الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DR. YAHIA FARES DE MEDEA

FACULTE DE TECHNOLOGIE

Département du Génie Électrique



Mémoire de Fin d'Études

Présenté pour l'obtention du Diplôme de Master en Génie Électrique

Option : Systèmes Intelligents et Communicants

THEME

Fenêtrage multiple pour la reconnaissance robuste de la parole

Proposé et dirigé par,

Mr : RAGGAI DJAHID

Préparé et présenté par,

Melle. KHELFI AMEL

Remerciements

Avant tout, nous remercions le Dieu tout puissant de nous
avoir donné la force et la volonté pour achever ce modeste travail. Nous
Nous remercions infiniment mes Parents pour tout ce qu'ils ont fait et ce qu'ils font
encore pour nous
Nous nous devons remercier le corps professoral et administratif de
la Faculté de Technologie, pour la richesse et la qualité
de leur enseignement et qui déploient de grands efforts pour assurer à leurs
étudiants une formation actualisée.

Nous nous adressons tous nos gratitudes et nos remerciements à
Mon promoteur Mr. REGGUAI DJAHID qui s'est toujours montré à
l'écoute et très disponible tout au long de la réalisation de ce mémoire,
ainsi pour l'inspiration, l'aide et le temps qu'il a bien voulu nous consacrer

Nous tenons également à remercier l'ensemble du corps enseignant du département de Génie éléctrique (UYF-Médéa), pour leurs concours dans la réussite de notre formation.

Nous remercions les membres du jury d'avoir accepter de juger notre travail.
Mr : ALIMOHAD ABDENOUR
Mr : DALI MOHAMED
En fin, nous tenons à remercier tous ceux qui ont contribué d'une façon ou d'une autre à la réalisation de ce travail.

Dédicaces

A la lumière de ma vie, mes très chères parents

Pour leur amour, leurs prières et leurs sacrifices, pour m'avoir donnée la vie et la joie de vivre. Tes bonne éducation, tes conseils et tes bénédictions n'ont jamais fait défaut, que Allah le tout puissant vous garde pour moi (amin).

A mon frère Mohamed et sa femme et ses petits : Ishak, Manel, Naim.

A ma adorable sœur Bahia qui n'a pas cessé de m'encourager dans les moments difficiles qu'Allah la protège. et son époux et ses deux petit Islam et Sara.

A mes chers frères : Karim, sidali et Abderrahmane.

Et toute ma famille

A Nassira, Aicha, Ibtissam, Khadidja, Samiha, Chahra.

A mes collègues

A touts mes amis.

A toute la promotion 2016-2017.

AMEL

Sommaire

NTRODUCTION GENERALE 1

Chapitre I : Généralité 3

1. Introduction 4

2. Production de la parole 4

3. Caractéristiques phonétiques 5

4. Les caractéristiques acoustiques 6

5. La reconnaissance 8

5.1. La reconnaissance du locuteur 8

5.2. La reconnaissance de la parole 8

6. Analyse et codage de la parole 8

6.1. Analyse temporelle 9

6.2. Analyse fréquentielle (spectrale) 10

6.2.a) Transformé de Fourier Discrète à Court Terme 10

6.2.b) Spectrogramme 12

6.3. Bancs de filtre passe bandes 14

6.4. Analyse cepstrale 14

6.5. Analyse prédictive linéaire(LPC) 16

7. Conclusion 17

Chapitre II : les modèles de Markov cachés 18

1. Introduction 19

2. Présentation desmodèles de Markov cachée 19

3. Reconnaissance des mots isolés 19

4. L'apprentissage des modèles de Markov cachés 23

5. Reconnaissance et décodage parl'algorithme de Viterbi 25

6. Conclusion 26

Chapitre III : estimation spectrale par fenêtrage multiple 27

1. Introduction 28

2. Analyse spectrale 28

3. Quelques fonctions de fenêtrages 29

3.1) Fenêtre rectangulaire (Box Car) 31

3.2) Fenêtre de Hann (Hanning) 31

3.3) Fenêtre de Hamming 32

4. L'utilisation de fenêtrage multiple 32

Chapitre IV : Résultats 39

1. Introduction 39

2. Description de la base de données utilisée 39

3. Description de l'outil de reconnaissance HTK 40

4. Préparation des fichiers de la base de données 41

4.a) Codage utilisant fenêtrage de Hamming 42

4.b) Codage utilisant fenêtrage de Multitaper 42

5. Choix de la topologie des modèles (prototypes) 44

6. Initialisation des modèles 47

7. Apprentissage 48

8. Définition du grammaire et dictionnaire 48

9. Reconnaissance 49

10. Tests de robustesse et interprétations 50

11. Conclusion 53

CONCLUSION GENERALE 54

Annexe 55

ملخص :

نقدم في هذا العمل تقنية جديدة لتجزئة الإشارة الصوتية إلى نوافذ بهدف تقليل تغيرية الطيف المقدر مما ينتج عنه تقدير ذو نوعية أحسن. الفكرة تكمن في استكشاف مفعول و وقع استعمال هذه التقنية على صلابة نظام التعرف على الكلام. نظام التعرف على الكلام في حد ذاته يعتمد على نماذج ماركوف المخفية. النتائج المحصلة عليها تبين بوضوح الإضافة الإيجابية و أفضلية التقنية الجديدة المقترحة مقارنة مع التقنية الكلاسيكية.

Résumé :

Ce travail introduit l'utilisation d'une nouvelle technique de fenêtrage du signal vocal dans le but de minimiser la variance du spectre estimé conduisant ainsi à une meilleure qualité d'estimation. L'idée est d'explorer l'effet et l'impact de l'utilisation de cette méthode sur la robustesse du système de reconnaissance. Le système de reconnaissance en question est basé sur les modèles de Markov cachés. Les résultats obtenus montrent clairement l'apport positif et la supériorité de la nouvelle méthode proposée par rapport à la méthode classique.

Abstract :

In this work we introduce the use of a new windowing technique of the vocal signal in order to minimize the variance of the estimated spectrum, hence leading to a better quality of estimation. The idea is to investigate the effect and the impact of using this technique on the robusteness of the recognition system. The recognition system used is based on hidden Markov models. The obtained results show clearly the superiority of the new proposed method over the classic method.

Glossaire

signal acoustique	signal analogique représentatif de l'onde sonore
signal acoustique	les propriétés statistiques de ce signal caractérisées par des espérances mathématiques sont indépendants du temps
Nyquist	ou la théorème de Shannon fréquence d'échantillonnage supérieure à 2 fois la fréquence maximale du signal
formant	endroit qui représente le maximum d'énergie sonore
Harmoniques	les multiples de la fréquence fondamentale F_0
Echel de Mel	est une échelle de fréquence basée sur la perception humaine. Elle se mesure en Mel

Sigle et abréviation :

TFCT	Transformé de Fourier à Court Terme
LPC	Linear Prediction Coefficients
TFD	Transformé de Fourier Discrète
DCT	Discrete Cosine Transform
MFCC	Mel Frequency Cepstral Coefficients
RAP	Reconnaissance Automatique de la Parole
HMM	Hidden Markov Models
MMC	Modèle de Markov Caché

Figure (IV.2): Codage par Multitapers 41

Figure (IV.3): Conversion des données d'apprentissage 42

Figure (IV.4): Fichier de configuration 42

Figure (IV.5): Topologie choisie pour les MMCs 44

Figure (IV.6): Exemple de fichier de description de HMM 44

Figure (IV.7): Fonctionnement de HCompv..... 46

Figure (IV.8): Fonctionnement de HRest 47

Figure (IV.9): Le fichier de la grammaire 48

Figure (IV.10): Le fichier du dictionnaire 48

Figure (IV.11): Le réseau de connexion entre les mots..... ...49

Figure (IV.12): Processus du codage et reconnaissance 50

Figure (IV.13): Taux de reconnaissance (Hamming) 51

Figure (IV.14): Taux de reconnaissance par multi-taper..... 52

Figure (IV.15): Comparaison de taux de reconnaissance des deux méthodes..... 53

Liste des tables

Tablea (I.1): Propriéte de quelque fenêtre connue……..11

Tableau (III.1): Valeur propre correspendent à 8 fénetre propre et la fénetre rectangulaire……..34

Tableau (IV.1): Résultat obtenue par méthode de Hamming(propre et bruité) 51

Tableau (IV.2): Résultat obtenu par méthode de fenétrage multiples(propre et bruité)..... 51

Introduction générale :

La parole est l'un des moyens de communication les plus naturels grâce à sa simplicité. Le traitement fait par le cerveau, de la production de la parole jusqu'à sa perception et sa compréhension, est un traitement très complexe, ce qui rend la parole difficilement automatisable pour une machine. L'avancement technologique et surtout de l'informatique a suscité le besoin de nouveau moyen de dialogue homme-machine (parole), limitant l'utilisation du clavier et autres périphériques qui rendait la communication avec la machine très difficile et très lente.

La parole est le principal support de la langue, comme pour l'écriture, le message donc doit être bien structuré selon des règles connues par tous ceux qui parlent la même langue (grammaire, syntaxe, vocabulaire).

En générale, le traitement de la parole utilise des sources de connaissance phonétiques, phonologiques, prosodiques, lexicales, syntaxiques et sémantiques. Cela dit que les connaissances nécessaires pour la compréhension de la parole sont vastes. On s'intéresse à étudier l'aspect reconnaissance de la parole qui a pour objectif de décoder le signal de la parole en unités de base (phonème, mot,…).

De nos jours, les systèmes de reconnaissance de la parole ont évolué et utilisent non-seulement des connaissances en linguistique (Phonétiques, Phonologiques, Prosodiques, Lexicales…) mais aussi des connaissances dans les domaines : Traitement du signal, reconnaissance des formes…

Les avantages de la reconnaissance automatique de la parole sont tellement importants que l'on trouve déjà sur le marché des dispositifs d'utilisation limités, mais néanmoins efficaces. Citons certaines applications qui ont déjà vu le jour : saisie vocale de donnée, chambres d'hôpital avec possibilité de commande vocales pour le malade, commande vocale de robot, commande vocale d'un téléphone portable ou une montre portable,…

La mise au point d'un système de reconnaissance automatique de la parole rencontre des difficultés comme les variations intra-locuteur et les variations interlocuteur. Ces variations sont dues à la complexité physique de l'appareil vocal qui présente une soixantaine de muscles pouvant participer à la phonation.

Les variations intra-locuteur concernent un même locuteur qui prononce le même mot mais de façons différentes d'une fois à l'autre.

Les variations interlocuteurs sont dues au fait que les dimensions et proportions de l'appareil vocal varient selon l'âge, le sexe et la taille,…

Les systèmes de reconnaissance actuels restent sensibles au bruit et au style d'élocution. Plusieurs solutions ont été proposées pour accroitre la robustesse des systèmes. Cependant, l'amélioration des performances de ces techniques pour un signal fortement dégradé représente toujours un vrai défi. Les techniques d'amélioration de la robustesse du système de reconnaissance interviennent en général au niveau de l'un des étages constituant la chaine de reconnaissance en commençant par l'étage d'entrée (front end) chargé de l'extraction des paramètres du signal, passant par l'étage de modélisation acoustique et arrivant à l'étage de décision.

Le présent travail focalise sur l'étage d'entrée et plus particulièrement sur l'estimation spectrale du signal via une nouvelle forme de fenêtrage.

Organisation du document :

Le présent travail est organisé comme suit :

Le **premier chapitre** expose brièvement les notions générales de base sur la parole notamment : le mécanisme de production de la parole, ses caractéristiques phonétiques et acoustiques, les méthodes utilisées pour l'analyse et le codage de la parole.

Le **deuxième chapitre** est une introduction aux **M**odèles de **M**arkov **C**achées (**MMC**) ; **H**idden **M**arkov **M**odels en anglais (**HMM**). Ce chapitre illustre l'utilisation des MMC dans la reconnaissance automatique de la parole.

Le **troisième chapitre** ayant pour titre : fenêtrage et estimation spectrale, l'accent sera notamment mis sur le rôle du fenêtrage dans l'estimation spectrale : les différents types de fenêtres, les propriétés et les caractéristiques de chacune, et comment nous pouvons obtenir une meilleure estimation en utilisant le type de fenêtre approprié.

Le **quatrième chapitre** présente tous les aspects pratiques de notre étude en commençant par la description de la base de données utilisée, la modélisation du vocabulaire de reconnaissance et arrivant aux résultats de reconnaissance obtenus.

Finalement, nous terminons ce mémoire par une **conclusion générale.**

Chapitre I : Généralités

1. Introduction :

Le traitement de la parole est aujourd'hui un domaine suscitant un grand intérêt de la part des chercheurs en traitement de signal. Il bénéfice d'efforts de recherche très importants, liés au développement des moyens et techniques de télécommunication.

Il s'agit d'un domaine pluridisciplinaire englobant une vaste étendue de connaissances (mathématiques, probabilités et statistiques, acoustique, ….). Loin de prétendre donc à l'exhaustivité, ce chapitre expose les notions de base en relation directe avec les besoins de notre étude.

2. Production de la parole :

La figure **(I.1)** montre l'appareil phonatoire de l'être humain. La parole est l'information portée par le signal de parole qu'on peut analyser de bien des façons. On en distingue généralement plusieurs niveaux de description non exclusifs : acoustique, phonétique, phonologique, morphologique, syntaxique, sémantique, et pragmatique. Les systèmes de description de la voix se basent sur le modèle de production acoustique source filtre.

Le signal de la parole est le codage et la simplification d'un message (parole) en fonction du temps, en associant à chaque instant t une valeur (amplitude).Il est provoquée par des mécanismes complexes issus de plusieurs sources. La vibration des cordes vocales correspond, en effet, à une onde périodique, de bruit provoqué par le souffle de friction, tandis que les différentes cavités de la gorge et de la bouche peuvent être assimilées à des filtres successifs. On peut avoir d'autres sources impulsives comme des claquements de langue ou l'ouverture rapide des lèvres après accumulation d'air dans la bouche **[1].**

Le conduit vocal, long de 17 cm en moyenne chez l'homme adulte, se compose de plusieurs cavités : pharynx, cavité orale, fosse nasale, forment autant de résonance (formant) et d'anti résonance (anti formant). Si ce corps n'agit pas sur la fréquence fondamentale il est en revanche extrêmement déformable, ou divisible. Agissant comme un filtre acoustique, il donne ainsi au son les différentes couleurs qui distinguent les voyelles ou les fricatives entre elles **[1].**

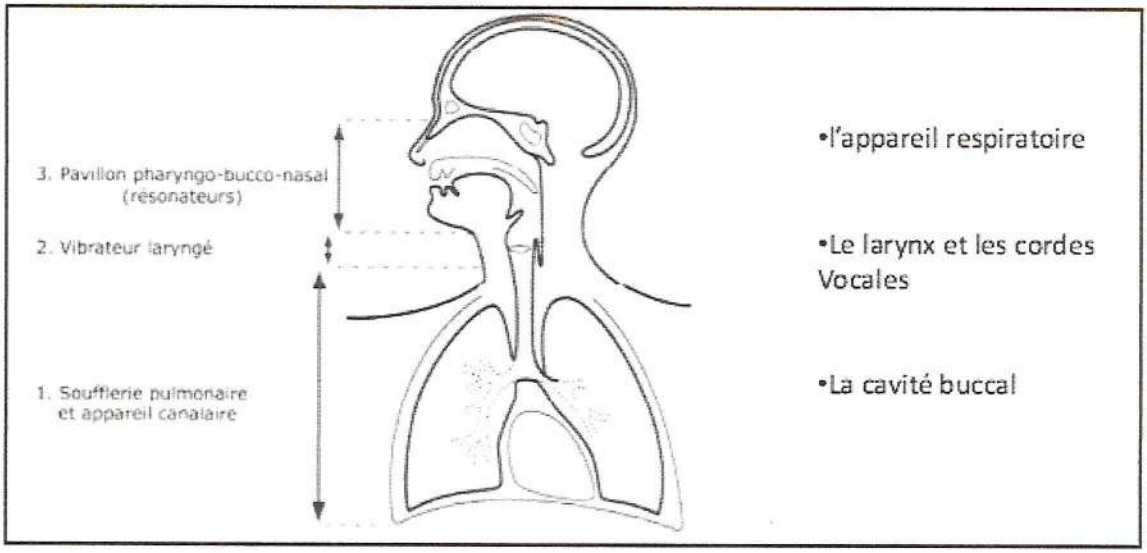


Figure (I.1) : les composantes principales de l'appareil phonatoire

Les parties du corps impliquées dans la production de la parole sont :

- ***La cavité buccale :*** les poumons et la cage thoracique, qui sont la source de l'air qui stimulera le système.
- ***Le larynx et les cordes vocales :*** la trachée conduit l'air des poumons aux cordes vocales qui vibrent et modulent le flot d'air.
- ***Le canal vocal*** : raffine la production du son.

Un signal parole est représenté phonétiquement comme un ensemble fini de symboles appelés « phonèmes».

3. Caractéristiques phonétiques :

Phonème :

C'est la petite unité de langage parlé dont l'association permet de constituer des énoncées et de les distinguer entre eux.

Les phonèmes sont regroupés par classes selon leurs modes de production ; le rapport bruit de friction, voisement, l'utilisation de la cavité nasale ou l'occlusion permettant l'éclat des plosives sont autant de descripteurs articulatoires donnant cette classification.

➢ **Les consonnes :**

- *Les plosives :* sont produite par occlusion totale du conduit vocal (prononcée avec une fermeture complète du canal vocal).

- *Fricatives :* prononcée avec une fermeture partielle du conduit vocal

  Les consonnes peuvent être **non-voisées** ; il n'y a pas de vibration des cordes vocales et le son est essentiellement produit par un bruit (bruit de friction, d'explosion ou de relâchement). Ou au contraire **voisées**

- *Nasales :* qui se prononcent en expirant de l'air par les fosses nasales, grâce à un abaissement du voile du palais

➢ **Les voyelles** : sont produites par la vibration des cordes vocales. Ces sons sont souvent considérés comme quasi-périodiques. Elles peuvent être nasales ou orales selon que l'air passe par la cavité nasale ou la cavité buccale.

4. Les caractéristiques acoustiques :[2]

La parole apparait physiquement comme une variation de la pression d'air causée et émise par l'appareil phonatoire. Cette variation est convertie dans un premier temps en signal électrique grâce au transducteur approprié : le microphone. Le signal électrique résultant est le plus souvent numérisé.

L'opération de **numérisation**, schématisée à la figure **(I.2)** requiert successivement : un filtre de garde (filtre anti-repliement), un **échantillonnage,** et une **quantification**.

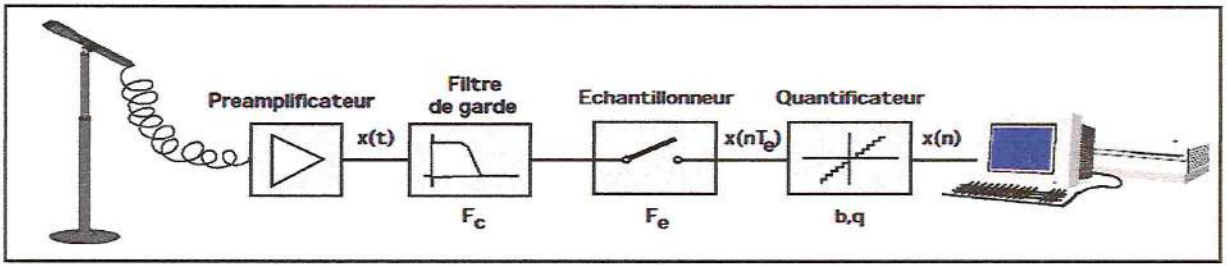


Figure (I.2) : enregistrement numérique d'un signal acoustique **[2]**

L'échantillonnage transforme le signal à temps continu **x(t)** en signal à temps discret **x(nTe)** défini aux instants d'échantillonnage, multiple entiers de la période d'échantillonnage (**Te**) ; pour ce qui concerne le signal vocal, le choix de **Fe** résulte d'un compromis. Son spectre peut s'étendre jusqu'à 12 KHz. Il faut donc en principe choisir une fréquence (**Fe**) égale à 24 KHz au moins pour satisfaite raisonnablement au théorème de Shannon.

La **quantification** ne retient qu'un nombre fini 2^bde valeurs (**b** étant le nombre de bits de quantification), espacées du pas de quantification **q**, le signal numérique résultant est noté **x(n).**

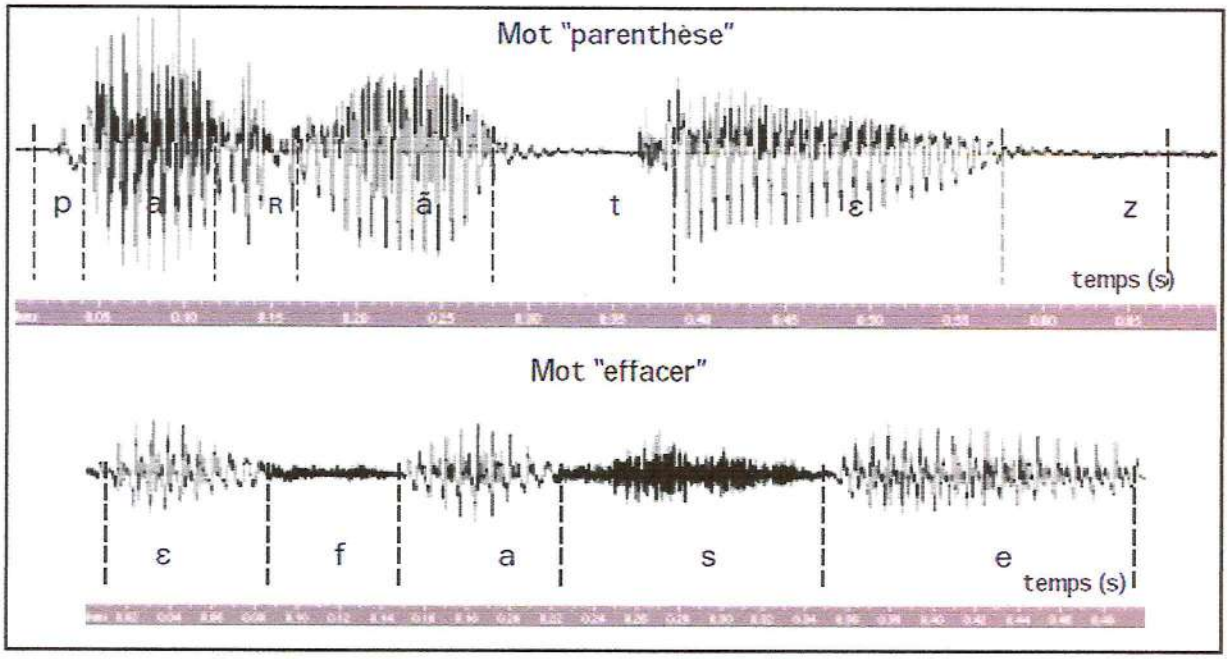


Figure (1.3): audiogramme des signaux représentant les mots «parenthèse» et «effacer» **[2]**

La figure **(1.3)** représente l'évolution temporelle, ou **audiogramme**, du signal vocal pour les mots « parenthèse », et « effacer ». On y constate une alternance de zones assez périodiques et de zones bruitées, appelées zones **voisées** et **non-voisées.**

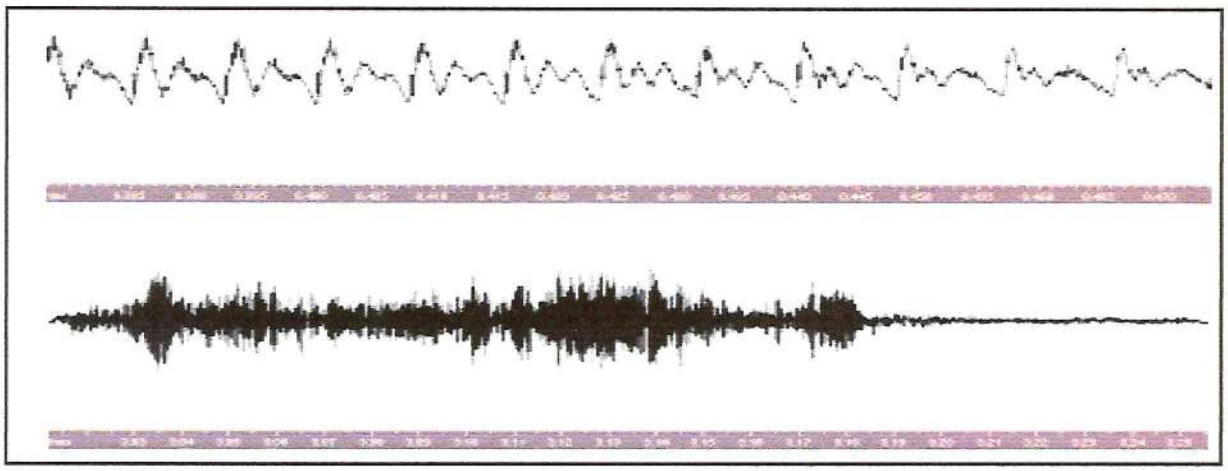


Figure (1.4) : exemple de son voisé (haut) et non-voisé (bas) **[2]**

La figure **(1.4)** donne une représentation plus fine de tranche de signaux **voisés** et **non-voisés.** L'évolution temporelle ne fournis cependant pas directement les trais acoustiques (transformation fréquentielles) du signal.

Le premier trait est la fréquence fondamentale F_0 fréquence de vibration de cordes vocales : ce mouvement vibratoire correspond à une succession de cycles d'ouverture et de

fermeture de la glotte. Le nombre de ces cycles par seconde correspond à la fréquence fondamentale F_0, elle dépend essentiellement de l'âge et du sexe de locuteur.**[3]**

Le deuxième trait est le spectre fréquentiel. Ce spectre résulte du filtrage dynamique du signal glottique par le conduit vocal. **[3]**

Le dernier trait acoustique est l'énergie correspond à l'intensité sonore qui liée à la pression de l'air en amont du larynx. **[3]**

5. La reconnaissance:

C'est le décodage de l'information portée par le signal vocal en séquence de mots par des algorithmes. On distingue deux types de reconnaissance ; reconnaissance de locuteur et reconnaissance de la parole.

5.1. La reconnaissance du locuteur :

Il s'agit d'identifier ou de vérifier l'identité d'un locuteur. On distingue les cas de reconnaissance dépendante du texte, avec texte dicté ou indépendant du texte.

5.2. La reconnaissance de la parole :

L'appellation « Reconnaissance de la Parole» (ASR en anglais) se réfère à plusieurs types de systèmes ; c'est une technique qui permet d'analyser et comparer une onde sonore (la voix humaine) captée au moyen d'un microphone à un ensemble des échantillons ; composé généralement de phonème ; et la transcrire sous la forme d'un texte exploitable par une machine.

De plus les cas d'applications sont toujours plus complexes que la théorie, dans la mesure où ils introduisent des bruits extérieurs, la contribution du matériel utilisée, les différences des locutions…

Parmi les techniques utilisées en reconnaissance on peut citer : les modèles de Markov cachés, les réseaux de neurones, la programmation dynamique (DTW).

6. Analyse et codage de la parole :

L'analyse des signaux de parole tient en considération les façons par lesquelles les sons de parole sont produits. On dira que les signaux de parole résultent de l'interaction de deux facteurs ; la source glottique et le conduit vocal. Par conséquent, le signal de parole est obtenu par une source passant dans un filtre linéaire à temps variable. Le filtre peut dériver des modèles de production de la parole basés sur la théorie acoustique. Dans cette théorie, la source représente le flux d'air au niveau des cordes vocales et le filtre représente les

résonnances du conduit vocal qui change dans le temps. La figure suivante illustre le processus source-filtre de la parole

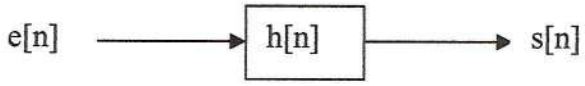


Figure (I.5):système d'obtention du signal parole.

La pression du flux d'air rayonné par la bouche et/ou le nez est convertie en un courant électrique à travers un microphone. Ce signal de parole électrique contient des fréquences allant de 100 Hz à quelques 8000 Hz et des amplitudes variant de 30 à 90 dB.

L'analyse des signaux de parole est le processus d'estimation des paramètres d'un modèle, variant dans le temps, dans le but d'extraire des informations sur la production de ce signal. Les méthodes modernes d'analyse des signaux sont basées sur le traitement numérique des signaux. La première raison de la numérisation est la facilitation des techniques de traitement sophistiquées qui sont très difficiles voir impossible à réaliser en analogique. La seconde raison est que le traitement numérique est beaucoup plus efficace et peut être effectué dans des circuits compacts.

La numérisation implique l'échantillonnage pour discrétiser le temps et la quantification pour discrétiser l'amplitude. Le taux avec lequel le signal analogique est échantillonné est appelé fréquence d'échantillonnage. Dans les réseaux de télécommunications les signaux de parole analogique sont limités entre 300 et 3400 Hz, la fréquence d'échantillonnage dans ce cas est de 8000 Hz conformément au théorème de Nyquist. Pour une qualité supérieure, la bande passante fréquentielle de la parole est limité entre 0 et 7000 Hz, dans ce cas la fréquence d'échantillonnage est choisie égale à 16 KHz.

Le signal échantillonné est en ensuite quantifié en amplitude en utilisant un convertisseur analogique- digital permettant de représenter chaque échantillon réel en un nombre limité de bits. 12 bits sont nécessaires en pratique pour assurer un rapport signal sur bruit supérieur à 35 dB.

6.1. Analyse temporelle :

Les fonctions constituant le modèle source – filtre de la parole sont toutes variables dans le temps. L'excitation est souvent considérée comme un bruit blanc, un signal périodique ou un mélange des deux. Un signal périodique résulte d'une excitation voisée

qui se produit lorsque les cordes vocales vibrent. Le bruit aléatoire se manifeste dans le cas d'une excitation non voisée au niveau d'une constriction dans le conduit vocal créant un flux d'air turbulent.

On appelle voisement le mélange de ces deux excitations. L'excitation périodique est caractérisée par une fréquence fondamentale appelée « pitch».

6.2. Analyse fréquentielle (spectrale):

6.2.a) Transformée de Fourier Discrète à Court Terme :

En traitement de la parole, la plupart des paramètres se trouvent dans le domaine spectral. Le signal de parole est plus facile à analyser et plus systématique en fréquence qu'en temps. Aussi, le spectre de sortie du modèle de production de la parole n'est rien d'autre que le produit de la réponse fréquentielle du conduit vocal et le spectre de l'excitation. Donc, il est évident que cette sortie spectrale reflète les propriétés des réponses fréquentielles de l'excitation et du conduit vocal. Un des processus les plus communément utilisée pour obtenir le spectre d'un signal temporel est la transformée de fourrier (TF). Mathématiquement, on définit la TF d'une séquence comme :

$$S(e^{j\omega}) = \sum_{n=-\infty}^{\infty} s(n)e^{-j\omega n} \qquad (1.1)$$

Où $s(n)$ représente l'échantillon de la séquence.

Comme il a été expliqué précédemment, vu la nature non stationnaire du signal de parole, due aux changements du système phonatoire dans le temps, il est impératif de considérer des séquences courtes en temps telle que les caractéristiques du conduit vocal restent inchangées. On montre qu'une durée de 25 à 30 ms permet de vérifier cette supposition, on dira que le signal est quasi stationnaire et on définit alors une nouvelle forme de la transformée de Fourier dépendant du temps par :

$$S_m(e^{j\omega}) = \sum_{n=-\infty}^{\infty} w(m-n).s(n)e^{-j\omega n} \qquad (I.2)$$

Où $w(n)$ est une fonction de fenêtrage qui est nulle sauf pour une durée finie. La variable m est utilisé pour sélectionner la partie de séquence à analyser. La fonction S_m est appelée Transformée de Fourier à court terme (TFCT).

Pour des raisons de calcul, on échantillonne la fréquence de l'équation **(I.2),** on obtient par conséquent, la transformé de Fourier discrète comme suit :

$$S_m(k) = \sum_{n=0}^{N-1} w[m-n].s[n]e^{-\frac{j2\pi nk}{N}}, \quad pour\ k = 0, \dots, N-1 \qquad (I.3)$$

L'amplitude du spectre du signal de parole, définie par la TFD, permet d'extraire plusieurs caractéristiques utiles comme la fréquence fondamentale d'un son voisé à travers la séparation des pics. Les pics les plus larges sont appelés «formant», ils résultent des résonances du conduit vocal.

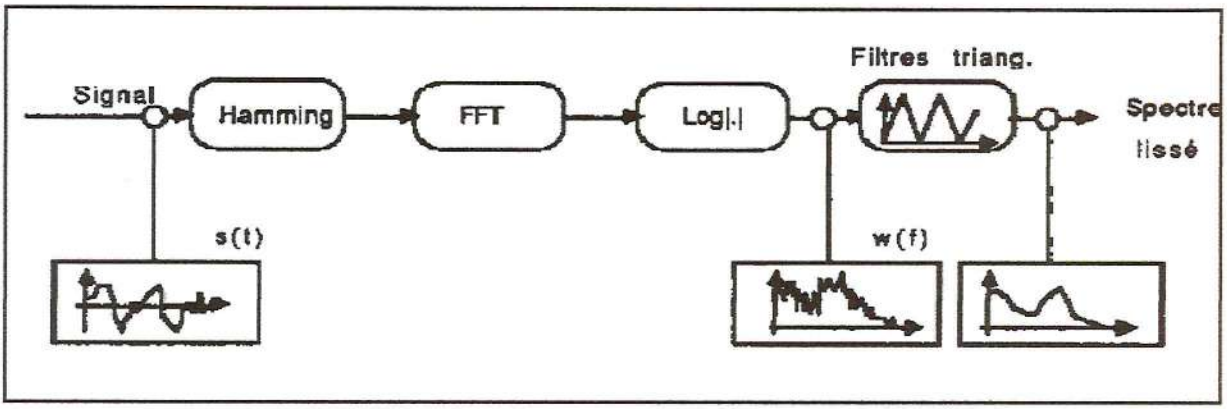


Figure (I.6) : du signal vers son spectre lissé **[4]**

Pour réduire la distorsion spectrale due au fenêtrage, la fonction **w(n)** doit satisfaire deux conditions. Premièrement, un lobe principal étroit et fort. Deuxièmement, une large atténuation des lobes secondaires. Comme ces deux conditions sont contradictoires, un compromis doit être fait, ce qui a donné lieu a plusieurs fenêtres. Le tableau suivant résume les propriétés de fenêtres utilisées dans la pratique.

nom de la fenêtre	W[n] dans $-\frac{N-1}{2} \le n \le \frac{N-1}{2}$	largeur du lobe principal (x(π/N))	niveau des lobes secondaire [dB]
rectangle	1	4	-13.3
Hanning	$\frac{1}{2}(1+\cos(\frac{2\pi n}{N}))$	8	-31.5
hamming	$0.54+0.46\cos(\frac{2\pi n}{N})$	8	-42.7
Blackman	$0.42+0.5\cos\left(\frac{2\pi n}{N}\right)+0.08\cos(\frac{4\pi n}{N})$	12	-58.1

Tableau (I.1): Propriétés de quelques fenêtres connues

Bien que la fenêtre rectangle soit la plus simple à construire, ses niveaux de lobes secondaires très élevés font qu'elle soit très peu utilisée. Dans le domaine des signaux de parole, la fenêtre la plus utilisée est celle de Hamming, d'autres méthodes de fenêtrage existent parmi lesquelles celle proposée dans ce travail à savoir le fenêtrage multiple.

6.2.b Spectrogramme :

La **Transformé de Fourier à Court Terme** permet d'obtenir une représentation temporelle et fréquentielle simultanée pour un signal : **le spectrogramme**. L'amplitude du spectre y apparait sous la forme de niveaux de gris **[3].**

- Une fenêtre plus courte permettra d'identifier **les sons fricatifs**. Ceux-ci seront caractérisés par des régions remplies en raison d la nature aléatoire de leur onde.

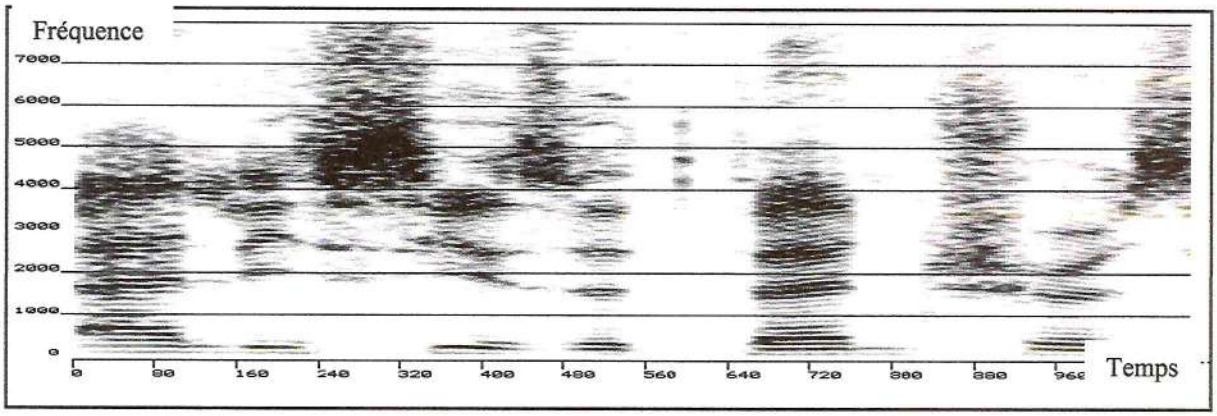


Figure (I.7) : spectrogramme à bande étroite (d'un son non voisé) **[5]**

- Une fenêtre plus large permettra d'identifier **les sons voisés**. Ceux-ci seront caractérisés par des rayures verticales.

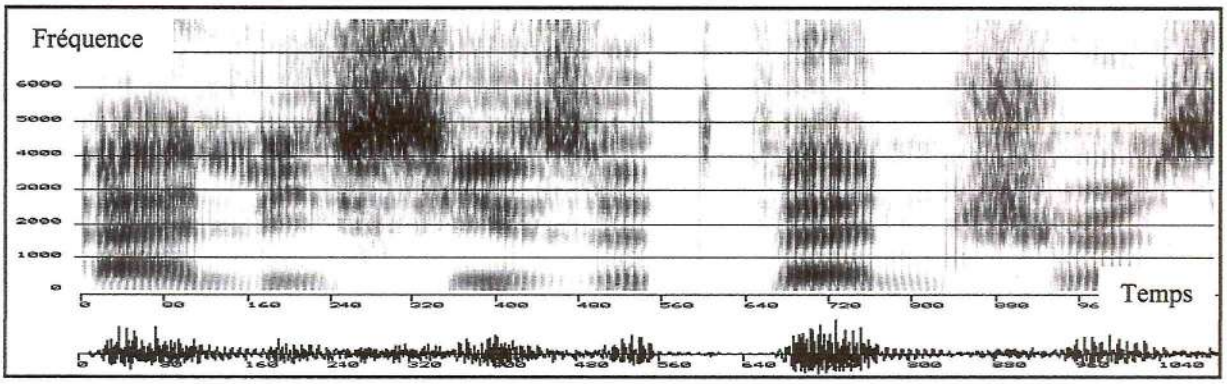


Figure (I.8) : spectrogramme à large bande (d'un son voisé) **[5]**

Les figures **(I.7)** et **(I.8)** illustrent la **Transformée de Fourier** d'une tranche **non-voisée** et d'une tranche **voisée** respectivement. Les parties **voisées** du signal apparaissent sous la

forme de successions de pics spectraux marqués et des creux correspondants aux résonnances et aux anti-résonnances du conduit vocal et sont appelés **formants** et **anti formants**, par contre, le spectre d'un signal non voisé ne présente aucune structure particulière. La forme générale de ces spectres, appelée **enveloppe spectrale**.

6.3. Bancs de filtre passe bandes :

Le signal électrique obtenu à la sortie du microphone (après avoir été amplifié et pré-accentué) passe à travers un banc de filtres passe-bande (figure 1.9). Le signal à la sortie de ces filtres passe ensuite à travers un circuit détecteur d'enveloppe, On obtient, ainsi, une approximation du module de la **TF** à court terme dans différentes bandes de filtres.

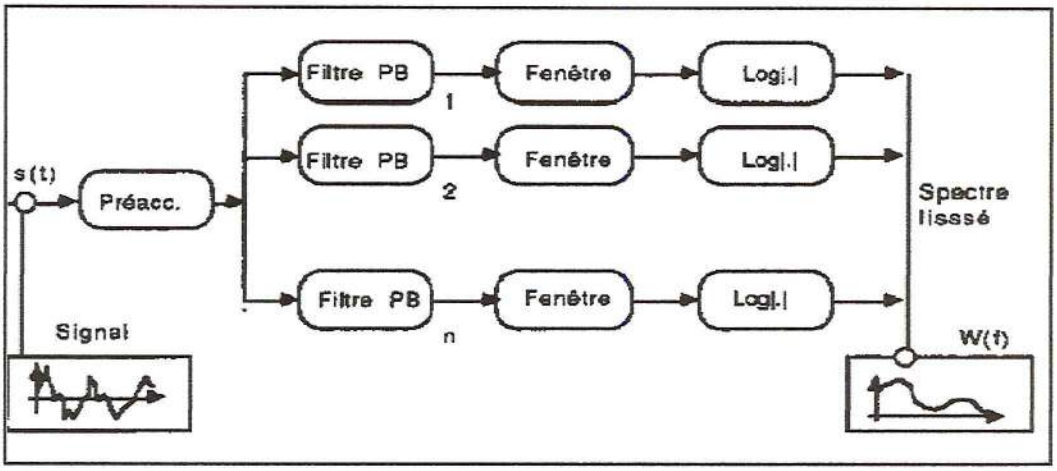


Figure (I.9) : analyse de bancs de filtre **[2]**

6.4. Analyse cepstrale :

Les éléments d'un vecteur spectral (contenant les spectres d'énergie des fréquences) ne sont pas indépendants et il existe par la suite une redondance d'information. Le cepstre est une manière d'obtenir une version non-corrélée du vecteur spectral corrélé.

Le cepstre est aussi une interprétation du modèle de production de la parole. On part de l'hypothèse que la suite S constituant un signal vocal est le résultat de la convolution du signal de la source par le filtre correspondant au conduit vocal :

$$\mathbf{s = u * h} \qquad \textbf{(I.4)}$$

Où **s**, **u** et **h** le signal de parole, le signal d'excitation (source) et la fonction de transfert du conduit vocal (filtre), le but du cepstre est de séparer ces deux contributions (source et filtre) par application de la dé-convolution à travers une **T**ransformée en **C**osinus **D**iscret

$$\mathbf{TFD(S) = U.H} \qquad \textbf{(I.5)}$$

Le logarithme de l'amplitude transforme le produit de la TFD en somme. On obtient alors :

$$\log | S(v) | = \log | U | + \log | H | \qquad (I.6)$$

Par transformations en cosinus discret (DCT), on obtient le cepstre. L'expression du cepstre réel est donc :

$$C = DCT(\log(TFD(s))). \qquad (I.7)$$

Les coefficients cepstaux permettent la séparation des effets du conduit vocal et de la source périodique. En effet, la contribution du conduit vocal est représentée par l'enveloppe du spectre de fréquence et affecte en conséquence les premiers coefficients cepstraux. Les caractères plus fins du spectre de fréquence, qui représentent la source périodique, sont décrits par les coefficients cepstraux d'ordre supérieur. Il est possible de définir la fréquence fondamentale de la source u en détectant les pics périodiques (harmoniques) au-delà d'un certain nombre N de coefficients. Cependant, déterminer la fréquence fondamentale d'un signal de parole reste encore un problème difficile. En effet, les algorithmes classiques manquent de robustesse quand le bruit est présent, quand la fréquence fondamentale change rapidement ou quand la valeur de celle-ci n'est pas assez élevée.**[3]**

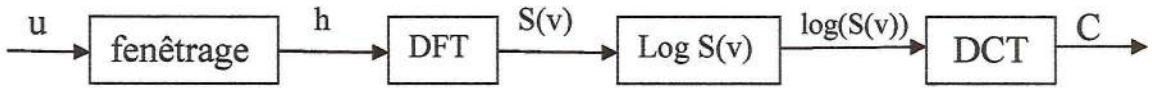


Figure (I.10): étapes de calcul du cepstre

L'analyse cepstrale est basée sur un calcul de coefficients dits coefficients cepstraux de **Mel**, soit en abrégé **MFCC(Mel Frequency Cepstral Coefficients)**.Le calcul est en effet basé sur une échelle fréquentielle de **Mel**(figure **1.11**). Cette échelle se rapproche de la perception fréquentielle de l'oreille. L'idée est de moyenner le spectre dans des bandes de fréquences correspondant grossièrement au filtrage effectué par la membrane basilaire. L'échelle **Mel** est approchée par un banc de 15 à 24 filtres triangulaire espacées linéairement jusqu'à 1 KHz,

puis espacés logarithmiquement jusqu'aux fréquences maximales. Elle a été conçue de telle façon que 1000 Hz corresponde à 1000 Mels.**[3].**

La formule de conversion Hertz en **Mel** est :**[1]**

$$Mel = 2595\ log_{10}(1 + \frac{fhz}{700}) \qquad (I.8)$$

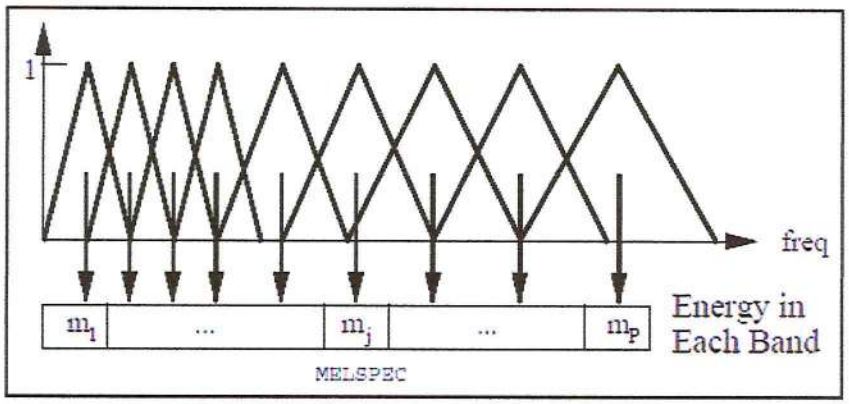


Figure (I.11) : banc de filtre à l'échèle de Mel **[6]**

Nous allons faire une description plus détaillée de la façon de calculer les MFCC

- Découper le signal en trames courtes.
- Pour chaque fenêtre renforcer(donner plus d'énergie) la contribution des hautes fréquences (préaccentuation) ; avec un filtre passe-haut puis multiplier les échantillons de la trame par une fonction fenêtrage (Hamming par exemple)
- Calculer la **TFD** sur chaque trame.
- Appliquer le banc de filtre triangulaire d'échelle de Mel aux spectres de puissance,
- Prenez le logarithme de toutes les énergies en sortie de la banque de filtres.
- Prenez la DCT^{-1} (**T**ransformé en **C**osinus **D**iscrète inverse) des énergies du banc de filtres.
- Conserver les 12 premiers coefficients DCT.

6.5. Analyse prédictive linéaire(LPC):

L'analyse prédictive linéaire appelée aussi modélisation auto régressive (AR) fait partie des techniques les plus puissantes pour l'analyse des signaux de parole. Elle est utilisée pour estimer les paramètres de la parole comme le **pitch,** les formants, ou les fonctions d'aire du conduit vocal. L'intérêt de cette méthode réside dans sa précision d'estimation et sa relative vitesse de calcul des différents paramètres de la parole. La méthode de codage prédictif linéaire (*LPC,Linear Predictive Coding*)repose principalement sur l'hypothèse que la

parole peut être modélisée par un processus linéaire. Il s'agit donc de prédire le signal à un instant *n* à partir des *p* échantillons précédents (équation **(I.11)**)

Une bonne approximation des signaux de parole est assurée par un filtre tout pôle avec un nombre suffisants de pôle. La réponse fréquentielle du système est donnée par :

$$H(z) = \frac{X(z)}{E(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k \ z^{-k}} = \frac{G}{A(z)'} \qquad \textbf{(I. 9)}$$

Où p est l'ordre des coefficients de l'analyse par prédiction linéaire,

A(z) est le prédicateur linéaire, et $\mathbf{a_k}$ sont les coefficients de prédiction.

$$A(z) = \sum_{k=1}^{p} a_k z^{-k} \qquad \textbf{(I. 10)}$$

X(z) est le signal de parole produit en sortie, E(z) est le signal d'excitation et G est un gain.

On représente le signal de parole x(n) à la sortie du système par la transformé en z inverse de l'équation (1.9) qui donne :

$$x[n] = \sum_{k=1}^{p} a_k x[n-k] + Ge[n] \qquad \textbf{(I. 11)}$$

Il permet de représenter la parole juste par ses paramètres pertinents, sans redondance. Signalons également que les coefficients sont choisis de façon à minimiser l'erreur quadratique de prédiction sur chaque trame d'analyse.

7. Conclusion :

Ont été exposées dans ce chapitre les notions de base en traitement de la parole. L'accent a été mis sur les aspects qui apparaitront souvent dans la suite de ce manuscrit. Le prochain chapitre abordera avec plus de détails la reconnaissance automatique de la parole par les modèles de Markov cachés MMC

Chapitre II : les modèles de Markov cachés

1. Introduction :

Dans ce chapitre nous exposons les aspects théoriques des modèles de Markov cachés (MMC) ainsi que leur utilisation dans le contexte de la reconnaissance automatique de la parole. On s'intéressera particulièrement à la reconnaissance de mots isolés.

Le noyau de tous les systèmes de reconnaissance vocale se compose d'un ensemble de modèles statistiques représentant les différents sons ou les différents mots de la langue à reconnaître. Puisque la parole a une structure temporelle et peut être codée en tant que séquence de vecteurs spectraux couvrant la gamme de fréquences audio, les **M**odèles de **M**arkov **C**achés(**MMC**) **[4]** fournissent un cadre naturel pour la construction (génération) de telles séquences. Les MMC sont pratiquement au cœur de tous les systèmes modernes de reconnaissance de la parole.

2. Présentation des modèles de Markov cachés:

Un modèle de Markov caché MMC (appelé aussi chaine de Markov cachée) est un modèle statistique dans lequel le système modélisé est supposé être un processus Markovien avec des paramètres inconnus. Il s'agit d'un ensemble d'états, chaque état est susceptible de générer un évènement (symbole) non-déterministe associé avec une probabilité d'émission d'observation propre à cet état **[7]**. En outre, la transition d'un état du modèle vers un autre état du modèle est, elle aussi, associée à une probabilité de transition .D'une manière formelle, les éléments d'un HMM sont:

1. Un ensemble fini d'états :

   $Q = 1, 2, 3, ..., N$; où N est le nombre d'états,

   Une séquence d'états de longueur temporelle T sera dénotée :

   $Q = q_1, q_2, ..., q_T$; où q_t est l'état à l'instant t.
2. Un ensemble de symboles d'observation distincts pour chaque état correspondant à la sortie physique du système en cours de la modélisation. Si le nombre de symboles possibles est discret, on parlera de MMCs discrets sinon les MMCs sont dits continus. La séquence de sortie observée de longueur T est :

   $O = O_1, O_2, ..., O_t, O_T$; où O_t est la sortie observée (vecteur de paramètres) à l'instant t, O_T est le dernier vecteur de paramètres (le mot contient donc T vecteurs)
3. La matrice de probabilités de transitions de l'état ***i*** à l'état ***j*** :

$$\mathbf{A} = \{\mathbf{aij}\}; \quad \text{Où} \quad a_{ij} = P(q_t = j | q_{t-1} = i) \qquad \textbf{(II. 1)}$$

La distribution des probabilités d'émission du symbole d'observation à l'état j (probabilités d'émission) : $\mathbf{B} = \{bj\ (O_t)\}$; où

$$b_j(O_t) = P(O_t | q_t = j) \qquad \textbf{(II. 2)}$$

Cette distribution peut être discrète (MMCs discrets) ou continue (MMCs continus), dans ce dernier cas la distribution d'usage est un mélange de gaussiennes [7]

4. La distribution initiale des états :

$$\pi = \{\pi i\}; \qquad \text{où} \quad \pi i = P\,(q1 = i) \qquad \textbf{(II. 3)}$$

La définition complète d'un MMC nécessite donc la spécification des paramètres cités plus haut.

3. Reconnaissance des mots isolés : [6]

Considérons le signal issu d'un mot prononcé par un locuteur donné. Comme expliqué dans le premier chapitre, ce signal subira certains prétraitements (subdivision en trames, puis calcul des coefficients MFCC pour chaque trame). Notre mot *(O)* sera finalement présenté sous forme d'une suite de vecteurs acoustiques ou bien vecteurs d'observations (vecteur de paramètres) *O.*

Le problème de reconnaissance de mots isolés revient à calculer

$$arg\ \max_i \{P(W_i/O)\} \qquad \textbf{(II. 4)}$$

Où Wi représente le $i^{ème}$ mot du vocabulaire.

Cette probabilité n'est pas calculable directement, mais en utilisant la formule de Bayes**[C]** on obtient :

$$P(W_i/O) = P(O/W_i)\,P(W_i)/P(O) \qquad \textbf{(II. 5)}$$

Donc pour un ensemble donné de probabilité à priori $P(W_i)$, le mot prononcé le plus probable dépend uniquement de la vraisemblance $P(O/W_i)$.

Ayant la dimensionnalité de la séquence d'observation O, l'estimation directe de la probabilité conditionnelle jointe $\mathbf{P(O_1, O_2, \ldots, O_T/W_i)}$ à partir d'exemples de mots prononcés n'est pas envisageable. Cependant si on dispose d'un modèle paramétrique de production du mot, tel qu'un MMC, l'estimation de $P(O/W_i)$ est remplacée par l'estimation des paramètres

du MMC. Dans la reconnaissance par MMC on considère que la séquence d'observation *O* correspondant à chaque mot prononcé est générée par un MMC.

La figure **(II.1)** montre un exemple de MMC à 6 états. Dans cet exemple la séquence d'observations ***(O=O1O2O3O4O5O6)*** est générée en passant par les séquences d'états **X=1,2,2,3,4,4,5,6** les états **1** et **6**, représentant respectivement l'état initial et final, ce sont des états non émetteurs.

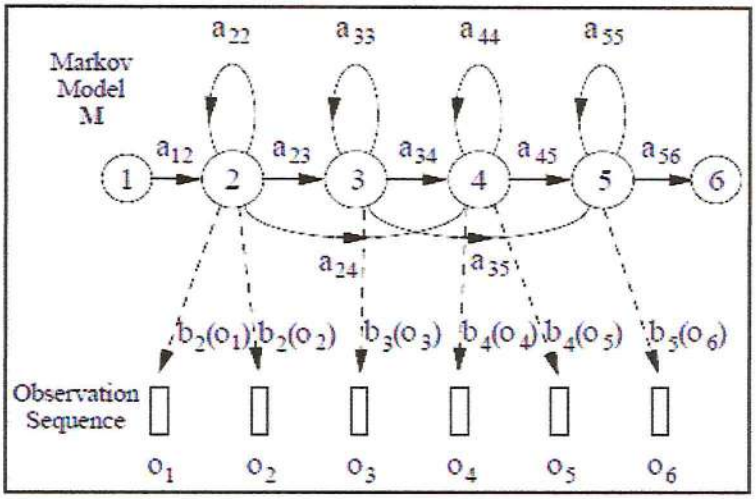


Figure (II.1) : exemple de modèle générateur de Markov

La probabilité jointe de génération de la séquence d'observations ***O=O1O2O3O4O5O6*** par le modèle de notre exemple en passant par la séquence d'états **X=1,2,2,3,4,4,5,6** est donnée par :

$$P(O,X/M) = a_{12}b_2(O_1)a_{22}b_2(O_2)a_{23}b_3(O_3)a_{34}b_4(O_4)a_{44}b_4(O_5)a_{45}b_5(O_6)a_{56} \ldots \quad \textbf{(II.6)}$$

Il est facile de remarquer qu'il y a plusieurs manières de transiter entre les états (et donc plusieurs chemins) pour générer notre séquence d'observations *O*. En pratique, seule la séquence d'observation *O* est connue, la séquence d'états X sous-jacente est cachées d'où le nom de Chaine de Markov **Cachée**.

Etant donné que **X** est inconnue, la vraisemblance est calculée en sommant sur toutes les séquences d'états possibles. **X= X(1), X(2), … X(T)**

C'est-à-dire :

$$P(O/M) = \sum_{X} a_{X(0)X(1)} \prod_{t=1}^{T} b_{X(t)}(O_t)a_{X(t)X(t+1)} \quad \textbf{(II.7)}$$

En disposant d'un ensemble d'exemples d'apprentissage correspondant à un modèle particulier, les paramètres de ce modèle peuvent être déterminés automatiquement par une procédure efficace de ré-estimation de paramètres .Ainsi donc, si l'on collecte un nombre

suffisant d'exemples représentatifs de notre mot, on pourra donc construire un MMC correspondant qui modéliser implicitement les sources de variabilité inhérente au signal vocal réel.

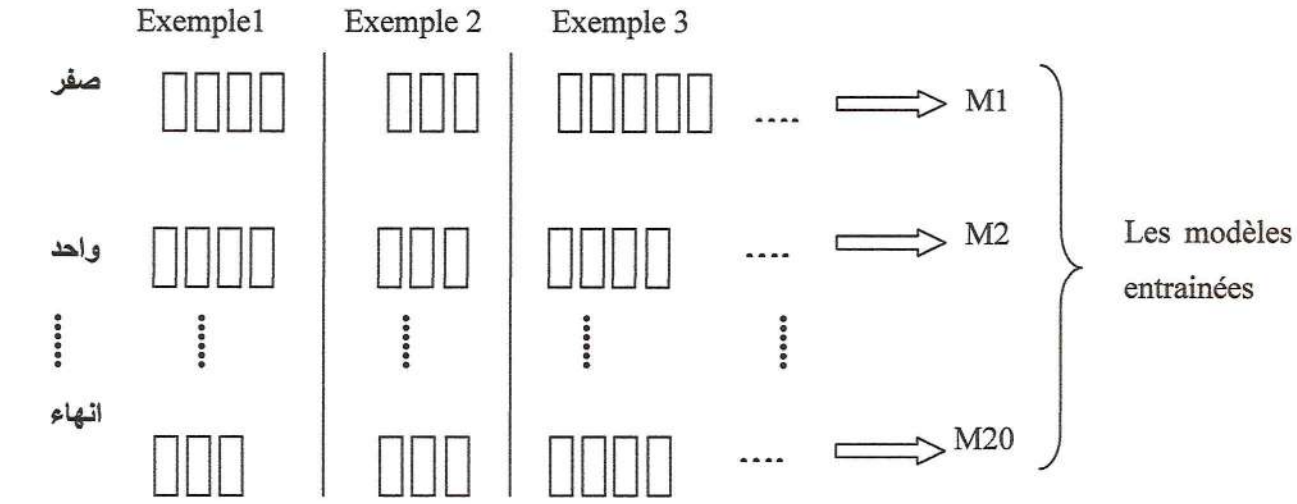
La figure **(II.2)** récapitule l'utilisation des MMC pour la reconnaissance de la parole. Nous allons nous placer dans le cadre de notre travail.

Le vocabulaire de notre système est composé de vingt (20) mots isolés représentant notre base de données prononcés en arabe.

Etant donné que chaque mot est représenté par un MMC on doit disposer de20 MMCs (un MMC pour chaque mot)

- Premièrement chaque MMC est entrainé par un ensemble d'exemples du mot lui correspondant. On obtiendra à la fin de cette phase les modèles entrainés (M1… M20)
- Deuxièmement ayant un mot inconnu, on calcule la vraisemblance de génération de ce mot par tous les MMC. On choisira le MMC ayant donné le maximum de vraisemblance.

A : apprentissage :



B : reconnaissance :

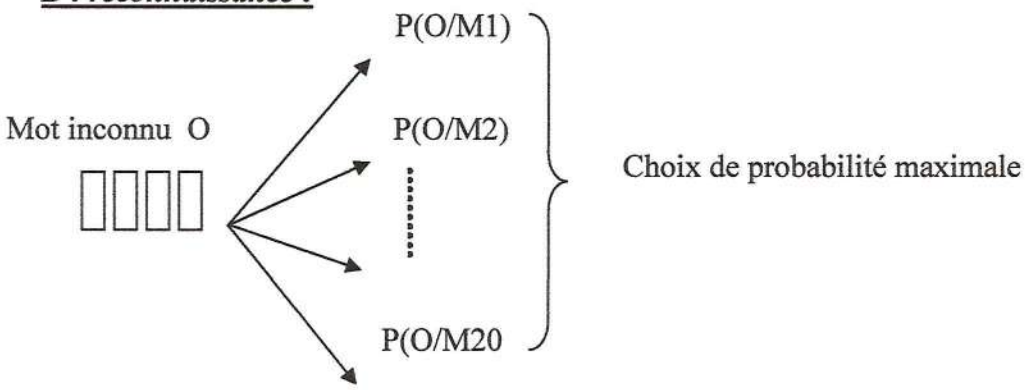


Figure (II.2) : utilisation des MMC dans la reconnaissance des mots isolés

4. L'apprentissage des modèles de Markov cachés :

La phase de la construction d'un modèle est le point crucial de tout système, l'apprentissage revêt donc une grande importance. Les modèles acoustiques nécessitent un important corpus sonore. Une transcription phonétique est associée à chaque échantillon sonore de sorte qu'à la fin, chaque modèle possède des représentants dans le corpus. Une fois l'ensemble des données prêt, la phase d'apprentissage des modèles s'effectue. Nous allons décrire dans cette partie l'algorithme communément utilisé dans la phase d'apprentissage des modèles de Markov cachés pour la reconnaissance de la parole. A partir d'exemples dont nous connaissons à la fois la séquence des modèles et la séquence d'observation engendrée, nous souhaitons déterminer les paramètres définissant les modèles de Markov cachés de chaque unité phonétique. Il nous faut donc estimer pour chaque modèle :

- Les probabilités initiales π_i.
- Les probabilités de transition a_{ij}.
- Les probabilités d'émission b_i **(Ot)** qui sont caractérisées par :
  - Les moyennes μ_i
  - Les matrices de covariance $\sum_i$
  - Les coefficients du mélange de gaussien c_i

Dans le cadre des modèles de Markov cachés, la méthode communément utilisée repose sur le critère du maximum de vraisemblance (Maximum Likelihood **ML**).

L'algorithme de Baum-Welch :

L'estimation de paramètres de model **M= (A, B, π)** est un peu difficile car il n'existe pas de méthode analytique connue pour maximiser la probabilité de la séquence d'observation. Toute fois, les paramètres peuvent être estimés par maximisation de **P(O/M)** en utilisant un algorithme itératif, comme l'algorithme de **Baum-Welch** (connu aussi comme l'algorithme FORWARD/BACKWARD)

Les formules de ré-estimation des paramètres d'un MMC entrainé par un nombre d'exemples égal à (R) sont données par **[8]:**

$$\hat{a}_{ij} = \frac{\sum_{r=1}^{R} \frac{1}{P_r} \sum_{t=1}^{T_r-1} \alpha_i^r(t) a_{ij} b_j(O_{t+1}^r) \beta_j^r(t+1)}{\sum_{r=1}^{R} \frac{1}{P_r} \sum_{t=1}^{T_r} \alpha_i^r(t) \beta_j^r(t)} \qquad \textbf{(II. 8)}$$

$$\hat{a}_{1j} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{P}\alpha_j^r(1)\beta_j^r(1) \qquad \textbf{(II. 9)}$$

$$\hat{a}_{iN} = \frac{\sum_{r=1}^{R}\frac{1}{P_r}\alpha_i^r(T)\,\beta_i^r(T)}{\sum_{r=1}^{R}\frac{1}{P_r}\sum_{t=1}^{T}\alpha_i^r(t)\beta_i^r(t)} \qquad \textbf{(II. 10)}$$

Pour les gaussiennes simples, il est plus efficace dans ce cas d'utiliser :

$$l_j^r(t) = \frac{1}{p_r}\alpha_j(t)\beta_j(t) \qquad \textbf{(II. 11)}$$

Où $\boldsymbol{Pr = P\,(O/M)}$

Les formules de ré-estimation peuvent maintenant être exprimées en terme de $l_j^r(t)$:

$$\hat{\mu}_J = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r} l_j^r(t)o_t^r}{\sum_{r=1}^{R}\sum_{t=1}^{T_r} l_j^r(t)} \qquad \textbf{(II. 12)}$$

$$\hat{\sum}_j = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r} l_j^r\left(o_t^r - \hat{\mu}_j\right)\left(o_s^t - \hat{\mu}_j\right)'}{\sum_{r=1}^{R}\sum_{t=1}^{T_r} l_j^r(t)} \qquad \textbf{(II. 13)}$$

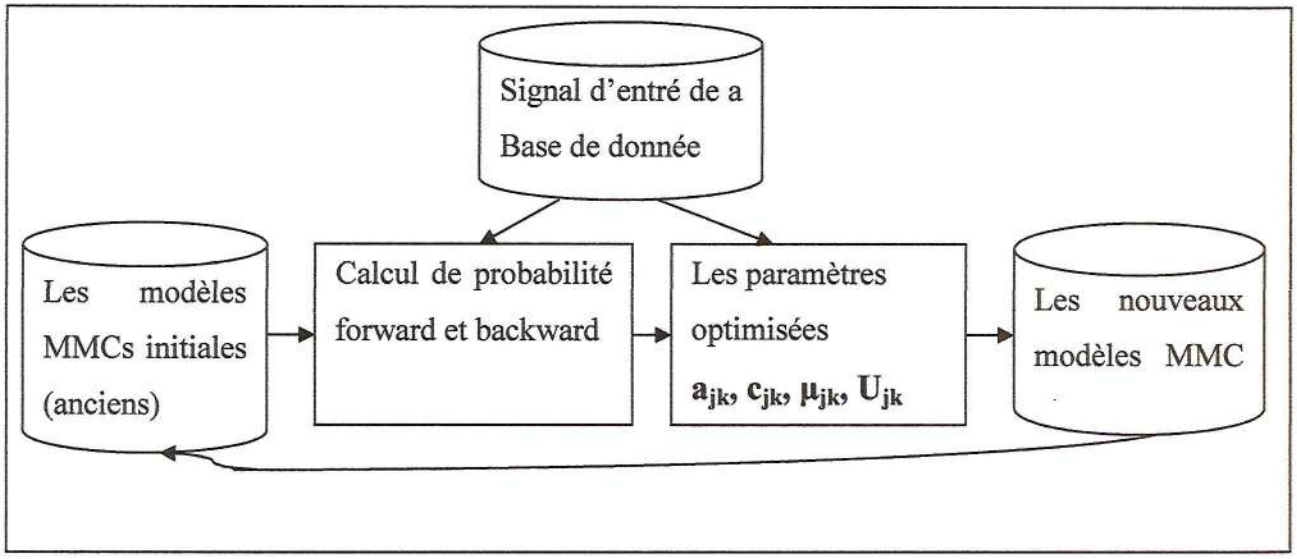


Figure (II.3) : procédure d'apprentissage de Baum Welch basée sur un ensemble d'apprentissage donné

5. Reconnaissance et décodage par l'algorithme de Viterbi :

La section précédente a décrit les idées de base qui sous-tendent ré-estimation des paramètres du MMC à l'aide de l'algorithme de Baum-Welch, il est à noter que cet algorithme récursif est efficace pour le calcul de la probabilité directe (Forward).

Ainsi, cet algorithme peut également être utilisé pour trouver le modèle qui donne la valeur maximale de P (O|M), et, par conséquent, il pourrait être utilisé pour la reconnaissance.

Cependant, il est préférable à la reconnaissance de se baser sur la séquence d'états qui a générer la probabilité du maximum de vraisemblance, Cette probabilité est calculée en utilisant essentiellement le même algorithme de Farward sauf que la sommation est remplacée par un opérateur de maximisation.

Pour un modèle M donné, Soit $\varphi_X(t)$le maximum de vraisemblance de la séquence O_1à O_t sachant qu'on se trouve dans l'état **X** au temps **t**, cette vraisemblance partielle peut être calculée comme suit :

$$\varphi_X(t) = \max_i\{\varphi_i(t-1)a_{iX}\}b_X(o_t) \quad \textbf{(II.14)}$$

Avec $$\varphi_x(1) = a_{1X}b_X(o_1) \quad \textbf{(II.15)}$$

Pour **1 < X < N,** le maximum de vraisemblance **P (O|M)** est alors donné par :

$$\varphi_N(t) = \max_i\{\varphi_i(T)a_{iN}\} \quad \textbf{(II.16)}$$

L'algorithme en relation avec le calcul de cette entité est appelé l'algorithme de Viterbi **[9]**

La figure **(II.4)** illustre la reconnaissance des mots isolés par l'algorithme de Viterbi :

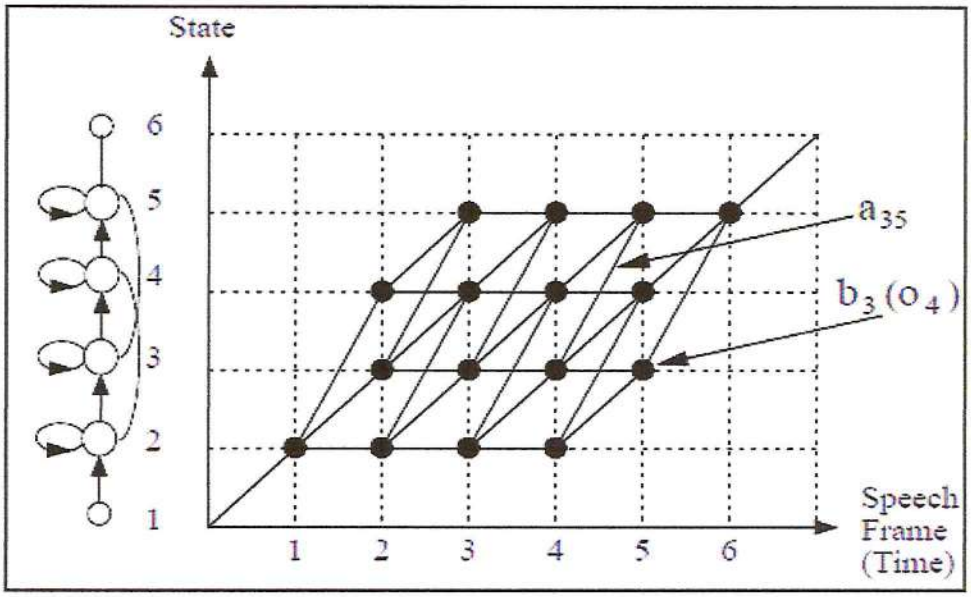


Figure (II.4) : algorithme de Viterbi pour la reconnaissance des mots isolés **[6]**

6. Conclusion :

Dans ce chapitre, nous avons exposé brièvement les modèles de Markov cachés et leurs applications dans le domaine de reconnaissance automatique de la parole, l'accent a été mis sur l'aspect apprentissage et reconnaissance (décodage) à travers les deux algorithmes (Baum-Welch et Viterbi).Les MMCs ont été utilisés dans notre cas comme des modèles générateurs de mots, ces mots en question se présentant sous forme d'une séquence de vecteurs spectraux qui ont été obtenus à partir du signal vocal original via l'opération d'extraction de paramètres

(anglais :feature extraction). Il va sans dire que les performances de la modélisation markovienne présentée dans ce chapitre dépendent de la fiabilité de l'étage d'extraction de paramètres, notamment concernant l'estimation du spectre du signal vocal. C'est précisément cet aspect qui sera abordé dans le prochain chapitre.

Chapitre III : estimation spectrale par fenêtrage multiple

1. Introduction :

Ce chapitre sera consacré au rôle du fenêtrage dans l'estimation du spectre du signal vocal. Il est important de noter qu'il n'est nullement question ici de développer tous les aspects théoriques de l'estimation spectrale qui, malgré son importance, ne peut faire l'objet de quelques lignes ; le lecteur intéressé pourra trouver plus d'informations dans **[3]**. Un intérêt spécial sera attaché à l'effet de la forme de la fenêtre sur la variance du spectre estimé. On va donc présenter quelques fonctions de fenêtrage communément employées. Aucune fenêtre n'étant la meilleure dans tous les aspects, il s'agit de chercher le type de fenêtre le plus approprié à l'application en question. L'idée principale de ce chapitre s'articule sur le fait qu'il existe une approche plus élégante se basant sur l'utilisation de plusieurs fenêtres (d'où le nom : multi-taper) simultanément sur la même tranche de signal afin de minimiser la variance de l'estimation.

2. Analyse spectrale :

Le spectre d'un signal représente son contenu fréquentiel (harmoniques). L'information complète se traduit par deux quantités (amplitude et phase de ces harmoniques). En traitement de parole on s'intéresse généralement à l'amplitude des harmoniques (spectre d'amplitude). L'obtention du spectre d'amplitude passe par le calcul de la transformée de Fourier du signal, toutefois la nature non stationnaire du signal de parole impose que le calcul de la TF se fasse sur des tranches temporelles limitées (typiquement 25 à 30 ms) pendant lesquelles le signal peut être considéré comme stationnaire. Ceci revient donc à multiplier le signal de parole original par une fenêtre (anglais : window ou Taper).On obtient ainsi une série temporelle échantillonnée. Obtenir le spectre de cette série relève d'un domaine appelé communément l'estimation spectrale.

La technique la plus souvent appliquée pour estimer un spectre de puissance d'une série temporelle échantillonnée de longueur finie N se base sur le calcul de la transformé de Fourier suivi par la multiplication par sa conjuguée complexe et divisée par la longueur N.

$$\widehat{S}(f) = X(2\pi f).\frac{X^*(2\pi f)}{N} \qquad \textbf{(III. 1)}$$

La fiabilité de ce spectre de puissance estimé est réduite significativement à la fois par la variance de l'estimation $S(f)$ pour chaque fréquence (les spectres sont habituellement très bruités) et par la fuite d'énergie à travers les fréquences, en créant le biais. La fuite est due au

fait que nous prenons une trame finie du signal, ce qui équivaut à multiplier le signal par une fenêtre rectangulaire. La transformé de Fourier de la fenêtre rectangulaire montre des lobes secondaires multiples ce qui cause le problème de la fuite d'énergie. A rappeler que la multiplication dans le domaine temporel est équivalente à une convolution dans le domaine fréquentiel, par conséquent, il y aura une convolution de la transformé de Fourier du signal et de la fenêtre rectangulaire conduisant à une fuite d'énergie. Une solution pour réduire les fuites dans le domaine fréquentiel consiste à multiplier le signal dans le domaine temporel avec une fenêtre non-rectangulaire caractérisée par une transformée de Fourier avec moins d'énergie dans ses lobes secondaires.

Le spectre estimé est donnée par :

$$\widehat{S}(f) = \left| \sum_{t=0}^{N-1} x(t)a(t)\, e^{-2\pi jft} \right|^2 \qquad \textbf{(III. 2)}$$

Si la fenêtre a(t) n'est explicitement pas présente dans l'équation **(III.2),** ou si elle est définie comme fenêtre rectangulaire, l'estimation spectrale $\widehat{S}(f)$ est appelé le **périodogramme**. On peut s'attendre dans ce cas à ce que le biais dû à la fuite soit élevé **[3].**

Il existe deux catégories pour les méthodes d'estimation spectrales :

- méthodes non-paramétriques.
- méthodes paramétriques.

- Méthodes paramétriques :

Ces méthodes consistent à estimer le spectre à partir des paramètres du modèle.

Alternative à l'analyse classique, l'analyse spectrale paramétrique présuppose donc une connaissance particulière du phénomène observé. Elle consiste en trois étapes. Dans un premier temps, on sélectionne un modèle : les paramètres du modèle sont alors estimés de façon à approximer au mieux la séquence de signal observée. Enfin, on réécrit la Densité Spectrale de Puissance DSP directement en fonction des paramètres.

- Méthodes non-paramétriques :

Dans ces méthodes on travaille directement sur le signal (et non sur le modèle qui aurait pu générer ce signal) pour trouver la Densité Spectrale de Puissance (DSP) estimée.

Périodogramme :

La densité spectrale de puissance est donnée par :

$$P_{xx}(e^{j\omega}) = \lim_{N\to\infty} \left\{ \frac{1}{2N+1} \left| \sum_{n=-N}^{N} x[n]e^{-j\omega n} \right|^2 \right\} \qquad \textbf{(III. 3)}$$

Etant donné qu'une multiplication dans le domaine temporel est équivalente à une convolution dans le domaine fréquentiel, le périodogramme est en fait la convolution du spectre du signal par une fenêtre en sinus cardinal. Le fait de tronquer le signal induit 2 phénomènes principaux :

- Élargissement du lobe principal ⇒ perte de résolution ;
- Apparition de lobes secondaires.

On montre que la résolution du périodogramme est de l'ordre de 1/N, c'est-à-dire que l'on peut discriminer 2 fréquences distantes de 1/N.

Pour remédier aux problèmes de troncature, on utilise en général des fenêtres permettant soit de réduire le lobe principal, soit d'atténuer les lobes secondaires. On montre que le périodogramme est un estimateur biaisé **[2]** :

$$E\{\widehat{S}(f)\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} A(f)\,(f-u)S_x(u)du \qquad \textbf{(III. 4)}$$

Où A(f) est la transformé de Fourier de la fenêtre rectangulaire. Le périodogramme est donc en moyenne la convolution du véritable spectre avec la transformée de Fourier de la fenêtre rectangulaire. Néanmoins, lorsque N $\rightarrow \infty$, le biais devient nul. On peut dire également que la variance est pratiquement indépendante de N et proportionnelle au spectre.

$$var\{\widehat{S}(f)\} \cong S_x(f)^2 \qquad \textbf{(III. 5)}$$

Le périodogramme n'est donc pas un estimateur consistant de la densité spectrale de puissance (DSP). Afin de diminuer la variance de cet estimateur, on peut utiliser un périodogramme moyenné. Ceci consiste à séparer le signal en K tranches (de longueur N/K), et à calculer le périodogramme sur chaque tranche et à faire la moyenne. Du fait des K moyennages, la variance est presque divisée par K : néanmoins, les tranches étant plus courtes, la résolution diminue.

3. Quelques fonctions de fenêtrages :[10]

Dans ce qui suit on va présenter brièvement quelques fonctions de fenêtrage connues dans le domaine de traitement de signal. Chaque fenêtre est décrite par sa forme dans le domaine temporel continue : **f(t),** et sa Transformé de Fourier (TF) : **F(jω).**

τ : représente la durée d'une fenêtre unilatérale dans le domaine temporel.

1) Fenêtre rectangulaire (Box Car) :

Appelée aussi fenêtre uniforme due à sa forme, elle est définie comme suit :

$$f(t) = \begin{cases} 1, & |t| \le \tau \\ 0, & ailleurs \end{cases} \qquad \text{(III. 6)} \qquad \xrightarrow{\text{TF}} \qquad F(j\omega) = \frac{2\tau \sin(\omega\tau)}{\omega\tau} \qquad \text{(III.7)}$$

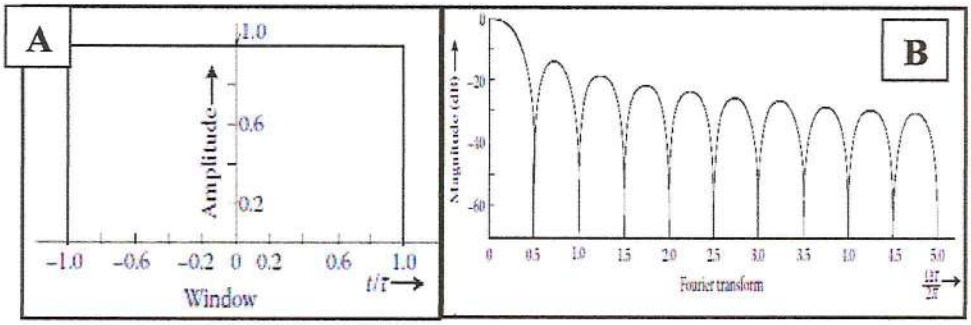


Figure (III.1): (A) : Fenêtre rectangulaire, **(B)** Son spectre d'amplitude

2) Fenêtre de Hann (Hanning) :

Elle est définit comme suit

$$f(t) = \begin{cases} 0.5 + 0.5\cos\left(\frac{\pi t}{\tau}\right), & |t| \le \tau \\ 0 & , ailleurs \end{cases} \qquad \text{(III. 8)}$$

$$\xrightarrow{\text{TF}} \qquad F(j\omega) = \frac{\sin(\omega\tau)}{\omega} + 0.5\left[\frac{\sin\left(\left(\omega+\frac{\pi}{\tau}\right)\tau\right)}{\left(\omega+\frac{\pi}{\tau}\right)} + \frac{\sin\left(\left(\omega-\frac{\pi}{\tau}\right)\tau\right)}{\left(\omega-\frac{\pi}{\tau}\right)}\right] \qquad \text{(III. 9)}$$

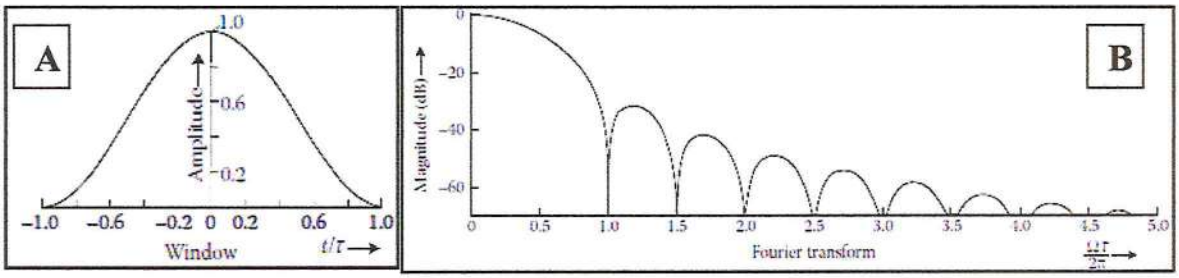


Figure (III.2): (A) Fenêtre de Hann, **(B)** Son spectre d'amplitude

3) Fenêtre de Hamming :

Cette fenêtre proposée par Hamming est une version (forme) optimisée de la fenêtre de Hann ; les coefficients de cette fenêtre sont optimisés de sorte qu'on obtient niveau du premier lobe secondaire (First Side-Lobe Level FSLL) minimal.

Elle est définit comme suit :

$$f(t) = \begin{cases} 0.54 + 0.46\cos\left(\frac{\pi t}{\tau}\right) & , \quad |t| \leq \tau \\ 0 & , \quad ailleurs \end{cases} \qquad \textbf{(III. 10)}$$

$$\xrightarrow{\text{TF}} \quad F(j\omega) = 1.08\left[\frac{\sin(\omega\tau)}{\omega}\right] + 0.46\left[\frac{\sin\left(\left(\omega+\frac{\pi}{\tau}\right)\tau\right)}{\left(\omega+\frac{\pi}{\tau}\right)} + \frac{\sin\left(\left(\omega-\frac{\pi}{\tau}\right)\tau\right)}{\left(\omega-\frac{\pi}{\tau}\right)}\right] \qquad \textbf{(III. 11)}$$

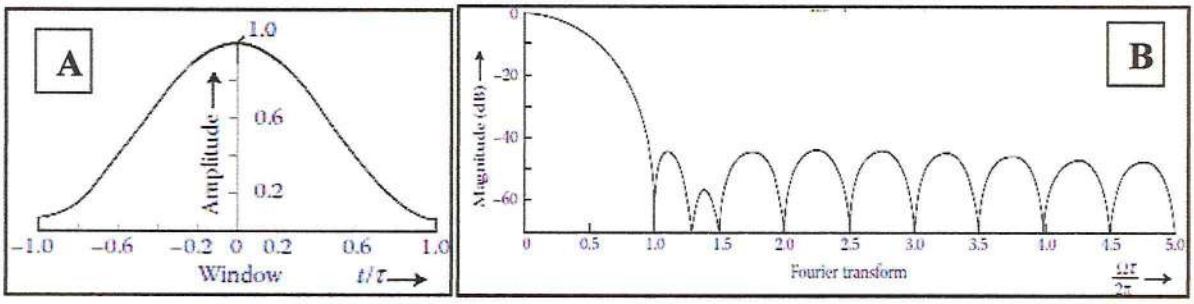


Figure (III.3): (A) : fenêtre de Hamming, **(B)** son spectre d'amplitude

Remarque : d'autres fonctions de fenêtrage sont données en annexe.

4. L'utilisation de fenêtrage multiple :

L'approche de fenêtrage multiple (Multi-Taper) améliore l'estimation spectrale en diminuant la variance du spectre estimé. L'idée de base est simple, utiliser plusieurs fenêtres au lieu d'une seule. Dans cette approche, une série de k fenêtres v_k (judicieusement choisies) est appliquée au signal. Les fenêtres sont orthogonales et elles sont utilisées pour fournir k échantillons orthogonaux (c à d : indépendants) des données x(t). Ces échantillons sont utilisés pour calculer une moyenne $\overline{S}_k(f)$ avec réduction de la variance.

$$\overline{S}(f) = \frac{1}{K}\sum_{k=1}^{K}\widehat{S}_k(f) \qquad \textbf{(III. 12)}$$

L'idée est que chaque fenêtre a(t) est associée à un spectre estimé qui est donné par l'équation **(III.2)** afin de maintenir des valeurs correctes pour l'énergie totale, on suppose que la fenêtre est normalisée tel que $\sum_{t=0}^{N-1}|a(t)|^2 = 1$ **(III. 13)**

L'énergie du spectre de la fenêtre $a(t)$ est $|A(f)|^2$ et ses propriétés spectrales sont importantes puisque elles déterminent, via la convolution, l'estimateur de l'énergie spectrale de notre série temporelle fenêtrée

$$\widehat{S}(f) = \int_{-1/2}^{1/2}|A(f')|^2 S(f-f')df' \qquad \textbf{(III. 14)}$$

On souhaite que l'énergie du spectre de la fenêtre soit concentrée au lobe principal. Autrement dit, ces fenêtres auront des lobes secondaires de faible énergie. Ce qui conduit à une meilleure estimation $\widehat{S}(f)$

Supposons qu'on veut estimer notre spectre avec une résolution de bande passante *W*, qui est nécessairement définit à une valeur entre la résolution du spectre 1/N et le maximum du spectre, la fréquence de Nyquist.

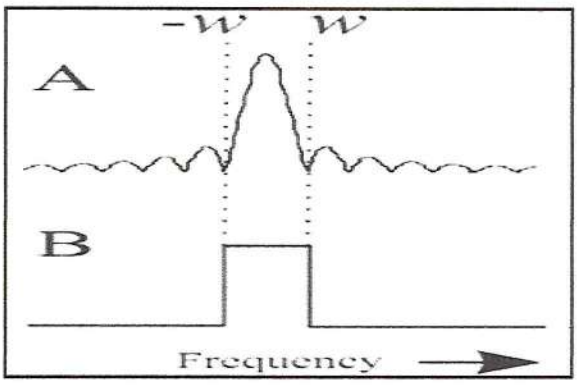


Figure (III.4) : (A) Spectre d'une fenêtre rectangulaire, **(B)** spectre d'une fenêtre idéale **[11]**

Le spectre de la fenêtre rectangulaire (boxcar) illustré dans la figure **(III.4)(A)** a un maximum d'énergie concentré dans la bande passante désirée indiqué entre les lignes en pointillés, mais aussi, on voit beaucoup d'énergie au niveau des lobes secondaires.

La fonction en **(III.4)(B)** est le spectre d'une fenêtre idéale où l'énergie totale se trouve dans l'intervalle entre **–w** et **w**. Quand un spectre est convolué avec les spectres de ces fenêtres,**(A)** génère des fuites importantes et donc un biais, alors que l'idéal sur **(B)** donne une estimation sans biais.

Pour simplifier la notation, on suppose que l'intervalle de l'échantillon est unitaire et la fréquence de Nyquist est normalisée à 1/2. La fraction **λ** de l'énergie de la fenêtre dans la bande de fréquence sélectionnée est donnée par :

$$\lambda(N,W) = \frac{\int_{-W}^{W} |A(f)|^2 \, df}{\int_{-1/2}^{1/2} |A(f)|^2 \, df} \qquad \textbf{(III. 15)}$$

L'idée de base c'est qu'on veut trouver la fenêtre avec une fuite minimale en maximisant **λ**, i.e. la fraction de l'énergie dans la bande passante W est optimisée (maximisée).

Il s'agit donc de maximiser $\boldsymbol{\lambda}$ en définissant la dérivé de l'expression dans l'équation **(III.15)** Cela est équivalent à résoudre le problème des valeurs propres d'une matrice **D** :

$$\boldsymbol{D.a = \lambda\, a} \qquad \textbf{(III. 16)}$$

Dans lequel la matrice D est de dimension N*N à composantes

$$\boldsymbol{D = \frac{\sin(2\pi W(t-t'))}{\pi(t-t')}} \qquad \textbf{(III. 17);}$$

D est une matrice symétrique.

La solution a **N** valeurs propre ($\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_{N-1}$) et N vecteurs propres orthogonaux ($v_0, v_1, v_2, \ldots, v_{N-1}$) (ils sont normalisés en unité $\sum_{t=0}^{N-1}|v_k(t)|^2$ **(III. 18)**)

Les vecteurs propres $\mathbf{v_k}$ sont les fenêtres (tapers) cherchées et les $\boldsymbol{\lambda_i}$ constituent le poids de chaque fenêtre.

Comme la dimension de **D** est **N × N**, il y a N valeurs propres λ_k et vecteurs propres v_k. L'ensemble des premiers composants capture la plupart des propriétés de cette fenêtre idéale. La première valeur propre est très proche de l'unité, et est associée au premier vecteur propre (fenêtre v_1) qui aura donc la contribution la plus importante. Puisque nous utilisons la Transformée de Fourier discrète pour l'obtention de l'estimation spectrale, le nombre de points **P** dans la largeur de bande **w** est un multiple de la résolution spectrale **1/N** de la série temporelle d'échantillonnage: donc **W = P / N** et **P = NW**. Comme le montre le tableau ci-dessous, pour les choix de **N = 128** et **NW = 4, 3** ou **2**, les **2NW-1** premières valeurs propres sont très proches de l'unité (tableau **III.1) [11].** Par conséquent, comme première approche, les **2NW-1**premiers vecteurs propres normalisés pourraient être considérés comme des bonnes fenêtres à utiliser pour réduire les fuites et aussi (en faisant la moyenne de leurs résultats) pour réduire la variance du spectre d'estimation.

	$P\pi$ Prolate		
	$P=4$	$P=3$	$P=2$
λ_0	0.9999999998	0.999999885	0.999948125
λ_1	0.999999978	0.999992014	0.997764652
λ_2	0.999999008	0.999750480	0.962155175
λ_3	0.999972984	0.995477689	0.733922358
λ_4	0.999500363	0.951033908	0.287339619
λ_5	0.993525891	0.725208760	*
λ_6	0.943750573	0.307789684	*
λ_7	0.721233936	0.060764834	*
Boxcar	0.974748450	0.966410435	0.949939339

$^*\lambda_k < .05$

Tableau (III.1) : valeurs propres correspondant à 8 fenêtres propres et la fenêtre rectangulaire.

Comme montré sur les figures **(III.5)** et **(III.6)** même la meilleur fenêtre optimisée n'est pas idéale comme dans **(III.4)(B),** mais certainement mieux que celui de la rectangulaire représenté dans **(III.4)(A).**

La procédure de fenêtrage multiple est illustrée aux figures **(III.5)** et **(III.6)** [11]. Dans cet exemple, il apparait que le nombre de fenêtres utilisées constitue un compromis entre la réduction des fuites et la variance. Dans cet exemple **NW = 4**, nous considérons jusqu'à sept (2NW -1) auto vecteurs. La figure **(III.5)** représente les propriétés des fenêtres individuelles (la 8ème est également affichée), y compris leurs spectres et les effets de leur application dans les domaines temps et fréquence. La figure **(III.6)** illustre l'utilisation de ces fenêtres pour obtenir le spectre moyen. Comme on le voit, la 1ère fenêtre est la meilleure fenêtre (presque pas d'énergie dans les lobes secondaires) mais produit un spectre d'estimation unique associée à une variance assez importante. L'addition du résultat d'une 2ème fenêtre ajoute un peu de lobes secondaires (et donc un peu plus de fuite) mais en moyenne avec le 1er réduit la variance car maintenant le spectre est moyenné à partir de deux sous-spectres. En incluant le résultat de chaque fenêtre supplémentaire dans l'estimation moyenne, il y a à la fois une augmentation des fuites et une réduction de la variabilité en raison de la procédure de calcul de la moyenne.

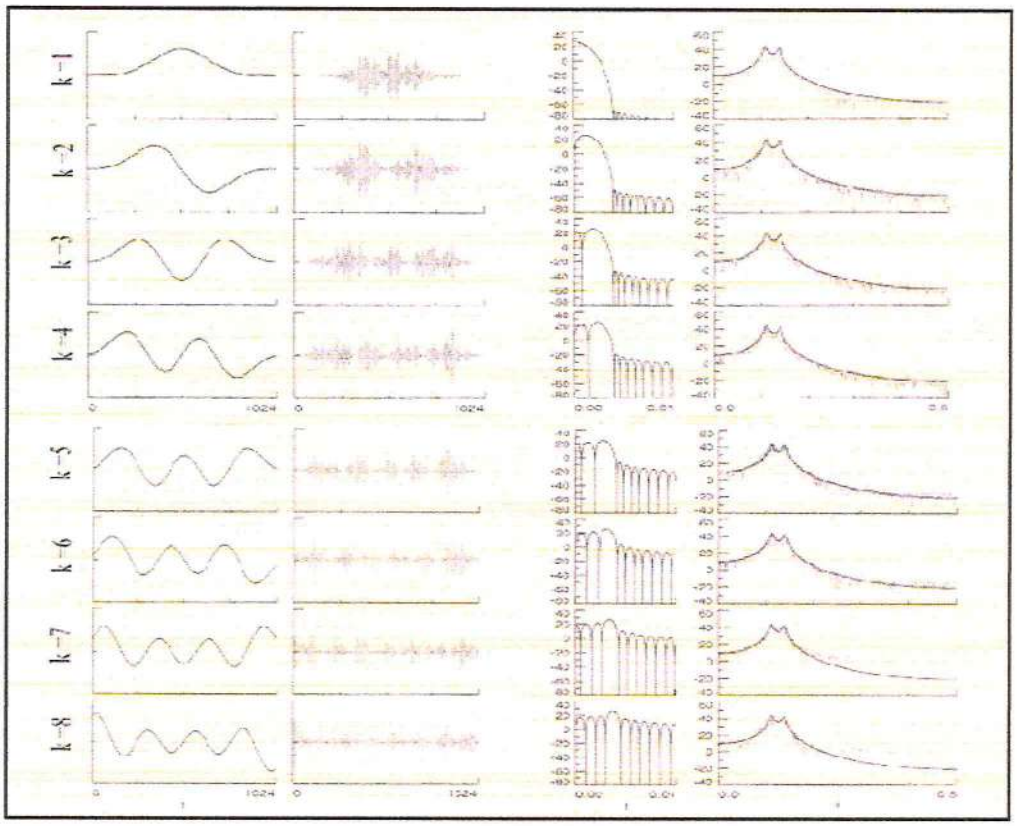


Figure (III.5) : propriétés des fenêtres individuelles (NW=4).

La colonne de gauche dans cet ensemble de graphe montre huit fenêtres sphéroïdales discrètes. Les 2NW, huit fenêtres (k = 1 - 8) sont représentés dans cette colonne. La 2ème colonne montre le signal vocal multiplié par ces fenêtres. La 3ème colonne montre les spectres des fenêtres. La dernière colonne à droite représente l'estimation du spectre du signal vocal fenêtré. La ligne noire est le spectre théorique et les lignes bruyantes grises sont les estimations des signaux fenêtrés. De toute évidence, l'énergie des lobes secondaires et les fuites associées (en particulier dans la gamme de 0,3 à 0,5) augmentent avec k.

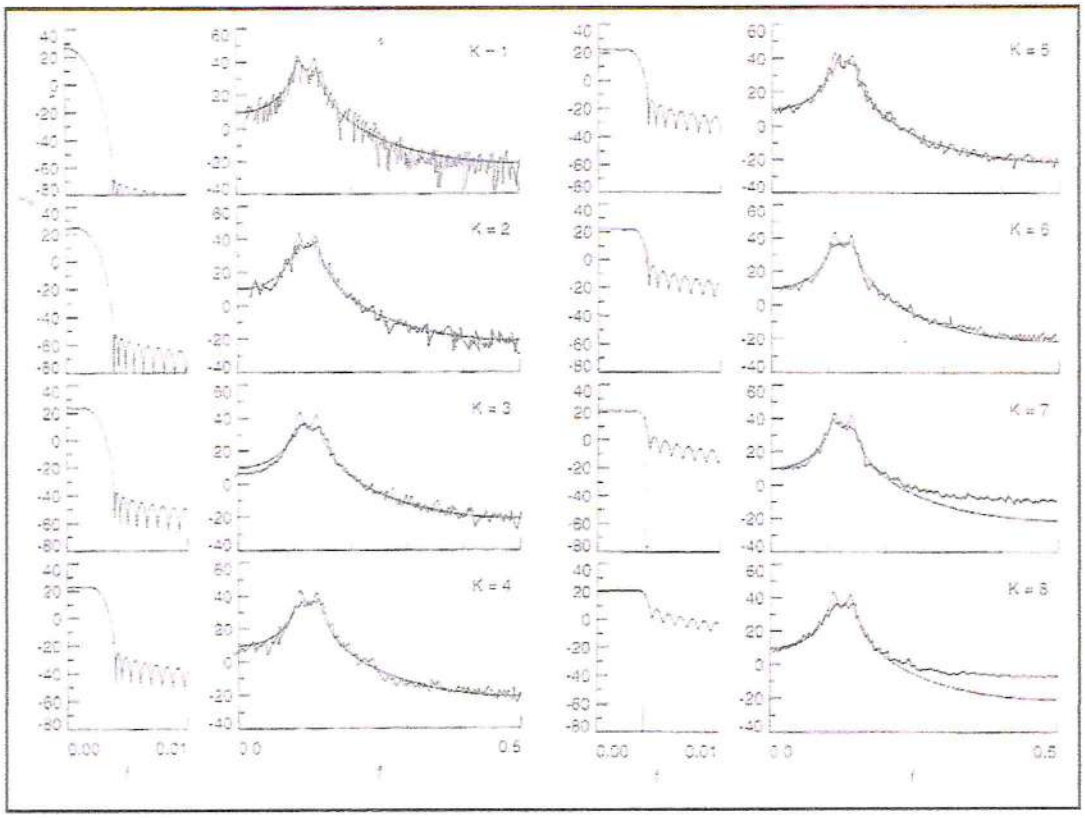


Figure (III.6) : Efficacité moyenne de l'application des fenêtres

La figure **(III.6)** illustre les estimations spectrales en utilisant les mêmes huit fenêtres représentées sur la figure **(III.5)**. Dans ce cas, la colonne de gauche et la troisième colonne montrent les spectres des fenêtres combinés. Notez que les lobes secondaires augmentent significativement avec k. La deuxième colonne et la dernière colonne montrent les estimations de fenêtrage multiples (moyennes) (lignes grises et bruyantes) superposées au spectre théorique (lignes noires). En augmentant k, le biais augmente (en raison de l'énergie dans les lobes secondaires et des fuites associées), en particulier dans la gamme de 0,3 à 0.5, mais la variance diminue.

Donc, on pourrait améliorer l'estimation en composant une moyenne pondérée des fenêtres (ce qui donnera de meilleurs résultats).

Le spectre vocal estimé dans les deux figures précédentes est présenté ci après avec plus de clarté. On observe sur cette figure les trois spectres : le périodogramme en bleu, le spectre estimé par la méthode de fenêtrage multiple en rouge et le spectre théorique du signal en noir.

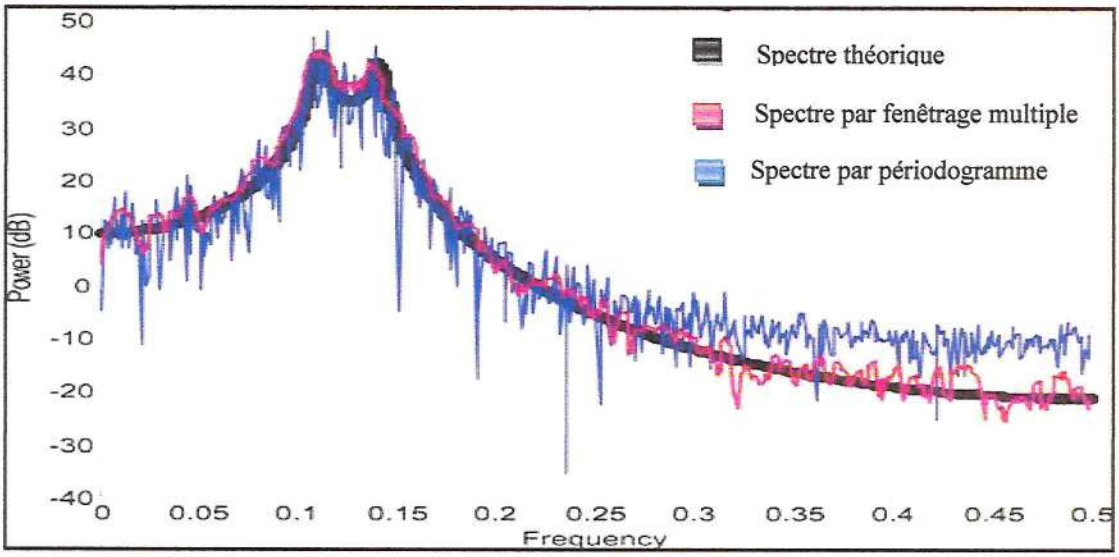


Figure (III.7): comparaison de méthode de fenêtrage multiple et rectangulaire

La figure **(III.7)** montre clairement la supériorité de la méthode de fenêtrage multiple par rapport au périodogramme. En effet le spectre en rouge (fenêtrage multiple) embrasse le mieux le spectre théorique (donc moins de biais) avec une variance moindre par rapport au périodogramme (en bleu)

5. Conclusion :

Nous avons exposé dans ce chapitre quelques notions d'estimation spectrale en relation directe avec l'objet de notre travail. L'accent a en effet été mis sur l'effet du fenêtrage sur la qualité de l'estimation du spectre, le but cherché étant d'avoir un spectre estimé avec un minimum de variance tout en espérant que ceci ait un impact positif sur la robustesse de notre système de reconnaissance ; aussi bien dans un environnement sans bruit que dans un environnement fortement bruité. Le prochain chapitre mettra donc à l'épreuve la technique de fenêtrage multiple pour illustrer sa supériorité par rapport aux techniques de fenêtrage classique.

Chapitre IV : Résultats

1. Introduction :

Ce chapitre expose en détails les travaux de reconnaissance réalisés dans ce projet. L'ordre chronologique de l'exposition est le suivant :

Dans un premier temps, la base de données utilisée pour la reconnaissance est décrite. Ensuite on décrira brièvement le logiciel HTK qui a été utilisé pour manipuler les modèles de Markov cachés et préparer la base pour les étapes ultérieures. Les travaux de reconnaissance seront ensuite présentés dans le contexte de différents niveaux de bruits et différentes méthodes de fenêtrage. Le chapitre sera clôturé par une conclusion générale.

2. Description de la base de données utilisée :

Nous utilisons une base de données de mots arabes isolés **[12]**.Le corpus contient environ 9992 énoncés de 20 mots parlés, chaque mot est répété 10 fois par 50 locuteurs natifs arabes masculins.

Le corpus a été enregistré avec une fréquence d'échantillonnage de 44100 Hz et une résolution de 16 bits.

Notons que notre base de données sera divisées en 2 ensembles ; un pour l'apprentissage: (20%) de la base et l'autre pour le test :(80%) de la base.

i. Vocabulaire : la liste des mots est la suivante :

1) Zero صفر
2) One واحد
3) Two اثنان
4) Three ثلاثة
5) Four أربعة
6) Five خمسة
7) Six ستة
8) Seven سبعة
9) Eight ثمانية
10) Nine تسعة
11) Activation تنشيط
12) Transfert تحويل
13) Balance الرصيد
14) Payement التسديد

15) Yes نعم
16) No لا
17) Funding التمويل
18) Data البيانات
19) Account الحساب
20) End انهاء

ii. Système d'étiquetage des fichiers de la base:

Chacun des fichiers a été étiqueté en utilisant le système de codage suivant :

S (numéro de locuteur) . (nombre de répétition) . (numéro du mot)

Exemple :

S01.01.01.wav représente le premier locuteur (parmi 50) prononce pour la première foi (parmi 10) le premier mot (parmi 20).

3. Description de l'outil de reconnaissance HTK [4] :

Le Hidden Markov Model Toolkit (HTK) est une boîte à outils portable pour créer et manipuler des modèles de Markov cachés. HTK est principalement utilisée pour de nombreuses autres applications, y compris la recherche en synthèse vocale, reconnaissance de caractères et séquençage de l'ADN.

HTK consiste en un ensemble de modules de bibliothèques et des outils disponibles sous langage source C. Les outils fournissent des fonctionnalités sophistiquées pour l'analyse de la parole, l'analyse de l'information, les tests et les résultats. HTK prend en charge les HMM à observations continues à densité de distribution avec mélange de gaussiennes **[13]** ainsi que les HMM discrets (distributions discrètes) et peut être utilisé pour créer des systèmes HMM complexes. HTK contient une documentation complète et des exemples. Il est à noter, que ce qui a contribué au succès de HTK, est qu'il est accompagné d'une assez bonne documentation.

On peut exploiter les fonctionnalités de HTK en tapant une simple ligne de commande. Les principaux outils que nous avons utilisés seront évoqués progressivement dans ce chapitre.

4. Préparation des fichiers de la base de données:

Cette étape est appelée extraction de paramètres (le codage des fichiers paroles), on utilise deux méthodes de fenêtrage Hamming et fenêtrage multiple :

a) Codage utilisant fenêtrage de Hamming :

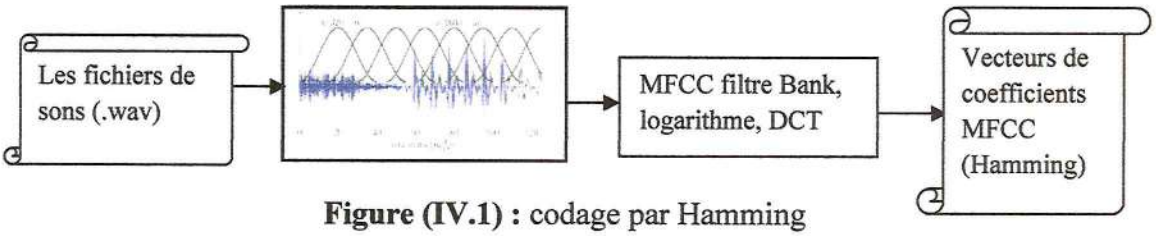


Figure (IV.1) : codage par Hamming

Consiste à multiplier chaque trame du signal parole par une fenêtre de Hamming puis à calculer le log du module de sa TF ; suivie par une DCT. Cette fonctionnalité est donnée par la fonction« HCopy » de HTK. On obtient ainsi les vecteurs de coefficients MFCC finaux

b) Codage utilisant fenêtrage de Multitaper :

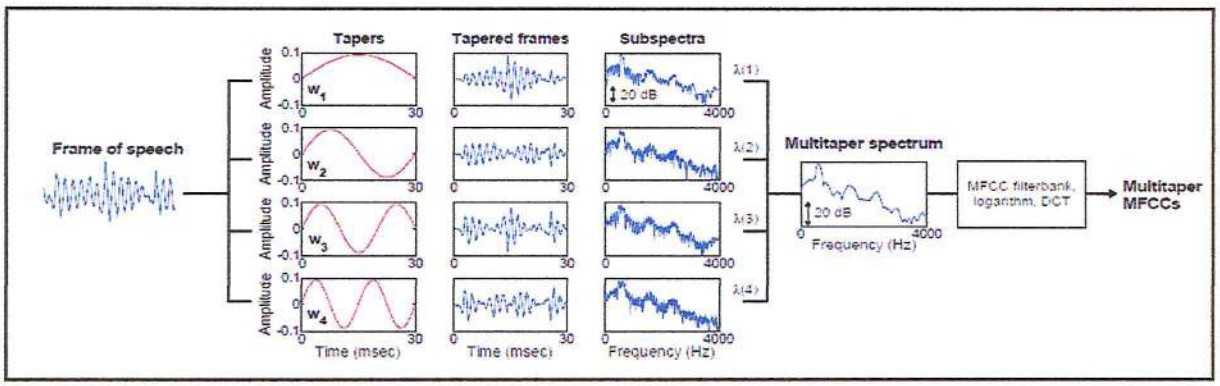


Figure (IV.2) : codage par Multitapers **[14]**

Ce type de fenêtrage multiple (Thomson) consiste à multiplier chaque trame du signal parole par plusieurs fenêtres indépendantes. On obtient ensuite un spectre pour chaque fenêtre. Le spectre estimé sera la moyenne de ces spectres. Le reste des étapes de calcul des MFCC est identique au cas précédent.

A noter que la transformation de l'onde acoustique originale en une série des vecteurs MFCC obtenus par HCopy peut être effectuée pour chaque fichier wav séparément ou, mieux encore, sur un groupe de fichiers wav dont les noms seront spécifiés dans un script. D'autres scripts seront nécessaires notamment celui spécifiant les paramètres de codage et celui spécifiant le chemin des fichiers cibles **(figure IV.3)**.

Syntaxe:HCopy –C codconf.txt –S targetlist.txt

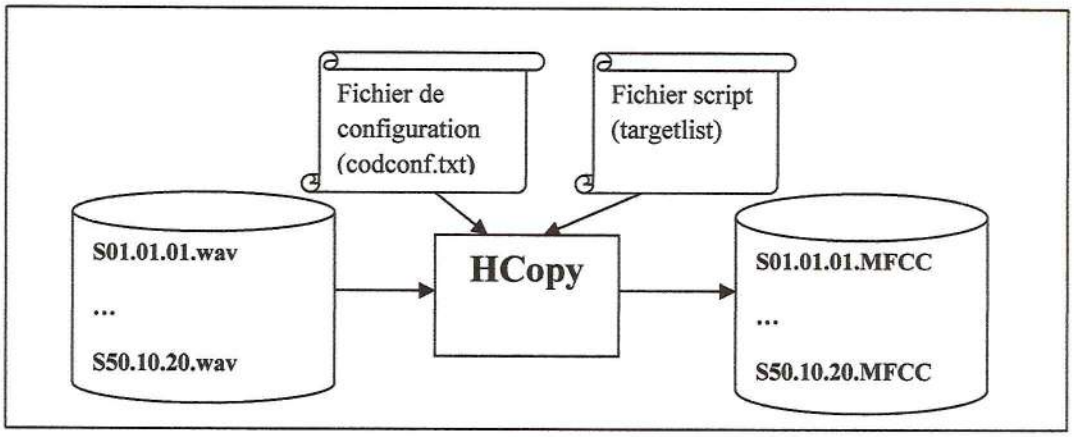


Figure (IV.3): Conversion des données d'apprentissage **[1]**

Tel que ; **codconf.txt** est un fichier de configuration définissant les paramètres de l'extraction des coefficients acoustiques.

codconf - Bloc-notes

Fichier Edition Format Affichage ?

```
SOURCEKIND = WAVEFORM
SOURCEFORMAT = NOHEAD
SOURCERATE = 899
TARGETKIND = MFCC_0_D_A     # Identifier of the coefficients to use
WINDOWSIZE = 250000.0       # = 25 ms = length of a time frame
TARGETRATE = 100000.0       # = 10 ms = frame periodicity
NUMCEPS = 12                # Number of MFCC coeffs (here from c1 to c12)
USEHAMMING = T              # Use of Hamming function for windowing frames
PREEMCOEF = 0.97            # Pre-emphasis coefficient
NUMCHANS = 26               # Number of filterbank channels
CEPLIFTER = 22              # Length of cepstral liftering
```

Figure (IV.4) : fichier de configuration

Avec un tel fichier de configuration, l'analyse MFCC (Mel Frequency Cepstre Coefficients) est effectuée (préfixe « MFCC » dans l'identifiant TARGETKIND). Pour chaque période de signal, les coefficients suivants sont extraits :

- Les 12 premiers coefficients MFCC [c1,…, c12] (depuis NUMCEPS=12).
- Les coefficients MFCC « nul » c0, indique que le cepstre 0 est ajouté (suffixe « _0 » dans TARGETKIND).
- 13 « coefficients Delta », estimant la dérivée du premier ordre de [c0, c1, …, c12] (suffixe « _D » dans TARGETKIND).

- 13 coefficients d'accélération, estimant le second ordre dérivatif de [c0, c1, ..., c12] (suffixe « _A » dans TARGETKIND).
- Le symbole # est utilisé pour les commentaires

Chaque vecteur contient donc 39 coefficients extraits de chaque période de signal.

targetlist.txt est le fichier texte spécifiant le nom et localisation de chaque fichier signal à traiter et du fichier des coefficients correspondant à générer.

S01.01.01.wav	S01.01.01.MFCC
S01.01.02.wav	S01.01.02.MFCC
.....	.....
S50.10.20.wav	S50.10.20.MFCC

Remarque : si le chemin des fichiers n'est pas spécifié dans le script, tous les fichiers (wav et MFCC) seront dans le répertoire en cours.

A noter aussi que l'opération illustrée sur la figure **(IV.3)** sera réalisée sur plusieurs versions de la base de données à savoir : la version propre (sans bruit), les versions bruitées obtenus par ajout d'un bruit blanc à différents niveaux de bruits : 20 dB, 15 dB, 10dB et 0 dB.

Nous avons maintenant préparé nos fichiers audio sous forme d'une suite de vecteurs de paramètres (fichier à une extension .MFCC).

5. Choix de la topologie des modèles (prototypes) :

On doit définir 1MMC pour chaque mot à reconnaitre. Chaque MMC est caractérisé par :

- un nombre d'états égal à 8 dont 6 états émetteurs. Les transitions possibles sont montrées **(voir figure IV.5)**. les probabilités de transition possibles sont données par la matrice de probabilité de transition.
- Chaque état émetteur est caractérisé par une loi d'émission continue à une seule gaussienne caractérisée par un vecteur moyenne μ (39 composantes) et une matrice de covariance $\sum$ diagonale.

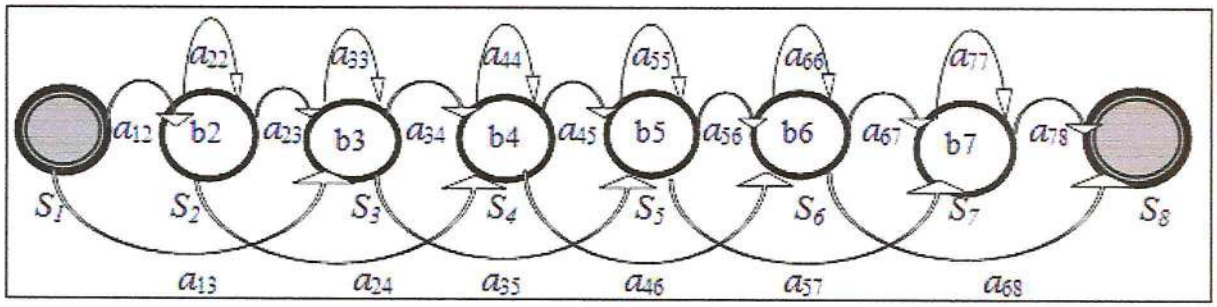


Figure (IV.5) : topologie choisie pour les MMCs

Remarque: généralement la topologie la plus communément adoptée est celle de **BAKIS.** Il n'y a pas une loi fixe pour le choix de la topologie.

Le fichier de description du HMM est un simple fichier texte comme le montre la figure **(IV.6).**

```
~o <VecSize> 39 <MFCC_0_D_A>
~h "zero"
<BeginHMM>
<NumStates> 6
<State> 2
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 3
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 4
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 5
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<TransP> 6
0.0 0.5 0.5 0.0 0.0 0.0
0.0 0.4 0.3 0.3 0.0 0.0
0.0 0.0 0.4 0.3 0.3 0.0
0.0 0.0 0.0 0.4 0.3 0.3
0.0 0.0 0.0 0.0 0.5 0.5
0.0 0.0 0.0 0.0 0.0 0.0
<EndHMM>
```

Figure (IV.6) : exemple de fichier de description de HMM

- ***~o <VecSize> 39 <MFCC_0_D_A>***: l'en-tête du fichier, contient la taille du vecteur de coefficient (39 vecteurs) et le type de ces coefficients (MFCC_0_D_A).
- ***~h "zero" <BeginHMM> (…) <EndHMM> :***

Renferme la description d'un HMM appelé « zero ».

- ***<NumStates> 8 :*** donne le nombre total d'états dans le HMM, y compris les 2 états non émetteurs (états 1 et 8).
- ***<State> 2 :*** introduit les paramètres de loi d'émission d'observation de l'état2. Ici, nous avons choisis d'utiliser une fonction d'observation gaussienne unique, avec des matrices d'auto-corrélation diagonales. Chaque fonction est décrite par un vecteur de moyenne et un vecteur de variance (les éléments diagonaux de la matrice d'auto-corrélation). Les états 1 et 8 ne sont pas décrits car ils n'ont pas des fonctions d'observations.
- ***<Mean> 39***

```
0.0 0.0 (...) 0.0 (x39)
```

Représente le vecteur de moyenne de l'observation courante, chaque élément est arbitrairement initialisé à 0 : le fichier ne donne que le ″ prototype ″ du MMC (sa topologie globale). Les valeurs de ces coefficients seront obtenues plus tard après l'apprentissage.

- ***<Variance> 39***

```
1.0 1.0(…) (x 39)
```

Représente le vecteur de variance de la fonction d'observation courante. Chaque élément est arbitrairement initialisé à 1, les vraies valeurs seront obtenues après l'apprentissage.

- ***<TransP> 8 :***

Représente la matrice de transition 8*8 du HMM, où a_{ij} est la probabilité de transition de l'état i à l'état j. les valeurs nulles indiquent que les transitions correspondantes ne sont pas autorisées. Les autres valeurs sont arbitrairement

initialisées (la somme sur chaque ligne de la matrice doit être égale à 1). Les valeurs seront modifiées plus tard au cours du processus d'apprentissage.

Chaque mot est généré par un prototype. Dans notre cas on doit écrire un prototype pour chacun des 20 MMCs (20 mots).

6. Initialisation des modèles :

L'initialisation des modèles se fait par l'outil htk "**HCompv**". Chaque état du HMM reçoit le même vecteur de variance et de moyenne : ceux-ci sont calculés globalement sur le corpus entier d'entrainement.

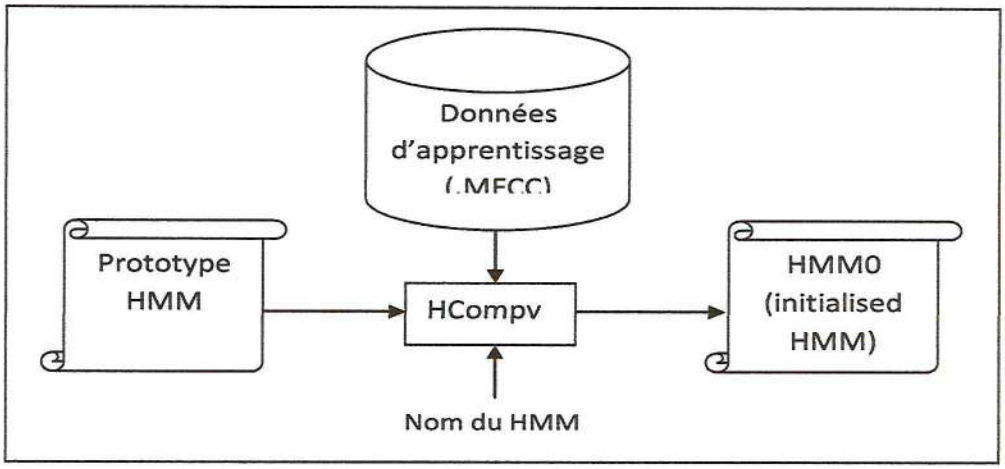


Figure (IV.7) : fonctionnement de HCompv

HCompv doit être invoqué comme suit :

HCompv –S trainlist.txt –M model/hmm0 –H model/proto/hmmfile nameofHMM

Tel que:

Trainlist.txt : représente la liste complète des fichiers MFCC formant le corpus d'apprentissage pour chacun des 20 mots de la base de données.

Model/hmm0 : est le nom du dossier (il doit être créé) où les descriptions de HMM résultants (après initialisation) seront placées.

Hmmfile : est un fichier de description contenant les prototypes HMMs (des 20 mots).

nameofHMM : est le nom du HMM à initialisé (zero, one, …, end).

Ce fichier de sortie HMM par **HCompv** a le même nom que le prototype d'entrée.

Cette procédure doit être répétée pour chacun des 20 modèles.

7. Apprentissage :

La ré-estimation des modèles est faite par l'outil htk HRest utilisant l'algorithme de **Baum-Welch** qui permet de calculer les valeurs optimales pour les paramètres HMM (probabilité de transition, plus les vecteurs de moyenne et de variance pour chacune des fonctions d'observation).

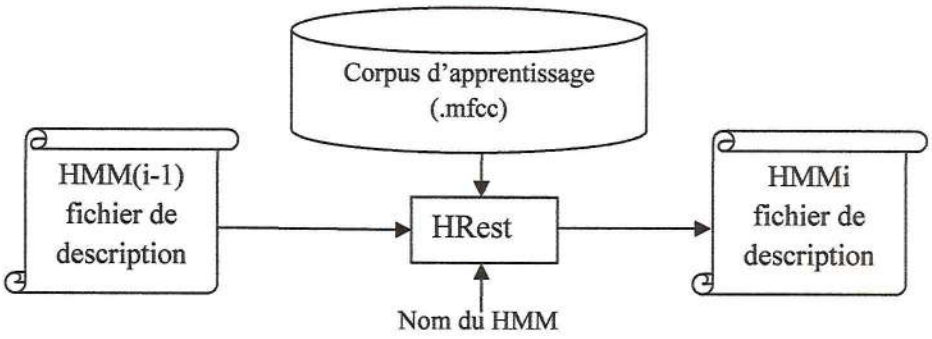


Figure (IV.8) : fonctionnement de HRest

La ligne de commande suivante effectuer une itération de ré-estimation avec **HRest** :

HRest –S trainlist.txt –M model/hmmi –H model/hmmi-1/hmmfile nameofhmm.

Tel que:

Model/hmmi : représente le répertoire de sortie, indique l'indice de l'itération courante i.

Hmmfile : est le fichier de description de HMM appelé nameofHMM. Il est stocké dans un répertoire dont le nom indique l'indice de la dernière itération (ici model/hmmi-1 par exemple).

Nameofhmm : représente le nom du HMM à apprendre (entrainer).

- 4 itérations de ré-estimation sont suffisantes pour obtenir les modèles finaux.

Finalement, on dispose de 20 modèles entrainés et prêts à être utilisés pour la reconnaissance.

8. Définition du grammaire et dictionnaire:

Avant d'utiliser nos modèles de mots, on doit définir l'architecture de base de notre reconnaissance (la grammaire). En **htk**, la grammaire est décrite dans un fichier texte, selon certaines règles syntaxiques.

Dans notre cas, la grammaire est assez simple :

```
$Word = zero | one | two | three | four | five | six | seven | eight | nine |
activation | transfert |balance | payement | yes | no | funding | data | account |
end ;
```

Figure (IV.9): le fichier de la grammaire

La variable **word** peut être remplacée par l'un des 20 mots du corpus (zero **ou** one **ou** two **ou** three **ou** four **ou** five **ou** .... **Ou** activation **ou** payement **ou** funding ...). Le système doit bien sûre savoir à quel HMM correspond chacune des variables de grammaire zero, one, ..., end. Cette information est stockée dans un fichier texte appelé **dictionnaire**. Dans une telle tâche simple, la correspondance est simple, et la tâche dictionnaire renferme tout simplement les 20 entrées

```
Zero          zero
One           one
Two           two
Three         three
Four          four
Five          five
Six           six
Seven         seven
Eight         eight
Nine          nine
activation    activation
transfert     transfert
balance       balance
payement       payement
yes           yes
no            no
funding       funding
data          data
account       account
end           end
```

Figure (IV.10) : le fichier du dictionnaire

Les éléments gauches se réfèrent aux noms des variables de grammaire. Les éléments droits réfèrent aux noms des MMCs (introduit par ~h dans les fichiers de description HMM).

Pour compléter la définition des tâches on a besoin de créer le réseau de connexion reliant les 20 mots **(voir figure IV.11)** par la compilation du fichier grammaire avec l'outil HParse :

Syntaxe :

HParse gram.txt wdnet.slf

Tel que :

gram.txt : est le fichier grammaire.

Wdnet.slf : est le fichier obtenu qui représente le réseau.

Maintenant, notre tâche de reconnaissance est complètement définie par son réseau, son dictionnaire et son ensemble de HMM (les 20 HMMs de la dernière itération).

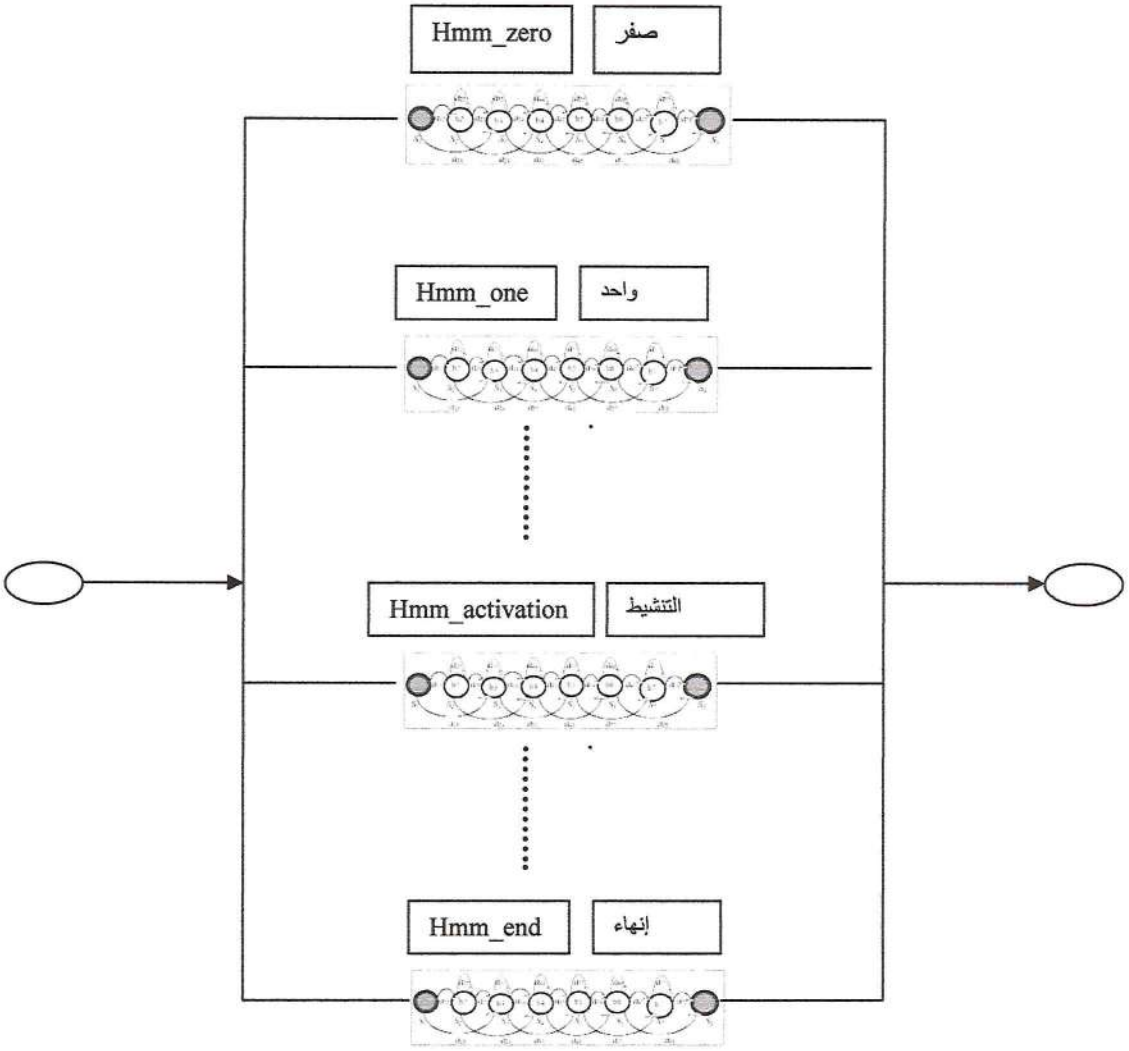


Figure (IV.11) : le réseau de connexion entre les mots

9. Reconnaissance :

Dans cette étape nous allons tester les performances de notre système à reconnaitre les mots inconnus à l'entrée du système par l'algorithme de **Viterbi**, cette fonction est invoquée en HTK comme suit :

HVite -Hhmmdef.mmf -i reco.mlf -w wdnet.slf dict.txt hmmlist.txt -S testlist.txt

Tel que :

Hmmdef.mmf : fichier texte contient la définition des MMCs. Il est obtenu en copiant chaque définition des MMCs (de la dernière itération de ré-estimation) dans un fichier unique, en évitant la répétition de l'information de l'entête.

Reco.mlf : c'est le fichier de sortie.

Testlist.txt : le fichier script qui contient la liste des fichiers (MFCC) de teste.

Hmmlist.txt : fichier texte contient les noms des 20 modèles à utiliser les éléments sont séparés par une nouvelle ligne.

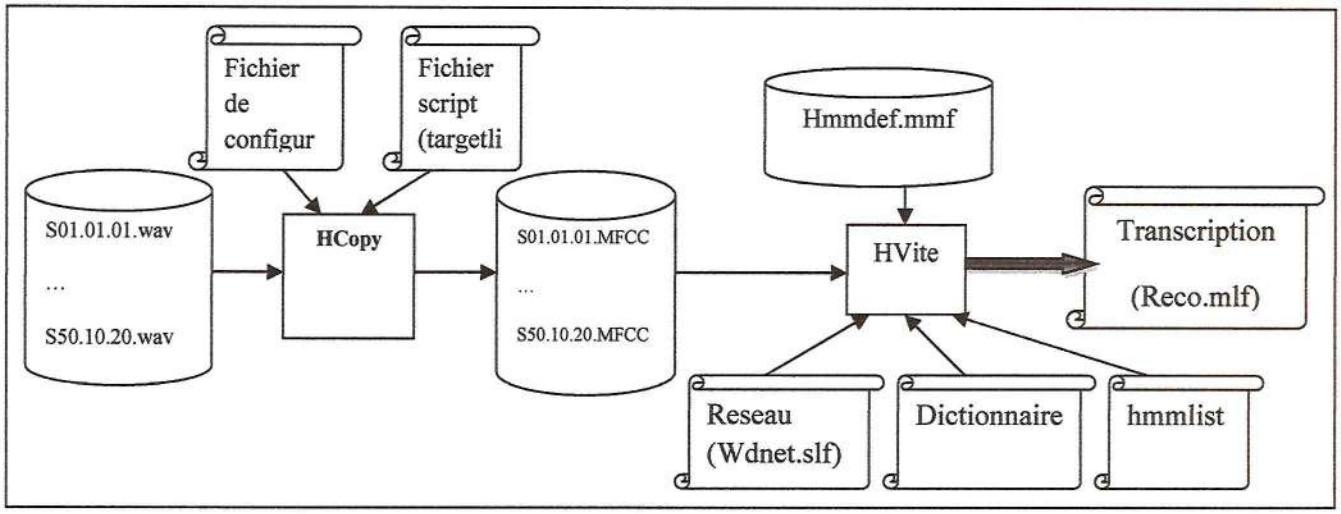


Figure (IV.12) : processus du codage et reconnaissance

10. Tests de robustesse et interprétations :

Nous avons effectué une série d'expérimentation s'inscrivant dans la chronologie suivante :

- Reconnaissance audio en absence de bruit (propre) utilisant le fenêtrage de Hamming
- Reconnaissance audio en présence de bruit (bruité) utilisant le fenêtrage de Hamming
- Reconnaissance audio en absence de bruit (propre) utilisant le fenêtrage multiple (multitaper).
- Reconnaissance audio en présence de bruit (bruité) utilisant le fenêtrage multiple (multitaper).
- Comparaison entre les deux méthodes de fenêtrage.
- Conclusion.

1. Taux de reconnaissance avec fenêtrage de Hamming :

RSB (dB) signal/bruit	propre (clean)	20	15	10	5	0
taux de reconnaissance (%)	86.8526	41.1058	18.5014	9.307	5.967	5.7168

Tableau (IV.1) : résultat obtenus par méthode de Hamming (clean et bruité).

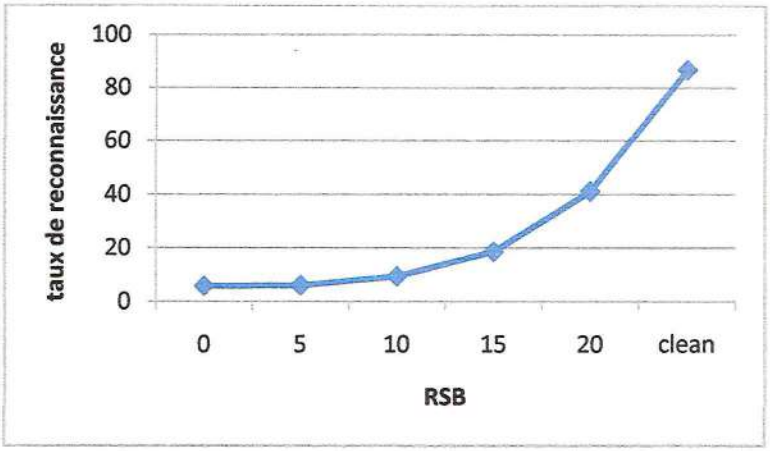


Figure (IV.13) : taux de reconnaissance (Hamming)

Interprétation :

La performance de reconnaissance est proportionnelle au rapport signal sur bruit RSB, l'optimum étant atteint pour un niveau RSB très élevé (équivalent à un environnement sans bruit). Ceci est en concordance avec la réalité des choses puisque en présence de bruit, le système n'arrive plus à discerner entre le signal utile et le bruit.

2. Taux de reconnaissance avec fenêtrage multiple :

RSB (dB)	propre	20	15	10	5	0
taux de reconnaissance (%)	86.8777	48.7991	26.5449	16.2997	10.5825	6.63

Tableau (4.2) : résultat obtenus par méthode de fenêtrage multiple (propre et bruité).

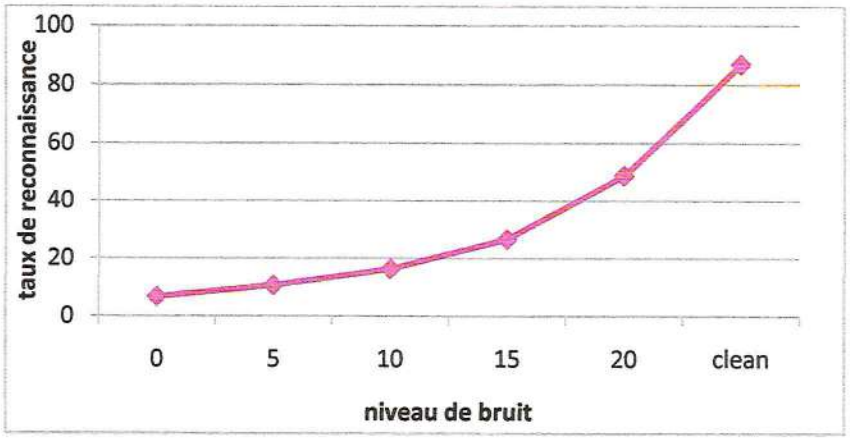


Figure (IV.14) : taux de reconnaissance par multi-taper

Interprétations :

- La performance de reconnaissance est proportionnelle au rapport signal sur bruit comme dans le cas de fenêtrage hamming, ce qui est tout à fait logique.
- Toutefois, on remarque clairement que les performances de la méthode de fenêtrage multiple dépassent celles du fenêtrage Hamming et ce quelque soit le niveau de bruit envisagé. Ceci est d'autant plus vrai plus les niveaux de bruit élevés (RSB faible). Ceci est du à la fiabilité de l'estimation spectrale obtenue par la méthode de fenêtrage multiple ; fiabilité qui se manifeste surtout par une faible variance du spectre estimé contrairement à la méthode de fenêtrage Hamming. Le tableau **(IV.3)** et la figure **(IV.15)** illustrent la comparaison entre les deux méthodes envisagées.
- Une observation plus minutieuse des résultats de reconnaissance montre que le taux de reconnaissance diminue de manière dramatique pour les niveaux élevés de bruit. L'explication à cela est que les modèles MMC obtenus après apprentissage en absence de bruit ont été aussi utilisés en milieu bruité, l'objectif étant d'observer le **seul** effet de l'estimation spectrale sur le taux de reconnaissance.

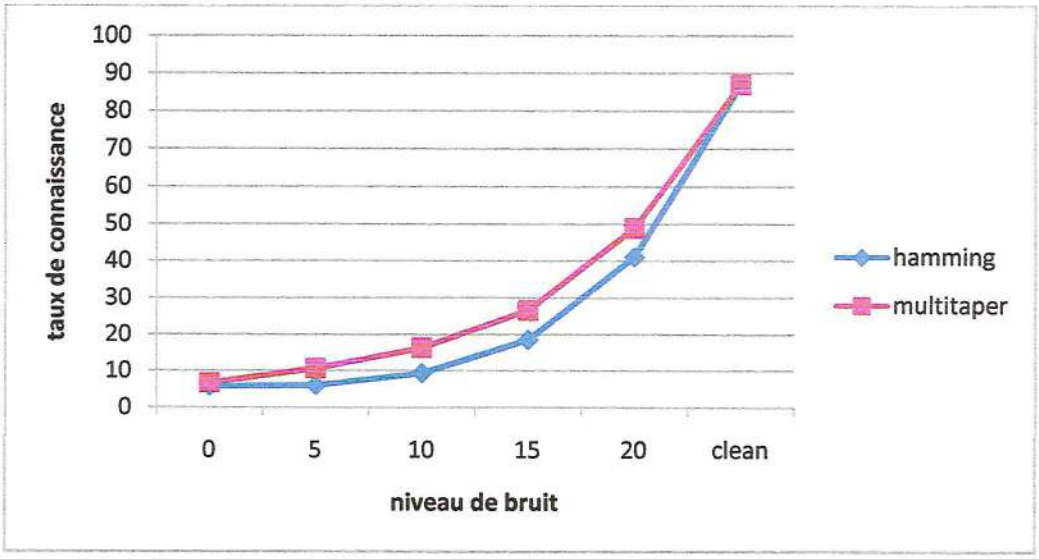


Figure (IV.15) : comparaison de taux de reconnaissance des deux méthodes

11. Conclusion :

Ont été exposés dans ce chapitre les résultats de reconnaissance automatique de la parole sur une base de données de mots arabes isolés. L'objectif était de mettre la lumière sur l'importance de l'estimation spectrale et son impact sur les performances de reconnaissance. Les résultats montrent la supériorité de la méthode de fenêtrage multiple par rapport à la méthode de fenêtrage classique surtout lorsqu'il s'agit d'un environnement bruité.

CONCLUSION GENERALE

Arrivé à terme de ce travail, il convient de consacrer ces quelques lignes ultimes pour souligner un certain nombre de points importants. Tout d'abord, le choix de ce thème comme projet de fin d'études nous a permis d'explorer un domaine fascinant, tant pour l'intérêt que lui portent les chercheurs actuellement que par l'étendue du champ de disciplines et de connaissances nécessaires à l'appréhension de ce domaine. Aussi, il nous a été possible, durant la réalisation de ce travail, de voir la concrétisation et l'illustration de tant de notions qui jusque là était pour nous de pures notions théoriques.

L'idée initiatrice de ce travail trouve son origine dans une nouvelle et élégante manière d'estimer le spectre du signal vocal à travers la méthode dite de fenêtrage multiple. On s'est donc proposé d'explorer l'apport d'une telle méthode pour la robustesse d'un système de reconnaissance. Apport confirmé et mis en évidence à travers les résultats exposés dans le chapitre précédent. Il va sans dire que la qualité de l'estimation spectrale ne suffit pas à elle seule pour concevoir un système de reconnaissance robuste, l'intervention sur les autres étages du système de reconnaissance étant nécessaire. On parle, entre autres techniques, d'amélioration et de débruitage (speech enhancement, denoising) ou encore de techniques d'adaptation de modèles. Ce qui pourra faire objet de futurs travaux de projets de fin d'études.

Références

[1] Antoine Chemardin, "un système de reconnaissance automatique de la parole" (la traduction française de HTK (v.3.1) Basic Tutoriel de Nicolas Meauro /02.02.2002), 30 juin 2006.

[2] Jean-Paul Haton, LORIA-INRIA, "la reconnaissance automatique de la parole", Université Henry Poincaré, Nancy 1 France, 2005.

[3] Sean A. Fulop, "Speech Spectrum analysis", Springer-Verlag Berlin Heidelberg, 2011

[4] Mark Gales, Steve Young, "The Application of Hidden Markov Models in Speech Recognition", now publishers Inc, 2008.

[5] Thierry Dutoit, "introduction au traitement automatique de la parole" (note de cours), Faculté Polytechnique de Mons, 2000.

[6] Steve Young, Dan Kershaw, Julian Odell, Dave Ollason, Valtcho Valtchev, Phil Woodland, "The HTK book", Microsoft Corporation, juillet 2000.

[7] Sakriani Sakti, Konstantin Markov, Satoshi Nakamura et Wolfgang Minker, "Incorporating Knowledge Sources into Statistical Speech Recognition", Springer Science Business Media, LLC, 2009

[8] Ferarsa Amel, "Reconnaissance audiovisuelle de la parole par les Modèles de Markov Cachés" (thèse), université Dr Yahia Fares Medea, 2012.

[9] Lawrence R.Rabiner et Ronald W.Schafer, "introduction to digital speech processing", now publishers inc, 2007.

[10] K.M.M PRABHU, "Window Functions and their applications in signal processing" , Taylor and Francis Group, LLC, 2014.

[11] Wim van Drongelen, "Multitaper Power Spectrum Estimation", 2014

[12] Abdulrahman Alalshekmubarak and Leslie S. Smith On Improving the Classification Capability of Reservoir Computing for Arabic Speech Recognition in Wermter, S., Weber, C., Duch, W., Honkela, T., Koprinkova-Hristova, P., Magg, S., Palm, G., Villa, A.E.P. (Eds.) , Artificial Neural Networks and Machine Learning-ICANN 2014, Lecture Notes in Computer Science 8681, Springer Heidelberg, 2014, pages 225-232.

[13] Barbra Resch, "Mixtures of Gaussians", Signal Processing and Speech Communications laboratory.

[14] Tomi Kinnunen, Rahim Saeidi, Johan Saudberg, Maria Hayzhou Li; "low variance Multitaper MFCC Features: a case study in Robust speaker verification, transaction on speech, audio and language processing".

Annexe

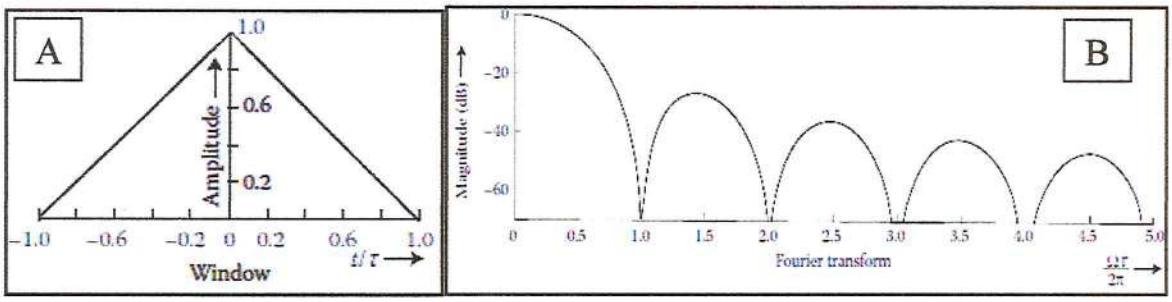
Quelques fonctions de fenêtrage:

1) Fenêtre triangulaire (Bartlett) :

Cette fonction est la dérivée par la convolution linéaire de deux rectangulaire d'une demi-durée (τ/2). Elle est définit comme suit :

$$f(t) = \begin{cases} 1 - \frac{|t|}{\tau} & , \ |t| \leq \tau \\ 0 & , ailleurs \end{cases} \quad (1) \qquad \xrightarrow{\textbf{TF}} \qquad F(j\omega) = \tau \left[\frac{\sin\left(\frac{\omega\tau}{2}\right)}{\left(\frac{\omega\tau}{2}\right)} \right]^2 , \quad (2)$$

$$-\infty < \tau < \infty$$



Figure(1) : (A) **fenetre triangulaire,** (B) **son spectre d'amplitude.**

2) Fenêtre tronquée de Taylor :

Les fonctions de Taylor sont obtenues par l'addition d'une série de cosinus pondérées à un constant appelé piédestal. Elle est définit par :

$$f(t) = \begin{cases} \frac{(1+k)}{2} + \frac{(1-k)}{2} \cos\left(\frac{\pi t}{\tau}\right) & , \ |t| \leq \tau \\ 0 & , \ ailleurs \end{cases} \quad (3)$$

$$\xrightarrow{\textbf{CTFT}} \quad F(j\omega) = \frac{[\pi^2(1+k) - 2k\omega^2\tau^2]\sin(\omega\tau)}{\omega(\pi^2 - \omega^2\tau^2)} \quad (4)$$

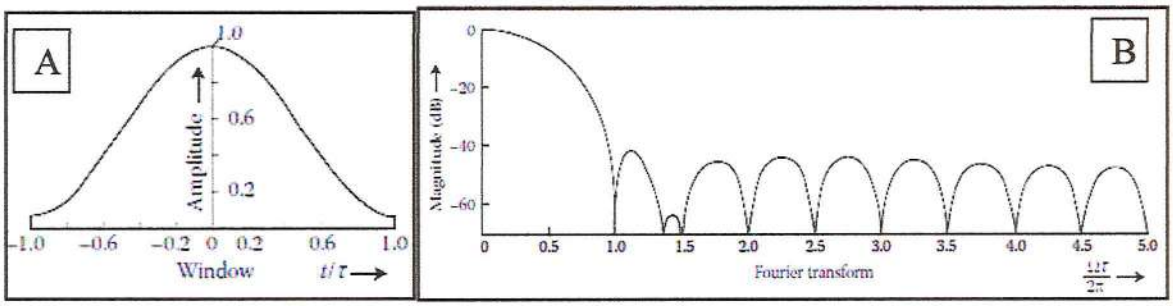


Figure (2): (A) : fenêtre de Taylor (k=0.072), **(B)** : la TF de la fenêtre de Taylor

3) Fenètre cos(x) :

Elle définit comme suit :

$$f(t) = \begin{cases} \cos\left(\frac{\pi t}{2\tau}\right), & |t| \leq \tau \\ 0 & , \ ailleurs \end{cases} \quad (5) \qquad \xrightarrow{CTFT} \quad F(j\Omega) = \frac{\sin\left(\left(\Omega+\frac{\pi}{2\tau}\right)\tau\right)}{\left(\Omega+\frac{\pi}{2\tau}\right)} + \frac{\sin\left(\left(\Omega-\frac{\pi}{2\tau}\right)\tau\right)}{\left(\Omega-\frac{\pi}{2\tau}\right)} \quad (6)$$

$$-\infty < \tau < \infty$$

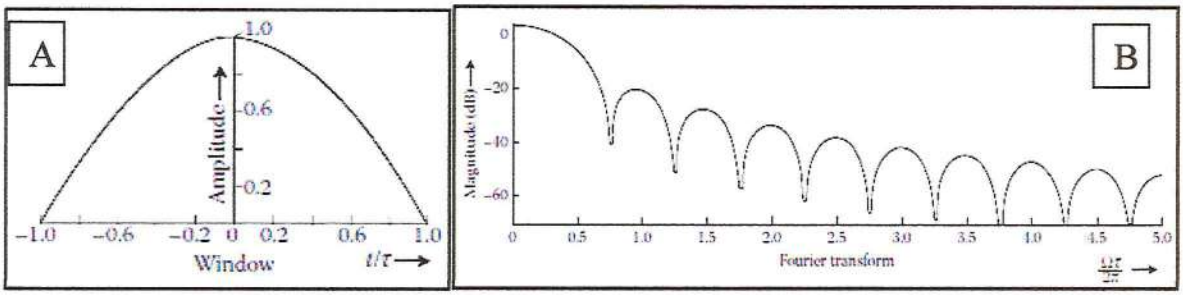


Figure (3) : (A) : fenêtre cos(x), **(B)** : la transformé de Fourier de la fenêtre cos(x)

4) Fenêtre de BlackMan optimisée :

Elle est exprimée comme suit :

$$f(t) = \begin{cases} A + 2B\cos\left(\frac{\pi t}{\tau}\right) + 2C\cos\left(\frac{2\pi t}{\tau}\right) & , \ |t| \leq \tau \\ 0 & , ailleurs \end{cases} \quad (7)$$

Pour rendre la fonction unitaire au point d'origine en gardant la valeur B=0.25

$$f(t) = (0.5 - 2C) + 0.5\cos\left(\frac{\pi t}{\tau}\right) + 2C\cos\left(\frac{2\pi t}{\tau}\right) \qquad (|t| \leq \tau) \qquad (8)$$

L'optimisation du constant C est faite de sorte que le peak du lobe principal $\left|\frac{F(0)}{F(j\omega_1)}\right| = Maximum.$

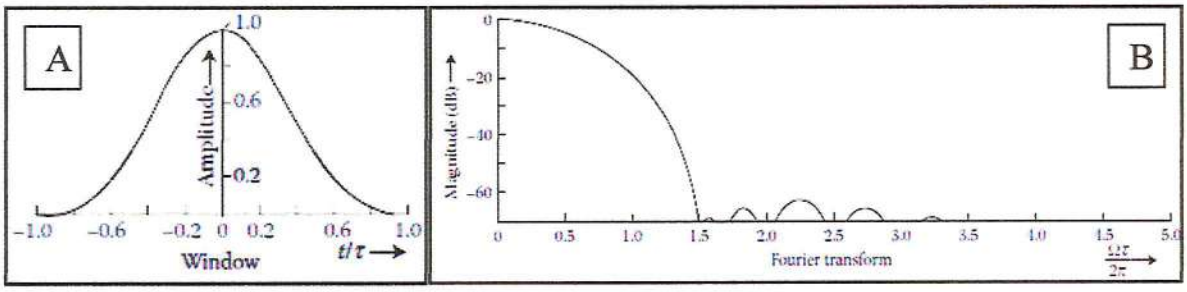


Figure (4) : (A): blackMan optimisé, **(B) : la TF de la fenêtre BlackMan optimisée**

5) Fenêtre parabolique :

Elle définit comme suit :

$$f(t) = \begin{cases} 1 - \left(\frac{|t|}{\tau}\right)^2 & ,|t| \leq \tau \\ 0 & ,ailleurs \end{cases} \quad (9) \quad \textbf{CTFT} \xrightarrow{\quad} F(j\omega) = \frac{4\tau}{(\omega\tau)^2}\left(\frac{\sin(\omega\tau)}{\omega\tau}\right) - \cos(\omega\tau) \quad (10)$$

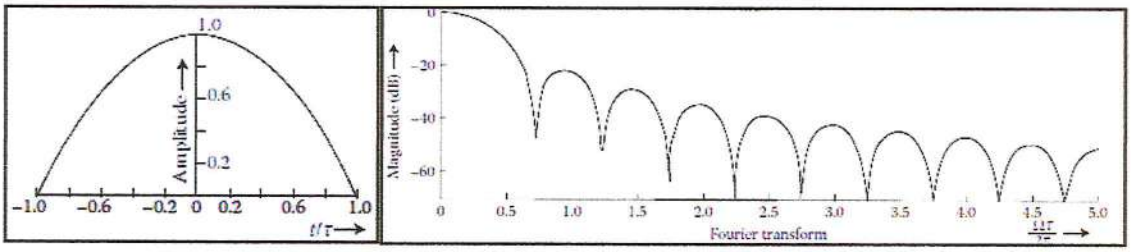


Figure (5): (A) fenêtre parabolique, **(B)** : la TF de la fenêtre parabolique

6) Fenêtre de BlackMan :

Définit par la fonction :

$$f(t) = \begin{cases} 0.42 + 0.5\cos\left(\frac{\pi t}{\tau}\right) + 0.08\cos\left(\frac{2\pi t}{\tau}\right), & |t| \leq \tau \\ 0 & ,ailleurs \end{cases} \quad (11)$$

TF ⟶

$$F(j\omega) = 0.84\left[\frac{\sin(\omega\tau)}{\omega}\right] + 0.5\left[\frac{\sin\left(\left(\omega+\frac{\pi}{\tau}\right)\tau\right)}{\left(\omega+\frac{\pi}{\tau}\right)} + \frac{\sin\left(\left(\omega-\frac{\pi}{\tau}\right)\tau\right)}{\left(\omega-\frac{\pi}{\tau}\right)}\right] + 0.08\left[\frac{\sin\left(\left(\omega+\frac{2\pi}{\tau}\right)\tau\right)}{\left(\omega+\frac{2\pi}{\tau}\right)} + \frac{\sin\left(\left(\omega-\frac{2\pi}{\tau}\right)\tau\right)}{\left(\omega-\frac{2\pi}{\tau}\right)}\right]$$

$$-\infty < \Omega < \infty \quad (12)$$

La fonction F(jω) peut s'écrire aussi comme suit :

$$F(j\omega) = \frac{(3.36\pi^4 - 0.36\pi^2\omega^2\tau^2)\sin(\omega\tau)}{\omega(\omega^2\tau^4 - 5\pi^2\omega^2\tau^2 + 4\tau\pi^4)} \quad (13)$$

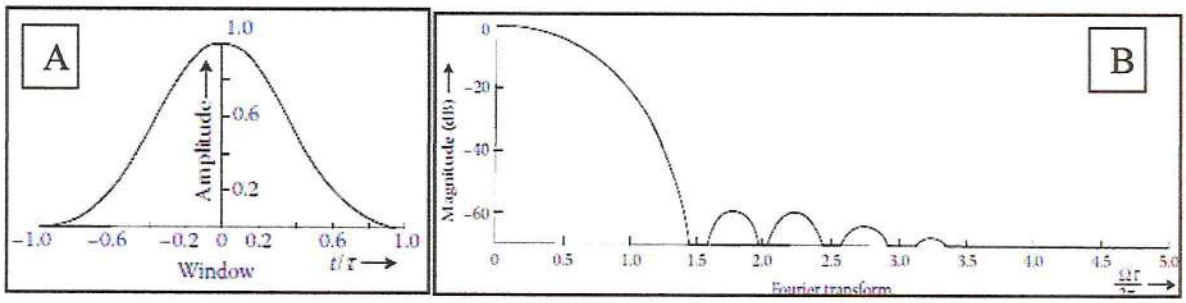


Figure (6) : (A) fenêtre de BlackMan, **(B)** son spectre d'amplitude

7) Fenêtre de Tukey :

Connue aussi comme la fenêtre cosinus fenêtré (cosine-tapered) et peut être exprimée comme un lobe de cosinus convolué avec une fenêtre rectangulaire.

Elle est définit comme suit :

$$f(t) = \begin{cases} 1 & , \ |t| \leq \beta\tau \\ 0.5 + 0.5\cos\frac{\pi(|t|-\beta\tau)}{(1-\beta)\tau} & , \beta\tau \leq |t| \leq \tau \\ 0 & , ailleurs \end{cases} \qquad (14)$$

TF ⟶

$$F(j\omega) = \frac{\sin[\omega(1+\beta)/2]\cos[\omega(1-\beta)/2]}{\omega[1-(1-\beta)^2(\omega/\pi)^2]} \qquad (15)$$

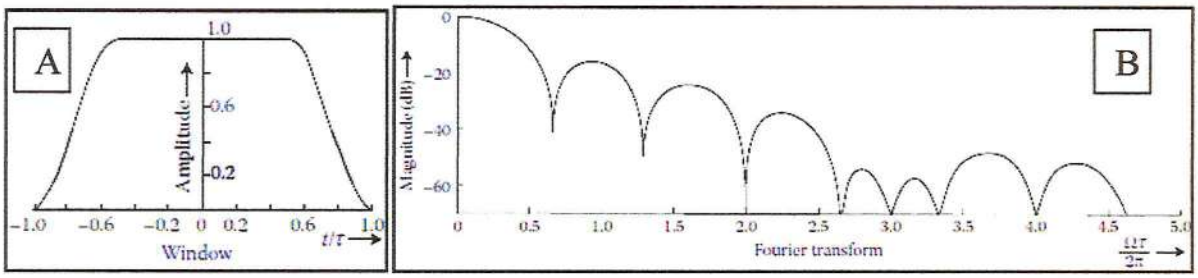


Figure (7) : (A) fenêtre de Tukey, **(B)** son spectre d'amplitude