0	0	0.0168
59	0.0847	0.0326	$1.7280\ (10^{-4})$
60	0	0.0109	$1.6866\ (10^{-4})$
62	0	0	0.0056
64	0	0	0.6308

- **Modèle de la consonne « B » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.9984 & 0.0016 & 0 & 0 \\ 0 & 0.8702 & 0.1298 & 0 & 0 \\ 0 & 0 & 0.9566 & 0.0397 & 0.0037 \\ 0 & 0 & 0 & 0.9256 & 0.0744 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
2	0.0310	0.0200	0
6	3.5889 (10^{-4})	0.0021	0.3012
7	0	0	0.0244
11	0.0130	0	0.0163
14	0	0.0022	0
20	0.2089	0.8314	0
22	0.0023	0.0105	0
25	0.2787	0.0991	0.0304
29	0.2378	0.0289	0.0655
30	0.0975	0	0
32	0.0130	0	0
35	0	0	0.0041
36	0	0	0.4966
39	0.0065	0	0
43	0.0065	0	0
49	0	0	0.0081
54	0	0	0.0041
59	0.1044	0.0021	0
60	0	0.0020	2.5954 (10^{-4})
62	0	0.0021	1.7622 (10^{-4})
63	0	0	0.0041
64	0	0	0.0448

- **Modèle de la consonne « C » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.9997 & 0.0003 & 0 & 0 \\ 0 & 0.9296 & 0.0703 & 0 & 0 \\ 0 & 0 & 0.9430 & 0.0427 & 0.0142 \\ 0 & 0 & 0 & 0.9332 & 0.0668 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
2	0.0106	0.0010	0
5	0	0	0.1646
7	0	0	0.0089
8	0	0	0.0934
9	0	0	0.0473
10	0	0	0.0044
11	0.0230	0.0661	0.0012
14	0	0	0.0890
16	0	0	0.0089
17	0	0	0.1290
20	0.3004	0.7909	0
22	0.0273	0.0064	0
23	0	0	0.0133
25	0.1719	0.0204	0
26	0	0	0.0133
29	0.2032	0.0377	0
30	0.1028	0.0279	0
31	0.0035	0	0.1378
32	0.0035	0	0
34	0.0035	0	0.1646
35	0	0	0.0133
45	0.0043	0.0022	0
49	0	0	0.0044
52	0	0	0.0089
54	0	0	0.0089
59	0.1425	0.0242	0
60	0.0035	0.0231	0.0618
64	0	0	0.0267

- **Modèle de la voyelle « I » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.9759 & 0.0241 & 0 & 0 \\ 0 & 0.9074 & 0.0926 & 0 & 0 \\ 0 & 0 & 0.9680 & 0.0242 & 0.0078 \\ 0 & 0 & 0 & 0.8789 & 0.1211 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
2	0.0221	0.0165	0
5	0	0	0.0240
7	0	0	0.0561
8	0	0	0.0080
11	0.0949	0	0.0079
20	0.0140	0.7328	0.0034
22	0.0302	0.0186	0
23	0	0	0.0240
25	0.2248	0.1803	$1.3354\ (10^{-4})$
26	0	0	0.2245
29	0.2656	0.0329	0.0037
30	0.0665	0	0.0159
31	0.0047	0	0
32	0	0.0032	0
34	0.0047	0	0.0721
35	0	0	0.0401
36	0.0047	0	0
39	0.0047	0	0
40	0.0051	0.0056	0.0031
41	0.0048	0.0032	0
45	0	0.0016	0
49	0	0	0.0240
51	0	0	0.3126
52	0	0	0.1443
55	$8.2600\ (10^{-4})$	0.0013	0
59	0.1101	0.0037	0.0120
60	0.0522	0	0.0159
64	0	0	0.0080

- **Modèle de la voyelle « O » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.8899 & 0.1101 & 0 & 0 \\ 0 & 0.9353 & 0.0638 & 0.0010 & 0 \\ 0 & 0 & 0.9676 & 0.0303 & 0.0020 \\ 0 & 0 & 0 & 0.7506 & 0.2494 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
2	0.0134	0.0215	0.0288
5	0.0073	0	0
6	0	0	0.0266
7	0..0037	0	0.1993
10	0	0	0.0133
11	0.0509	0	0.0133
14	0.0036	0	0
17	0.0073	0	0
20	0.0587	0.6832	0.1756
22	0.0255	0.0489	0.0019
25	0.1721	0.1873	0.0332
26	0	0	0.0532
29	0.4537	0.0358	0.0187
30	0.0543	0.0036	0.0246
31	0.0218	0	0
32	0.0073	0.0033	0
34	0.0036	0	0
35	0	0	0.1063
36	0	0	0.0133
39	0.0036	0	0
40	0.0070	0.0042	0.0066
41	0	0.0066	0
44	0	0	0.0399
45	0	0.0014	0.0019
46	0	0	0.0133
58	0	0	0.2127
59	0.0879	0.0041	0.0043
60	0.0182	0	0.0133

TESTS DE RECONNAISSANCE :

Les taux de reconnaissance sont donnés ci-après dans le tableau (4-9), le résultat pour le cas de centroides de dimension 10 est inclus pour des fins de comparaison.

	Voyelle A	Consonne B	Consonne C	Voyelle I	Voyelle O	Taux global
Centroides de dimension 10	63.33 %	73.33 %	63.33 %	40 %	56.66 %	59.33 %
Centroiodes de dimension 16	80 %	66.66 %	56.66 %	43.33 %	43.33 %	57.99 %

Tableau(4-9) : tableau comparatif pour un apprentissage par 20 exemples et un quantificateur à 64 centroides

COMMENTAIRES :

- Le taux de reconnaissance a augmenté pour certains mots et a diminué pour d'autres, le taux global de reconnaissance a légèrement diminué par rapport au cas précédent.

- Une mauvaise interprétation de ce dernier résultat conduirait à dire que l'augmentation du nombre de coefficients cepstraux n'a pas d'effet positif sur le taux de reconnaissance; en fait il n'en est rien. S'il y a une conclusion à tirer de ce résultat c'est que le nombre de coefficients cepstraux utilisé avant cela (10 coefficients) était suffisant si bien que lorsqu'on a augmenté ce nombre à 16 il n'y a pas eu d'amélioration du taux de reconnaissance. Mais alors le taux de reconnaissance aurait due au moins rester inchangé par rapport au cas de centroides de dimension 10; une explication plausible de cette légère diminution du taux de reconnaissance peut être le fait que les coefficients cepstraux s'adonnent mal à une quantification si bien qu'il constituent une source d'erreur.

Le tableau suivant récapitule tous les résultats précédents :

Vocabulaire ⟶	A	B	C	I	O	Taux global
Corpus d'apprentissage de taille 5 Quantificateur de 128 centroides de dimension 10	73.33 %	100 %	96.66 %	83.33 %	96.66 %	90.73 %
Corpus d'apprentissage de taille 10 Quantificateur de 128 centroides de dimension 10	90 %	100 %	100 %	93.33 %	93.33 %	95.33 %
Corpus d'apprentissage de taille 20 Quantificateur de 128 centroides de dimension 10	96.66 %	96.66 %	100 %	96.66 %	96.66 %	97.33 %
Corpus d'apprentissage de taille 20 Quantificateur de 64 centroides de dimension 10	63.33 %	73.33 %	63.33 %	40 %	56.66 %	59.33 %
Corpus d'apprentissage de taille 20 Quantificateur de 64 centroides de dimension 16	80 %	66.66 %	56.66 %	43.33 %	43.33 %	57.99 %

Tableau( ) : Récapitulation de tous les résultats trouvés pour le vocabulaire (**A,B ,C ,I ,O**)

On remarque que les meilleurs résultats correspondent au cas d'un apprentissage par vingt exemples et d'un quantificateur de 128 centroides de dimension 10. C 'est ce cas qui sera considéré dans ce qui suit pour le nouveau vocabulaire constitué par les chiffres de un à dix prononcés en langue française.

IV-5-4-EFFET DE LA TAILLE DU VOCABULAIRE :

Tous les paramètres sur lesquels on avait joué jusqu'à maintenant sont indépendants du vocabulaire en question . Comme dernière étape dans ce travail, nous avons donc changé le vocabulaire utilisé pour les tests; le nouveau vocabulaire est constitué par les chiffres de un à neuf prononcés en langue française. 120 versions (dont 20 sont utilisées pour l'apprentissage et 100 pour les tests de reconnaissance) de chaque chiffre ont été enregistrées, ce qui donne un corpus total de 1080 mots. Le quantificateur vectoriel aura un dictionnaire de 128 centroides de dimension 10. Les modèles HMM utilisés sont toujours ceux de la figure (4-3)
Les paramètres des modèles obtenus après apprentissage sont les suivants :

- **Modèle du chiffre « 1 » :**

$$A=\{a_{ij}\}=\begin{bmatrix} 0 & 0.9943 & 0.0057 & 0 & 0 \\ 0 & 0.9432 & 0.0568 & 0 & 0 \\ 0 & 0 & 0.9634 & 0.0351 & 0.0016 \\ 0 & 0 & 0 & 0.7023 & 0.2977 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0.0029	0	0.0156
7	0	0	0.0156
11	0	0	0.0933
16	0.0029	0	0
18	0	0	0.0467
22	0.0028	0	0
27	0.0114	0	0
35	0	0	0.0156
36	0	0	0.0467
37	0.0085	0	0
38	0	0	0.0467
50	0.2537	0.3311	0.1165
52	0.2678	0.1385	0.1201
63	0	0	0.0622
65	0	0	0.0467
69	$3.7754\ (10^{-4})$	0.0016	0
72	0.0027	0.0019	0
74	0	0	0.0156
75	0.0029	0	0
85	0.1107	0.2134	0.1056
86	0.0062	0.0193	0.0040
97	0.3018	0.2781	0.1344
99	0	0	0.0156
101	0	0	0.0467
113	$3.6741\ (10^{-4})$	0.0016	0
120	0.0057	0	0
125	0.0193	0.0144	0.0061
127	0	0	0.0467

- **Modèle du chiffre « 2 » :**

$$A=\{a_{ij}\}=\begin{bmatrix} 0 & 0.9635 & 0.0365 & 0 & 0 \\ 0 & 0.9230 & 0.0770 & 0 & 0 \\ 0 & 0 & 0.9581 & 0.0357 & 0.0061 \\ 0 & 0 & 0 & 0.9195 & 0.0805 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0.0080	0	0.0047
2	0.0026	0.0091	0
3	0	0	0.0094
12	0	0.0021	0
15	0	0	0.0047
22	0	0	0.0802
26	0	0.0042	0
27	0.0080	0	0
29	0	0	0.0142
33	0	$1.8565\ (10^{-4})$	0.0043
37	0	0.0021	0
43	0.0054	0.0202	0
47	0	0.0021	0
50	0.2468	0.0822	0.0469
52	0.1629	0.0109	0.0047
67	0	0.0063	0
70	0	0.0021	0
72	0	0	0.5754
79	0.0040	0	0
85	0.3408	0.4972	0.0434
86	0.0795	0.3310	0
93	0	0	0.0283
94	0.0039	0	0
95	0	0	0.1132
97	0.0839	0.0157	0.0401
99	$3.1981\ (10^{-4})$	0.0019	0
100	0	0	0.0047
102	0	0	0.0189
116	0	0.0021	0
117	0	0	0.0047
125	0.0539	0.0085	0.0021
126	0	0.0021	0

- **Modèle du chiffre « 3 » :**

$$A=\{a_{ij}\}=\begin{bmatrix} 0 & 0.9998 & 0.0002 & 0 & 0 \\ 0 & 0.9348 & 0.0644 & 0.0008 & 0 \\ 0 & 0 & 0.9570 & 0.0365 & 0 \\ 0 & 0 & 0 & 0.9122 & 0.0878 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0.0066	0.0041	0.0313
2	0.0029	0.0024	0
7	0	0	0.0723
12	1.5487 (10^{-4})	0.0021	0
15	0	4.3340 (10^{-4})	0.0454
16	0.0033	3.0220 (10^{-4})	0.0096
18	0	0	0.0465
21	0	0	0.0052
26	0.0032	0	0
27	0.0163	0	0
37	0.0033	0.0043	0
40	0	0	0.1394
43	0.0033	0	0
46	0	0	0.0052
50	0.1826	0.2066	1.1522 (10^{-4})
51	0.0033	0	0
52	0.1873	0.0208	0
56	0	0	0.1961
60	0.0095	0.0023	0
63	0	0	0.0826
65	0	0	0.0413
70	0	0.0022	0
72	0	0.0021	0
85	0.3136	0.4631	0
86	0.0189	0.1723	0
90	0	0	0.0102
97	0.1711	0.0991	0
100	0	0	0.0671
102	0	0	0.0258
103	0	0	0.1858
114	0	0	0.0103
120	0.0033	0	0
123	0.0033	0	0
125	0.0681	0.0089	0.0052
128	0	0	0.0206

- **Modèle du chiffre « 4 » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.8339 & 0.1661 & 0 & 0 \\ 0 & 0.9449 & 0.0546 & 0.0005 & 0 \\ 0 & 0 & 0.9368 & 0.0581 & 0.0051 \\ 0 & 0 & 0 & 0.9464 & 0.0536 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0	0	0.0029
2	0.0226	0.0693	0.0332
3	0	0	0.0058
12	0	0.0031	0
13	0.0033	0.0031	0
15	0	0	0.0175
16	0.0033	0	0.0058
26	0.0250	0.0300	0.0117
37	0.0066	0	0
43	0.1422	0.4794	0.7236
50	0.0475	0.0146	0
52	0.0131	0.0064	0
59	0	0.0032	0
67	0.0129	0.0066	0
70	0.0031	0.0127	0.0090
75	0.0033	0	0
76	0.0059	0.0038	0
85	0.2284	0.0637	0.0141
86	0.4284	0.2841	0.1379
87	0.4244	0	0.0117
89	$3.0070\ (10^{-4})$	0.0029	0
90	0	0	0.0058
97	0.0297	0	0.0058
114	0	0	0.0117
125	0.0033	0	0
126	0.0249	0.0169	0.0034

- **Modèle du chiffre « 5 » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9624 & 0.0374 & 0.0002 & 0 \\ 0 & 0 & 0.9029 & 0.0920 & 0.0051 \\ 0 & 0 & 0 & 0.9150 & 0.0850 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission:

J	I1	I2	I3
1	0	0	0.0134
2	0.0667	0.1490	0
11	0	0	0.0045
12	0.0062	0.0034	0
15	0	0	0.0045
18	0	0	0.0045
21	0.0019	0	0.0980
24	0	0	0.0224
26	0.0305	0.0039	0
35	0	0	0.0179
36	0	$8.3212\ (10^{-4})$	0
37	0	0	0.2503
38	0	0	0.0134
43	0.4768	0.4457	0
46	0	0	0.0045
50	0.0188	0	0
52	0.0123	0.0433	0.0161
55	0	0	0.2152
59	0.0019	0	0
60	0.0038	0	0.0672
63	0	0	0.0045
67	0.0020	0.0142	0
70	0.0092	0.0102	0
72	0.0019	0	0
74	0	0	0.0045
75	0	0	0.0179
76	0.0102	0.0125	0
77	0.0019	0	0
85	0.0448	0.0896	0.0082
86	0.2837	0.1808	0
92	0	0	0.1031
97	0.0197	0.0166	$6.2055\ (10^{-4})$
100	0	0	0.0090
106	0	0	0.0269
109	0	0	0.0090
120	0	0	0.0672
123	0	0.0033	0.0015
124	0.0019	0	0
125	0.0022	0.0267	0.0151
126	0.0038	0	0

- **Modèle du chiffre « 6 » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0.9380 & 0.0620 & 0 & 0 \\ 0 & 0.9471 & 0.0413 & 0.0116 & 0 \\ 0 & 0 & 0.9631 & 0.0231 & 0.0138 \\ 0 & 0 & 0 & 0.9198 & 0.0802 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0	0	0.0057
2	0.0081	0.0119	0
3	0	0	0.0057
5	0	0.0023	0
21	0	0	0.1142
24	0	0	0.0114
26	9.9653 (10^{-4})	0.0480	0
27	0.0169	0	0
37	0.0057	0	0.3023
43	0.0068	0.2966	0
50	0.1880	0.0186	0.0019
52	0.1186	0.0033	0.0773
55	0	0	0.1884
60	0	0	0.0057
67	0.0011	0.0200	0
70	0	0.0070	0
75	0	0	0.0571
77	0	0.0093	0
85	0.3423	0.1502	0
86	0.1693	0.3812	0
92	0	0	0.1028
93	0	0	0.0571
97	0.1113	0.0147	0.0013
100	0	0	0.0057
102	0	0	0.0140
104	0	0	0.0057
116	0	0.0070	0
120	0	0	0.0341
125	0.0263	0.0014	0.0119
126	0.0047	0.0286	0

- **Modèle du chiffre « 7 » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9580 & 0.0418 & 0.0001 & 0 \\ 0 & 0 & 0.9439 & 0.0367 & 0.0194 \\ 0 & 0 & 0 & 0.8978 & 0.1022 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
1	0	0	0.0156
2	0.0342	0.0611	0
12	0.0057	0.0092	0
16	0	1.2390 (10^{-4})	0.0075
21	0	0	0.0312
24	0	0	0.0546
26	0.0389	0.0520	0
37	0	0	0.3277
43	0.1709	0.2155	0
50	0.0219	0.0061	0.0184
52	0.0020	0.0084	0.0865
55	0	0	0.0702
60	0.0021	0	0.1014
66	0.0021	0	0
67	0.0165	4.1518 (10^{-4})	0
70	0.0088	0.0050	0
75	0	0	0.0234
76	0.0022	0.0139	0
77	0.0041	0.0030	0
85	0.1993	0.1513	0.0017
86	0.4859	0.4373	3.1375 (10^{-4})
92	0	0	0.1561
93	0	0	0
97	0	0.0268	0.0039
99	0.0023	0.0025	0
100	0	0	0
102	0	0	0
104	0	0	0
116	0	0	0
120	0	0	0.0546
123	0	0.0028	0
124	0.0021	0	0
125	0.0021	0	0.0468
126	8.8169 (10^{-4})	0.0044	0

- **Modèle du chiffre « 8 » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9649 & 0.0347 & 0.0005 & 0 \\ 0 & 0 & 0.8941 & 0.0792 & 0.0267 \\ 0 & 0 & 0 & 0.9266 & 0.0734 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
2	0.0111	0.0572	0
3	0	0	0.1369
7	0	0	0.0440
22	0	0.0075	0.0127
26	0.0311	0.0176	$1.0791\ (10^{-4})$
27	0.0035	0	0
37	0.0018	0	0
40	0	0	0.0880
43	0.7728	0.2429	0
50	0.0162	0.0684	0
52	0.0158	0	0
56	0	0	0.2053
65	0	0	0.0244
67	0.0107	0.0263	0
70	0	0.0322	0
72	0	0.0069	0.1208
76	$1.7684\ (10^{-4})$	0.0048	0
77	0	0.0054	0
85	0.0324	0.1962	0
86	0.0808	0.2789	0
90	0	0	0.0929
93	0	0	0.0489
95	0	$1.6790\ (10^{-4})$	0.1074
97	0.0079	0.0391	0.0011
103	0	0	0.0440
111	0.0018	0	0
116	0.0018	0	0
117	0	0	0.0684
125	0.0035	0.0053	0.0049
126	0.0086	0.0112	0

- **Modèle du chiffre « 9 » :**

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9688 & 0.0310 & 0.0002 & 0 \\ 0 & 0 & 0.6992 & 0.2706 & 0.0302 \\ 0 & 0 & 0 & 0.9288 & 0.0712 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrice des probabilités d'émission :

J	I1	I2	I3
2	0.0130	0.0101	0
3	0	0	0.0356
11	0	0	0.0396
12	0.0016	0	0
17	0	0	0.0396
21	0.0035	0.0113	0
22	0	0.0150	0
26	0.0420	0.0166	0
27	0.0140	0	0
29	0	0	0.0079
37	0.0031	0	0
43	0.1785	0.0391	0
44	0	0	0.1780
46	0	0	0.0435
50	0.0648	0.0222	0
52	0.0286	0.0555	0
60	$1.2064\ (10^{-4})$	0.0140	0
67	0.0122	0.0180	0
72	0	0.2414	$1.4672\ (10^{-4})$
76	$1.9765\ (10^{-4})$	0.0132	0
81	0.0016	0	0
85	0.2546	0.1829	0
86	0.2931	0.1373	0
87	0	0	0.1662
90	0	0	0.0277
93	0	0	0.0198
95	0	0.1609	0.0291
97	0.0272	0.0234	0
99	0.0016	0	0
100	0	$3.6869\ (10^{-4})$	0.0355
101	0	0	0.0040
102	0	0	0.0079
114	0	0	0.0079
116	0.0047	0	0
117	0	0.0249	0.3100
120	0.0016	0	0
125	0.0051	0.0115	0
126	0.0091	0.0024	0
128	0	0	0.0475

TESTS DE RECONNAISSANCE :

Les résultats de reconnaissance sont présentés ci-après dans le tableau (4-10) :

	Taux de reconnaissance (%)
Un	77
Deux	79
Trois	80
Quatre	44
Cinq	28
Six	40
Sept	42
Huit	70
Neuf	97

Tableau((4-10) taux de reconnaissance de chaque chiffre du vocabulaire

- Le taux de reconnaissance des chiffres (1, 2, 3, 8, 9) este assez bon alors que pour le reste des chiffres ce taux n'a pas atteint le 50 %.
- Le taux global de reconnaissance est de 53.33 %, cette baisse dans le taux de reconnaissance était prévisible car la taille de notre nouveau vocabulaire est supérieure à celle de l'ancien vocabulaire (A, B, C, I, O). L'amélioration du taux de reconnaissance nécessite l'augmentation de la taille du corpus d'apprentissage. N'ayant pour but dans cette dernière expérience que la mise en évidence de l'effet de la taille (et la nature) du vocabulaire sur le taux de reconnaissance il serait superflue d'augmenter le nombre d'exemples d'apprentissage, opération déjà entreprise pour le premier corpus.

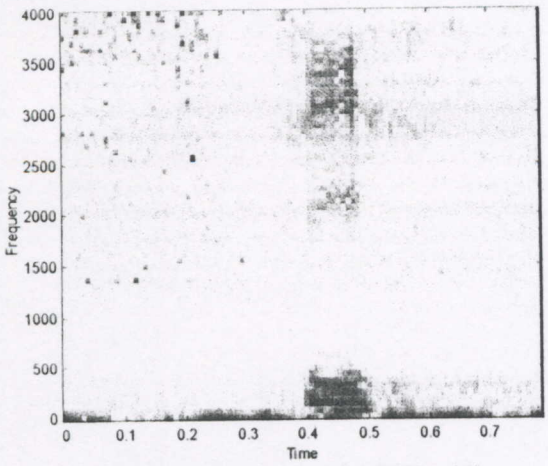
REMARQUES :

- Pour l'apprentissage, le temps de calcul est proportionnel à $N \times T \times L \times Q \times I$ où :
  N est le nombre d'états du modèle, T la longueur du mot (nombre de fenêtres), L le nombre de centroides du quantificateur, Q est le nombre d'exemples d'apprentissage. et I le nombre d'itérations de l'algorithme. Pour (N=5,T=50,L=128,Q=20 ,I=5) le temps de calcul est d'environ 48 heures sur un Pentium-2-.
- Pour la reconnaissance, le temps de calcul est proportionnel à $N \times T \times L \times M$, où V est la taille du vocabulaire. Par exemple, pour notre premier vocabulaire(A, B, C, I, O) ce temps est d'environ 4 minutes. Il va sans dire que l'utilisation d'une machine série telle qu'un PC n'est pas à envisager pour une application dans le monde réel; encore faudrait-il rappeler que ce n'est nullement le but de ce travail. Un exemple de parallélisation des algorithmes de reconnaissance et d'apprentissage est donné dans [8]

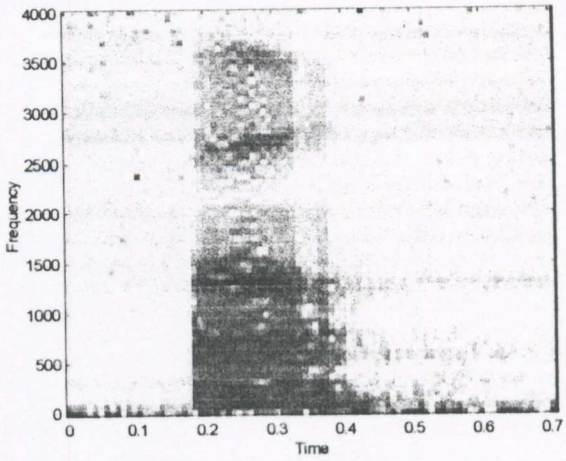
IV-5-5-CONCLUSION :

D'après les résultats obtenus , les performances de reconnaissance peuvent être améliorées sur plusieurs plans à savoir :

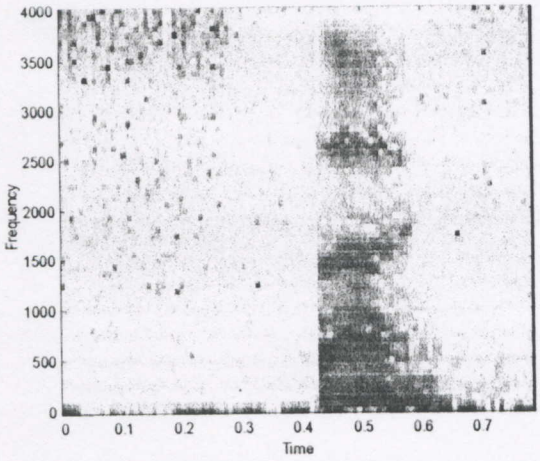
- les paramètres acoustiques.
- La quantification vectorielle
- Le choix du vocabulaire (lorsqu'il s'agit d'une application bien spécifique).



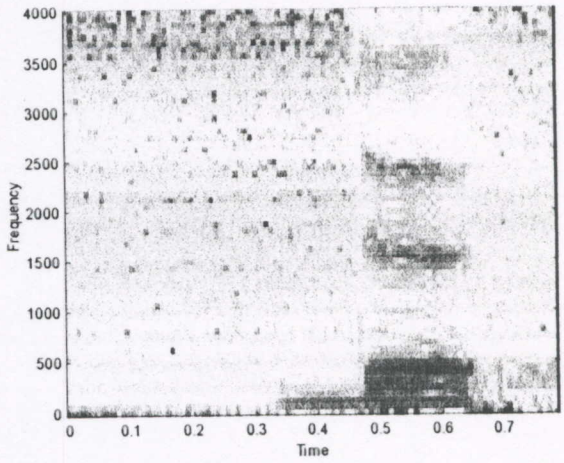
a-Spectrogramme du chiffre '6'



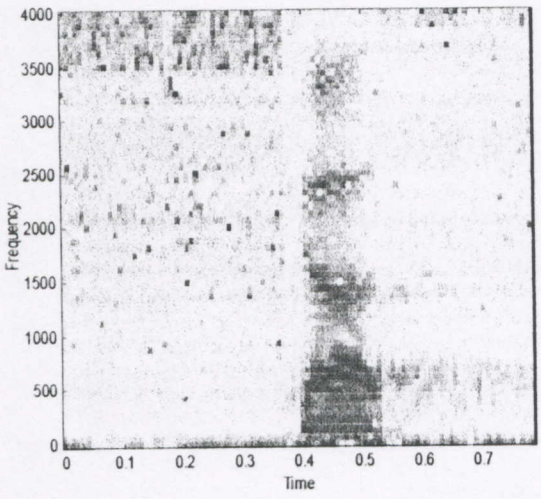
b-Spectrogramme du chiffre '1'



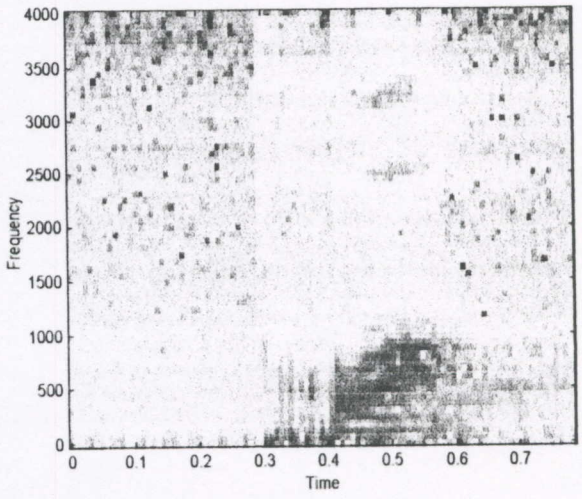
c-Spectrogramme du chiffre 5'



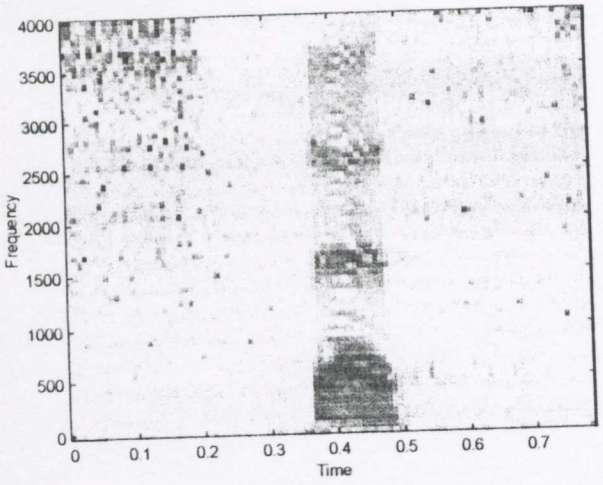
d-Spectrogramme du chiffre '2'



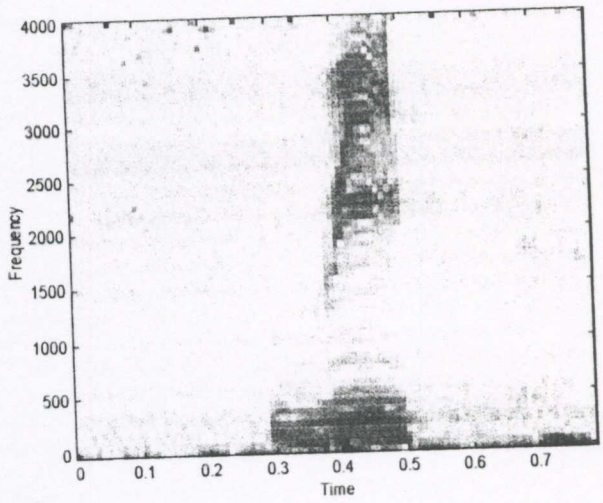
e-Spectrogramme du chiffre '4'



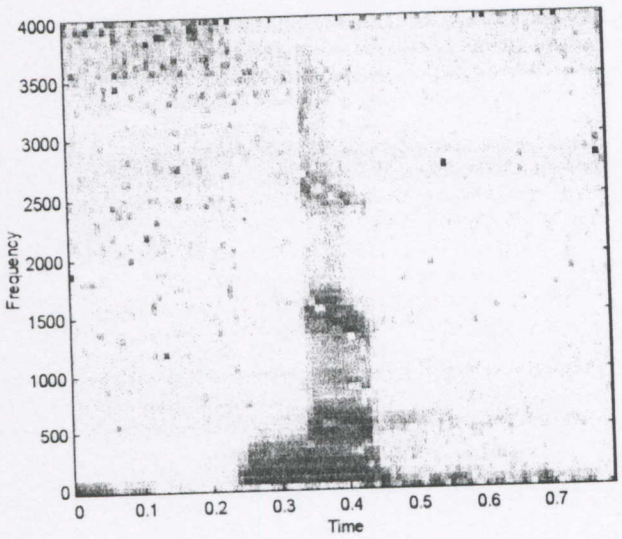
f-Spectrogramme du chiffre '3'



g-Spectrogramme du chiffre '7'



h-Spectrogramme du chiffre '8'



i-Spectrogramme du chiffre '9'

Figure(4-11) :spectrogrammes des chiffres de un à neuf

La confusion souvent remarquée entre le chiffre '2' et le chiffre '9' est due à la partie voisée des deux chiffres et qui est bien visible sur les spectrogrammes des deux chiffres, en effet on observe des similitudes entre les positions des formants (maximas du spectre –niveau de gris élevé-) qui se situent pour les deux chiffres aux environs de 500 , 1500 et 2500 Hz
Des confusions semblables ont été remarquées entre le chiffre '5' et le chiffre '7' , là encore la ressemblance entre les deux chiffres est visible sur les spectrogrammes leurs spectrogrammes ; en effet pour les deux chiffres on observe des formants aux environs de 500 1500 , 2500 et 3500 Hz

REMARQUES :

- Pour l'apprentissage, le temps de calcul est proportionnel à $N \times T \times L \times Q \times I$ où : N est le nombre d'états du modèle, T la longueur du mot (nombre de fenêtres), L le nombre de centroides du quantificateur , Q est le nombre d'exemples d'apprentissage. et I le nombre d'itérations de l'algorithme. Pour (N=5,T=50,L=128,Q=20 ,I=5) le temps de calcul est d'environ 48 heures sur un Pentium-2-.
- Pour la reconnaissance ,le temps de calcul est proportionnel à $N \times T \times L \times M$, où M est la taille du vocabulaire. Par exemple , pour notre premier vocabulaire(A, B, C, I, O) ce temps est d'environ 4 minutes. Il va sans dire que l'utilisation d'une machine série telle qu'un PC n'est pas à envisager pour une application dans le monde réel ; encore faudrait-il rappeler que ce n'est nullement le but de ce travail. Un exemple de parallélisation des algorithmes de reconnaissance et d'apprentissage est donné dans [8]

IV-5-5-CONCLUSION :

D'après les résultats obtenus , les performances de reconnaissance peuvent être améliorées sur plusieurs plans à savoir :

- les paramètres acoustiques.
- La quantification vectorielle
- Le choix du vocabulaire (lorsqu'il s'agit d'une application bien spécifique).

Il est très important de noter aussi que les performances de reconnaissance sont intimement liées à la méthode choisie , en effet les modèles de Markov dit standards et qui se basent se le principe du maximum de vraisemblance présentent un aspect peu discriminant , cette particularité inhérente à cette approche a été soulignée dans la section III-6-3 du chapitre –III-

CONCLUSION GENERALE

Ce travail est avant tout une tentative visant à confirmer le succès de la modélisation Markovienne dans la reconnaissance de la parole.

La tache n'était pas facile notamment pour l'enregistrement de la base de données, opération qui nécessitait le choix des moments convenables(ambiance du laboratoire). Toutefois les difficultés rencontrées ne sont nullement comparables avec la satisfaction de voir ce travail aboutir à des résultats positifs.

A travers ce travail, il a été question d'appliquer les modèles de Markov cachés dans la reconnaissance de la parole et d'étudier les performances de cette modélisation en spéculant sur plusieurs paramètres à partir des prétraitements jusqu'au vocabulaire utilisé.

N'ayant donc guère la prétention d'avoir fait un travail parfait, nous donnons ci-après quelques propositions comme perspectives de travail :

- Utiliser les coefficients LPC ou les bancs de filtres comme paramètres acoustiques.
- Dans ce mémoire nous avons travaillé avec les HMMs discrets, il serait intéressant de travailler avec les HMMs continus et de faire une comparaison de performance.
- Implanter l'algorithme de reconnaissance élaboré , pour une application en temps réel.

Références bibliographiques

[01]- Abderrahmane Menacer , "Reconnaissance de la parole en mode mulitilocuteur par des méthodes globales (mots isolés)", thèse de docteur ingénieur.
devant l'université de Rennes 1, UER mathématiques et informatique 1985.

[02]- Leon Bottou, "Approche théorique de l'apprentissage connectionniste ; applications à la reconnaissance de la parole". , thèse de doctorat.Université de Tours, école d'ingénieurs en informatique pour l'industrie 1990.

[03]- Creaney and R.N.Gorgui-Naguib , "A scaly artificial neural network for speaker independant isolated word recognition using nonlinear time alignement. M.J. "
departement of electrical and electronic engineering , university of new castle Upon Tyne U.K 94 .

[04]- K.Hassanein , Lee Deng and M.I Elmasry, " A neural predictive HMM for speech and speaker recognition. " Electrical and computer engineering departement , university of waterloo , ont , Canada.92

[05]- Yve Normandin , Regis cardin , Renato Demori, "Hight performance connected digit recognition using maximum mutuel information estimation ".IEEE trans. on speech and audio processing , april 94.

[06]- Li .Deng , Mike Aksmanovic , Xiao.Dong.Sun , C.F.Jeff Wu, "Speech recognition using HMM with polynomial regression functions as non stationary states".IEEE trans on speech and audio processing vol.2 , n 4. October 1994.

[07]- Jean luc gauvain , Chin Hui Lee; "Maximum a posteriori estimation for multivariate gaussian mixture observations of Markov chains". IEEE trans on speech and audio processing vol.2 , n2,april 94.

[08]- Carl D.Mitchell Mary P.Harper , Leah H.Jamieson, "A parallel implementation of A hidden Markov model with duration modeling for speech recognition".Digital signal processing 5.43,57. 1995.

[09]-Murat Kunt , René Boite , "traitement de la parole".Presse polytechniques Romandes, CH 1015-Lausanne 1987.

[11]- Sandro Ridella , Stefano Rovetta , Rodolfo Zunino, "Circular backpropagation networks for classification",IEEE trans on neural networks , vol.8 , january 97.

[12]- Jorg Kindermann , Christoph Windheuser; "unsupervised sequence classification",German national research center , Carnegie Mellon university for computer science(GMD), school of computer science(USA), 1992.

[13]- Simon Haykin ,Mc Mastr university; "Neural networks , a comprehensive fondation",by Macmillan college publishing company1994.

[14]- S.Y Kung , digital neural networks ,department of electrical engineering , Princeton university, by PTR printice hall,Inc 1993.

[15]- Benrachi.B , Sellami.M , Noui.A , Benzerrouk.S, "reconnaissance automatique de la parole; système VOCNET : pour la reconnaissance de quelques voyelles françaises et sylabes arabes". Université de Constantine, inst d'informatique ; route Ain El Bey 25000 Constantine , Algérie.SSA 99

[16]- Ait Akkache mustapha, "les reseaux de neurones , application aux modèles de Markov". these de magister ,université de Blida, institut des sciences exactes, département de mathémathiques, 1996.

[17]-Yves THOMAS , "signaux et systèmes linéaires", MASSON Paris , 1995.

[18]- Calliope , "La parole et son traitement automatique ", Masson 1989.

[19]- Pierre Demartines and Jeany Herault, "Curvilinear component analysis ; Aself-organising neural network for nonlinear maping of data sets",IEEE trans on neural networks , vol.8 , n1 , january 1997.

[20]-Allen Gersho ,Robert M.Gray ," vector quantization and signal compression".Kluwer academic publishers. Sixth printing 1997.

[21]-John Makhoul, Salim Roucos , Herbert Gish, "vector quantisation in speech coding".Proceedings of the IEEE , vol 73.n 11,november 1985.

[22]-Robert M.Gray, vector quantisation.IEEE ASSP magazine APR 1984.

[23]-Robert M.Gray and David L.Neuhoff , "Quantization".IEEE transactions on information theory, vol.44,n.6, october 1998.

[24]- Stamatios V0. Kartalopoulos , "Understanding neural networks and fuzzy logic".Basic concepts and applications,IEEE press-1996 .

[25]- Seyed A Rizvi , Nasser M.Nasrabadi, "Residual vector Quantization using a multilayer competitive neural network".IEEE journal on selected areas in communications , vol.12 , n9, december 94.