

**ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC**

**MÉMOIRE PRÉSENTÉ À
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE**

**COMME EXIGENCE PARTIELLE
À L'OBTENTION DE LA**

**MAÎTRISE EN TECHNOLOGIE DES SYSTÈMES
M.Ing.**

**PAR
NABIL BADRI**

**UTILISATION DE LA TRANSFORMÉE DE FOURIER ET DE LA TRANSFORMÉE
EN ONDELETTES POUR LA RECONNAISSANCE DU LOCUTEUR**

MONTRÉAL, LE 27 NOVEMBRE 2002

**CE MÉMOIRE A ÉTÉ ÉVALUÉ
PAR UN JURY COMPOSÉ DE :**

**M. Chakib Tadj, professeur et directeur de mémoire
Département de génie électrique à l'École de technologie supérieure**

**M. Christian Gargour, professeur et codirecteur
Département de génie électrique à l'École de technologie supérieure**

**Mme. Rita Noumeir, professeure et présidente du jury
Département de génie électrique à l'École de technologie supérieure**

**M. Mohamed Cheriet, professeur
Département de génie de production automatisée à l'École de technologie
supérieure**

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 31 OCTOBRE 2002

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

**UTILISATION DE LA TRANFORMÉE DE FOURIER
ET DE LA TRANFORMÉE EN ONDELETTES
POUR LA RECONNAISSANCE DU LOCUTEUR**

Nabil Badri

SOMMAIRE

Ce travail porte sur l'utilisation de la transformée en ondelettes pour la vérification du locuteur en mode de texte dépendant. Nous avons étudié deux approches dans le cadre de ce sujet : (a) La première est basée sur l'utilisation de la méthode MFDWC (Mel Frequency Discrete Wavelet Coefficients) faisant usage des filtres de Daubechies d'ordres fixe, et d'ordres adaptés au sens de certains critères d'entropies. (b) La deuxième approche est basée sur l'utilisation de l'algorithme de sélection de la meilleure base d'ondelettes au sens du critère d'entropie de Shannon non normalisé. Nous avons proposé dans le cadre de cette approche un algorithme de construction du meilleur arbre du locuteur. Nous avons nommé cet algorithme MAL (Meilleur arbre du locuteur). Cette méthode permet de construire une librairie d'arbres admissibles à partir des phrases d'entraînement. Chaque arbre admissible correspond à un locuteur. Nous avons testé les performances de reconnaissance en effectuant des expériences sur (60) locuteurs. Ces derniers sont extraits de la base de données de Yoho. Finalement, les résultats des deux méthodes ont été comparés avec ceux de la méthode MFCC (Mel Frequency Cepstral Coefficient).

**UTILISATION DE LA TRANFORMÉE DE FOURIER
ET DE LA TRANFORMÉE EN ONDELETTES
POUR LA RECONNAISSANCE DU LOCUTEUR**

Nabil Badri

ABSTRACT

In this work, we study the use of the Wavelet Transform in text-dependent speaker verification mode. An experimental framework has been conducted using two approaches: (a) The use of MFDWC (Mel Frequency Discrete Wavelet Coefficients) with Daubechies wavelet filters of fixed and adapted orders. These adapted orders have been determined according to several criteria. (b) We have used the best basis wavelet selection algorithm according to the non-normalized Shannon entropy criterion. We have named the proposed algorithm MAL (Meilleur Arbre du Locuteur). We have then used it to build a speaker dependent library tree, corresponding to the best structure for given speakers. Experiments have been conducted using 60 speakers, extracted from the Yoho Database and using a phone-based HMM speaker verification system. Finally, the results of the two approaches have been compared with those given by the MFCC method.

REMERCIEMENTS

Je tiens à remercier mon directeur de recherche, monsieur Chakib Tadj, professeur à l'école de technologie supérieure, pour le soutien technique et moral qu'il m'a accordé tout au long de cette recherche. Je le remercie aussi pour sa disponibilité, ses conseils judicieux, et son aide lors des travaux expérimentaux.

J'aimerais aussi remercier mon codirecteur de recherche, monsieur Christian Gargour, professeur à l'École de technologie supérieure, pour ses directives pertinentes et son aide lors de la recherche et de la révision du mémoire.

Je remercie aussi monsieur Daniel Sales, responsable technique du laboratoire des calculs avancés (LCA) à l'École de technologie supérieure, pour son assistance technique.

Enfin, j'aimerais remercier mes parents Mustapha et Sakina, mes frères, mes sœurs, mes oncles, mes tantes et tous mes amis de près et de loin, tout particulièrement Omar.

TABLE DES MATIÈRES

	Page
SOMMAIRE	i
ABSTRACT	ii
REMERCIEMENTS	iii
TABLE DES MATIÈRES	iv
LISTE DES TABLEAUX	vii
LISTES DES ABRÉVIATIONS ET SIGLES	x
INTRODUCTION	1
CHAPITRE 1 TRANFORMÉE EN ONDELETTES	4
1.1 Introduction	4
1.2 Transformée en ondelettes	4
1.3 Analyse multirésolution	7
1.4 L'algorithme pyramidal	9
1.4.1 Coefficients d'approximation	9
1.4.2 Coefficients de détail	10
1.5 Propriétés des filtres h et g et construction des fonctions d'échelle φ et des ondelettes ψ	12
1.6 Interprétation de l'algorithme récursif d'analyse multirésolution de S. Mallat en banc de filtres :	15
1.7 Les paquets d'ondelettes	20
1.7.1 Définition	20
1.7.2 Arbre admissible	22
1.7.3 Interprétation en terme de segmentation fréquentielle	24
1.8 Ondelettes de Daubechies	25
1.9 Méthode de sélection de la meilleure base d'ondelettes	27
1.10 Conclusion	28

CHAPITRE 2 QUELQUES MÉTHODES D'EXTRACTION DES PARAMÈTRES	
ACOUSTIQUES.....	29
2.1 Introduction.....	29
2.2 MFCC	29
2.2.1 Filtre de préaccentuation.....	30
2.2.2 Fenêtres de Hamming	30
2.2.3 Coefficients MFCC.....	31
2.3 MFDWC	34
2.3.1 Introduction.....	34
2.3.2 Arbre admissible approximant l'échelle de Mel.....	34
2.3.3 Coefficients MFDWC.....	38
2.4 Méthode basée sur la sélection de la meilleure base.....	39
2.4.1 Introduction.....	39
2.4.2 Stratégie locale de recherche d'un optimum global.....	40
2.4.3 Critère d'entropie	43
2.4.4 Extraction des paramètres	44
2.5 Conclusion	44
CHAPITRE 3 MODÉLISATION ET RECONNAISSANCE DU LOCUTEUR.....	45
3.1 Introduction.....	45
3.2 Estimation du modèle du locuteur et du modèle global.....	47
3.3 Règle de décision et seuil.....	48
3.4 Modèles de Markov cachés.....	49
3.5 Probabilité d'observation d'une séquence sachant un modèle.....	52
3.6 Réestimation des paramètres d'un modèle pour l'apprentissage.....	53
3.7 Algorithme de Viterbi pour la reconnaissance d'une séquence.....	55
3.8 Conclusion	58
CHAPITRE 4 MODIFICATIONS APPORTÉES AUX MÉTHODES EXISTANTES	59
4.1 Introduction.....	59
4.2 Schème de prétraitement.....	59

4.3	Première partie du travail.....	61
4.3.1	Performance de reconnaissance en fonction de l'ordre de Daubechies...	61
4.3.2	Performance de reconnaissance en fonction de l'ordre adapté au sens de quelques critères d'entropie	62
4.4	Deuxième partie du travail.....	64
4.4.1	Algorithme de construction d'arbre admissible d'un locuteur.....	64
4.4.2	Quelques exemples d'application de l'algorithme proposé	69
4.5	Conclusion	77
CHAPITRE 5 RÉSULTATS EXPERIMENTAUX		78
5.1	Introduction.....	78
5.2	Description de la base de donnée.....	78
5.3	Estimation du modèle de chaque locuteur et du modèle global.....	79
5.4	Performances de reconnaissance.....	79
5.5	Résultats expérimentaux	80
5.5.1	Première partie du travail.....	81
5.5.2	Deuxième partie du travail.....	88
5.6	Conclusion	91
CONCLUSION.....		93
BIBLIOGRAPHIE.....		95

LISTE DES TABLEAUX

	Page
Tableau I	Description de l'algorithme MAL (Meilleur Arbre du Locuteur) 65
Tableau II	Performances de reconnaissance – MFDWC à ordre fixe (5 locuteurs, phrases d'entraînement) 82
Tableau III	Performances de reconnaissance – MFDWC à ordre fixe (5 locuteurs, phrases de vérification) 83
Tableau IV	Performances de reconnaissance – MFDWC à ordre fixe (60 locuteurs, phrases d'entraînement) 84
Tableau V	Performances de reconnaissance – MFDWC à ordre fixe (60 locuteurs, phrases de vérification) 84
Tableau VI	Performances de reconnaissance – MFDWC à ordre adapté (5 locuteurs, phrases d'entraînement) 85
Tableau VII	Performances de reconnaissance – MFDWC à ordre adapté (5 locuteurs, phrases de vérification) 86
Tableau VIII	Performance de reconnaissance – MFDWC à l'ordre adapté (60 locuteurs, phrases d'entraînement) 87
Tableau IX	Performance de reconnaissance – MFDWC à ordre adapté (60 locuteurs, phrases de vérification) 87
Tableau X	Performance de reconnaissance – SMB_MAL (60 locuteurs, phrases d'entraînement) 88
Tableau XI	Performance de reconnaissance – SMB_MAL (60 locuteurs, phrases de vérification) 89
Tableau XII	Performances de reconnaissance – SMB_MAL (60 locuteurs, phrases de vérification bruitées) 90
Tableau XIII	Performances de reconnaissance pour les différentes méthodes utilisées (60 locuteurs, phrases de vérification) 91

LISTE DES FIGURES

		Page
Figure 1	Représentation temps-échelle de la TO et de la STFT	6
Figure 2	Algorithme récursif d'analyse multirésolution de S. Mallat.....	11
Figure 3	Banc de filtres de l'analyse multirésolution à 3 niveaux d'échelles.....	15
Figure 4	Propriété d'équivalence entre une décimation (un sous-échantillonnage par un facteur M suivie d'un filtrage avec $H(z)$ et un filtrage avec $H(z^M)$ suivi d'une décimation par un facteur M.	16
Figure 5	Schéma équivalent du banc de filtre d'analyse de la figure 3.....	17
Figure 6	Réponses fréquentielles en module des filtres $H(z)$ et $G(z)$	17
Figure 7	Réponses fréquentielles en module des filtres $H(z^2)$ et $G(z^2)$	18
Figure 8	Réponses fréquentielles en module des filtres $H(z^4)$ et $G(z^4)$	18
Figure 9	Réponses en fréquence des filtres du banc illustré sur la figure 5	19
Figure 10	Largeur de bande des détails $d_j, j = 1, \dots, 3$ et de la dernière.....	19
Figure 11	Arbre binaire de la décomposition totale en paquets d'ondelettes à trois niveaux d'échelle	22
Figure 12	Exemple d'arbre admissible C : les paquets $C_2^0 C_2^1 C_3^4 C_3^5 C_2^3$ constituent les feuilles de l'arbre admissible C	23
Figure 13	Segmentation fréquentielle obtenue à partir de l'arbre admissible illustré sur la figure 12	24
Figure 14	Spectres des réponses impulsionnelles des filtres passe-bas h pour l'ordre de Daubechies 2, 4, 6, et 8	26
Figure 15	Spectres des réponses impulsionnelles des filtres passe-haut g pour l'ordre de Daubechies 2, 4, 6, et 8	26
Figure 16	Fenêtres d'analyse et vecteurs paramètres.....	31
Figure 17	Schéma d'extraction des coefficients MFCC.....	32
Figure 18	Banc de filtres à l'échelle de Mel.....	33

Figure 19	Décomposition du nœud $C_j^p(n)$ d'un seul niveau d'échelle.....	36
Figure 20	Arbre admissible proposé par Farooq [7] approximant l'échelle de Mel	37
Figure 21	Schéma d'extraction des coefficients MFDWC.....	38
Figure 22	Entropies des nœuds et des feuilles marquées en cercle de l'arbre initial ...	41
Figure 23	Sélection des nouveaux nœuds feuilles retournés par l'algorithme.....	42
Figure 24	Feuilles finales du nouveau arbre optimal	43
Figure 25	Courbe de variation des FA en fonction des FR.....	47
Figure 26	Modèle Markov Caché à six états [25]	51
Figure 27	Représentation des mixtures multiples par état [25].....	51
Figure 28	Algorithme itératif de la méthode EM	54
Figure 29	Sélection d'un chemin dans un treillis entre deux instants successifs	57
Figure 30	Reconstruction du chemin correspondant à la séquence d'états optimale ...	57
Figure 31	Schéma de prétraitement utilisé dans les deux parties du travail.....	60

LISTES DES ABRÉVIATIONS ET SIGLES

a	Facteur de changement d'échelle.
b	Facteur de translation.
CWT	Transformée en ondelettes continue.
DWT	Transformée en ondelettes discrète.
FA	Pourcentage des fausses acceptations
FR	Pourcentage des faux rejets
MAL	Meilleure Arbre du locuteur.
SC	Pourcentage des scores
STFT	Transformée de Fourier à fenêtre glissante.
TF	Transformée de Fourier.
a'_j	Coefficients d'approximation à l'échelle j
d'_j	Coefficients du détail à l'échelle j
C_j^p	Nœud de l'arbre admissible à l'échelle j et à la position p .
E_j^p	Entropie du nœud C_j^p .
O_j^p	Nombre d'occurrence du nœud C_j^p .
V_j	Espace d'approximation à l'échelle j
W_j	Espace de détail à l'échelle j
Δt	Résolution temporelle
$\Delta \omega$	Résolution fréquentielle
φ	Fonction d'échelle
ψ	Fonction d'ondelette
Φ	Transformée de Fourier de φ
Ψ	Transformée de Fourier de ψ
φ_j^{ip}	Fonction d'échelle au niveau d'échelle j et à la position p
ψ_j^p	Fonction d'ondelette au niveau d'échelle j et à la position p

INTRODUCTION

Les systèmes de reconnaissance du locuteur sont généralement constitués de trois principaux modules (module de prétraitement et de paramétrisation, module de modélisation et d'apprentissage, et finalement module de reconnaissance). Dans le module de prétraitement, les vecteurs de paramètres acoustiques issus de l'analyse du spectre à court terme du signal (par exemple, les coefficients de banc de filtres, ou les coefficients cepstraux) sont extraits à partir des signaux de parole. Dans le module de modélisation, ces paramètres sont utilisés pour créer les modèles de références des différents locuteurs. Le module de reconnaissance permet soit de vérifier ou d'identifier le locuteur qui a parlé. Les paramètres qui sont extraits à partir des signaux de parole, permettent de représenter aux mieux les traits acoustiques du locuteur. Dans le cadre de ce sujet, nous présenterons trois méthodes d'analyse acoustique et de paramétrisation. La première, appelée MFCC [8] (Mel Frequency Cepstral Coefficients) fait usage à la transformée de Fourier et à un banc de filtres dont les espacements des fréquences centrales sont réglés à l'échelle de Mel [3]. Ce dernier a modélisé la perception auditive de l'oreille humaine. Quoiqu'elle est souvent utilisée dans les systèmes de reconnaissance du locuteur. Certes elle a un inconvénient majeur, celui de l'uniformité de ses résolutions temporelle et fréquentielle. En effet, cette méthode a la difficulté à détecter les brusques variations qui se produisent à l'intérieure d'un signal lent. La seconde méthode, dite MFDWC (Mel Frequency Discret Wavelet Coefficients) [7] est basée sur la transformée en ondelettes, elle fait usage à un banc de filtres qui approxime l'échelle de Mel. Quoique cette méthode est récente, toutefois, il n'en demeure pas moins intéressante. En effet, les ondelettes bénéficient de la propriété de localisation en temps et en fréquence. Elles ont la propriété de se contracter et de se dilater, leur permettant ainsi d'avoir des résolutions variables dans le plan temps-fréquence. En se contractant, elles détectent mieux les hautes fréquences, et mieux les basses fréquences en se dilatant. La troisième méthode, utilise l'algorithme de sélection de la meilleure base d'ondelettes proposé initialement par Coifman et Wickerhauser [4], et l'algorithme

MAL (Meilleur Arbre du locuteur) que nous avons proposé dans le cadre de ce sujet. L'algorithme de sélection de la meilleure base d'ondelettes se base sur la recherche de la base d'ondelettes sur laquelle l'information du signal analysé est la plus concentrée. Cet algorithme est défini à la fois par un critère, et par une stratégie de recherche de la base optimisant celui-ci. Nous avons proposé l'algorithme MAL dans le but de créer pour chaque locuteur son meilleur arbre admissible. Ce dernier permettra d'analyser chaque signal de parole dans ses bandes de fréquences où l'information est la plus concentrées.

Le signal de la parole appartient à la classe des signaux multicomposantes stationnaires par morceaux. À ce type de signaux, nous pouvons associer en particulier la notion d'analyse en sous-bandes (ou de banc de filtres), qui permet d'analyser l'évolution du contenu spectral du signal de la parole, et par conséquent extraire les paramètres acoustiques pertinents. L'analyse en sous-bandes (ou en banc de filtres) isolera les différentes composantes du signal les unes des autres, ainsi on peut extraire séparément la contribution fréquentielle de chaque composante en terme d'énergie.

Le travail présenté dans le cadre de ce sujet est divisé en deux parties principales. Dans la première, nous utilisons la méthode MFDWC. Nous testerons les performances de reconnaissance de cette méthode en fonction de l'ordre d'ondelette de Daubechies. Cet ordre peut être fixe ou adapté au sens d'un critère d'entropie, en l'occurrence le critère d'énergie. En maximisant ce dernier au niveau de chaque base d'ondelettes de l'arbre admissible proposé par Farooq [7], nous déterminons l'ordre d'ondelette de Daubechies adapté à cette maximisation qui en fait peut être interprétée comme étant la meilleure corrélation entre l'ondelette de Daubechies adapté à cet ordre et le signal analysé en terme d'énergie.

Dans la deuxième partie du travail, nous utiliserons la méthode basée sur l'algorithme de sélection de la meilleure base d'ondelettes au sens du critère d'entropie de Shannon. Dans le cadre de la classification ce critère a été proposé par Saito et Coifman [21]. Il

permet de mesurer le degrés de concentration de l'information du signal analysé dans chacune de ses bandes de fréquences. Pour chaque locuteur, cet algorithme sera appliqué pour toutes ses phrases d'entraînements afin d'obtenir pour chacune d'elles le meilleur arbre admissible qui lui correspond. L'algorithme MAL (Meilleur Arbre du Locuteur) que nous avons proposé permet de construire le meilleur arbre du locuteur à partir de tous ses arbres admissibles d'entraînement optimaux. L'objectif fixé dans cette deuxième approche est de maximiser la séparabilité entre les classes d'arbres d'apprentissage de tous les locuteurs et d'évaluer le pouvoir discriminant de chacun de ces arbres représentant un locuteur. Dans cette deuxième partie de notre travail, nous testerons les performances de reconnaissance en utilisant une base de donnée non bruitée et une autre bruitée.

Le premier chapitre traite quelques notions théoriques sur la transformée en ondelettes. Le second présente quelques méthodes d'extraction des paramètres acoustiques. Le troisième chapitre donne un aperçu sur la modélisation et la reconnaissance du locuteur. Le quatrième, présente les principaux travaux que nous avons effectués dans le cadre de ce sujet. Finalement, le cinquième chapitre, est celui des résultats expérimentaux.

CHAPITRE 1

TRANSFORMÉE EN ONDELETTES

1.1 Introduction

Dans ce chapitre, nous commencerons par présenter la différence entre la transformée de Fourier et la transformée en ondelettes, tout particulièrement en termes de résolution temps-fréquence. Nous présenterons ensuite les concepts de base de la théorie de l'analyse multirésolution. Cette dernière présente les liens qui existent entre bases d'ondelettes et bancs de filtres. Nous présenterons ensuite un bref aperçu sur la décomposition en paquets d'ondelettes qui est une généralisation de l'analyse multirésolution. Un algorithme de sélection de la meilleure base d'ondelettes sera défini à la fois par un critère, et par une stratégie de recherche de la base optimisant celui-ci.

1.2 Transformée en ondelettes

L'analyse d'un signal consiste souvent à déterminer son contenu fréquentiel. Ce dernier peut être obtenu en calculant sa transformée de Fourier (TF), définie pour un signal $x(t)$ d'énergie finie, par :

$$TF_x(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt = \langle x, e^{j2\pi ft} \rangle \quad (1.1)$$

La TF calcul un coefficient de similarité entre le signal $x(t)$ et une collection de sinusoïdes (exponentielles complexes) de longueurs infinies. Cette transformation donne la contribution de chaque sinusoïde de fréquence f au signal $x(t)$. Toutefois l'information reliée aux variations temporelles de ce signal, n'est pas représentée. En effet, la transformée de Fourier ne donne pas la distribution temporelle des fréquences formant le signal.

La transformée de Fourier à fenêtre glissante (STFT : Short Term Fourier Transform) calcule la TF sur un intervalle de temps limité par une fenêtre d'analyse h centrée sur un temps τ ; elle est définie par :

$$STFT_x(f, \tau) = \int_{-\infty}^{\infty} x(t)h^*(t - \tau)e^{-j2\pi ft} dt \quad (1.2)$$

La STFT consiste à comparer le signal $x(t)$ à une collection d'ondes sinusoïdales limitées dans le temps et dont les durées dépendent de celle de la fenêtre d'analyse h qui subit une translation temporelle τ . Ainsi la STFT mesure la contribution de la fréquence f au signal analysé dans un intervalle de temps centré à l'instant τ .

La transformée continue en ondelettes (CWT : Continuous Wavelets Transform), calcule la projection du signal $x(t)$ sur une base de fonctions analysantes $\psi_{a,b}$ appelées ondelettes. L'indice a s'appelle facteur d'échelle, et l'indice b s'appelle opérateur de translation temporelle. La transformée continue en ondelettes d'un signal $x(t)$ est définie par l'équation (1.3) [5] :

$$CWT_x(a, b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}^*(t)dt \quad (1.3)$$

où les ondelettes $\psi_{a,b}$ sont obtenues à partir d'une ondelette mère ψ conformément à l'équation (1.4).

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad (1.4)$$

La transformée continue en ondelettes est inversible, si l'ondelette mère ψ vérifie les conditions d'admissibilité [1]. Ainsi la transformée en ondelettes inverse de $x(t)$ est obtenue par :

$$x(t) = C_{\psi}^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} CWT_x(a, b)\psi_{a,b}(t) \frac{dad b}{a^2} \quad (1.5)$$

où C_{ψ} est un coefficient calculé en fonction de $\Psi(\omega)$ (Transformée de Fourier de $\psi(t)$) comme suit [5] :

$$C_{\psi} = 2\pi \int_{-\infty}^{\infty} |\Psi(\omega)|^2 \frac{d\omega}{\omega} \neq 0 \quad (1.6)$$

La résolution temporelle Δt ainsi que fréquentielle $\Delta\omega$ de l'ondelette ψ dépendent de l'opérateur d'échelle a [24]. En effet on montre que pour a petit (c'est à dire pour les fréquences hautes), la résolution temporelle est bonne mais la résolution fréquentielle est mauvaise, par contre pour a grand, donc pour les fréquences basses, la résolution temporelle est mauvaise, cependant celle fréquentielle est bonne. Ceci s'explique du fait que les pavés élémentaires de la représentation temps-fréquence de la TO : Transformée en ondelettes (voir figure 1) s'élargissent et se contractent selon les variations de l'opérateur d'échelle a . L'inconvénient majeur de la STFT, est qu'elle ne peut donner de résolutions temporelles et fréquentielles variables. Par conséquent les pavés élémentaires de la représentation temps-fréquence de la STFT (voir figure 1) sont identiques.

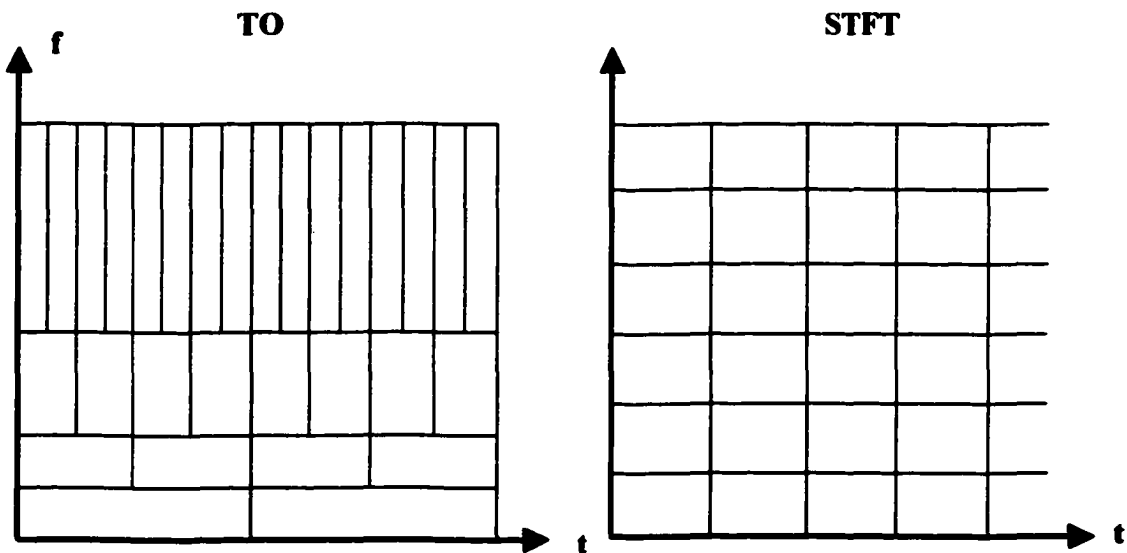


Figure 1 Représentation temps-échelle de la TO et de la STFT

Le principe d'incertitude d'Heisenberg met en évidence que la localisation temps-fréquence idéale ($\Delta t=0, \Delta \omega=0$) est impossible et que le produit $\Delta t \Delta \omega$ est borné inférieurement par :

$$\Delta t \Delta \omega \geq \frac{1}{4\pi} \quad (1.7)$$

Ceci s'explique que si nous augmentons la résolution temporelle (c'est à dire que si nous diminuons Δt), alors nous nous retrouvons à diminuer la résolution fréquentielle (c'est à dire à augmenter $\Delta \omega$) et vice-versa.

La transformée continue en ondelettes présente un caractère de redondance, tant du point de vue du temps que du facteur d'échelle. La transformée en ondelettes discrète se base sur la discrétisation du facteur d'échelle a et de l'opérateur de translation b , et l'on pose habituellement [6] $a = 2$ et $b=1$, pour une transformée en ondelettes dyadique.

1.3 Analyse multirésolution

La théorie de l'analyse multirésolution a été mise au point vers la fin de l'année 1986 par Mallat [12,13] et Meyer [16,17]. Elle constitue un outil nécessaire à la compréhension et à la réalisation de bases d'ondelettes. On utilise les bases orthonormées d'ondelettes comme un outil pour décrire le passage d'une représentation grossière du signal à son approximation plus fine. L'analyse multirésolution consiste en une séquence d'espaces d'approximations successives V_j . Ces derniers doivent satisfaire les conditions suivantes :

$$\dots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \dots \quad (1.8)$$

$$\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}) \quad (1.9)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \quad (1.10)$$

$L^2(\mathbf{R})$ est l'espace vectoriel des signaux $f(t)$ continus à énergie finie où t appartient à l'ensemble des réels \mathbf{R} . Notons que dans la notation employée, plus j est petit plus l'espace V_j est grand. Les trois conditions (1.8), (1.9) et (1.10), expriment la propriété de complétude des sous espaces V_j . La propriété de multirésolution est, entre autres, une conséquence d'une condition supplémentaire qui exprime l'invariance par changement d'échelle:

$$f(t) \in V_j \Leftrightarrow f\left(\frac{t}{2}\right) \in V_{j+1}, \quad j \in \mathbf{Z} \quad (1.11)$$

Ainsi, tous les espaces sont des versions à différentes échelles de l'espace central V_0 . Une autre propriété nécessaire à la réalisation d'une analyse multirésolution est l'invariance de V_0 par translation :

$$f(t) \in V_0 \Rightarrow f(t-n) \in V_0 \quad \forall n \in \mathbf{Z} \quad (1.12)$$

Une dernière condition, nécessaire à une analyse multirésolution, est l'existence d'une fonction ϕ , appelée fonction d'échelle, appartenant à V_0 , telle que :

$$\{\phi_{0,n}; n \in \mathbf{Z}\}, \text{ est une base orthonormée de } V_0 \quad (1.13)$$

Par conséquent, les relations (1.11) et (1.13) réunies impliquent que pour tout j, n appartenant à \mathbf{Z} , $\phi_{j,n}$ est une base orthonormée de V_j tel que [5] :

$$\{\phi_{j,n}; j, n \in \mathbf{Z}\}, \quad \phi_{j,n}(t) = 2^{-j/2} \phi(2^{-j} t - n) \quad (1.14)$$

Pour tout $j \in \mathbf{Z}$, nous définissons W_j comme le complément orthogonal de V_j dans V_{j-1} . L'espace vectoriel W_j représente les détails à ajouter aux signaux représentés dans V_j pour les représenter dans V_{j-1} . Nous avons dans une transcription mathématique :

$$V_{j-1} = V_j \oplus W_j \quad (1.15)$$

L'idée fondamentale d'une telle analyse est l'existence d'une fonction appelée ondelette, notée ψ qui par dilatations et translations engendre une base orthonormée d'ondelettes de W_j tel que [5] :

$$\{\psi_{j,k}; j, k \in \mathbf{Z}\}, \quad \psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j} t - k) \quad (1.16)$$

L'approximation à l'échelle immédiatement plus fine pourra donc être reconstruite en utilisant les détails du signal fournis par sa projection sur la base de W_j telle que, pour

tout f de $L^2(\mathbb{R})$ on a la relation (1.24) où T_j est l'opérateur de projection orthogonale de f sur V_j .

$$T_{j-1} f = T_j f + \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k} \quad (1.17)$$

1.4 L'algorithme pyramidal

Les coefficients d'approximation et de détail s'obtiennent grâce à l'algorithme pyramidal proposé par S. Mallat [14]. Cet algorithme se base sur les relations à deux échelles qui seront succinctement présentées dans ce qui suit.

1.4.1 Coefficients d'approximation

Pour un signal f d'énergie finie, un coefficient d'approximation à un niveau d'échelle j est défini par :

$$a'_n = \langle f, \varphi_{j,n} \rangle \quad (1.18)$$

Sachant que φ est une fonction de V_0 , et comme $V_0 \subset V_{-1}$ on peut décomposer $\varphi(x)$ sur la base de V_{-1} . Et donc, il existe une suite numérique h_n avec $n \in \mathbb{Z}$ telle que [11] :

$$\varphi(t) = \sum_n h_n \varphi_{-1,n}(t) \quad (1.19)$$

conformément à (1.14), on a :

$$\varphi_{-1,n}(t) = 2^{\frac{1}{2}} \varphi(2t - n) \quad (1.20)$$

la suite h_n sera considérée comme étant la réponse impulsionnelle d'un filtre numérique, et on a :

$$h_n = \langle \varphi, \varphi_{-1,n} \rangle \quad (1.21)$$

Remarquons que $\varphi(x)$ étant par construction normée (en énergie), et on a [24] :

$$\|\varphi(t)\|^2 = \langle \varphi, \varphi \rangle = 1 \quad (1.22)$$

et donc [24], la norme se conservant à travers les échelles :

$$\sum_n h_n^2 = 1 \quad (1.23)$$

Le calcul récursif des coefficients d'approximation est effectué en décomposant les coefficients a'_n en fonction des coefficients $a_{n-1}^{j-1} = \langle f, \varphi_{j-1,n} \rangle$

Étant donnée les relations (1.19) et (1.20) on a :

$$\varphi(x) = \sum_n h_n 2^{\frac{1}{2}} \varphi(2t - n) \quad (1.24)$$

donc d'après [24]:
$$a'_n = \sum_l h_{2n-l} \langle f, \varphi_{j-1,l} \rangle \quad (1.25)$$

et finalement on a l'équation récursive suivante :

$$a'_n = \sum_l h_{2n-l} a_l^{j-1} \quad (1.26)$$

Les coefficients d'approximation a'_n au niveau d'échelle j sont calculés à partir des coefficients d'approximation a_l^{j-1} au niveau d'échelle inférieur $j-1$. L'expression (1.26) montre que le calcul des a'_n peut être interprété comme le filtrage de la séquence a_l^{j-1} par le filtre de réponse impulsionnelle h_n suivi par un sous-échantillonnage de rapport 2.

1.4.2 Coefficients de détail

Pour un signal f d'énergie finie, un coefficient de détail à un niveau d'échelle j est défini par :

$$d'_n = \langle f, \psi_{j,n} \rangle \quad (1.27)$$

Sachant que ψ est une fonction de W_0 , et comme $W_0 \subset V_{-1}$ on peut décomposer $\psi(x)$ sur la base de V_{-1} . Et donc, il existe une suite numérique g_n avec $n \in \mathbf{Z}$ telle que [11]:

$$\psi(t) = \sum_n g_n \varphi_{-1,n}(t) \quad (1.28)$$

la suite g_n sera considérée comme étant la réponse impulsionnelle d'un filtre numérique, et on a :

$$g_n = \langle \psi, \varphi_{-1,n} \rangle \quad (1.29)$$

Le calcul récursif des coefficients de détail est effectué en décomposant les coefficients d'_n en fonction des coefficients $a_{n-1}^{j-1} = \langle f, \varphi_{j-1,n} \rangle$. Étant donnée les relations (1.16), (1.28) et

(1.29) on a :

$$\psi(t) = \sum_n g_n 2^{\frac{1}{2}} \varphi(2t - n) \quad (1.30)$$

donc d'après [24] :

$$d'_n = \sum_l g_{2n-l} \langle f, \varphi_{l-1,l} \rangle \quad (1.31)$$

et finalement on a l'équation réursive suivante :

$$d'_n = \sum_l g_{2n-l} a_{l-1}^{j-1} \quad (1.32)$$

Les coefficients de détail d'_n au niveau d'échelle j sont calculés à partir des coefficients d'approximation a_{l-1}^{j-1} au niveau d'échelle inférieur $j-1$. L'expression (1.32) montre que le calcul des d'_n peut être interprété comme le filtrage de la séquence a_{l-1}^{j-1} par le filtre de réponse impulsionnelle g_n suivi par un sous-échantillonnage de rapport 2.

Les deux équations (1.26) et (1.32) décrivent l'algorithme récurif de l'analyse multirésolution de S. Mallat [14], ainsi nous représentons schématiquement sur la figure 2, le bloc de calcul issu de ces deux équations.

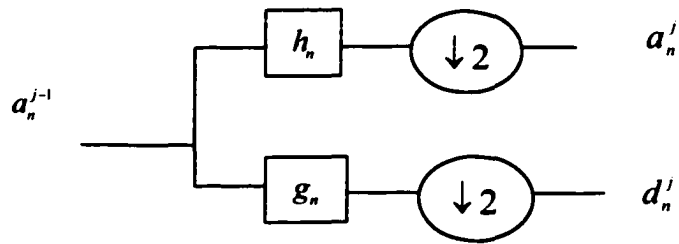


Figure 2 Algorithme récurif d'analyse multirésolution de S. Mallat

1.5 Propriétés des filtres h et g et construction des fonctions d'échelle φ et des ondelettes ψ

Dans cette section, nous étudierons les propriétés des filtres dont les réponses impulsionnelles sont h_n et g_n pour en déduire les règles de construction des bases d'ondelettes ψ et de fonctions d'échelle φ liées à une analyse multiresolution. Nous suivrons ici le traitement de [24] pour présenter la construction et quelques propriétés des filtres h et g associés à une analyse multiresolution orthogonale.

Les contraintes d'orthogonalité et d'orthonormalité des fonctions d'échelles avec leurs translatées de même niveau de résolution, détermineront la nature du filtre h et ses propriétés. Nous nous baserons ici sur l'analyse donnée par [24]. La transformée de Fourier de la relation (1.19) donne l'équation suivante :

$$\Phi(\omega) = \frac{1}{\sqrt{2}} H\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right) \quad (1.33)$$

sachant que $H(\omega)$ est la réponse fréquentielle du filtre h . $H(\omega)$ est périodique, sa période est de $\omega_p = 2\pi$. Le produit scalaire de la fonction d'échelle $\varphi(x)$ par une de ses translatées, peut être vu comme une fonction d'autocorrélation $r(\tau)$ telle que :

$$r(\tau) = \langle \varphi(t), \varphi(t + \tau) \rangle = \int_{-\infty}^{\infty} \varphi(t) \varphi(t + \tau) dx \quad (1.34)$$

La transformée de Fourier de $r(\tau)$ s'écrit donc tel que:

$$R(\omega) = \Phi(\omega) \Phi^*(\omega) = |\Phi(\omega)|^2 \quad (1.35)$$

La fonction échantillonnée de la fonction d'autocorrélation $r(\tau)$ s'écrit comme suit :

$$r_e(\tau) = r(\tau) \cdot \sum_n \delta(\tau - n) \quad (1.36)$$

La transformée de Fourier de $r_e(\tau)$ s'écrit tel que:

$$R_e(\omega) = \frac{1}{2\pi} R(\omega) * 2\pi \sum_n \delta(\omega - 2\pi n) \quad (1.37)$$

$$R_e(\omega) = \sum_n R(\omega - 2\pi n) \quad (1.38)$$

La fonction d'échelle engendre par translation une famille orthonormée, telle que :

$$r(\tau) = \langle \varphi(t), \varphi(t + \tau) \rangle = \delta(\tau) \quad (1.39)$$

La transformée de Fourier du Dirac $\delta(\tau)$ égale à l'unité, donc les relations (1.39),(1.35), et (1.38) donne l'expression (1.40) qui exprime l'orthonormalité de la fonction d'échelle dans Fourier.

$$\sum_n |\Phi(\omega + 2\pi n)|^2 = 1 \quad (1.40)$$

En reportant la relation (1.40) dans l'équation (1.33), à la pulsation double, et sachant que $H(\omega)$ est périodique de $\omega_p=2\pi$, on retrouve la relation (1.41) telle que :

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 = 2 \quad (1.41)$$

Etant donné que la fonction d'échelle $\varphi(t)$ n'est pas à moyenne nulle [1] donc on a :

$$\int_{-\infty}^{\infty} \varphi(t) dt = \Phi(0) \neq 0 \quad (1.42)$$

L'expression (1.42) s'explique du fait que la fonction d'échelle permet de lisser le signal lors du passage d'une résolution fine à une résolution plus grossière. Ayant la condition de la relation (1.42), l'équation (1.41) conduit à :

$$H\left(\frac{\omega}{2}\right) = \sqrt{2} \frac{\Phi(\omega)}{\Phi\left(\frac{\omega}{2}\right)} \quad (1.43)$$

Si on calcule l'équation (1.43) à la pulsation $\omega=0$, et on la combine avec la condition (1.42), on obtient :

$$\begin{cases} H(0) = \sum_n h_n = \sqrt{2} \\ H(\pi) = 0 \end{cases} \quad (1.44)$$

Ce résultat nous indique que le filtre h est de la famille des filtres passe-bas. Il nous permettra d'avoir une approximation du signal en éliminant les détails.

Les fonctions d'ondelettes doivent former une base orthonormée globale et être orthogonales à travers les échelles aussi bien que pour une échelle donnée. Selon le traitement de [24], la transformée de Fourier de la relation (1.28) donne l'équation (1.45) :

$$\Psi(\omega) = \frac{1}{\sqrt{2}} G\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right) \quad (1.45)$$

L'orthonormalité des fonctions d'ondelettes vis à vis de leurs translatées donne une relation analogue à celle trouvée pour les fonctions d'échelle dans l'équation (1.40) :

$$\sum_n |\Psi(\omega + 2\pi n)|^2 = 1 \quad (1.41)$$

par conséquent avec le même traitement que précédemment, on trouve [24] la relation suivante :

$$|G(\omega)|^2 + |G(\omega + \pi)|^2 = 2 \quad (1.42)$$

La fonction d'ondelette est à moyenne nulle, car elle doit être admissible [1] tel que :

$$\int_{-\infty}^{\infty} \psi(t) dt = \Psi(0) = 0 \quad (1.43)$$

Si l'on combine la relation (1.43) avec (1.42) on en déduit que :

$$\begin{cases} G(0) = \sum_n g_n = 0 \\ G(\pi) = \sqrt{2} \end{cases} \quad (1.44)$$

Ces résultats pourront nous indiquer que le filtre g est de la famille des filtres passe-haut, il nous permettra d'avoir le détail du signal en éliminant les approximations. Selon le traitement que l'on retrouve dans [24], l'équation (1.45) est une conséquence d'orthogonalité de la famille de la fonction d'ondelette ψ à travers les fonctions d'échelle, et on a :

$$\sum_n \Psi(\omega + 2\pi n) \cdot \Phi(\omega + 2\pi n) = 0 \quad (1.45)$$

En remplaçant Ψ et Φ par leurs expressions déduites de (1.33) et (1.45) et sachant que les filtres h et g sont périodiques de $\omega_p = 2\pi$, on obtient la relation suivante :

$$G(\omega)H^*(\omega) + G(\omega + \pi)H^*(\omega + \pi) = 0 \quad (1.46)$$

si $H^*(\omega) \neq 0$, on déduit de l'expression suivante :

$$G(\omega) = -\frac{G(\omega + \pi)}{H^*(\omega)} H^*(\omega + \pi) \quad (1.47)$$

Smith et Barnwell [23] ont posé la relation (1.48) tel que :

$$\lambda(\omega) = -\frac{G(\omega + \pi)}{H^*(\omega)} = e^{-j\omega} \quad (1.48)$$

Si l'on combine les relations (1.47) et (1.48), et que l'on les exprime par leurs réponses impulsionnelles on trouve :

$$g_n = (-1)^n h_{1-n} \quad (1.49)$$

La relation (1.49) permet de construire le filtre de détail g (filtre passe-haut) connaissant le filtre d'approximation h (filtre passe-bas). Ces filtres numériques sont appelés filtres miroirs en quadrature (QMF : Quadrature mirror filter). Ils sont la base de la théorie de la décomposition des signaux en sous-bandes. La réponse en fréquence du filtre passe bas h est 'l'image miroir' de celle du passe haut g par rapport à la pulsation $\pi/2$.

1.6 Interprétation de l'algorithme récursif d'analyse multirésolution de S. Mallat en banc de filtres :

Un banc de filtres est un ensemble de filtres passe-bande qui permettent d'extraire les sous-bandes d'un signal donné. Supposons que l'on veuille analyser un signal $x(n)$ suivant 3 niveaux de résolution, et cela en utilisant l'algorithme récursif d'analyse multirésolution de S. Mallat, donc suivant les équations (1.37) et (1.43) précédemment définies, on pourra obtenir la représentation – en banc de filtres – correspondante à cette analyse tel que représenté par la figure 3.

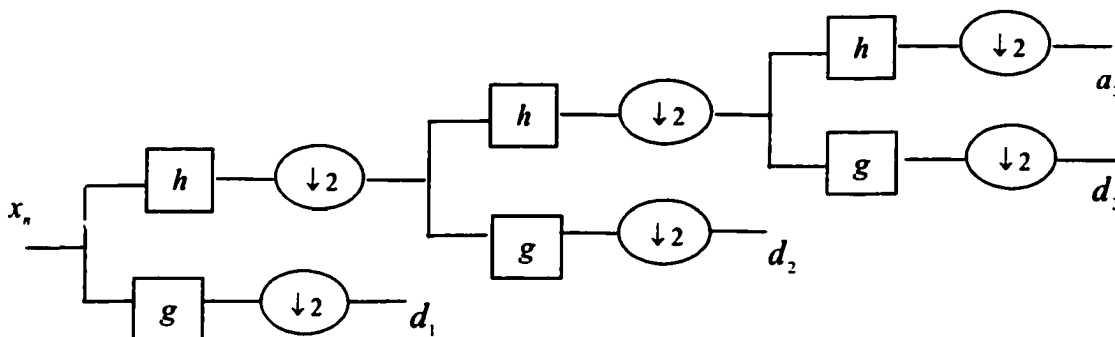


Figure 3 Banc de filtres de l'analyse multirésolution à 3 niveaux d'échelles.

Les transformées en Z des filtres h et g , sont respectivement $H(z)$ et $G(z)$. Nous savons que l'identité de Noble donne la propriété d'équivalence entre une décimation (un sous-échantillonnage) par un facteur M suivie d'un filtrage avec $H(z)$ et un filtrage avec $H(z^M)$ suivi d'une décimation par un facteur M . cette propriété est illustré sur la figure 4.

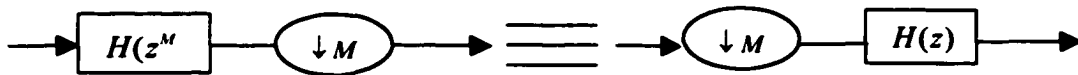


Figure 4 Propriété d'équivalence entre une décimation (un sous-échantillonnage) par un facteur M suivie d'un filtrage avec $H(z)$ et un filtrage avec $H(z^M)$ suivi d'une décimation par un facteur M .

En nous basant sur cette propriété, nous présentons sur la figure 5 le schéma équivalent [22] du banc de filtres de l'analyse multirésolution à 3 niveaux d'échelles illustré sur la figure 3. Les réponses fréquentielles des filtres miroirs à quadrature h et g sont représentées par la figure 6 en supposant qu'ils soient idéaux.

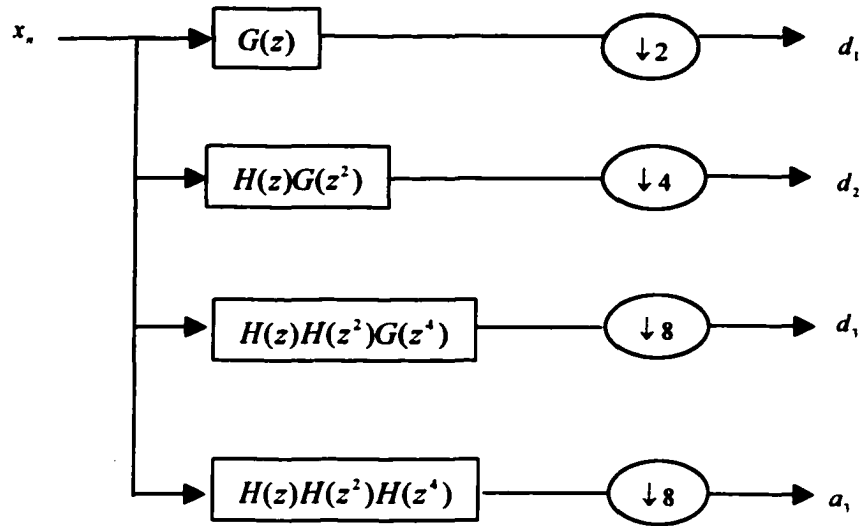


Figure 5 Schéma équivalent du banc de filtre d'analyse de la figure 3

La figure 7 et 8 illustre les réponses fréquentielles des filtres h et g respectivement sous-échantillonnés par 2 et par 4.

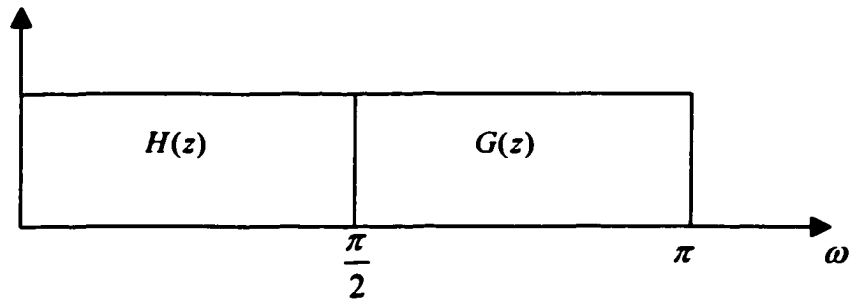


Figure 6 Réponses fréquentielles en module des filtres $H(z)$ et $G(z)$

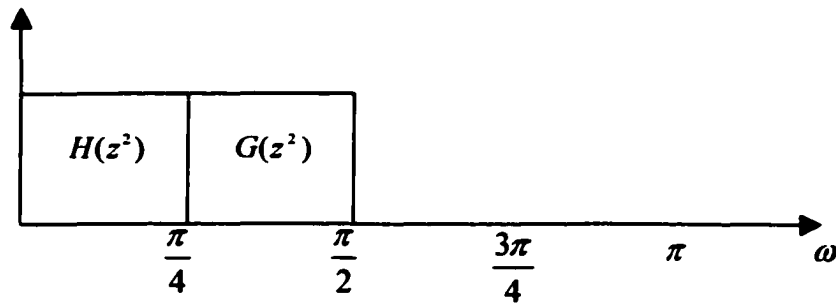


Figure 7 Réponses fréquentielles en module des filtres $H(z^2)$ et $G(z^2)$

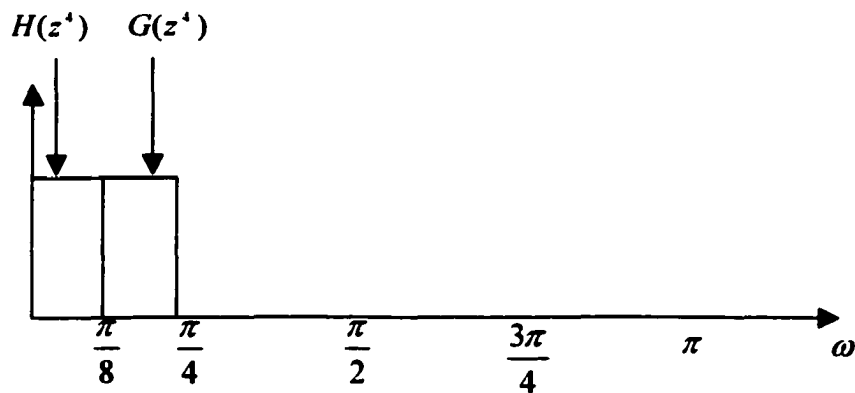


Figure 8 Réponses fréquentielles en module des filtres $H(z^4)$ et $G(z^4)$

Les opérations successives de filtrage et de décimation par un facteur de 2 réitérées au long de la décomposition de la figure 3 réalisent une partition de l'axe de fréquence sous la forme d'un banc de filtres à structure dyadique comme l'indique la figure 9. Selon la figure 10, chaque détail $d^j, j = 1, \dots, 3$, contient de l'information relative à tout le signal dans une bande de fréquence particulière indiquée par j et dont la largeur de bande est égale à $\Delta\omega_j = \frac{\pi}{2^j}$. L'approximation a_j^1 correspond à la bande de fréquences la plus basse du signal, sa largeur est égal à $\frac{\pi}{8}$.

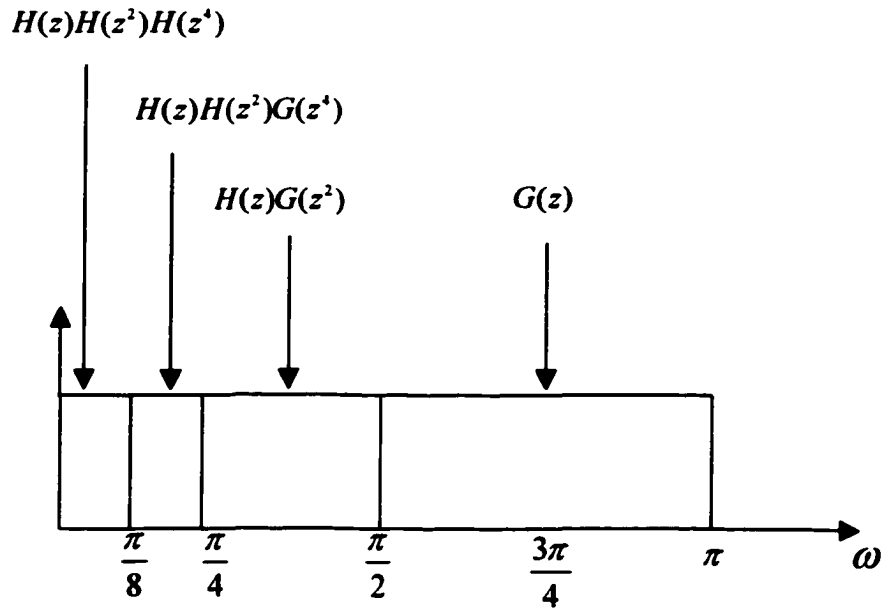


Figure 9 Réponses en fréquence des filtres du banc illustré sur la figure 5

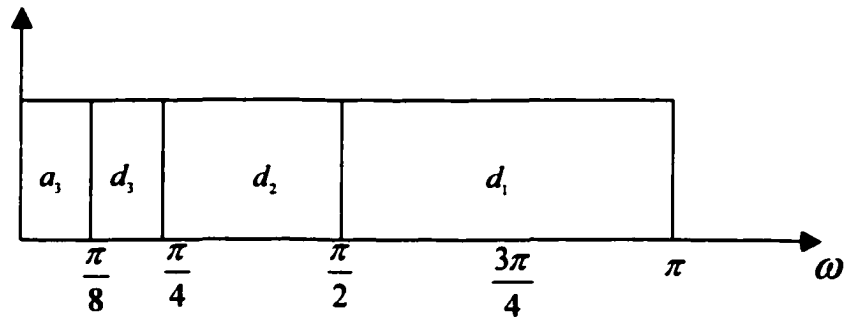


Figure 10 Largeur de bande des détails $d_j, j = 1, \dots, 3$ et de la dernière approximation a^1

1.7 Les paquets d'ondelettes

À partir d'une ondelette initiale donnée, on peut générer une multitude de bases possibles [12]. Toutefois la nécessité de disposer d'une organisation qui exploite toutes ces bases potentielles est cruciale. En effet cela nous permettra de pouvoir sélectionner une base. C'est ce que réalise la décomposition en paquets d'ondelettes. Elle permet d'offrir une librairie d'ondelettes organisées selon leurs propriétés d'analyse et de localisation temps-fréquence. Cette librairie d'ondelettes conduit à une segmentation fréquentielle du signal analysé par le biais des filtres passe-bande associés, et ceci en fonction d'une architecture d'arbre binaire.

1.7.1 Définition

La décomposition en paquets d'ondelettes [5] est considérée au sens structurel comme étant une généralisation de l'analyse multirésolution et de la transformée en ondelettes discrète. Elle permet d'étendre la construction de nouvelles bases non seulement à partir de tous les sous-espaces d'approximation, mais aussi à partir de tous les espaces de détails. Ayant le cas de l'analyse multirésolution, nous avons vu que chaque espace d'approximation V_j peut être décomposé en deux sous-espaces de résolution inférieure V_{j+1} et W_{j+1} , et ceci en transformant la base orthogonale $\phi_{j,k}$ de V_j , en deux sous-bases orthogonales : $\phi_{j+1,k}$ de V_{j+1} et $\psi_{j,k}$ de W_{j+1} . Dans le cadre d'une décomposition en paquets d'ondelettes, on montre [14] que cette décomposition peut être appliquée aussi aux espaces de détails W_j , et qu'on peut par conséquent représenter l'ensemble des bases de la décomposition sous la forme d'un arbre binaire, que l'on note \mathcal{W} . Chaque nœud de cet arbre correspond à un sous espace W_j^p ayant une base orthogonale $\psi_j^p = \{\psi_j^p(t - 2^j n)\}_{n \in \mathbb{Z}}$ avec :

$$\begin{aligned}
W_0^0 &= V_0 \\
\psi_0^0(t) &= \phi_{0,0}(t) \\
\psi_{j+1}^{2^p}(t) &= \sum_n h_n \psi_j^p(t - 2^j n) \\
\psi_{j+1}^{2^{p+1}}(t) &= \sum_n g_n \psi_j^p(t - 2^j n) \\
W_j^p &= W_{j+1}^{2^p} \oplus W_{j+1}^{2^{p+1}}
\end{aligned} \tag{1.50}$$

Nous désignerons par C l'arbre qui correspond à la projection du signal f dans tous les sous-espaces de l'arbre binaire W . Il est bien évident que l'arbre des projections C ait la même structure que celui des sous-espaces engendrés W . Nous définissons un paquet d'ondelettes $C_j^p = \{C_j^p(n)\}_{n \in \mathbb{Z}}$ comme étant la projection du signal f dans le sous-espace W_j^p ayant est une base orthonormée ψ_j^p et on a :

$$C_j^p(k) = \langle f(t), \psi_j^p(t - 2^j n) \rangle \tag{1.51}$$

La décomposition en paquets d'ondelettes est obtenue à l'aide de l'algorithme pyramidal étendu (voir exemple à la figure 11), où h_n et g_n sont les réponses impulsionnelles des filtres passe-bas et passe-haut préalablement définis, ainsi le paquet $C_{j+1}^{2^p}$ (respectivement $C_{j+1}^{2^{p+1}}$) est obtenu par filtrage passe-bas (respectivement passe-haut) de C_{j+1}^p par le filtre h_n (respectivement g_n), suivi d'une décimation par deux qui projette le signal sur le sous-espace $W_{j+1}^{2^p}$ (respectivement $W_{j+1}^{2^{p+1}}$).

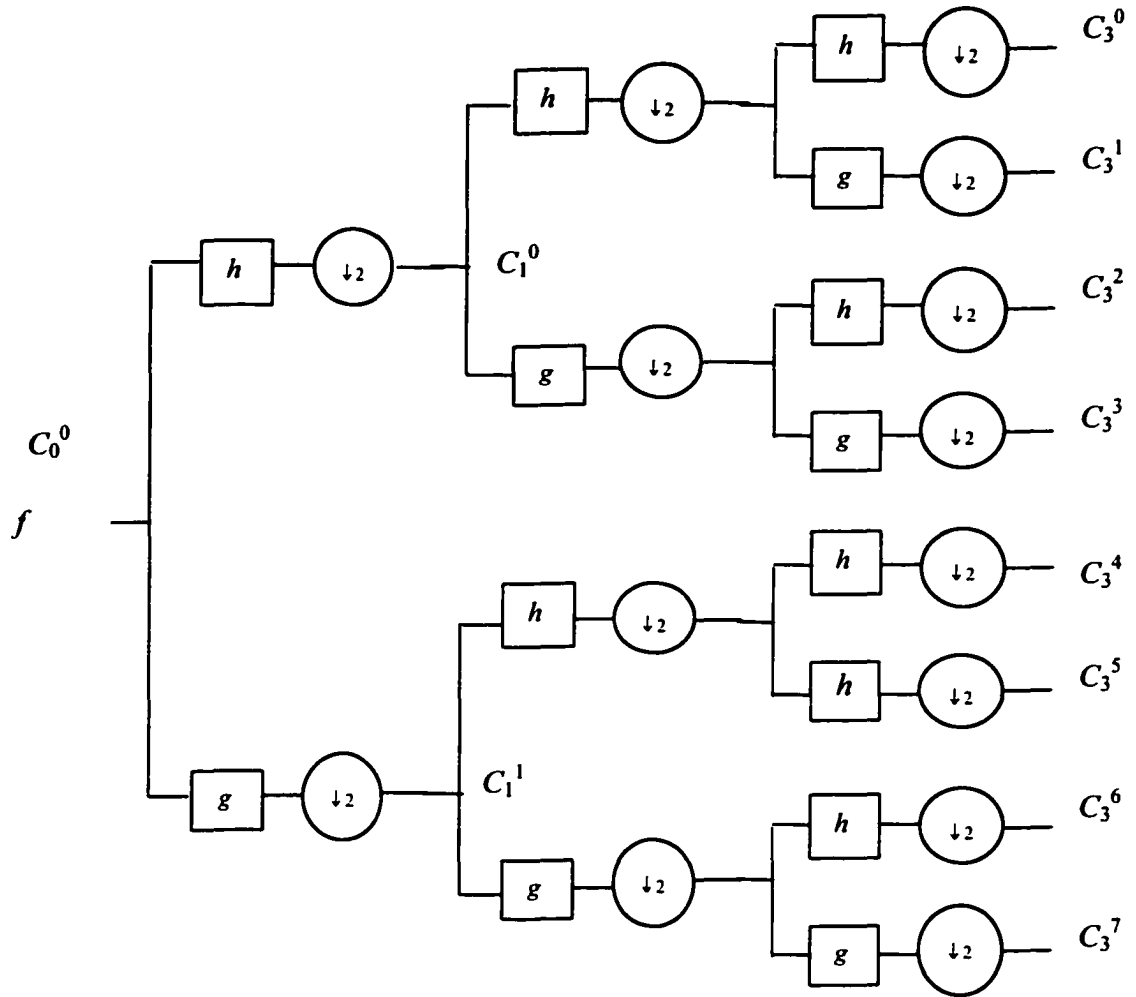


Figure 11 Arbre binaire de la décomposition totale en paquets d'ondelettes à trois niveaux d'échelle

1.7.2 Arbre admissible

La décomposition complète d'un signal en paquets d'ondelettes, est celle qui correspond à toutes les bases de l'arbre \mathcal{W} , elle fournit une représentation redondante. L'ensemble composé de toutes les bases d'ondelettes de chaque niveau de la décomposition contient de l'information du signal. Cet ensemble constitue une

représentation non redondante du signal analysé. Les espaces correspondants sont définis à partir de la notion d'arbre admissible [14]. Un arbre admissible est un arbre binaire tel que chacun de ses nœuds a exclusivement un ou deux fils. L'arbre W que nous avons défini est un arbre admissible à structure complète, à partir duquel on peut construire plusieurs arbres admissibles potentiels. Ces derniers sont obtenus en effectuant l'élagage de l'arbre W en partant du bas vers le haut de l'arbre et en regroupant les nœuds fils ayant le même père (voir l'exemple de la figure 12).

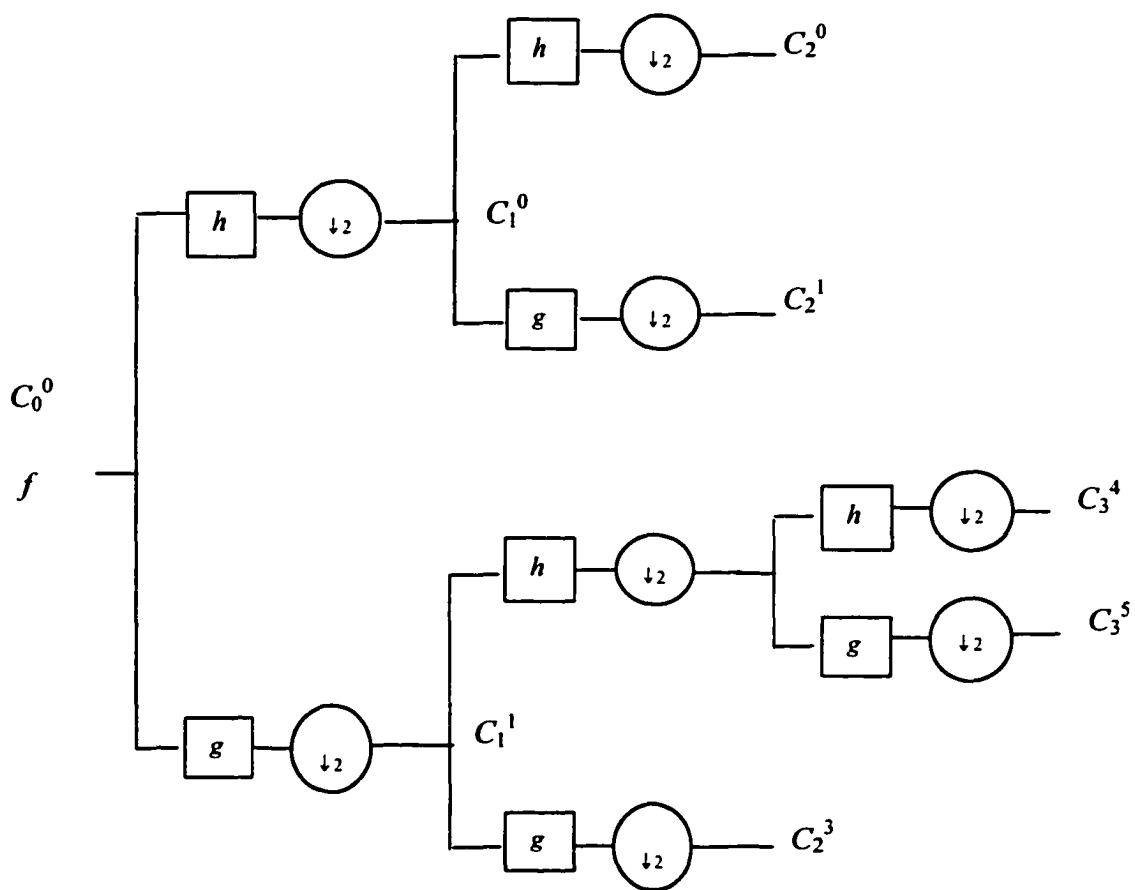


Figure 12 Exemple d'arbre admissible C : les paquets $C_2^0 C_2^1 C_3^4 C_3^5 C_2^3$ constituent les feuilles de l'arbre admissible C

1.7.3 Interprétation en terme de segmentation fréquentielle

La nature demi-bande des filtres h_n et g_n qui correspondent aux filtres QMF (Quadrature Mirror Filter) de la transformée discrète en ondelettes sont utilisés dans la décomposition en paquets d'ondelettes induisant une organisation de l'information selon une segmentation fréquentielle. En effet, à chaque niveau d'échelle de l'arbre C , l'information relative au signal analysé est répartie en bandes de fréquences de largeurs égales, l'ensemble des filtres associés constitue un banc de filtres qui couvre la largeur de bande de tout le signal. La largeur de bande est divisée par deux en passant d'un nœud père à ses deux nœuds fils. Nous pouvons constater et déduire de ce qui précède que les feuilles de n'importe quel arbre admissible (exemple : l'arbre admissible indiqué sur la figure 12) réalisent une partition de l'axe fréquentiel comme l'indique la figure 13.

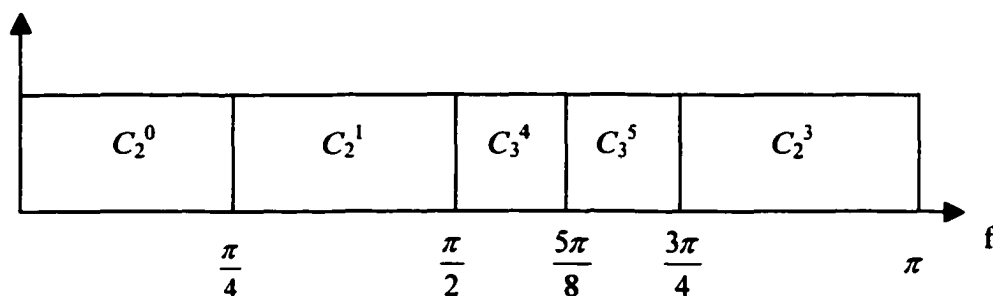


Figure 13 Segmentation fréquentielle obtenue à partir de l'arbre admissible illustré sur la figure 12

1.8 Ondelettes de Daubechies

Daubechies a construit ses fonctions d'échelle φ et d'ondelettes à supports compacts ψ en fonction de leurs ordres de régularité. Ensuite, elle a calculé pour chaque ordre donné, les coefficients des filtres h et g associés à ces fonctions. Elle a démontré aussi que pour que la fonction d'échelle et d'ondelette mères soient régulières et ceci pour au moins jusqu'à l'ordre N , il fallait satisfaire une contrainte principale qui exige que le zéro de la réponse fréquentielle du filtre h soit d'ordre $N \geq 1$ à la pulsation $\omega = \pi$. Ceci peut s'écrire comme suit [5] :

$$H(\omega) = \left[\frac{1 + e^{j\omega}}{2} \right]^N P(\omega) \quad (1.52)$$

où $P(\omega)$ est un polynôme de degrés fini (support compact). Daubechies a combiné l'expression 1.52 avec l'équation d'orthonormalité des fonctions d'échelle φ définie telle que :

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 = 2 \quad (1.53)$$

Par conséquent Daubechies a calculé pour chaque ordre de régularité N les coefficients du filtre passe-bas h et a déduit les coefficients du filtre passe-haut g en se basant sur l'équation 1.54 qui exprime la relation entre les filtres miroirs en quadrature tel que défini par :

$$g_n = (-1)^n h_{1-n} \quad (1.54)$$

Nous présentons sur les figures 14 et 15, pour différents ordres de régularité $N = \{2, 4, 6, 8, 20\}$, les spectres des filtres h et g associés aux fonctions d'échelles et d'ondelettes de Daubechies.

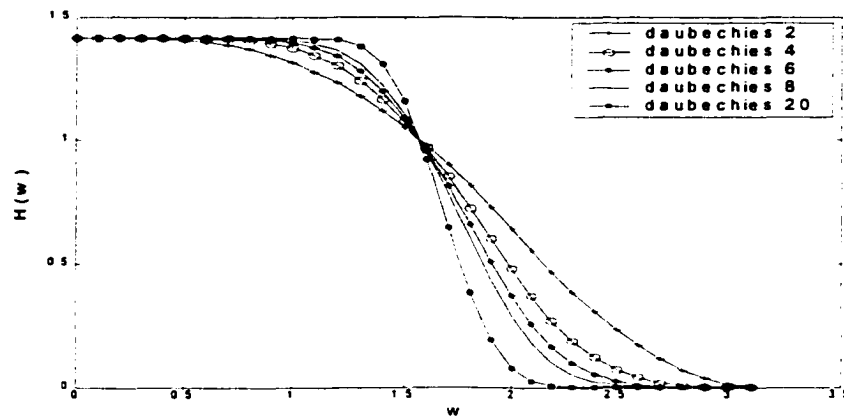


Figure 14 Spectres des réponses impulsionnelles des filtres passe-bas h pour l'ordre de Daubechies 2, 4, 6, et 8

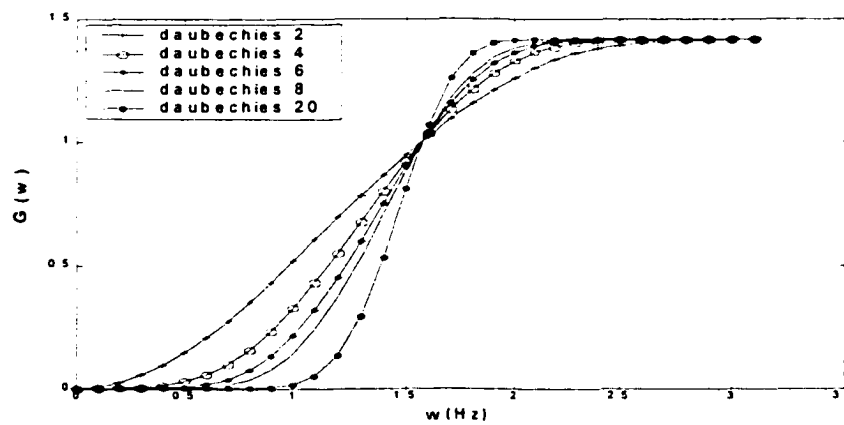


Figure 15 Spectres des réponses impulsionnelles des filtres passe-haut g pour l'ordre de Daubechies 2, 4, 6, et 8

1.9 Méthode de sélection de la meilleure base d'ondelettes

Indépendamment des structures des paquets d'ondelettes qui existent, extraire une base d'ondelettes se traduit implicitement par la sélection d'un ensemble de fonctions de base ayant la signification de segmentation fréquentielle. La méthode de la sélection d'une meilleure base d'ondelettes a été introduite initialement par Coifman et Wickerhauser [4], elle se base fondamentalement sur deux principaux facteurs : le premier détermine le choix de la stratégie de recherche sur laquelle s'appuie la sélection de la base optimale, le deuxième détermine le critère de sélection utilisé pour celui-ci. Ce critère est souvent adapté aux objectifs du traitement, par exemple, dans un cadre de classification, il aura la fonction d'optimiser la représentation des signaux de parole analysés et de discriminer les classes des arbres admissibles. Cette méthode s'applique indifféremment aux structures des paquets d'ondelettes utilisés. L'idée fondamentale de la méthode de sélection de la meilleure base d'ondelettes est de chercher la base optimale sur laquelle l'information est la plus concentrée. Le critère utilisé pour ce traitement est souvent caractérisé par une fonction additive qui mesure ce degré de concentration. L'algorithme de sélection de la meilleure base s'explique comme suit : à partir d'une décomposition complète en paquets d'ondelettes, la recherche de la base optimale se fait en partant du bas vers le haut de l'arbre admissible, alors en comparant la somme des entropies (fonctions additives) de deux sous bases possibles (nœuds fils), avec l'entropie de la sous base correspondant à leur nœud père. Si l'entropie résultante de ces deux nœuds fils est supérieure à celle de leur nœud père, alors on élimine les nœuds fils, et on garde leur nœud père, sinon on garde les nœuds fils.

Dans un cadre d'une classification, l'essentiel objectif est de chercher la base de représentation du signal traité qui maximisera la séparabilité entre les classes d'arbres d'apprentissage. Par conséquent, le critère utilisé ici doit être capable d'évaluer le pouvoir de discriminer chacun des nœuds représentant une sous-base optimale, et cela vis à vis des classes à séparer. Saito et Coifman proposent une mesure d'entropie relative

à ce traitement, qui correspond à l'entropie de Shannon [21], elle permet de mesurer le degrés de concentration du signal analysé dans chacune de ses bandes de fréquences. Si $x(n)$ correspond aux échantillons d'un nœud donné de l'arbre, alors son coût suivant le critère d'entropie de Shannon normalisé est défini par :

$$E(x) = -\sum p_i \log(p_i) \quad (1.55)$$

$$\text{où } p_i = \frac{x_i^2}{\|x\|} \quad (1.56)$$

1.10 Conclusion

Ce chapitre nous a permis de présenter la différence entre la transformée de Fourier et la transformée en ondelettes, en particulier en termes de résolution temps-fréquence. Nous avons constaté que les résolutions temporelle et fréquentielle des ondelettes sont variables, alors qu'elles sont constantes pour les fenêtres d'analyse de la transformée de Fourier. Par conséquent, la localisation temps-fréquence des signaux non stationnaires sera mieux représentée par analyse en ondelettes. On a vu que le calcul de la transformée en ondelettes discrète se fait d'une manière récursive et cela en se basant sur l'algorithme pyramidal proposé par S. Mallat. L'analyse des signaux en paquets d'ondelettes offre une possibilité supplémentaire à l'analyse multirésolution, car les paquets d'ondelettes permettent d'étendre et de généraliser la construction de nouvelles bases à partir de tous les sous-espaces de détail. Ce type d'analyse offre une grande souplesse pour l'utilisateur et lui permet de s'adapter au signal à analyser. Il conduit de même à une décomposition en sous-bandes de fréquences du signal. La méthode de sélection de la meilleure base d'ondelettes a pour but d'obtenir la meilleure structure d'arbre, c'est à dire celle qui minimise au mieux le coût de représentation des signaux. De plus elle devrait permettre de déterminer les bandes de fréquences où le signal est le plus représentatif et le plus concentré. Le chapitre suivant passe en revue les principales méthodes utilisées, mais relative à notre travail de recherche, pour extraire les paramètres acoustiques d'un signal de parole.

CHAPITRE 2

QUELQUES MÉTHODES D'EXTRACTION DES PARAMÈTRES ACOUSTIQUES

2.1 Introduction

Dans ce chapitre, nous allons présenter quelques méthodes de paramétrisation des signaux de parole. Nous commencerons par donner un bref aperçu sur la méthode MFCC [8](Mel Frequency Cepstral Coefficients), qui fait usage à la transformée de Fourier, et à un banc de filtres dont la répartition des fréquences centrales est réglée à l'échelonnement de Mel. Par la suite nous présenterons l'application de la méthode MFDWC [7] (Mel Frequency Discret Wavelet Coefficients), basée sur la transformée en ondelettes, et faisant usage aux paquets d'ondelettes à structure d'arbre fixe qui approxime l'échelonnement de Mel. La troisième méthode de paramétrisation utilise la méthode de sélection de la meilleure base d'ondelettes.

2.2 MFCC

La méthode MFCC retient les caractéristiques perspectives grâce aux bancs de filtres qui donnent la distribution de l'énergie liée à chaque bande de fréquence. La fréquence centrale et la largeur de bande de chaque filtre sont réglées suivant l'échelonnement de Mel. Ce dernier a modélisé les bandes critiques et de perceptions de l'oreille. Dans ce qui suit nous montrerons les principaux blocs du schème de prétraitement et de paramétrisation de la méthode MFCC.

2.2.1 Filtre de préaccentuation

Étant donné que les sons voisés (hautes fréquences) chutent de 6db/octave [3], ces derniers sont mal représentés par rapport aux fréquences basses. Pour les accentuer, on applique un filtre de préaccentuation souvent non récursif du premier ordre et dont la fonction de transfert est tel que :

$$H(z) = 1 - \alpha z^{-1}, \quad (2.1)$$

avec le coefficient de préaccentuation α généralement borné entre 0.9 et 1.

2.2.2 Fenêtres de Hamming

Quoique les signaux de parole sont des signaux non stationnaire, toutefois ils peuvent l'être dans des intervalles de temps bien limités, d'où la nécessité d'utiliser les fenêtres d'analyse qui divisent ainsi le signal de la parole en segments successifs et stationnaires. Les durées D_f de ces fenêtres sont semblables. Elles permettent d'extraire les vecteurs paramètres du signal (voir figure 16). La période P_f et la durée de recouvrement D_r entre fenêtres sont fixées à travers tout le signal. Chaque segment est multiplié par une fonction d'analyse $z(n)$, qui pourra avoir la forme triangulaire, ou rectangulaire, ou autres formes. Dans ce sujet, nous avons utilisé la fonction de Hamming telle que :

$$z(n) = \left\{ 0.54 - 0.46 \cos\left(\frac{2\pi(n-1)}{N-1}\right) \right\}, \quad 0 \leq n \leq N \quad (2.2)$$

La fonction analysante de Hamming permet d'atténuer les discontinuités au niveau des extrémités des segments. Le nombre total des fenêtres n_f est exprimé (2.3) en fonction de la durée du signal D_s , de la durée de recouvrement entre fenêtres D_r , et de la durée de la fenêtre d'analyse D_f .

$$n_f = \frac{D_s - D_r}{D_f - D_r} \quad (2.3)$$

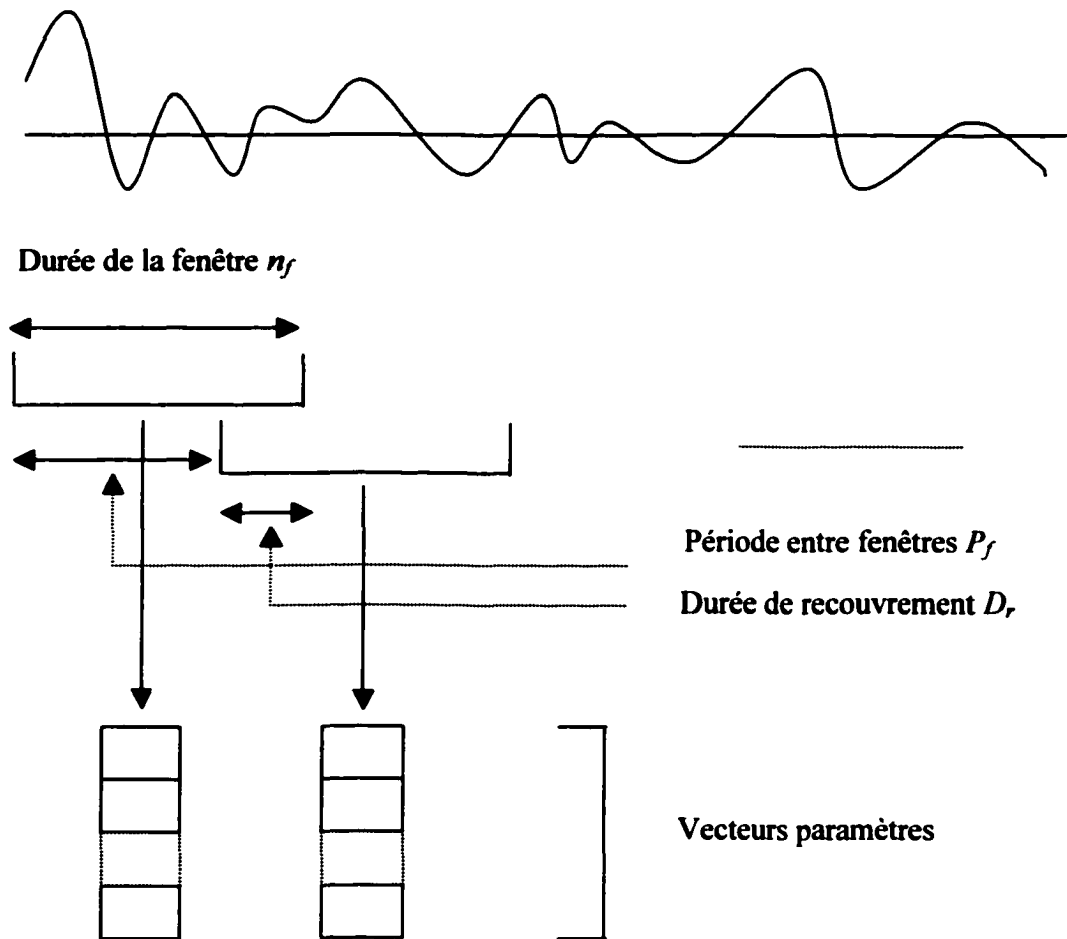


Figure 16 Fenêtres d'analyse et vecteurs paramètres

2.2.3 Coefficients MFCC

Dans ce sujet, les coefficients MFCC sont utilisés comme des paramètres d'apprentissage et de test pour la reconnaissance du locuteur. La figure 17 illustre le schème de pré-traitement et de paramétrisation de la méthode MFCC.

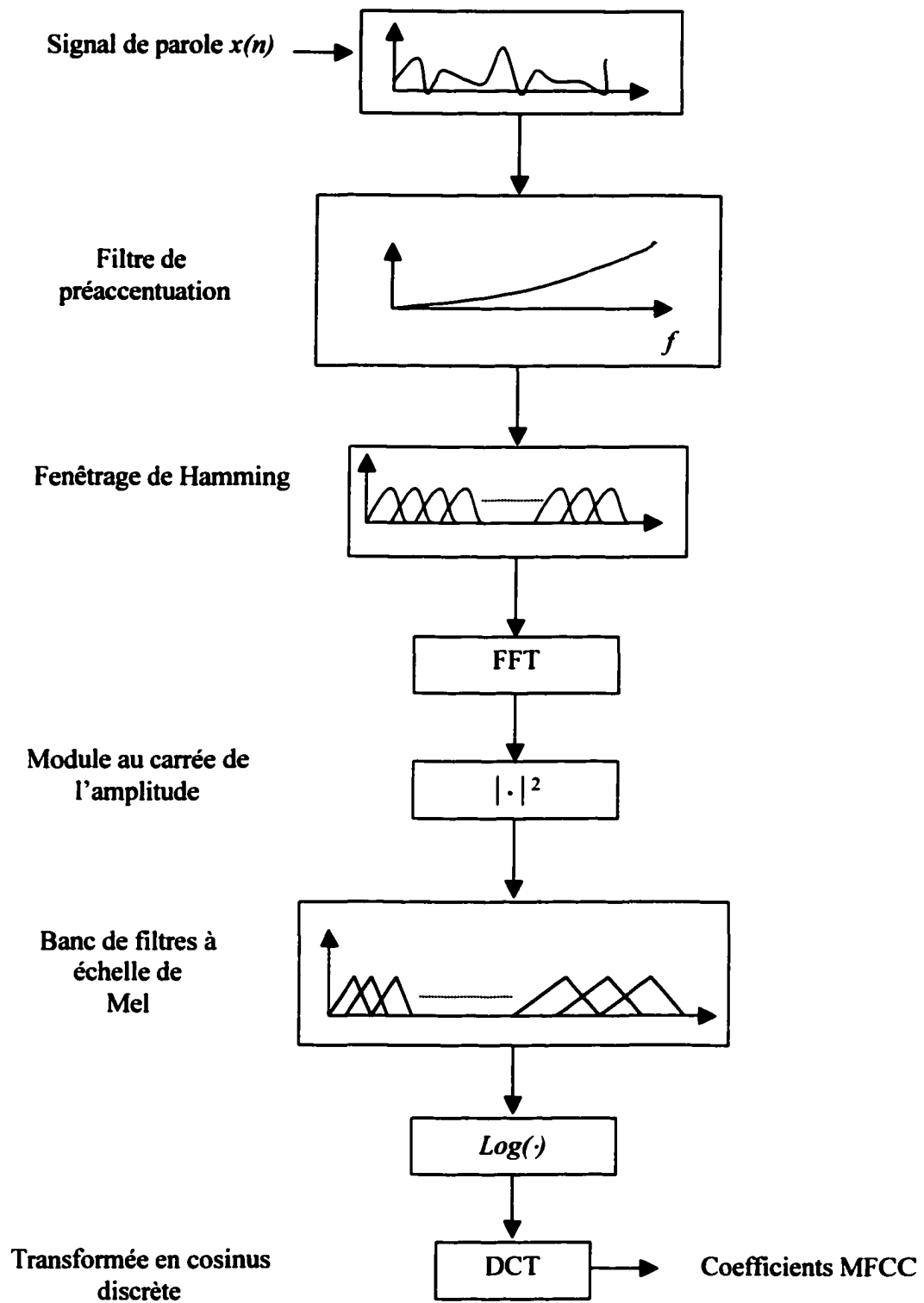


Figure 17 Schème d'extraction des coefficients MFCC

Pour chaque fenêtre d'analyse, on calcule l'amplitude de son spectre, obtenu par la transformée de Fourier rapide FFT (Fast Fourier Transform), et on retient son module au carré. Le résultat passe à travers un banc de filtres. Ces derniers sont triangulaires, uniformément espacés selon l'échelle de Mel [3]. L'échelle de Mel utilisée est :

$$Mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (2.4)$$

tel que les largeurs de bandes sont linéaires en dessous de 1000 Hz et logarithmique au delà de 1000 Hz (figure 18).

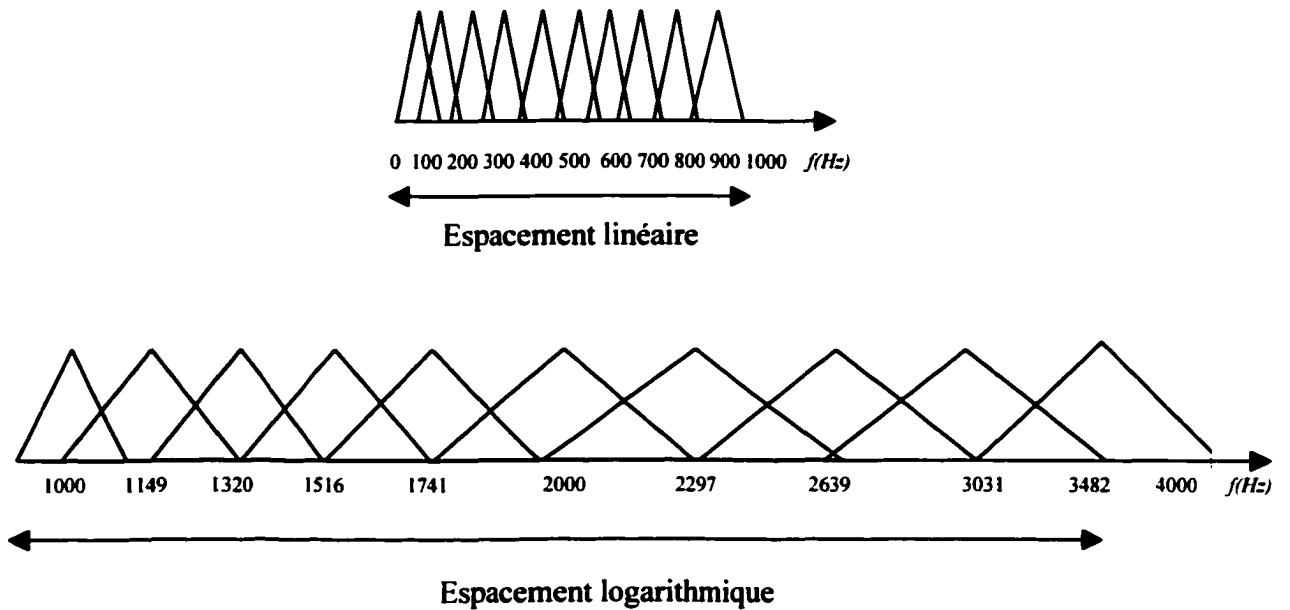


Figure 18 Banc de filtres à l'échelle de Mel

Les coefficients MFCC sont définis comme étant les résultats de la transformée de Fourier inverse appliquée au vecteur des logarithmes des énergies des signaux qui correspondent aux sorties du banc des filtres. Ce vecteur sera noté V_E . Les coefficients MFCC peuvent être calculés en utilisant la transformée en cosinus discrète appliquée au vecteur V_E . Si on note nb_f comme étant le nombre de filtres du banc, et nb_p comme

étant le nombre de paramètres ou de coefficients MFCC, alors l'expression (2.5) calcule ces derniers tel que :

$$MFCC(i) = \sqrt{\frac{2}{nbf}} \sum_{j=1}^{mf} V_{\epsilon}(j) \cos\left(i \left(j - \frac{1}{2}\right) \frac{\pi}{nbf}\right), \quad i = 1, \dots, nbp \quad (2.5)$$

2.3 MFDWC

2.3.1 Introduction

La méthode MFCC a été largement utilisée dans les systèmes de reconnaissance. Elle présente cependant un inconvénient majeur, celui de l'uniformité de sa résolution dans le plan temps-fréquence. En effet, elle a la difficulté à détecter les brusques variations qui se produisent à l'intérieure d'un signal lent. Quoique la méthode MFDWC est récente, toutefois, il n'en demeure pas moins intéressante, vu qu'elle utilise la transformée en ondelettes. Ces dernières bénéficient de la propriété de localisation en temps et en fréquence. Elles ont la propriété de se contracter et de se dilater, leur permettant ainsi d'avoir des résolutions variables dans le plan temps-fréquence. En se contractant, elles détectent mieux les hautes fréquences, et mieux les basses fréquences en se dilatant. Farooq et Datta [7] ont proposé une famille de bases d'ondelettes d'un arbre admissible. Ce dernier permet d'avoir une segmentation fréquentielle qui approxime l'échelonnement de Mel, utilisé dans la méthode MFCC. Cet arbre admissible est interprété par un banc de filtres lequel ses ondelettes associées sont des ondelettes à support compact de Daubechies d'ordre 6 [5].

2.3.2 Arbre admissible approximant l'échelle de Mel

L'arbre admissible proposé par [7] est illustré sur la figure 20, cette dernière montre que les décompositions successives en paquet d'ondelettes permettent de segmenter la bande de fréquences du signal de la parole de manière à approximer l'échelonnement de Mel.

La première décomposition complète en paquets d'ondelettes à trois niveaux d'échelle divise l'axe des fréquences en huit bandes, telles que chacune a une largeur de 500 Hz. Les trois premières bandes du niveau trois qui sont respectivement; 0-500 Hz, 500-1000 Hz, et 1000-1500 Hz représentant les basses fréquences, sont chacune à nouveau décomposées en deux niveaux d'échelle, ainsi on obtient au cinquième niveau, douze bandes de même largeurs égales à 125 Hz. Ces bandes approximent les neuf premières bandes de l'échelle de Mel qui ont les mêmes largeurs égales à 100 Hz, et aussi celles de largeurs 124 Hz, 160 Hz, et 184 Hz du banc de filtres de Mel. Les bandes du troisième niveau d'échelle de l'arbre, qui sont; 1.5-2 KHz, 2-2.5 KHz, et 2.5-3Khz sont chacune décomposées en un seul niveau d'échelle, par conséquent, on obtient au quatrième niveau, six bandes de même largeurs égales à 250 Hz, ces dernières correspondent approximativement aux bandes de largeurs égales respectivement à 211Hz, 242Hz, 278Hz, 320 Hz, 367 Hz, et 422 Hz du banc de filtre de Mel. Les deux dernières bandes du niveau trois qui sont; 3-3.5 KHz, et 3.5-4 KHz ne sont plus décomposées, elles correspondent approximativement aux bandes de largeurs égales respectivement à 484 KHz et 556 KHz du banc de filtres de Mel. L'arbre admissible de la figure 20 donne au total 20 bandes de fréquences qui correspondent à ses nœuds feuilles.

La décomposition d'un nœud donné de l'arbre, noté C_j^p , tel que j est le niveau d'échelle ou se trouve C_j^p dans l'arbre, et p sa position à cet échelle, donne ses deux nœuds fils droite et gauche, respectivement C_{j+1}^{2p} et C_{j+1}^{2p+1} . Ces derniers sont déduits par l'algorithme pyramidal de Mallat [13], et on a les expressions suivantes telles que :

$$C_{j+1}^{2p}(n) = \langle f, \psi_{j+1}^{2p}(t - 2^{j+1}n) \rangle = \downarrow 2[C_j^p * h](n) \quad (2.6)$$

$$C_{j+1}^{2p+1}(n) = \langle f, \psi_{j+1}^{2p+1}(t - 2^{j+1}n) \rangle = \downarrow 2[C_j^p * g](n) \quad (2.7)$$

Les coefficients $C_{j+1}^{2p}(n)$ correspondent à la projection du signal d'entrée f dans le sous-espace W_{j+1}^{2p} dont ψ_{j+1}^{2p} est une base orthonormée. L'équation (2.6) montre que le calcul

des $C_{j+1}^{2p}(n)$ peut être interprété comme le filtrage de la séquence $C_j^p(n)$ par le filtre de réponse impulsionnelle h suivi par un sous-échantillonnage de rapport 2. Par contre les coefficients $C_{j+1}^{2p+1}(n)$ correspondent à la projection du signal d'entrée f dans le sous-espace W_{j+1}^{2p+1} dont ψ_{j+1}^{2p+1} est une base orthonormée. L'équation (2.7) montre que le calcul des $C_{j+1}^{2p+1}(n)$ peut être interprété comme le filtrage de la séquence $C_j^p(n)$ par le filtre de réponse impulsionnelle g suivi par un sous-échantillonnage de rapport 2. La figure 19 montre la décomposition du nœud C_j^p suivant un seul niveau d'échelle.

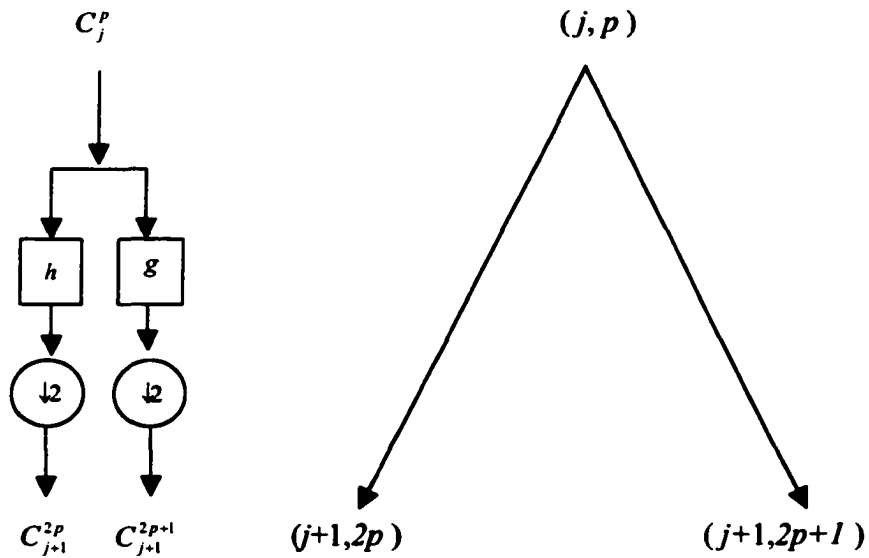


Figure 19 Décomposition du nœud $C_j^p(n)$ d'un seul niveau d'échelle

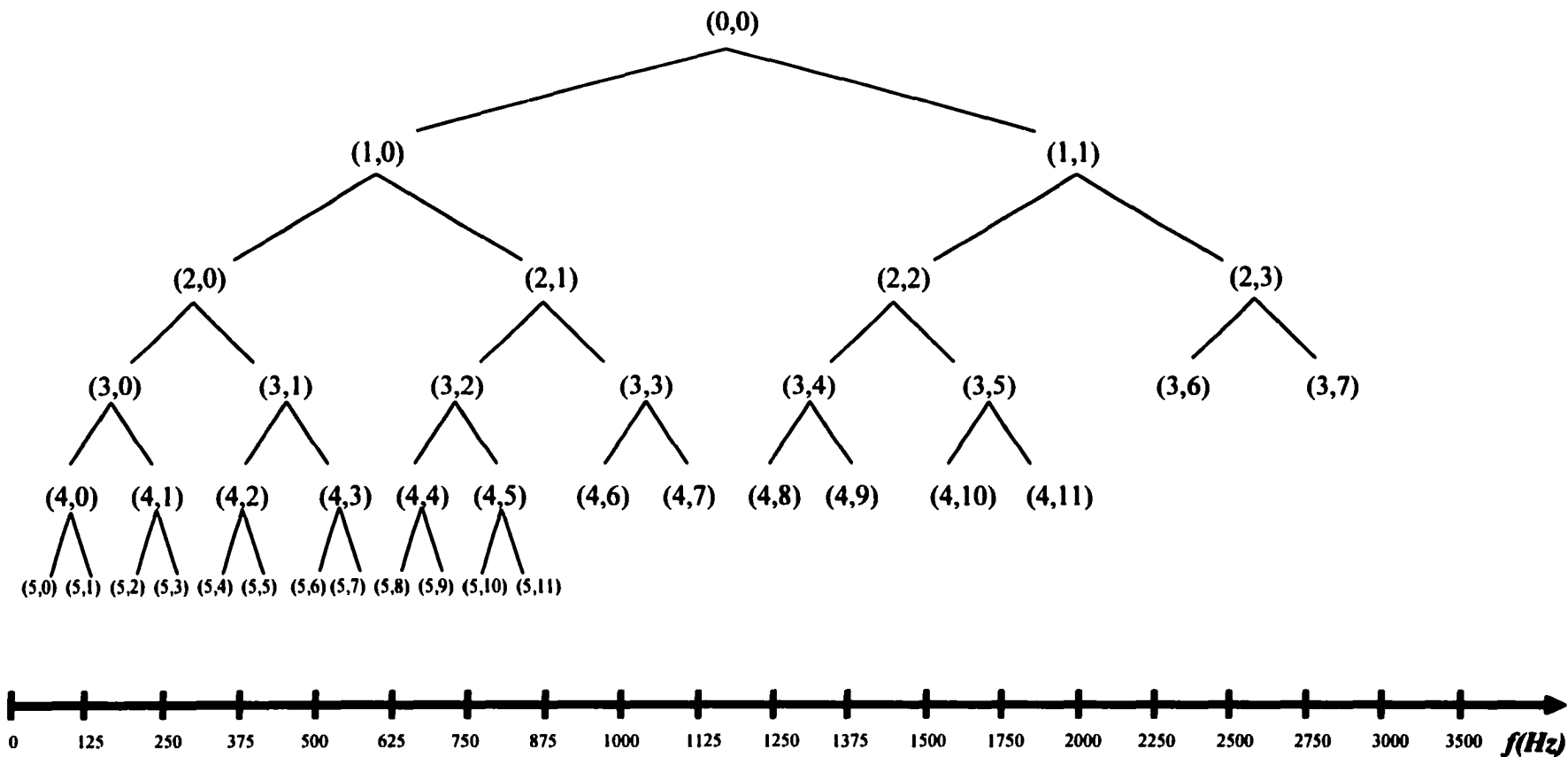


Figure 20 Arbre admissible proposé par Farooq [7] approximant l'échelle de Mel

2.3.3 Coefficients MFDWC

Les coefficients MFDWC sont définis de la même manière que ceux de la méthode MFCC, et cela en étant les résultats de la transformée en cosinus discrète appliquée au vecteur V_E , mais cette fois ci, les éléments du vecteur V_E représentent les logarithmes des énergies des signaux qui correspondent aux nœuds feuilles de l'arbre admissible indiqué sur la figure 20. Le schème d'extraction des coefficients où des paramètres MFDWC est présenté sur la figure 21 ci-dessous.

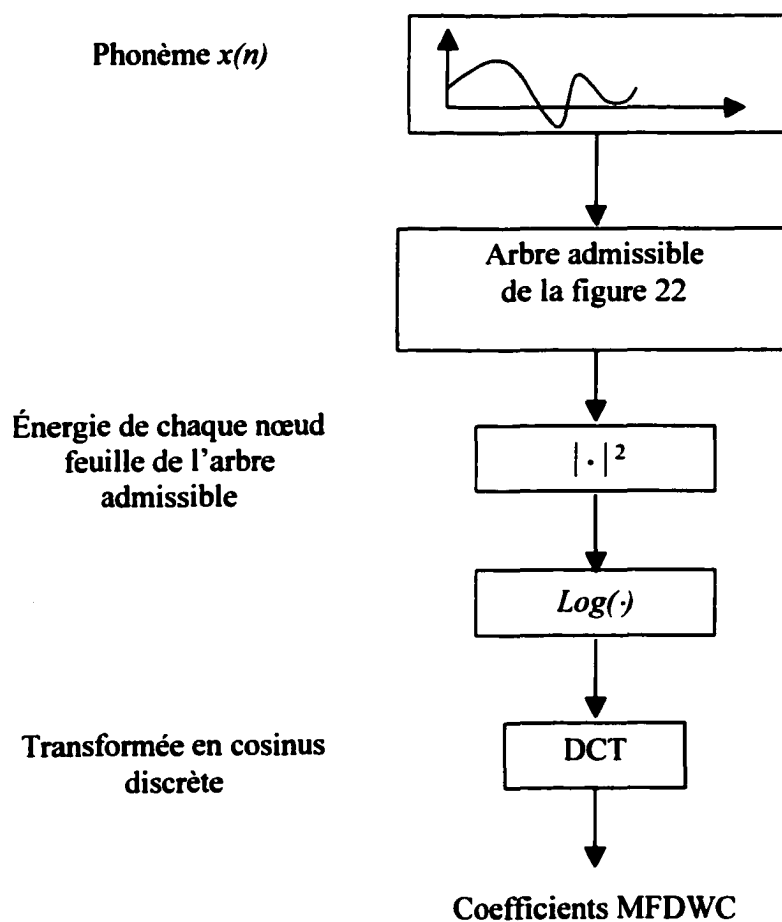


Figure 21 Schème d'extraction des coefficients MFDWC

2.4 Méthode basée sur la sélection de la meilleure base

2.4.1 Introduction

Afin de pouvoir offrir une représentation paramétrique plus efficace des signaux de parole, quelques systèmes de reconnaissance - exemple celui de Long et Datta [10] -, font usage de la méthode de sélection de la meilleure base d'ondelettes. Elle a été introduite initialement par Coifman [4]. En effet, dans un contexte de classification, l'objectif principal est de trouver la meilleure base de représentation des signaux de parole, qui maximise au mieux la séparabilité entre les classes d'apprentissage. La méthode de sélection de la meilleure base d'ondelettes permet entre autres de construire une librairie de structures d'arbres (bases orthonormales d'ondelettes) à partir d'une collection de signaux, qui sont dans notre cas, soit les signaux qui correspondent aux fenêtres d'analyses temporelles propre à chaque phrase prononcée par locuteur, soit les signaux correspondants aux phrases complètes prononcées par locuteur. L'apport principal et fondamental de la méthode de sélection de la meilleure base d'ondelettes, est d'obtenir la segmentation fréquentielle qui représente au mieux la distribution de l'énergie dans toute les bandes de fréquences du signal. En d'autres termes, cette méthode nous permettra de trouver la base sur laquelle l'information de l'énergie est la plus concentrée. Le critère d'entropie est un critère additif. Il est défini pour mesurer ce degré de concentration. Cette méthode se base sur la recherche d'un optimum global à partir d'une stratégie de recherche locale. Elle présente un caractère de généralité car elle définit cette stratégie de recherche qui est en fait applicable à n'importe quel critère d'entropie ayant une mesure additive. Dans le cadre de la classification, la méthode de sélection de la meilleure base d'ondelettes aura donc, la fonction de discriminer les classes d'arbres des paquets d'ondelettes pour chaque locuteur. L'application de l'algorithme à un signal donné fournira la meilleure représentation de ce signal au sens d'un critère d'entropie. Dans ce qui suit, nous présenterons la stratégie sur laquelle est basée la méthode de sélection de la meilleure base d'ondelettes, en l'occurrence la

stratégie de recherche d'optimum global à partir d'une stratégie de recherche locale. Par la suite, nous présenterons brièvement un critère d'entropie lié aux problèmes de classification.

2.4.2 Stratégie locale de recherche d'un optimum global

Nous désignons par E , un critère d'entropie additif, x un vecteur dans un sous-espace V , B une base orthogonale de V , et x_B la séquence de projection de x dans B . La meilleure base du vecteur $x \in V$ relativement à l'entropie E est définie comme étant la base pour laquelle $E(x_B)$ est optimal. Notons que selon la signification du critère d'entropie, son optimalité peut être considérée comme étant, soit sa minimisation ou soit sa maximisation. A partir d'une décomposition complète en paquets d'ondelettes de profondeur maximale J , la recherche est effectuée du bas vers le haut de l'arbre admissible à structure complète, en comparant l'entropie de deux sous-bases possibles, une qui correspond à la sous-base d'un nœud père, et l'autre correspondante à la résultante de l'union de ses deux nœuds fils. Si l'entropie est minimale au niveau du nœud père, on garde celui-ci, sinon on le remplace par ses deux nœuds fils. Dans ce qui suit, nous désignerons l'entropie d'un nœud C_j^p de l'arbre par E_j^p . Ainsi nous pouvons résumer l'algorithme de la façon suivante :

_ $\{C_j^p\}_{p=0, \dots, 2^j-1}$ sont les nœuds feuilles de l'arbre admissible qui correspondent aux meilleures bases initiales.

_ En partant du niveau d'échelle $j = J - 1$ jusqu'à $j=0$, et pour chaque nœud de ce niveau $p=0$ à $2^j - 1$, la sous-base W_j^p qui correspond au nœud C_j^p est sélectionnée comme étant la meilleure sous-base, si et seulement si on a :

$$E_j^p \leq E_{j+1}^{2p} + E_{j+1}^{2p+1} \quad (2.8)$$

sinon la sous-base W_j^p est remplacée par l'union de ses deux sous-bases fils W_{j+1}^{2p} et W_{j+1}^{2p+1} , qui sont dans ce cas sélectionnées comme étant deux meilleures sous-bases.

Deux raisons fondamentales font que l'algorithme de sélection de la meilleure base d'ondelettes converge vers l'obtention d'un arbre admissible optimale [4]. La première raison réside du fait que l'on garde exclusivement comme sous-base les deux nœuds fils ou leur père, la deuxième est que l'on débute l'algorithme en partant du bas vers le haut de l'arbre initial, et ceci en prenant comme base initiale toutes les feuilles du dernier niveau d'échelle.

Dans ce qui suit, nous illustrerons l'algorithme sur les figures 22, 23 et 24. Nous prenons comme exemple, un arbre complet de 3 niveaux d'échelle, tel que chacun de ses nœuds est matérialisé par la valeur de son entropie correspondante. Nous commençons l'algorithme en marquant les feuilles du dernier niveau d'échelle avec des cercles (voir figure 22).

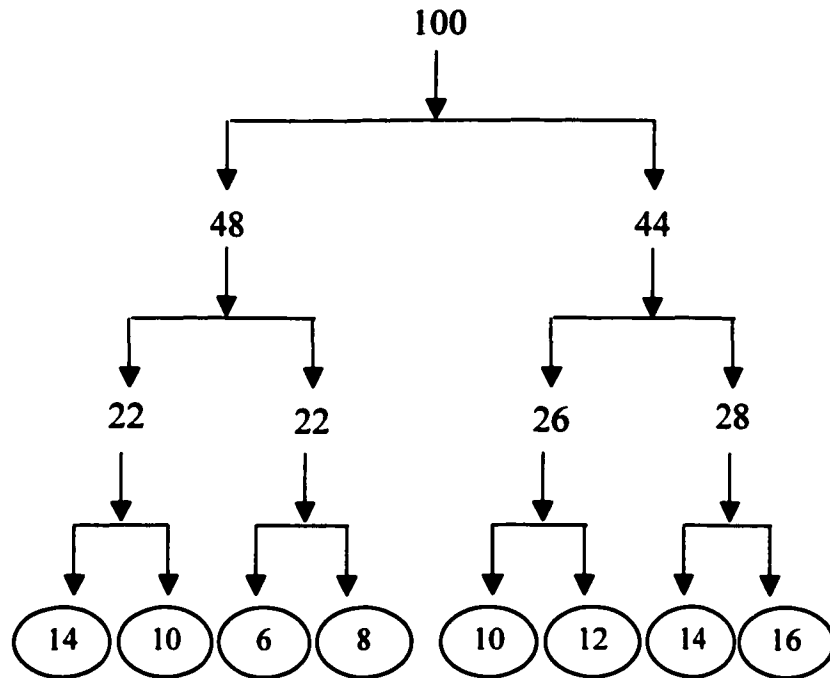


Figure 22 Entropies des nœuds et des feuilles marquées en cercle de l'arbre initial

La meilleure base d'ondelettes initiale correspond aux feuilles du dernier niveau d'échelle de l'arbre complet (figure 22). Ainsi en commençant l'algorithme du bas vers le haut de l'arbre, on compare l'entropie de chaque nœud aux entropies de ses deux nœuds fils. Si le nœud père a une entropie inférieure à la somme des entropies de ses deux nœuds fils, on marque leur père par un cercle l'indiquant ainsi comme une feuille. Sinon on ne marque pas le nœud père par un cercle (figure 23) et on garde ses nœuds fils. L'algorithme converge car l'arbre est de profondeur finie, ainsi la meilleure base sera définie par les nœuds marqués par des cercles (figure 24). Rappelons que l'on choisisse exclusivement le nœud père ou ses deux nœud fils. Par conséquent, si un nœud père est choisi comme feuille, alors ses deux nœuds fils et leurs descendants ne figureront plus dans l'arbre final (figure 24). Le nombre de nœuds du nouveau arbre retourné par l'algorithme, sera toujours inférieur ou égale au nombre de nœuds de l'arbre initial.

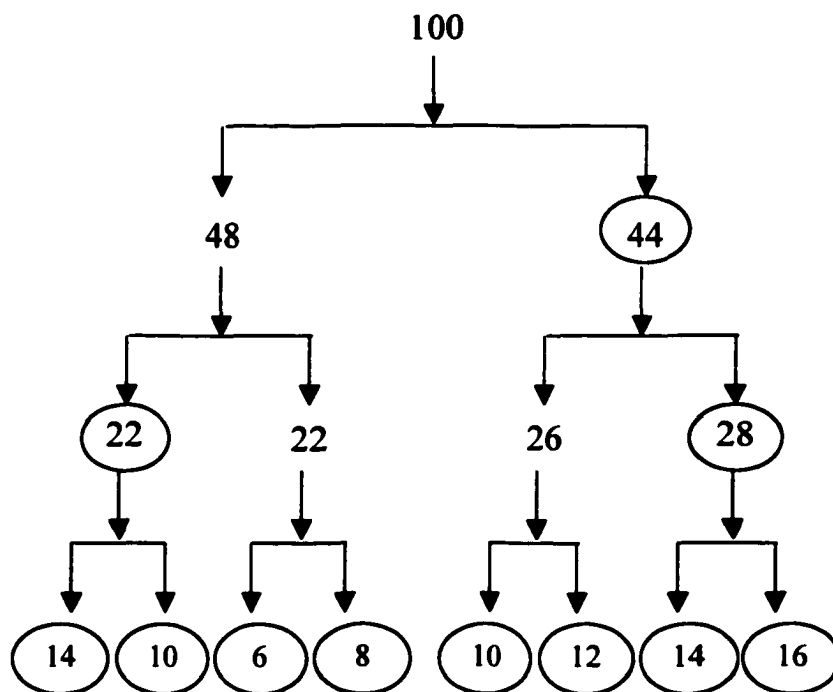


Figure 23 Sélection des nouveaux nœuds feuilles retournés par l'algorithme

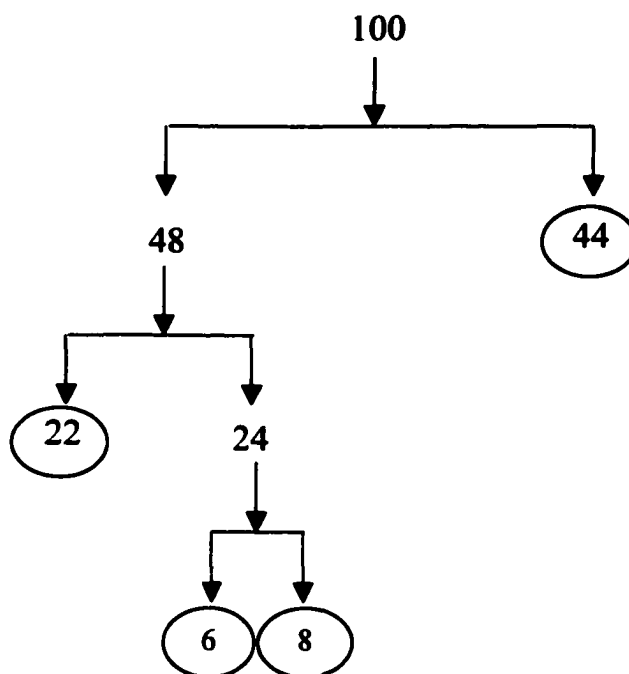


Figure 24 Feuilles finales du nouveau arbre optimal

2.4.3 Critère d'entropie

Dans un contexte de classification, l'objectif fondamental est de chercher la base de représentation optimale qui maximise la séparabilité entre les classes d'apprentissage. Par conséquent le choix du critère d'entropie doit donc être en mesure de discriminer les classes d'arbres admissibles à séparer. Saito et Coifman [4] ont proposé une mesure d'entropie, en l'occurrence l'entropie de Shannon normalisé, ce critère est appliqué à chacun des nœuds C_j^p de l'arbre. Sachant que E_j^p représente l'entropie du nœud C_j^p , l'entropie de Shannon normalisé de ce nœud est exprimé sous la forme suivante :

$$E_j^p = -\sum \alpha_i \text{Log}(\alpha_i) \quad (2.9)$$

avec

$$\alpha_i = \frac{|C_j^p(i)|^2}{\|C_j^p\|} \quad (2.10)$$

2.4.4 Extraction des paramètres

Le schème d'extraction des coefficients paramètres utilisé ici, est similaire à celui de la méthode MFDWC (figure 21), à l'exception d'un seul bloc qui est inséré dans ce schème. Celui-ci traite l'algorithme de sélection de la meilleure base d'ondelettes. Cet algorithme nous permet d'avoir l'arbre admissible optimal à partir d'un arbre à structure complète. Les paramètres sont les résultats de la transformée en cosinus discrète appliquée au vecteur V_E , dont les éléments représentent les logarithmes des énergies des signaux qui correspondent aux nœuds feuilles de l'arbre admissible optimal retourné par l'algorithme de sélection de la meilleure base.

2.5 Conclusion

En utilisant indépendamment chacune des trois méthodes, en l'occurrence, la méthode MFCC, la méthode MFDWC, et celle basée sur l'algorithme de sélection de la meilleure base d'ondelettes, nous pourrions étudier et analyser l'apport de chacune de ces trois méthodes en terme de performances du système de vérification du locuteur (voir chapitre 5).

Dans le prochain chapitre, on présente l'application qui nous intéresse à savoir la vérification du locuteur (VL). On décrira brièvement les différentes techniques utilisées pour concevoir un système de VL.

CHAPITRE 3

MODÉLISATION ET RECONNAISSANCE DU LOCUTEUR

3.1 Introduction

La reconnaissance du locuteur permet de reconnaître l'identité de la personne qui parle, à partir de son empreinte vocale. Il existe deux types fondamentaux de systèmes de reconnaissance du locuteur : Les systèmes de vérification du locuteur et les systèmes d'identification du locuteur. La vérification du locuteur consiste à vérifier l'adéquation du message vocal dont la référence acoustique appartient au locuteur, ainsi, le système de vérification du locuteur fait une décision par tout ou rien, soit qu'il accepte la personne qui a parlé, soit qu'il la rejette. Par contre l'identification du locuteur permet au système de comparer un message vocal prononcé par une personne, avec un ensemble de références acoustiques correspondantes à plusieurs locuteurs. Le système d'identification du locuteur doit alors identifier la personne qui a parlé.

On distingue deux types de textes prononcés par les locuteurs : Les textes dépendants, ils sont déterminés et imposés par le système. La personne doit impérativement prononcer le même texte qui lui est attribué pour qu'il soit reconnu par le système. La vérification du locuteur utilise souvent ce genre de textes. Par exemple si un client voudrait être autorisé par le système à faire une transaction bancaire par téléphone, il est supposé être coopératif et doit prononcer son mot de passe. Les textes indépendants sont souvent utilisés dans les systèmes d'identification du locuteur. La personne à reconnaître, prononce des phrases qui ne sont pas obligatoirement dans la base de données du système, ou qui ne sont pas des phrases à partir desquelles le système a construit la référence et le modèle du locuteur correspondant. Par exemple, les textes indépendants sont utilisés pour identifier l'identité d'un locuteur, à partir d'une conversation téléphonique.

Dans les systèmes de vérification du locuteur, lorsqu'une personne inconnue prononce une phrase de test, les paramètres de cette dernière sont comparés avec le modèle d'un locuteur connu par le système, si le résultat de la comparaison est supérieur à un certain seuil θ , la personne est acceptée par le système, sinon elle est rejetée. Ce seuil θ est déterminé par le modèle global de tous les locuteurs connus par le système. Quand le seuil est élevé, cela rend difficile aux imposteurs d'être acceptés, mais le risque de rejeter la personne devient important. Contrairement à ce que nous venons de dire, si le seuil est faible, l'acceptation de la personne par le système devient plus aisée, sauf que le risque d'accepter les imposteurs augmente. Le seuil peut être ajusté de manière à augmenter les fausses acceptations (FA) tout en diminuant les faux rejets (FR), et vice-versa. La courbe de variation des fausses acceptations en fonction des faux rejets est similaire à celle présentée sur la figure 25. Plus la distance D indiquée sur la figure 25, est petite plus le système de vérification du locuteur est performant.

Dans nos travaux, le modèle global a été obtenu à partir de toutes les données d'entraînement de tous les locuteurs connus par le système, ainsi, il couvre tous les espaces acoustiques de tous les locuteurs. Pour que le système de vérification du locuteur ne soit pas vulnérable aux imposteurs, il est préférable que le modèle global soit estimé à partir d'une base de données d'entraînement assez large.

Le modèle global ainsi que le modèle de chaque locuteur utilisés dans le cadre de ce sujet, sont des modèles de Markov cachés. Nous allons donner dans ce qui suit, un aperçu sur ce modèle, ensuite nous allons présenter l'étape d'estimation des paramètres du modèle de Markov caché lors de la phase d'apprentissage, et par la suite nous présenterons l'algorithme de viterbi utilisé dans la phase de reconnaissance.

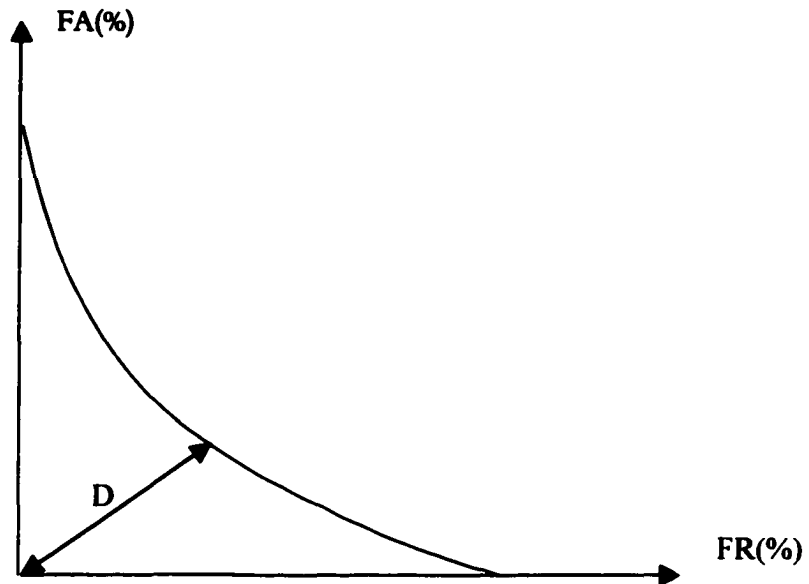


Figure 25 Courbe de variation des FA en fonction des FR

3.2 Estimation du modèle du locuteur et du modèle global

En vérification du locuteur, les modèles des locuteurs ainsi que le modèle global permettent de rendre compte la définition de mesures de ressemblance entre ces derniers et une observation. Cette mesure est obtenue en calculant la fonction de vraisemblance $P(O|\lambda)$ tel que λ représente soit le modèle du locuteur ou le modèle global et O représente une observation. Cette fonction peut s'exprimer sous la forme d'un modèle paramétrique tel que le Modèle de Markov Caché. La détermination des paramètres du modèle λ passe le plus souvent par une étape d'estimation consistant à déterminer le modèle optimale λ_{op} , c'est-à-dire celui qui maximise la fonction de vraisemblance $P(O|\lambda)$ pour un ensemble de phrases d'apprentissage prononcées par un locuteur dans le cas du modèle d'un locuteur ou bien pour un ensemble de phrases

d'apprentissage prononcées par tous les locuteurs dans le cas du modèle global. Dans certains cas, on utilise le critère du Maximum de Vraisemblance (MV) [18] :

$$\lambda_{op} = \arg \max_{\lambda} P(O | \lambda) \quad (3.1)$$

et dans d'autres, le critère du Maximum a Posteriori (MAP) [18] :

$$\lambda_{op} = \arg \max_{\lambda} P(O | \lambda) \cdot P(\lambda) \quad (3.2)$$

Quel que soit le critère considéré (MV ou MAP), l'estimation du modèle optimal λ_{op} s'effectue par l'intermédiaire de l'algorithme EM (Expectation-Maximization), qui fournit une solution correspondant à un maximum global de la fonction de vraisemblance.

3.3 Règle de décision et seuil

Pour un système de vérification du locuteur, la règle de décision se pose sous forme d'un test binaire, celui-ci consiste à décider si l'observation appartient au locuteur dont le modèle est $\lambda = \lambda_l$ ou si elle ne lui appartient pas (c'est-à-dire qu'elle appartient à un imposteur ou à un autre locuteur), c'est le cas où $\lambda \neq \lambda_l$. La décision est du type acceptation ou rejet. Théoriquement, elle peut être prise dans le cadre de la décision Bayésienne en calculant le rapport de probabilité R_{λ} et en le comparant à un seuil de décision S tel que [9] :

$$R_{\lambda}(O) = \frac{P(O | \lambda = \lambda_l)}{P(O | \lambda \neq \lambda_l)} \quad (3.3)$$

Si $R_{\lambda}(O) \geq S$, il s'agit d'une acceptation, par contre si $R_{\lambda}(O) < S$, la décision prise est un rejet. Le seuil de décision optimal S dépend aux conditions de fonctionnement du système par rapport aux fausses acceptations et aux faux rejets. En pratique, Matsui et Furui [15] ont proposé une règle de décision qui dépend d'un seuil optimal $\Theta(\lambda_g)$ ajusté en fonction du modèle global λ_g . Dans ce cas, le problème de l'estimation du seuil optimal $\Theta(\lambda_g)$, peut être interprété comme une normalisation du rapport de probabilité R_{λ} qui ramènerait le seuil de décision optimal $\Theta(\lambda_g)$ sur le seuil théorique S . Il s'avère

que cette deuxième approche donne de bons résultats lorsque le modèle global est correctement estimé. Ce dernier est réestimé dans la phase d'apprentissage à partir de toutes les données d'entraînement de tous les locuteurs connus par le système.

3.4 Modèles de Markov cachés

Parmi les techniques de modélisation qu'on développe dans la théorie et qu'on applique dans la pratique on trouve le modèle stochastique dit modèle de Markov caché (Hidden Markov Model, HMM) [19]. La reconnaissance du locuteur moyennant ce modèle se base sur la mesure de la probabilité d'avoir une observation (vecteur de paramètres) sachant le modèle du locuteur concerné. L'utilisation des modèles de Markov cachés dans notre cas se fait au niveau phonémique. Le modèle HMM est défini par l'ensemble (Ne, No, A, B, π) tel que :

e : séquence d'états, $e=[e_1, e_2, \dots, e_T]$ aux instants $t=1, 2, \dots, T$.

Ne : Nombre d'états dans le modèle.

No : Nombre de symboles d'observation.

A : Matrice des probabilités de transition entre états, $A=\{a_{ij}\}$.

$$a_{ij} = p[e_{t+1} = j | e_t = i] \quad 1 \leq i, j \leq Ne \quad \forall t \quad (3.4)$$

a_{ij} : probabilité qu'on soit à l'état j à l'instant $t+1$ sachant qu'on était à l'état i à l'instant t .

B : Matrice des probabilités d'observation, $B=\{b_j(O_t=k)\}$.

$$b_j(O_t = k) = p[O_t = k | e_t = j], \quad 1 \leq j \leq Ne, \quad 1 \leq k \leq No \quad (3.5)$$

$b_j(O_t = k)$: probabilité qu'on ait l'observation k à l'instant t , sachant qu'on est à l'état j .

π : Distribution des probabilités initiales, $\pi=\{\pi_j\}$.

$$\pi_j = p[e_1 = j], \quad 1 \leq j \leq Ne \quad (3.6)$$

π_j : probabilité qu'à l'instant initial $t=1$ on soit à l'état j .

Le modèle HMM est une machine à états fini qui change d'état i à l'instant t à un autre état j à l'instant $t+1$. À chaque état j on a une observation $O_t = k$ générée avec une probabilité $b_j(O_t = k)$. De la même manière, la transition d'un état i à un autre j est une probabilité gouvernée par une probabilité discrète a_{ij} . La figure 26 montre un modèle de Markov à 6 états. Il est à noter que dans notre cas, tous les modèles de tous les locuteurs ont la même structure que celle indiquée dans la figure 26. Les états 1 et 6 ne produisent aucune observation. Nous avons utilisé tout au long de nos travaux des modèles phonémiques pour chaque locuteur. Un état de chaque modèle est composé de 8 mixtures tel que chacun de ces derniers a son propre poids qui lui ait attribué.

La figure 27 nous montre la représentation de ces 8 mixtures. Les différents poids c_{jm} sont équiprobables. Comme nous avons 8 mixtures, nous aurons la valeur de chacun de leur poids égale à 0.125. Notons que le nombre de streams utilisé est égale à 2 : un représentant les paramètres statiques et l'autre représentant les paramètres dynamiques. Cependant la multiplicité de streams n'influence ni l'étape de l'estimation initiale ni celle de la re-estimation itérative. Cela est dû au fait que les streams sont statistiquement indépendants.

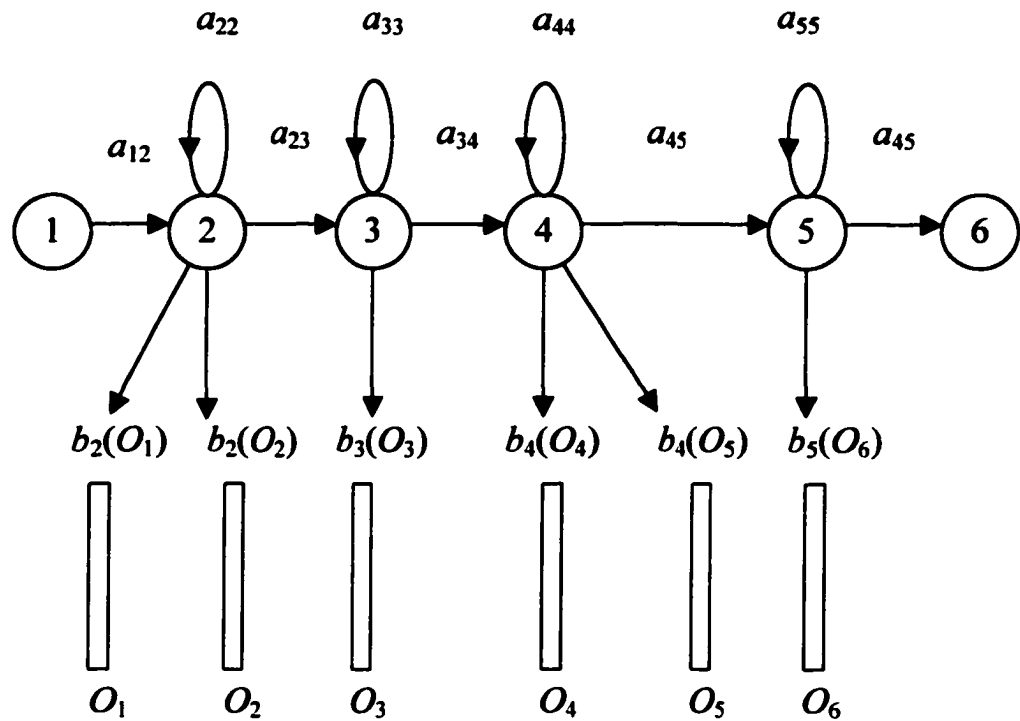


Figure 26 Modèle Markov Caché à six états [25]

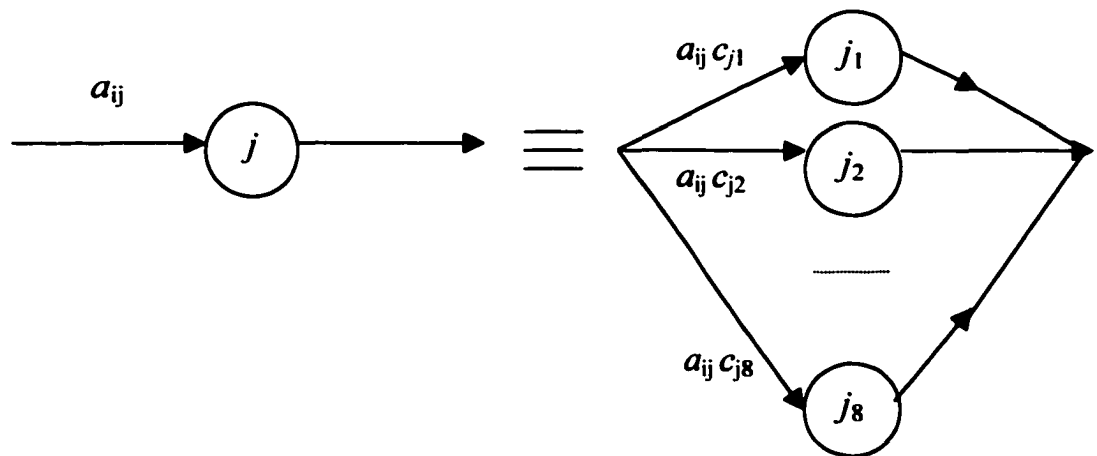


Figure 27 Représentation des mixtures multiples par état [25]

Les vecteurs d'observations émis par chaque état suit une distribution de probabilité continue. Dans notre cas, cette distribution suit à une loi normale, l'équation (3.7) [25] donne l'expression de la probabilité d'observer une séquence à l'état j à l'instant t .

$$b_j(O_t) = \prod_{s=1}^S \left[\sum_{m=1}^{M_s} C_{jm} N(O_{ts}, \mu_{jm}, \Sigma_{jm}) \right]^{\gamma_s} \quad (3.7)$$

M_s : Nombre de mixtures par état ($M_s=8$).

C_{jm} : Poids de chaque mixture ($C_{jm}=0.125$).

γ_s : Poids d'un stream ($\gamma_s = 1$).

S : Nombre de stream par état ($S = 2$).

$N(O_{ts}, \mu_{jm}, \Sigma_{jm})$: Fonction de densité de probabilité continue et multivariées avec un vecteur moyen et une matrice de covariance, son expression est donnée par l'équation (3.8) [25].

$$N(o, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^r |\Sigma|}} e^{-\frac{1}{2}(o-\mu)^T \Sigma^{-1}(o-\mu)} \quad (3.8)$$

3.5 Probabilité d'observation d'une séquence sachant un modèle

Un modèle de Markov caché génère une séquence d'observation. Il est complètement défini par l'ensemble $\lambda=\{A,B,\pi\}$ [19]. La génération des séquences commence par le choix de l'état initial suivant une distribution π_i , puis génère une ou plusieurs observations au niveau de chaque état j à chaque instant t suivant une probabilité $b_j(O_t = k)$ jusqu'à l'instant final $t=T$. Pour avoir un excellent modèle de chaque phonème, l'ajustement adéquat de chaque distribution probabilistique est primordial. Nous allons maintenant voir comment peut-on calculer la probabilité d'avoir la séquence d'observation sachant un modèle λ . Soit la séquence d'observation $O = [O_1 O_2 O_3 \dots O_t$

$O_{t+1} O_{t+2} \dots O_T$]. La probabilité qu'on ait la séquence d'observation $O_1 O_2 O_3 \dots O_t$ à l'instant t et qu'on soit à l'état j sachant un modèle λ est définie par :

$$\alpha_j(t) = p[O_1 O_2 \dots O_t, e_t = j | \lambda], 1 \leq j \leq Ne, 1 \leq t \leq T \quad (3.9)$$

La probabilité d'avoir la séquence d'observation $O_{t+1} O_{t+2} \dots O_T$ à partir de l'instant $t+1$ et qu'on soit à l'état j sachant un modèle λ est définie par :

$$\beta_j(t) = p[O_{t+1} O_{t+2} \dots O_T, e_t = j | \lambda], 1 \leq j \leq Ne, 1 \leq t \leq T \quad (3.10)$$

La probabilité d'avoir la séquence d'observation $O = [O_1 O_2 O_3 \dots O_t \dots O_{t+1} O_{t+2} \dots O_T]$ sachant un modèle λ est définie par :

$$p(O | \lambda) = \sum_{j=1}^{N_e} \alpha_j(t) \beta_j(t) \quad (3.11)$$

la probabilité «avant» $\alpha_j(t)$ se calcule d'une façon récursive :

$$\alpha_j(t) = \left[\sum_{i=2}^{N_e-1} \alpha_i(t-1) a_{ij} \right] b_j(O_t) \quad (3.12)$$

avec l'état initial tel que : $\alpha_i(1) = 1$ (3.13)

la probabilité «arrière» $\beta_j(t)$ se calcule d'une façon récursive :

$$\beta_j(t) = \sum_{i=2}^{N_e-1} \beta_i(t+1) a_{ij} b_i(O_{t+1}) \quad (3.14)$$

avec l'état initial tel que : $\beta_j(T) = a_{jN_e} = 1$ (3.15)

3.6 Réestimation des paramètres d'un modèle pour l'apprentissage

La réestimation des paramètres d'un modèle propre à un locuteur, permet d'effectuer la phase d'apprentissage au système de reconnaissance requis. Cela se fait en ajustant d'une manière itérative les paramètres du modèle du locuteur concerné jusqu'à ce que l'on maximise la probabilité que le système reconnaisse le locuteur sachant son modèle. L'algorithme EM (Expectation Maximisation) est souvent utilisé pour la réestimation

des modèles HMM. Il est basée sur la méthode de Baum-Welch. Supposons qu'on a une séquence de vecteurs paramètres X des signaux de parole d'apprentissage d'un locuteur. L'algorithme EM (voir figure 28) commence avec un modèle initial λ_0 , puis il estime le nouveau modèle λ_1 tel que si $p(X|\lambda_1) \geq p(X|\lambda_0)$, alors le modèle λ_1 devient le modèle du locuteur. Ce traitement est répété pour différents $\lambda_{1=2\dots op}$, jusqu'à ce que l'on atteigne la convergence de λ vers λ_{op} , tel que ce dernier représente le point d'inflexion (maximum global) de $p(X|\lambda)$.

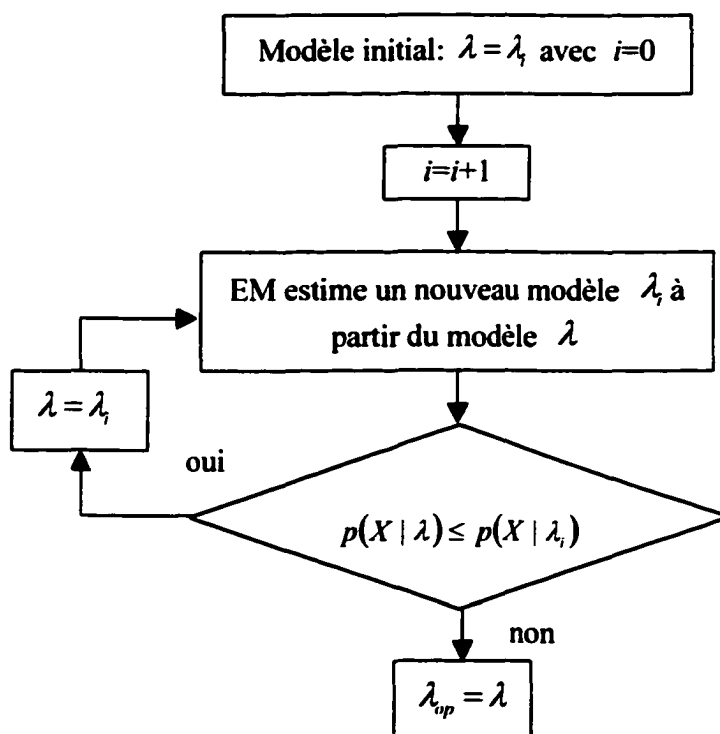


Figure 28 Algorithme itératif de la méthode EM

La réestimation des paramètres du modèle de Markov caché pour chaque locuteur, se fait en utilisant la formule de réestimation de Baum-Welch [2]. Notons que l'utilisation de "streams multiples" n'influence pas le calcul des paramètres puisque les "streams" sont considérés statistiquement indépendants. Le poids c_{jm} de chacune des mixtures

(voir figure 27) représente la probabilité de transition de chacune d'elles. Chaque modèle λ est caractérisé par deux principaux paramètres, l'un représentant le vecteur moyen μ_j et l'autre la matrice de covariance Σ_j tel que [25] :

$$\bar{\mu}_j = \frac{\sum_{t=1}^T L_j(t) o_t}{\sum_{t=1}^T L_j(t)} \quad (3.16)$$

et

$$\bar{\Sigma}_j = \frac{\sum_{t=1}^T L_j(t) (o_t - \mu_j)(o_t - \mu_j)^T}{\sum_{t=1}^T L_j(t)} \quad (3.17)$$

avec

$$L_j(t) = \frac{\alpha_j(t) \beta_j(t)}{p(o_t / \lambda)} \quad (3.18)$$

Le paramètre o_t représente la séquence d'observation générée par le modèle à l'instant t . Cette séquence représente en d'autre terme un vecteur paramètre d'apprentissage du signal de parole d'un locuteur donné.

3.7 Algorithme de Viterbi pour la reconnaissance d'une séquence

L'algorithme de Viterbi [25] est appliqué dans la phase de reconnaissance, afin de trouver la séquence d'états $e=[e_1, e_2, \dots, e_T]$ la plus probable ayant engendré la séquence d'observation $O=[O_1, O_2, \dots, O_T]$. A chaque instant t , l'algorithme calcule par récurrence la fonction $V_j(t)$ pour chacun des états e_t telle que :

$$V_j(t) = \max_j [p(e_1, e_2, \dots, e_{t-1}, e_t = j, O_1, O_2, \dots, O_t)] \quad (3.19)$$

Le maximum étant calculé sur toute les séquences d'états possibles $e=[e_1, e_2, \dots, e_{t-1}]$.

A l'instant initial $t=1$ on a :

$$V_j(1) = \pi_j b_j(O_1) \quad (3.20)$$

Etant à l'instant $t-1$, le calcul de $V_j(t-1)$ est effectué pour chacun des N_e états. Par récurrence on a alors :

$$V_k(t) = \max_j [V_j(t-1) \cdot a_{jk} \cdot b_k(O_t)] \quad (3.21)$$

L'état j est l'état tel que $V_j(t-1) \cdot a_{jk} \cdot b_k(O_t)$ est maximum, car c'est l'état le plus probable à l'instant $t-1$. Pour chacun des états k , on calcule $V_k(t)$; chacun de ces états a un prédécesseur $w_k(t)$, ce dernier servira à retrouver la séquence d'état la plus probable ayant engendré la séquence d'observation $O = [O_1 O_2 \dots O_T]$. L'état e_T retenu à l'instant T est celui pour lequel $V_j(T)$ est maximum. $V_j(T)$ représente la probabilité maximale pour laquelle le modèle ait généré la séquence d'observation. On peut retrouver la séquence des états en retrouvant récursivement les prédécesseurs de e_T tel que :

$$e_{T-1} = w_{e_T}(T-1)$$

et récursivement

$$e_{T-2} = w_{e_{T-1}}(T-2)$$

$$e_{T-3} = w_{e_{T-2}}(T-3)$$

...

$$e_2 = w_{e_3}(2)$$

$$e_1 = w_{e_2}(1)$$

La figure 29 représente la sélection d'un chemin dans un treillis entre deux instants successifs $t-1$ et t . La figure 30 montre la reconstruction du chemin correspondant à la séquence d'état optimale. C'est ainsi qu'on peut calculer pour chacun des modèles de Markov de chaque locuteur la probabilité que la séquence observée soit générée par lui (ou elle).

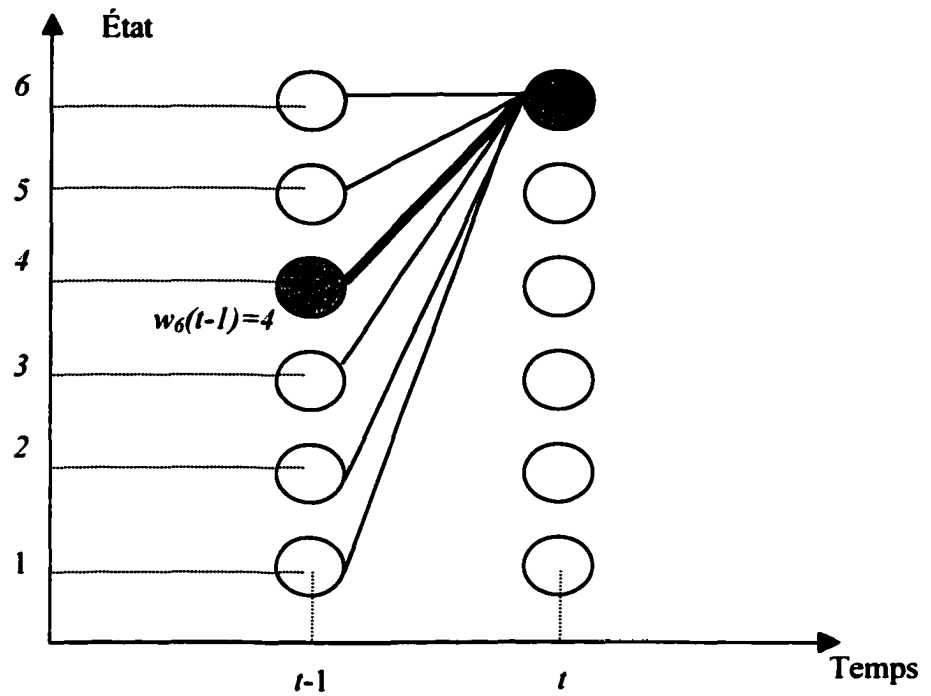


Figure 29 Sélection d'un chemin dans un treillis entre deux instants successifs

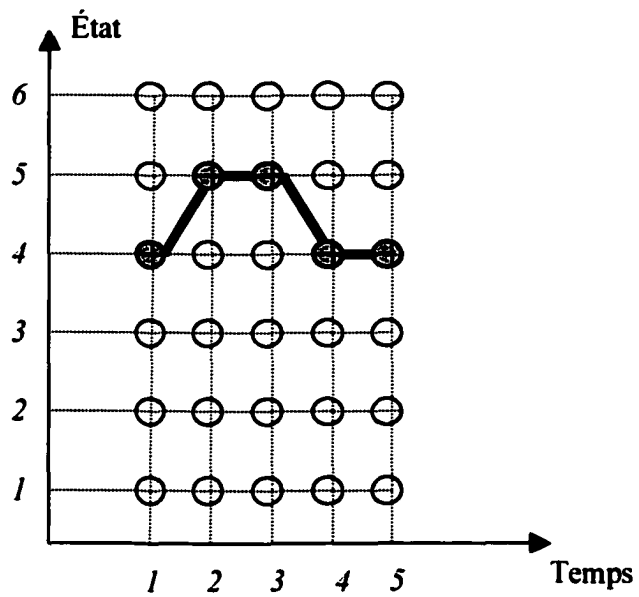


Figure 30 Reconstruction du chemin correspondant à la séquence d'états optimale

3.8 Conclusion

Nous avons montré initialement dans ce chapitre une brève présentation sur les modèles de Markov cachés utilisés dans notre projet. Nous avons vu que l'algorithme EM était appliqué à ces modèles de manière à estimer leurs paramètres dans la phase d'apprentissage, afin d'avoir pour chaque locuteur son modèle optimal. La phase de vérification est basée fondamentalement sur l'algorithme de Viterbi qui calcule pour chaque locuteur la probabilité que la séquence observée soit générée par lui (ou elle).

Dans le prochain chapitre, qui est le cœur de notre travail, on présentera nos différentes et principales contributions au niveau de l'analyse et la paramétrisation du signal de la parole.

CHAPITRE 4

MODIFICATIONS APPORTÉES AUX MÉTHODES EXISTANTES

4.1 Introduction

Dans ce chapitre, nous allons présenter les travaux effectués dans le cadre de ce sujet. Nous commencerons en premier lieu par présenter les différentes modifications que nous avons apporté à la méthode MFDWC. Nous avons suggéré ces modifications dans le but de tester les performances de reconnaissance en fonction de l'ordre de l'ondelette de Daubechies fixe et adapté dans le banc de filtre proposé par Farooq [7]. En deuxième lieu, nous présenterons la seconde partie des travaux que nous avons mené. Nous avons utilisé la méthode de sélection de la meilleure base d'ondelettes, et avons proposé un algorithme de construction du meilleur arbre du locuteur. Nous avons appelé cet algorithme MAL qui signifie (Meilleur Arbre du Locuteur). La deuxième approche permet de construire une librairie d'arbres admissibles. Chacun de ces derniers correspond à un locuteur donné. Cette deuxième étape est basée sur la discrimination des classes d'arbres admissibles d'apprentissage de tous les locuteurs.

4.2 Schème de prétraitement

Nous avons utilisé le schème de prétraitement de la méthode MFDWC (figure 21), dans lequel nous avons inséré un bloc de fenêtrage de Hamming. Pour le bloc représentant la structure d'arbre admissible, nous avons utilisé dans la première partie du travail, la structure d'arbre proposé par Farooq [7]. Dans la seconde partie, cette structure dépend du locuteur concerné.

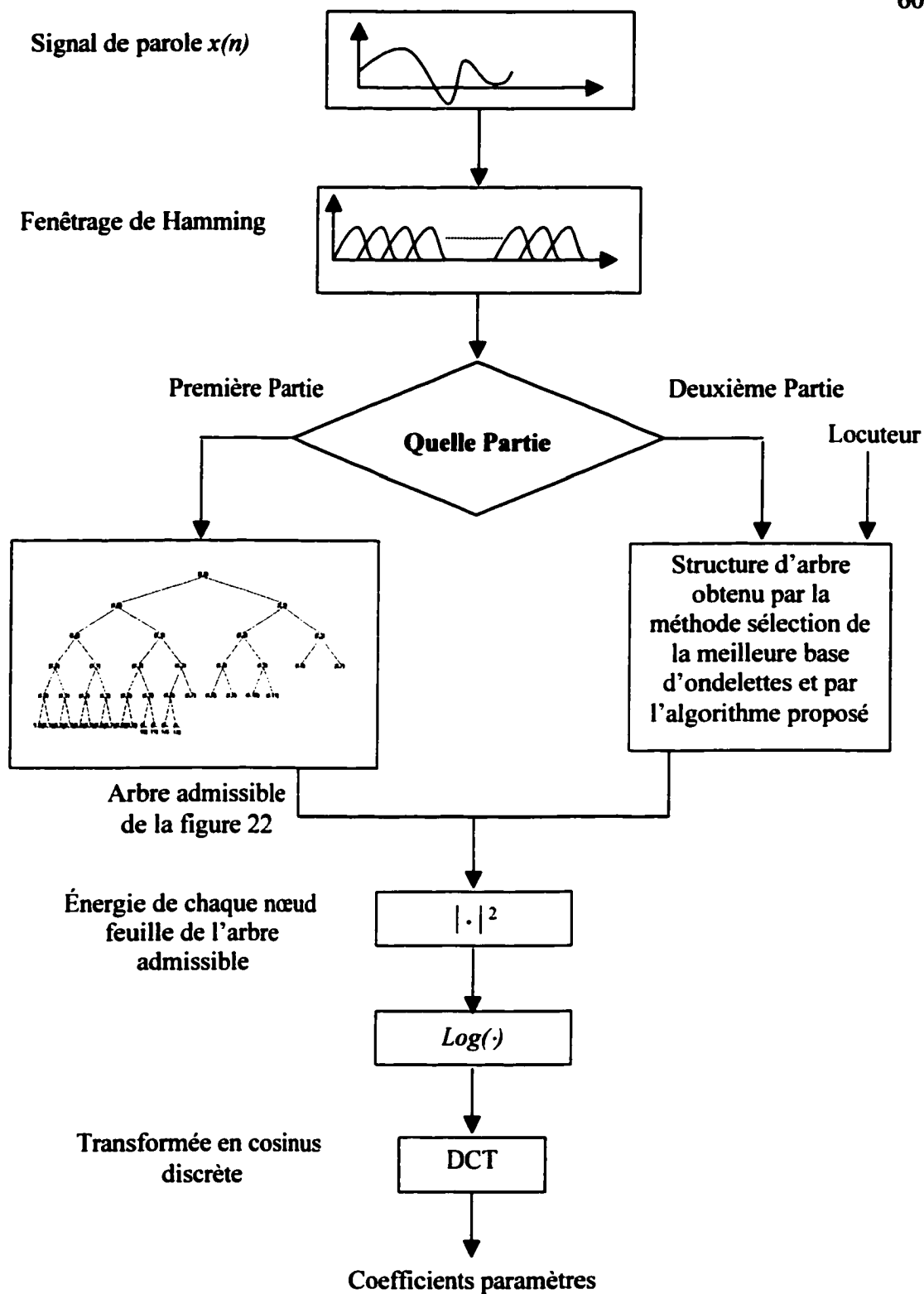


Figure 31 Schéma de prétraitement utilisé dans les deux parties du travail

Elle est obtenue par l'algorithme de sélection de la meilleure base d'ondelettes et par l'algorithme MAL que nous avons proposé. La figure 31 présente le schéma de prétraitement que nous avons utilisé, dans les deux parties de notre travail.

4.3 Première partie du travail

Dans cette étape, nous testerons les performances de reconnaissance en fonction de l'ordre d'ondelette de Daubechies. Cet ordre sera fixe ou adapté au sens d'un critère d'entropie. Le banc de filtre correspondant à l'arbre admissible utilisé est celui proposé par Farooq [7] comme l'indique la figure 20.

4.3.1 Performance de reconnaissance en fonction de l'ordre de Daubechies

Dans cette partie, nous avons testé les performances de reconnaissance en fonction de l'ordre de Daubechies. Comme nous l'avons vu au paragraphe 1.8 du chapitre 1, Daubechies a calculé pour chaque ordre de régularité des fonctions d'échelle et d'ondelettes, les coefficients de leurs filtres associés h et g . À chaque fois que nous fixons un ordre de Daubechies commun pour tous les filtres h et g du banc de filtres de l'arbre admissible proposé par Farooq [7], nous obtenons les vecteurs paramètres et nous testons les performances de reconnaissance. Alors en procédant ainsi, nous avons en premier lieu testé les performances pour 5 locuteurs tout en effectuant nos tests pour chaque ordre $N=2$ jusqu'à $N=10$. En fonction des résultats obtenus, nous avons délimité l'ordre entre un minimum $N_{\min}=2$, et un maximum $N_{\max}=8$. Nous avons procédé de la sorte afin de borner l'ordre adapté tel qu'expliqué au paragraphe 4.3.2. Nous avons ensuite testé les performances sur 60 locuteurs, et cela pour chaque ordre $N=\{2,4,6,8\}$. Le fait de tester les performances sur un nombre plus important de locuteurs (dans notre cas, 60 locuteurs), nous permet de vérifier plus sûrement l'amélioration des performances en fonction de l'ordre de Daubechies.

4.3.2 Performance de reconnaissance en fonction de l'ordre adapté au sens de quelques critères d'entropie

Dans cette partie, l'ordre des filtres de Daubechies ne sera ni fixe ni commun pour tous les filtres h et g . En effet pour chaque nœud C_j^p de l'arbre admissible de la figure 20, exceptés les feuilles, nous procéderons à la recherche de l'ordre adapté au sens d'un critère d'entropie comme suit :

- Étape 0.** Initialisation de l'ordre de Daubechies à l'ordre minimal que nous avons fixé dans le paragraphe 4.3.1.
- Étape 1.** Calcul des coefficients des filtres h et g à l'ordre de Daubechies courant.
- Étape 2.** Décomposition du nœud C_j^p en ses deux nœuds fils C_{j+1}^{2p} et C_{j+1}^{2p-1} .
- Étape 3.** Calcul de la somme des entropies (fonctions additives) des deux nœuds fils.
- Étape 4.** Sauvegarde du résultat de la somme et l'ordre courant en mémoire.
- Étape 5.** Incrémentation de l'ordre courant.
- Étape 6.** Vérification si l'ordre courant est strictement supérieur à l'ordre maximal que nous avons fixé dans le paragraphe 4.3.1
 - Si oui, on passe à l'étape 7.
 - Sinon, on revient à l'étape 1.
- Étape 7.** Recherche parmi les résultats stockés en mémoire à l'étape 4, celui qui est optimal (minimal ou maximal, cela dépend du critère d'entropie), et ressortir l'ordre adapté à ce résultat optimal au sens du critère d'entropie utilisé.
- Étape 8.** Décomposition du nœud C_j^p en ses deux nœuds fils C_{j+1}^{2p} et C_{j+1}^{2p-1} , tel que les coefficients des filtres h et g correspondent à l'ordre adapté obtenu à l'étape 7.
- Étape 9.** Fin

Nous avons commencé par tester les performances de reconnaissance pour l'ordre adapté suivant quatre critères d'entropie, et cela pour 5 locuteurs. Les entropies liés à ces quatre critères sont des fonctions additives, leurs expressions sont définies comme suit :

Sachant que l'on a déjà noté E_j^p l'entropie du nœud C_j^p , alors l'entropie liée au critère d'énergie est définie par :

$$E_j^p = \sum_n |C_j^p(n)|^2 \quad (4.10)$$

Pour le critère d'énergie normalisée, l'entropie s'exprime par :

$$E_j^p = \sum_n \frac{|C_j^p(n)|^2}{\|C_j^p\|} \quad (4.11)$$

Le troisième critère que nous avons utilisé est celui de Shannon non normalisé [4], son entropie est exprimée par :

$$E_j^p = -\sum_n |C_j^p(n)|^2 \log(|C_j^p(n)|^2) \quad (4.12)$$

Finalement le quatrième et dernier critère est celui de Shannon normalisé dont l'expression de son entropie est définie par :

$$E_j^p = -\sum_n \alpha_n \log(\alpha_n) \quad (4.13)$$

$$\alpha_n = \frac{|C_j^p(n)|^2}{\|C_j^p\|}$$

L'ordre d'ondelette de Daubechies adapté au sens du critère d'énergie est calculé pour maximiser l'énergie du signal analysé dans chacune de ses bandes de fréquences. Par contre le calcul de l'ordre adapté au sens du critère de Shannon est basé sur la minimisation de celui-ci, et cela dans chaque sous-bande. En effet, pour chaque bande de fréquences, plus ce critère est minimal plus le signal est concentré et par conséquent mieux représenté. Par la suite, nous avons testé les performances pour 60 locuteurs, avec un ordre adapté au sens du critère d'énergie comme présenté dans le chapitre 5.

4.4 Deuxième partie du travail

Le principal objectif que nous sommes fixé dans la deuxième partie du travail, a été de trouver pour chaque locuteur l'arbre admissible qui réalise la meilleure répartition de l'information dans sa bande de fréquences. En maximisant la séparabilité entre les structures d'arbres admissibles des différents locuteurs, nous pourrions améliorer les performances de reconnaissance.

4.4.1 Algorithme de construction d'arbre admissible d'un locuteur

À partir d'une collection de M signaux S_1, S_2, \dots, S_M (Signaux qui correspondent aux phrases d'entraînement prononcées par un locuteur), nous effectuons pour chacun de ces signaux la décomposition complète en paquets d'ondelettes à 5 niveaux d'échelle, avec des filtres de Daubechies d'ordre $N=8$. Nous nous retrouvons avec M arbres à structures complètes. Nous appliquerons ensuite à chacun de ces derniers l'algorithme de sélection de la meilleure base d'ondelettes en utilisant le critère d'entropie de Shannon non normalisé. Par conséquent nous aurons M arbres admissibles. Nous nous créons ainsi un arbre abstrait initial à structure complète et à 5 niveaux d'échelle. Ainsi nous calculons pour chaque nœud C_j^p de cet arbre abstrait, le nombre de fois où il apparaît dans les M arbres admissibles en tant que feuille. Nous désignons ce nombre par nombre d'occurrences du nœud C_j^p , et nous le noterons O_j^p . L'indice j représente le niveau d'échelle et p la position de ce nœud à l'échelle j . Donc nous nous retrouvons avec un arbre abstrait à structure complète et à 5 niveaux d'échelle, tel que chaque nœud C_j^p a son nombre d'occurrence O_j^p . L'algorithme MAL (Meilleur Arbre du Locuteur) que nous avons proposé ici permet d'extraire un arbre admissible à partir de l'arbre abstrait initial, en se basant sur la sélection des nœuds C_j^p qui ont les nombres d'occurrence les plus importants.

Nous utilisons comme nous l'avons indiqué au chapitre 3, un nombre égale à 12 de paramètres statiques. Ces derniers sont obtenus à partir de la DCT à 13 points. Nous ignorons la première raie de la DCT et prenons comme paramètres les 12 dernières. Ceci justifie que le seuil du nombre de feuilles requis dans un arbre admissible est égal à $L=13$. L'algorithme MAL (Meilleur Arbre du Locuteur) peut être résumé par la table I :

Tableau I

Description de l'algorithme MAL (Meilleur Arbre du Locuteur)

Étape	Description
Étape 0	Initialisation de l'arbre abstrait initial, on attribue à chaque nœud son nombre d'occurrences calculé au préalable.
Étape 1	Sélection de nouvelles feuilles à partir de l'arbre abstrait initial à structure complète. On recherche les nœuds dont les nombres d'occurrences, sont supérieurs à ceux de leurs descendants.
Étape 2	Vérification si le seuil prédéfini ($L=13$) du nombre de paramètres (le nombre qui correspond au nombre de feuilles sélectionnées) est atteint. <ul style="list-style-type: none"> • Si oui, on passe à l'étape 6. • Sinon, on passe à l'étape 3.
Étape 3	Recherche des sous-feuilles parmi les nœuds descendants de quelques ou toutes les feuilles.
Étape 4	Remplacement partiel ou total des feuilles par leurs sous-feuilles, pour atteindre le seuil prédéfini ($L=13$).
Étape 5.	Fin

Le détail des différentes étapes de l'algorithme est donné ci-dessous :

Étape 0. **Étape initiale.**

- $\{C_j^p\}_{\substack{j=0,\dots,5 \\ p=0,\dots,2^j-1}}$ sont les nœuds de l'arbre abstrait initial.
- $\{O_j^p\}_{\substack{j=0,\dots,5 \\ p=0,\dots,2^j-1}}$ sont les nombres d'occurrence de tous les nœuds de

l'arbre abstrait initial.

Le niveau d'échelle $J=5$ est le dernier niveau l'arbre abstrait initial.

Étape 1. **Sélection des nouvelles feuilles à partir de l'arbre abstrait initial à structure complète. La recherche de ces nouvelles feuilles est effectuée comme suit :**

Pour chaque niveau d'échelle j , tel que $j=1$ à $j=J-1$. Tout nœud C_j^p de ce niveau, tel que $p=0$ à $p=2^j-1$, sera considéré comme une feuille si et seulement si les deux conditions suivantes sont vérifiées.

1. Le nombre d'occurrence O_j^p du nœud C_j^p est strictement supérieur à tous les nombres d'occurrences de ses nœuds descendants.
2. Aucun de ses nœuds ascendants n'est une feuille.

Le nœud C_0^0 sera considéré comme feuille si et seulement si la première condition est satisfaite.

Tout nœud C_j^p du dernier niveau d'échelle $j=J$ sera considéré comme feuille si et seulement si la deuxième condition est vérifiée.

Lorsqu'un nœud C_j^p est retenu comme une feuille, ses nœuds descendants sont automatiquement rejetés, par conséquent ils ne figureront plus dans le nouvel arbre.

Remarque : Le fait d'effectuer ce traitement à partir du niveau $j=0$ jusqu'au niveau $j=J$ de l'arbre abstrait initial, et de retenir exclusivement le nœud C_j^p ou ses nœuds descendants, garantit la convergence de l'algorithme vers l'obtention d'un arbre admissible.

Étape 2. Vérification si le seuil prédéfini du nombre de paramètres $L=13$ est atteint.

- Si le nombre de feuilles du nouvel arbre calculé à l'étape 1, est inférieur au seuil fixé $L=13$, alors la DCT donnera des valeurs redondantes, elle est appliquée au vecteur de paramètres tel que chacun de ces derniers représente le logarithme de l'énergie des échantillons d'une feuille. Pour remédier à ce phénomène, il faut alors décomposer à nouveau soit quelques-unes ou toutes les feuilles jusqu'à ce que l'on obtienne au moins 13 feuilles. On passe dans ce cas à l'étape 3.
- Sinon, le nombre de feuilles est suffisant pour l'application de la DCT. Et l'on passe à l'étape 5.

Étape 3. Recherche des sous-feuilles parmi les nœuds descendants de quelques ou toutes les feuilles. Étant donné que le nombre de feuilles est inférieur au seuil fixé $L=13$, nous procédons à la recherche de ces sous-feuilles comme nous allons l'expliquer dans ce qui suit :

Pour toute feuille, exceptées celles qui appartiennent aux niveaux $j=J-1$ et $j=J$, on considère son fils gauche (droit) et les descendants de ce dernier comme étant un sous-arbre, duquel tout nœud sera considéré comme une sous-feuille, si et seulement si les deux conditions de l'étape 1 sont satisfaites. Il en est de même pour le fils droite (gauche). Les fils gauche et droit de chaque feuille du niveau $j=J-1$ sont considérés sans aucune

condition comme étant des sous-feuilles. Par conséquent pour toute feuille, nous obtenons ses sous-feuilles, et l'on passe à l'étape 4.

Étape 4. Remplacement partiel ou total des feuilles par leurs sous-feuilles.

Le nombre de feuilles étant inférieur au seuil requis $L=13$, cela nous oblige à remplacer quelques ou toutes les feuilles par leurs sous-feuilles, de manière à atteindre le seuil requis. Cela s'explique du fait qu'une feuille doit avoir deux sous-feuilles ou plus. Par conséquent, en remplaçant une feuille par ses sous-feuilles, on augmente par un ou plus le nombre de paramètres. La procédure du remplacement des feuilles par leurs sous-feuilles se fait comme suit :

Étape 4.1. Remplacement de la feuille dont le nombre d'occurrences est le plus proche de la moyenne des nombres d'occurrences de ses sous-feuilles, par celles-ci. Passer à l'étape 4.2.

Étape 4.2. Vérification si le seuil prédéfini du nombre de paramètre est atteint :

- Si oui, alors les sous-feuilles obtenues à l'étape 4.1 deviennent des nouvelles feuilles. Et nous passons à l'étape 5.
- Sinon, on élimine la feuille qui a été remplacée par ses sous-feuilles à l'étape 4.1, et l'on passe à l'étape 4.3

Étape 4.3. Vérification si toutes les feuilles ont été remplacées par leurs sous-feuilles à l'étape 4.1 :

- Si oui, les sous-feuilles deviennent des nouvelles feuilles, et on retourne à l'étape 3.
- Sinon, on revient à l'étape 4.1.

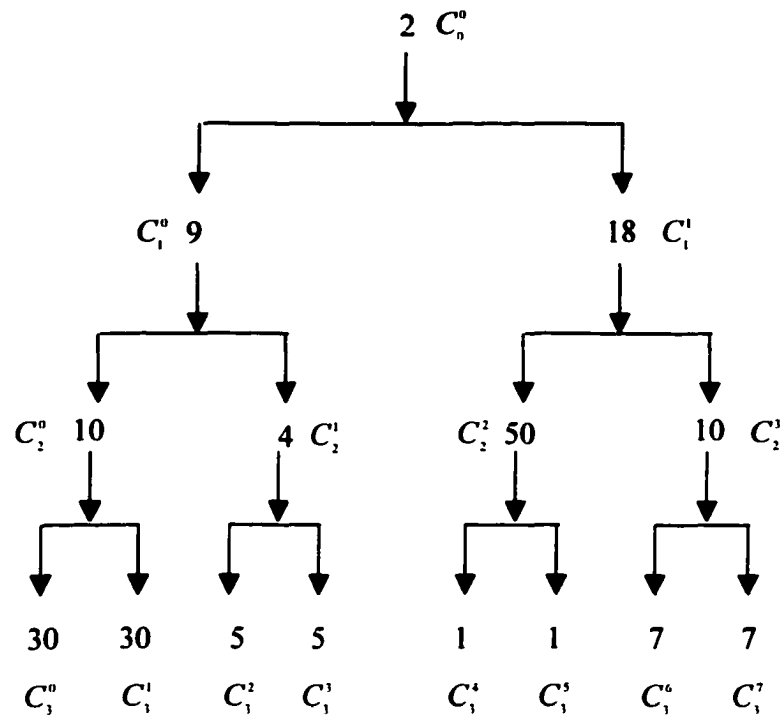
Étape 5. Fin

4.4.2 Quelques exemples d'application de l'algorithme proposé

Afin de faciliter la compréhension de l'algorithme de construction d'arbre admissible, que nous avons proposé, nous donnerons quelques exemples d'application. Toutefois, nous considérons seulement trois niveaux d'échelle, et nous supposons que le seuil prédéfini du nombre de paramètres est égal à $L=5$. Notons aussi, que chaque nœud de l'arbre abstrait initial est matérialisé par la valeur de son nombre d'occurrence correspondant.

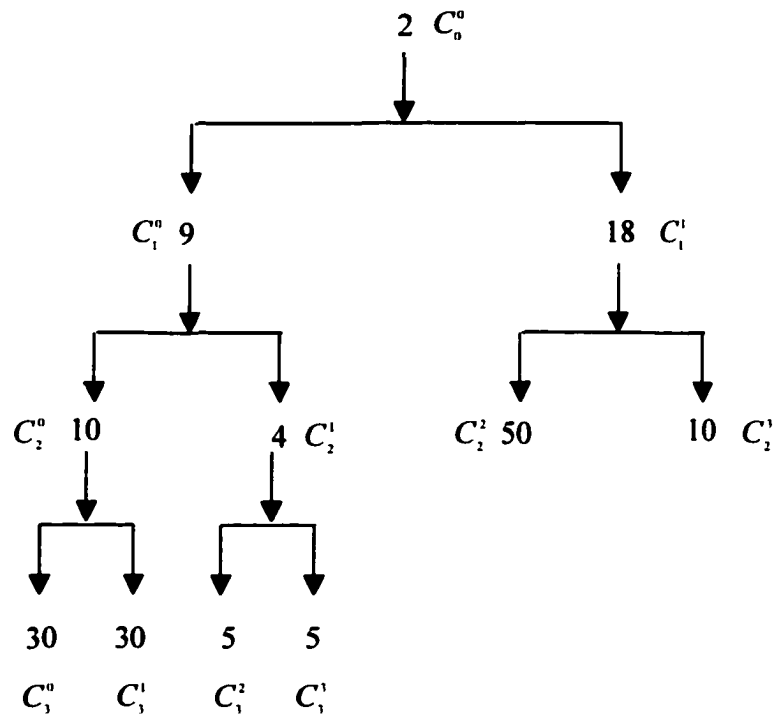
Exemple 1.

Étape 0. L'arbre abstrait initial est illustré ci-dessous.



Étape 1. Est-ce que le nombre d'occurrences du nœud C_0^0 est strictement supérieur aux nombres d'occurrences de tous ses nœuds descendants? La réponse est, non. Donc le nœud C_0^0 ne sera pas considéré comme une feuille. On passe au niveau d'échelle $j=1$, et de la même manière, on effectue le test sur le nœud C_1^0 . Est-ce que le nombre d'occurrences du nœud C_1^0 est strictement supérieur aux nombres d'occurrences de tous ses nœuds descendants? La réponse est, non. Donc on a déjà la première condition n'est pas satisfaite, par conséquent le nœud C_1^0 ne sera pas considéré comme une feuille. On continue le même traitement pour tous les nœuds du niveau d'échelle $j=1$ et $j=2$. Prenons l'exemple du nœud C_2^2 , son nombre d'occurrences est strictement supérieur à tous les nombres d'occurrences de ses nœuds descendant, donc, la première condition est satisfaite. Étant donné qu'aucun des nœuds ascendants du nœud C_2^2 n'est une feuille, la deuxième condition est satisfaite. Par conséquent le nœud

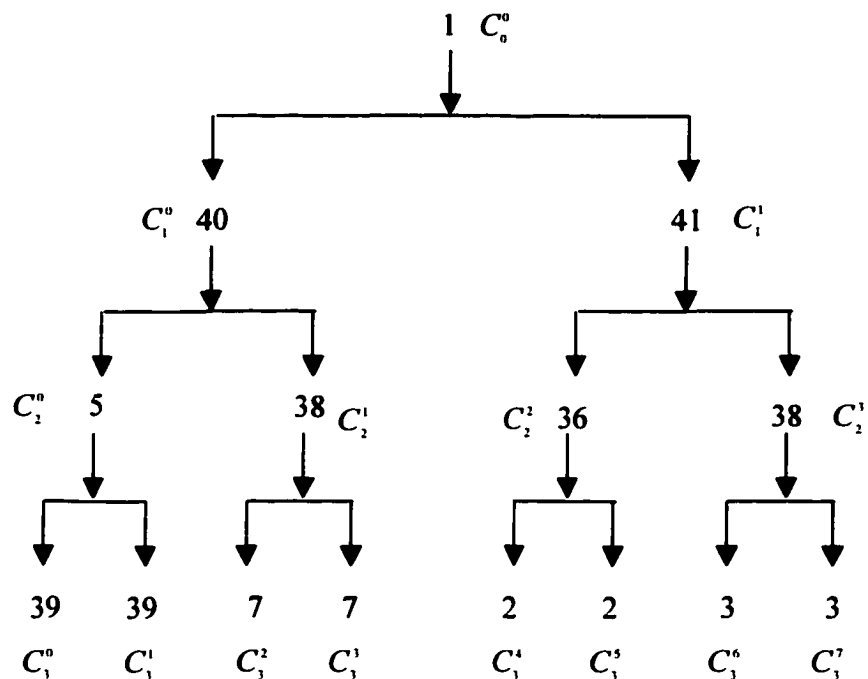
C_2^2 sera considéré comme une feuille du nouvel arbre. Finalement nous passons au dernier niveau $j=3$, et prenons comme exemple, le nœud C_3^0 . Il satisfait lui aussi la deuxième condition. Donc le nœud C_3^0 sera considéré comme une feuille. À la fin de l'étape 1, nous obtenons le nouvel arbre présenté ci-dessous.



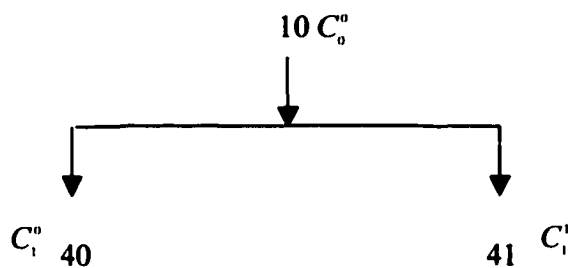
Les nœuds $C_3^0, C_3^1, C_3^2, C_3^1, C_3^2$, et C_3^1 , sont les nouvelles feuilles du nouvel arbre.

Étape 2. Le seuil prédéfini ($L=5$) du nombre de paramètres est atteint ici. Car on a obtenu 6 feuilles du nouvel arbre. On passe donc à l'étape 5.

Étape 5. Fin

Exemple 2.**Étape 0.** Arbre abstrait initial

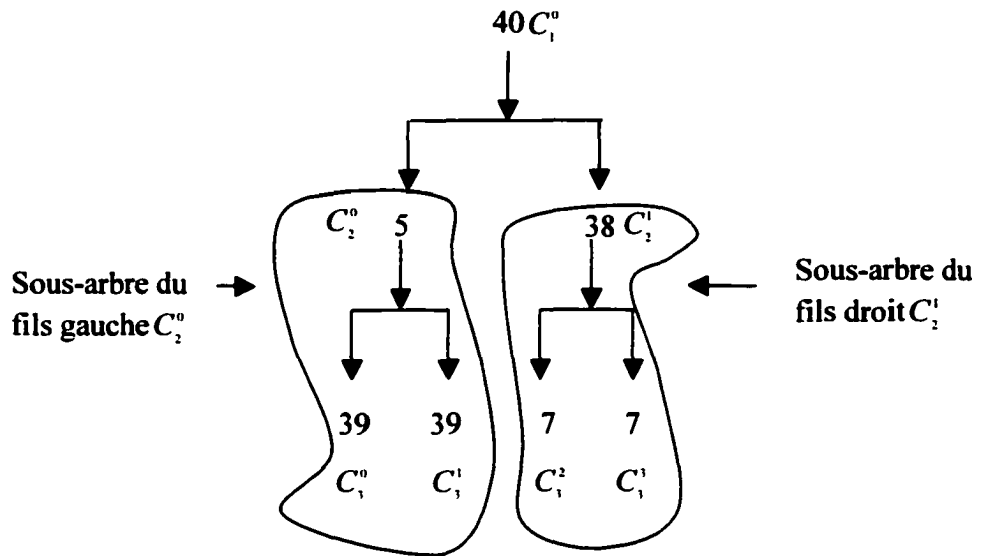
Étape 1. En appliquant l'étape 1 de l'algorithme, nous obtenons le nouvel arbre tel qu'il est présenté ci-dessous.



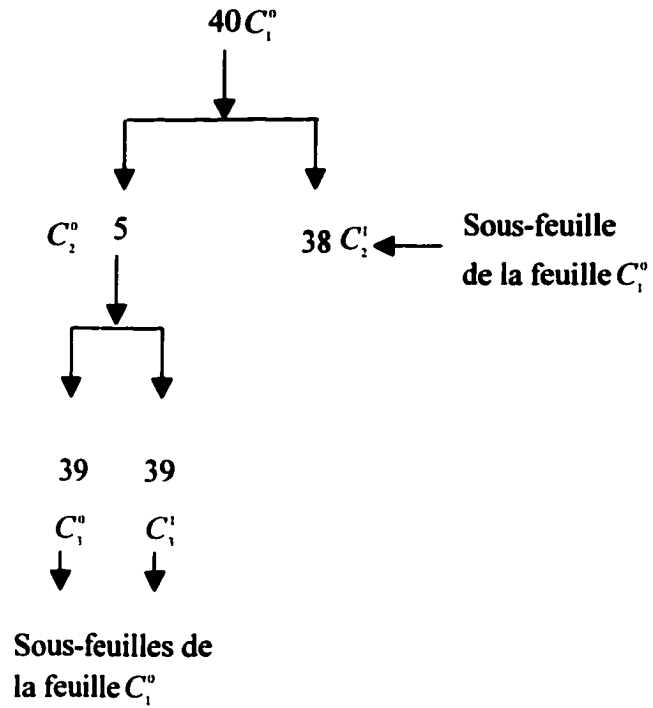
Les nœuds C_1^0 , et C_1^1 , sont les nouvelles feuilles du nouvel arbre.

Étape 2. Le seuil prédéfini ($L=5$) du nombre de paramètres n'est pas atteint, car on a obtenu 2 feuilles du nouvel arbre. On passe donc à l'étape 3.

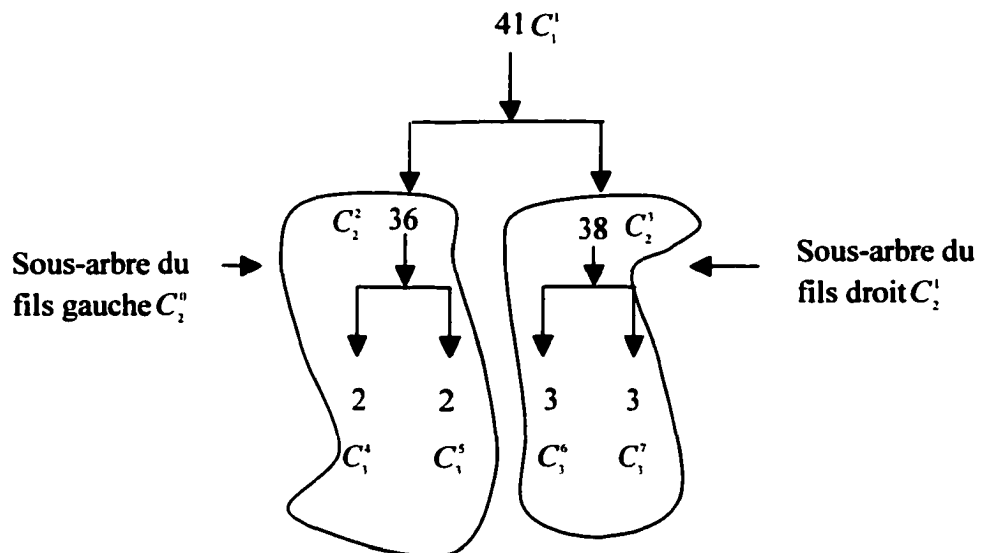
Étape 3. Pour la feuille C_1^0 , on considère son fils gauche (C_2^0) (droit (C_2^1)) et les descendants de ce dernier comme étant un sous-arbre. Tout nœud de celui-ci sera considéré comme une sous-feuille, si et seulement si les deux conditions de l'étape 1 sont satisfaites. Il en est de même pour le fils droit (C_2^1) (gauche (C_2^0)).



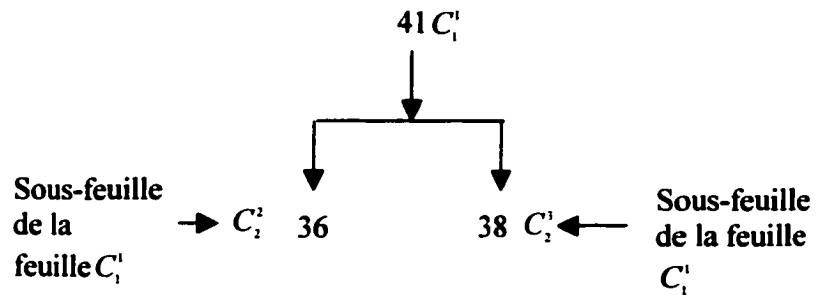
Ainsi, on obtient les sous-feuilles de la feuille C_1^0 , tel qu'illustré ci-dessous.



On répète ensuite la même procédure pour la feuille C_1^1 .



Donc, on obtient les sous-feuilles de la feuille C_1^1 , tel qu'illustré ci-dessous.



Par conséquent nous avons ainsi obtenu les sous-feuilles des feuilles C_1^0 et C_1^1 , et nous passons à l'étape 4.

Étape 4.

Étape 4.1. Remplacement de la feuille qui a le nombre d'occurrence le plus proche de la moyenne des nombres d'occurrences de ses sous-feuilles. Donc, dans ce cas, on remplace la feuille C_1^0 par ses sous-feuilles C_1^0 , C_1^1 et C_1^2 . On augmente ainsi le nombre de paramètre à 4. On passe à l'étape 4.2.

Étape 4.2. Vérification si le seuil prédéfini du nombre de paramètre est atteint.

Le nombre de paramètre étant égal à 4, donc il est inférieur au seuil prédéfini ($L=5$). On élimine alors la feuille qui a été remplacée par ses sous-feuilles à l'étape 4.1 (c'est à dire la feuille C_1^0), et l'on passe à l'étape 4.3.

Étape 4.3. Vérification si toutes les feuilles ont été remplacées par leurs sous-feuilles à l'étape 4.1.

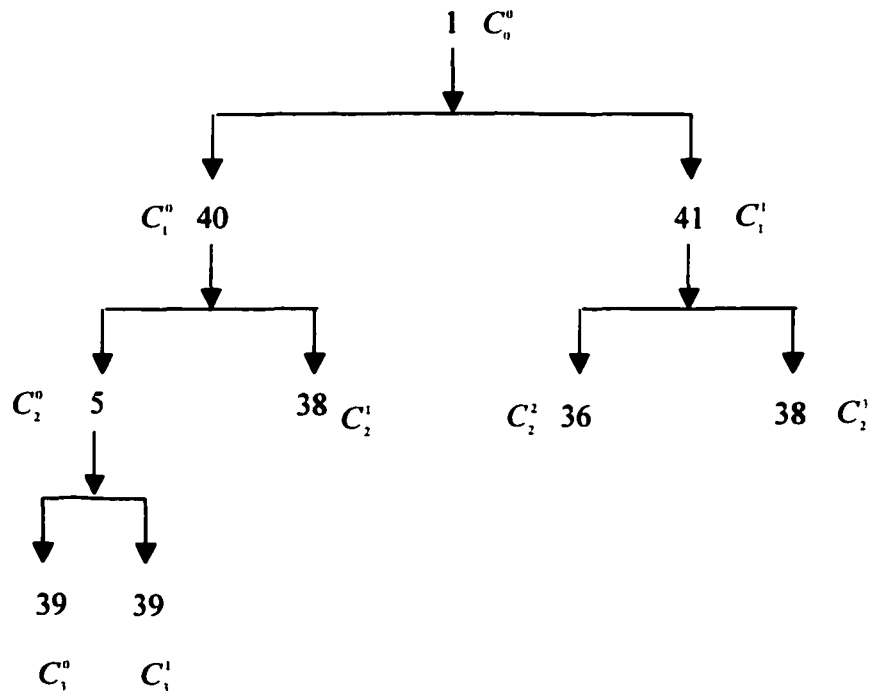
La feuille C_1^1 n'a pas été remplacée par ses sous-feuilles, donc, on retourne à l'étape 4.1.

Étape 4.1. Remplacement de la feuille C_1^1 par ses sous-feuilles C_2^2 et C_2^1 . Par conséquent, on augmente le nombre de paramètre à 5. On passe à l'étape 4.2.

Étape 4.2. Vérification si le seuil prédéfini du nombre de paramètre est atteint.

Le nombre de paramètre étant égal à 5. Le seuil est atteint. Les sous-feuilles C_2^0 , C_2^1 , C_2^2 et C_2^1 deviennent de nouvelles feuilles. On passe à l'étape 5.

Étape 5. Fin



Les nœuds C_2^0 , C_2^1 , C_2^2 et C_2^1 sont les nouvelles feuilles du nouvel arbre.

4.5 Conclusion

L'utilisation de la méthode MFDWC à l'aide des filtres d'ondelettes de Daubechies d'ordres fixes et adaptés, permet de déterminer l'ordre N le plus adapté dans la plage de valeurs $N \leq 8$, aux signaux de parole des différents locuteurs. Ainsi nous avons évalué les performances de reconnaissance de cette méthode pour différents ordres, et avons déduit celui qui donne le meilleur résultat. Toutefois la méthode MFDWC a une structure d'arbre fixe. Nous avons donc étudié une méthode basée sur la sélection de la meilleure base d'ondelettes, et avons ensuite proposé l'algorithme MAL (Meilleur Arbre du Locuteur) pour obtenir une structure d'arbre adaptée à chaque locuteur. Les tests et résultats expérimentaux obtenus à l'aide de ces méthodes seront exposés au chapitre 5.

CHAPITRE 5

RÉSULTATS EXPERIMENTAUX

5.1 Introduction

Nous présenterons dans ce chapitre les résultats expérimentaux que nous avons obtenu. Nous commencerons par donner un bref aperçu sur la base de données que nous avons utilisée. Nous montrerons comment le modèle global a été obtenu, et présenterons le calcul des performances de reconnaissance. Par la suite, nous présenterons les résultats de performances de la première approche considérée qui est la méthode MFDWC faisant usage des ondelettes de Daubechies d'ordres fixes et adaptés au sens de certains critères d'entropie. Nous présenterons ensuite les résultats des performances de la deuxième approche que nous avons abordé, qui est basée sur l'algorithme de sélection de la meilleure base d'ondelettes et sur l'algorithme MAL (Meilleur Arbre du Locuteur) que nous avons proposé.

5.2 Description de la base de donnée

Nous avons utilisé pour toutes les expériences, un ensemble de soixante (60) locuteurs (47 hommes et 13 femmes). Ces derniers sont extraits de la base de donnée non bruitée de Yoho. Chaque locuteur prononce 96 phrases dans la phase d'entraînement et 40 phrases dans la phase de vérification. Une phrase prononcée a approximativement une durée de 6 secondes. Dans la deuxième partie du travail, Nous avons testé les performances de reconnaissance avec ces phrases de vérification, mais cette fois ci bruitées à trois niveaux SNR (0dB,10dB, et 20dB). Ces bruits ont été obtenu en enregistrant des échantillons provenant d'un microphone de condensor de 1/2" B&k sur la bande sonore numérique (DAT). Le bruit d'usine a été enregistré près d'un "plat de

forme-de découpage" et d'un appareil à souder électrique. Les bruits ont été ajoutés à la base de données non bruité de Yoho, avec des niveaux de bruit de 0 dB, 10 dB et 20 dB.

5.3 Estimation du modèle de chaque locuteur et du modèle global

Le modèle initial de chaque locuteur est estimé dans la phase d'apprentissage moyennant une modélisation Markovienne (voir chapitre 3) en utilisant ses 96 phrases d'entraînement. Ce modèle est réestimé deux fois pour obtenir le modèle sur lequel nous effectuons la phase de vérification. En ce qui concerne le modèle global, il est obtenu initialement à partir de toutes les phrases d'entraînement de tous les locuteurs que l'on veut traiter. De la même manière le modèle global initial est réestimé deux fois.

5.4 Performances de reconnaissance

Les performances de reconnaissance sont évaluées en calculant le pourcentage des faux rejets (FR), le pourcentage des fausses acceptations (FA), et le pourcentage des scores (SC). Supposons que l'on a un ensemble de N locuteurs (ces N locuteurs sont ceux que l'on veut traiter pour tester les performances de reconnaissance), tel que chacun d'eux prononce M phrases (ces M phrases peuvent être des phrases d'entraînement ou de vérification). Les pourcentages suivants des FA(%), des FR(%), et des scores SC(%) sont calculés comme suit :

$$FR(\%) = \frac{\sum_{n=1}^n fr(n)}{M \times N} \times 100 \quad (5.1)$$

avec $fr(n)$ qui représente le nombre de faux rejets du $n^{\text{ème}}$ locuteur.

$$FA(\%) = \frac{\sum_{n=1}^n fa(n)}{M \times N} \times 100 \quad (5.2)$$

$fa(n)$ représente le nombre de fausses acceptations du $n^{\text{ème}}$ locuteur.

$$SC(\%) = \frac{\sum_{n=1}^N sc(n)}{M \times N} \times 100 \quad (5.3)$$

$sc(n)$ détermine le nombre de fois où le système reconnaît le $n^{\text{ème}}$ locuteur.

Dans tout ce qui suit, le seuil pour évaluer les fausses acceptations FA et les faux rejets FR a été fixé à 0.

5.5 Résultats expérimentaux

Comme nous l'avons décrit dans le chapitre précédent, nous avons effectué nos travaux expérimentaux en deux parties. Dans la première partie, nous avons testé les performances du système de vérification du locuteur, en utilisant la méthode MFDWC. En premier lieu, le test a été effectué pour évaluer la variation des performances de reconnaissance en fonction de l'ondelette de Daubechies d'ordre fixe utilisé dans le banc de filtre proposé par Farooq [7]. En second lieu, nous avons évalué les performances pour un ordre d'ondelette de Daubechies adapté, au sens de certains critères d'entropie. Dans la deuxième partie du travail, nous avons évalué les performances en utilisant la méthode de sélection de la meilleure base d'ondelette, et l'algorithme MAL (Meilleur Arbre du Locuteur) que nous avons proposé. Dans cette deuxième partie, nous avons effectué les tests sur une base de données non bruitée, et ensuite sur des données de vérification bruitées à trois niveaux SNR (0dB, 10dB, et 20 dB).

5.5.1 Première partie du travail

Nous avons testé les performances de reconnaissance en réalisant les expériences suivantes :

- Méthode MFDWC, pour cinq (5) locuteurs, pour chaque valeur de l'ordre fixe de Daubechies, allant de $N=2$ jusqu'à $N=10$. Les résultats sont présentés dans les tables II et III.
- Méthode MFDWC, pour soixante (60) locuteurs, pour chaque valeur de l'ordre fixe égal à 2,4,6, et 8. les résultats sont illustrés sur les tables IV et V.
- Méthode MFDWC, pour cinq (5) locuteurs, pour un ordre adapté au sens de quatre critères d'entropie énuméré comme suit :
 - Critère d'énergie
 - Critère d'énergie normalisée
 - Critère Shannon non normalisé
 - Critère Shannon normalisé).

Les résultats de performance de reconnaissance sont indiqués sur les tables VI et VII

- Méthode MFDWC, pour 60 locuteurs, pour un ordre adapté au sens du critère d'énergie. Voir les résultats dans les tables VIII et IX.

Tableau II

**Performances de reconnaissance – MFDWC à ordre fixe
(5 locuteurs, phrases d'entraînement)**

Ordre de Daubechies	FA(%)	FR(%)	SC(%)
2	0.42	2.29	97.29
3	0	3.13	96.88
4	0	1.67	98.33
5	0	1.46	98.54
6	0	1.25	98.75
7	0	0.83	99.17
8	0	0.42	99.58
9	0	0.83	99.17
10	0	0.63	99.38

Tableau III**Performances de reconnaissance – MFDWC à ordre fixe
(5 locuteurs, phrases de vérification)**

Ordre de Daubechies	FA(%)	FR(%)	SC(%)
2	0.5	6.5	93
3	0	8.5	91.5
4	0	6.5	93.5
5	0	3.5	96.5
6	0	4.5	95.5
7	0	3.5	96.5
8	0	3.5	96.5
9	0	5	95
10	0	6.5	93.5

Les résultats montrent que les pourcentages des fausses acceptations pour les différents ordres d'ondelette de Daubechies sont nuls, autant que pour les phrases d'entraînement que pour les phrases de vérifications. Nous constatons aussi que les meilleurs scores obtenus, pour les phrases de vérification, correspondent aux ordres suivants : 5,6,7, et 8. Cependant ces résultats ne mènent pas à une conclusion définitive concernant la variation des performances de reconnaissance de la méthode MFDWC en fonction de l'ordre fixe de l'ondelette de Daubechies. D'où la nécessité d'augmenter le nombre des locuteurs. Ainsi pour 60 locuteurs, nous obtenons les résultats indiqués sur les tables IV et V, pour les ordres 2,4,6, et 8.

Tableau IV**Performances de reconnaissance – MFDWC à ordre fixe
(60 locuteurs, phrases d'entraînement)**

Ordre de Daubechies	FA(%)	FR(%)	SC(%)
2	5.69	0.1	94.2
4	2.53	0.17	97.29
6	1.91	0.16	97.93
8	1.65	0.16	98.19

Tableau V**Performances de reconnaissance – MFDWC à ordre fixe
(60 locuteurs, phrases de vérification)**

Ordre de Daubechies	FA(%)	FR(%)	SC(%)
2	25.42	0.08	74.5
4	16.79	0.58	82.63
6	13.79	0.88	85.33
8	12.08	1	86.92

Nous constatons à partir de ces résultats, que le pourcentage des fausses acceptations diminue lorsque l'ordre de l'ondelette de Daubechies augmente. Par contre les pourcentages des faux rejets sont assez faibles. Nous pouvons les considérer comme étant négligeables devant les pourcentages des fausses acceptations et les pourcentage des scores. Que se soit pour les phrases d'entraînement ou pour les phrases de

vérification, nous constatons que le pourcentage des scores augmente lorsque l'ordre de l'ondelette de Daubechies augmente aussi.

Les résultats des performances de reconnaissance pour l'ordre adapté au sens des quatre critères d'entropie sont présentés dans les tables VI et VII. Ces résultats sont obtenus pour 5 locuteurs.

Tableau VI

**Performances de reconnaissance – MFDWC à ordre adapté
(5 locuteurs, phrases d'entraînement)**

Critère d'entropie	FA(%)	FR(%)	SC(%)
Énergie	0	0.42	99.58
Énergie normalisée	0	2.29	97.71
Shannon	0	1.46	98.54
Shannon normalisé	0.21	1.67	98.12

Tableau VII**Performances de reconnaissance – MFDWC à ordre adapté
(5 locuteurs, phrases de vérification)**

Critère d'entropie	FA(%)	FR(%)	SC(%)
Énergie	0	3.5	96.5
Énergie normalisée	0	5.5	94.5
Shannon	0	8	92
Shannon normalisé	0	6	94

Nous remarquons que le pourcentage des fausses acceptations est nul pour les quatre critères d'entropie utilisés, et cela, autant que pour les phrases d'entraînement que pour les phrases de vérification. Nous constatons aussi, que le meilleur pourcentage des scores que nous avons obtenu, est celui qui correspond au critère d'énergie. Notons aussi, que les pourcentages des scores correspondants aux phrases d'entraînement et de vérification, selon le critère d'énergie, sont les mêmes obtenus pour l'ordre fixe et maximal égal à 8. Par la suite, nous avons testé les performances sur 60 locuteurs, et cela, pour l'ordre adapté au sens du critère d'énergie. Nous obtenons les résultats indiqué dans les tables VIII et IX.

Tableau VIII

**Performance de reconnaissance – MFDWC à l'ordre adapté
(60 locuteurs, phrases d'entraînement)**

Critère d'entropie	FA(%)	FR(%)	SC(%)
Énergie	2.03	0.12	97.85

Tableau IX

**Performance de reconnaissance – MFDWC à ordre adapté
(60 locuteurs, phrases de vérification)**

Critère d'entropie	FA(%)	FR(%)	SC(%)
Énergie	12.33	0.84	86.83

En comparant les pourcentages des résultats indiqués ci-dessus (table VIII et IX), avec ceux qui correspondent à l'ordre fixe et maximal égal à 8 (table IV et V), nous constatons de la même manière, que les pourcentages des résultats pour l'ordre adapté au sens du critère d'énergie, convergent vers ceux qui correspondent à l'ordre maximal fixe égal à 8.

5.5.2 Deuxième partie du travail

Dans cette deuxième étape, nous avons utilisé l'algorithme de sélection de la meilleure base d'ondelettes, ainsi que l'algorithme MAL que nous avons proposé, pour construire pour chaque locuteur, son arbre admissible, et cela à partir de toutes ses phrases d'entraînement, extraites de la base de données non bruitée. Nous avons testé les performances de reconnaissance sur 60 locuteurs, avec un ordre de Daubechies fixe égale à 8 (car, c'est l'ordre qui nous a donné la meilleure performance de reconnaissance dans la première partie du travail). Nous avons obtenu les résultats indiqués dans les tables X et XI. Nous désignerons par (SMB_MAL), la méthode basée sur l'algorithme de sélection de la meilleure base d'ondelettes, et sur l'algorithme MAL proposé.

Tableau X

**Performance de reconnaissance – SMB_MAL
(60 locuteurs, phrases d'entraînement)**

Base de données	FA(%)	FR(%)	SC(%)
Non bruitée	0.07	0	99.93

Tableau XI

**Performance de reconnaissance – SMB_MAL
(60 locuteurs, phrases de vérification)**

Base de donnée	FA(%)	FR(%)	SC(%)
Non bruitée	0.96	0	99.04

Nous remarquons que ces résultats sont bien meilleurs comparativement à ceux obtenus dans la première partie du travail (c'est à dire, les résultats qui correspondent à la méthode MFDWC, en fonction de l'ordre fixe et adapté de l'ondelette de Daubechies). La comparaison entre ces résultats se fait naturellement sur ceux qui correspondent aux tests effectués sur 60 locuteurs.

Afin de mieux évaluer les performances de la méthode de sélection de la meilleure base d'ondelettes et de l'algorithme MAL, nous avons effectué nos tests sur une base de données bruitée à trois niveaux SNR (0 dB, 10 dB, et 20 dB). Ces tests sont réalisés sur 60 locuteurs. La table XII montre les résultats des performances obtenus pour ces trois niveaux de bruit.

Tableau XII**Performances de reconnaissance – SMB_MAL
(60 locuteurs, phrases de vérification bruitées)**

SNR (db)	FA(%)	FR(%)	SC(%)
0	44.92	1.33	53.75
10	17	0.75	82.25
20	7.25	0	92.75

Nous présentons dans la table XIII, les pourcentages des scores correspondant aux phrases de vérification, pour les différentes méthodes que nous avons utilisé, en l'occurrence la méthode MFCC, la méthode MFDWC à ordre fixe égale à 2,4,6,et 8, et à ordre adapté au sens du critère d'énergie, et finalement la méthode basée sur l'algorithme de sélection de la meilleure base d'ondelettes, et sur l'algorithme proposé (SMB_MAL). Ces résultats correspondent aux tests effectués sur la base de données non bruitée, et sur 60 locuteurs. Les résultats présentés dans la table XIII montrent que les pourcentages des scores de la méthode MFDWC à ordre de Daubechies fixe augmente lorsque l'ordre fixe augmente aussi. Nous remarquons que le pourcentage des scores de la méthode MFDWC à ordre adapté converge vers le pourcentage des scores de la méthode MFDWC à ordre fixe maximal (dans notre cas, l'ordre fixe maximal est égal à 8). Le meilleur pourcentage des scores que nous avons obtenu est celui qui correspond à la méthode basée sur l'algorithme de sélection de la meilleure base d'ondelettes et sur l'algorithme proposé (SMB_MAL).

Tableau XIII

**Performances de reconnaissance pour les différentes méthodes utilisées
(60 locuteurs, phrases de vérification)**

Méthodes		FA(%)	FR(%)	SC(%)
MFDWC Ordre fixe	2	25.42	0.08	74.5
	4	16.79	0.58	82.63
	6	13.79	0.88	85.33
	8	12.08	1	86.92
MFDWC Ordre adapté		12.33	0.84	86.83
MFCC		1.79	0.54	97.67
SMB_MAL		0.96	0	99.04

5.6 Conclusion

Les résultats expérimentaux que nous avons obtenu montrent que les performances de la méthode MFDWC à ordre fixe s'améliore lorsque ce dernier croît. Nous pouvons constater aussi que le résultat des performances de la méthode MFDWC à ordre adapté au sens du critère d'énergie converge vers celui de la méthode MFDWC à ordre fixe et maximal. Cependant le résultat de performance de la méthode MFCC est meilleur par rapport à la méthode MFDWC à ordre fixe et à ordre adapté. Ceci est du au fait que la structure du banc de filtres utilisé par la méthode MFCC, sépare mieux les composantes fréquentielles que la structure du banc de filtres utilisée dans la méthode MFDWC. Les résultats des performances que nous avons obtenu en utilisant la méthode de sélection de la meilleure base d'ondelettes, et de l'algorithme MAL, sont les meilleurs. Ceci est du au

fait que la deuxième approche que nous avons utilisé permet de créer pour chaque locuteur, son arbre admissible d'apprentissage. Celui-ci représente mieux la séparabilité de ses composantes fréquentielles. Dans la conclusion générale de ce mémoire, nous exposerons avec plus de détails les différentes interprétations que nous pouvons tirer des résultats expérimentaux obtenus. Nous présenterons aussi nos analyses et discussions.

CONCLUSION

Le travail que nous avons présenté dans le cadre de ce sujet a été divisé en deux parties. Dans la première, nous avons utilisé la méthode MFDWC faisant usage de l'ondelette de Daubechies d'ordre fixe et d'ordre adapté au sens de certains critères d'entropie. Dans la deuxième partie du travail, nous avons utilisé la méthode basée sur l'algorithme de sélection de la meilleure base d'ondelettes ainsi que l'algorithme MAL (Meilleur Arbre du Locuteur) que nous avons proposé. Nous avons comparé les résultats des performances de reconnaissance de ces méthodes avec ceux obtenus par la méthode MFCC.

À partir des résultats obtenus, nous avons constaté que les performances de reconnaissance de la méthode MFDWC à ordre fixe s'améliorent lorsque nous augmentons l'ordre de l'ondelette de Daubechies utilisée. Du point de vue temporel, cela s'explique du fait que les fonctions d'échelle et d'ondelettes générées par les filtres de Daubechies sont de plus en plus lisses quand l'ordre croît, et que par conséquent ces fonctions s'adaptent mieux au signal de la parole. Du point de vue fréquentiel, nous expliquons cette amélioration du fait que les réponses fréquentielles des filtres de Daubechies deviennent de plus en plus sélectives quand l'ordre croît. Nous atténuons donc le problème de chevauchement entre les filtres h et g et par conséquent nous améliorons la segmentation fréquentielle du signal de la parole sur toute sa bande de fréquences.

Les résultats des performances obtenus par la méthode MFDWC à ordre adapté au sens du critère d'énergie convergent vers ceux obtenus par la MFDWC à ordre fixe et maximal. Cela s'explique du fait que chaque ordre adapté à chaque décomposition de l'arbre admissible, est souvent égal à l'ordre maximal fixé. Nous avons calculé l'ordre

adapté pour maximiser l'entropie d'énergie à chaque décomposition de l'arbre. Nous pouvons dire que l'ordre maximal fixé donne souvent l'entropie d'énergie maximale.

Les performances de reconnaissance obtenus par la méthode MFDWC à ordre fixe et à ordre adapté sont moins bons que ceux obtenus à partir de la méthode MFCC. La raison est que le banc de filtres réglé à l'échelle de Mel et utilisé dans la MFCC sépare mieux les composantes fréquentielles comparativement au banc de filtres utilisé dans la méthode MFDWC.

Finalement nous avons constaté que la méthode basée sur l'algorithme de sélection de la meilleure base d'ondelettes et sur l'algorithme de construction du meilleur arbre du locuteur (MAL) que nous avons proposé, permet d'obtenir les meilleures performances par rapport à la MFCC et à la MFDWC à ordre fixe et à ordre adapté. Ainsi nous pouvons conclure que l'approche basée sur ces algorithmes permet d'obtenir pour chaque locuteur, son meilleur arbre admissible qui sépare au mieux les composantes fréquentielles où les traits acoustiques sont les plus concentrés.

Les travaux futurs dans ce domaine pourraient se porter sur la deuxième approche que nous avons développée ici. Ainsi l'algorithme de sélection de la meilleure base d'ondelettes pourrait être utilisé avec différents critères d'entropie, autres que celui qui a été utilisé dans ce travail. En effet, nous pourrions penser à utiliser le critère d'énergie, et cela en cherchant la meilleure base d'ondelettes qui maximise l'énergie du signal dans chacune de ses bandes de fréquences. Par conséquent nous pourrions tester les performances de reconnaissance pour ce critère. De même nous pourrions penser à utiliser d'autres types d'ondelettes biorthogonales (exemple celui de Spline [16]) qui garantie la linéarité de phase, chose qui est avantageux pour la reconstruction.

BIBLIOGRAPHIE

- [1] Akansu, A. N., & Smith, M. J. T., (1996), *Subband and wavelet transforms : design and applications*, Boston, Mass.: Kluwer Academic.
- [2] Baum L., Petrie T., Soules G., and Weiss N., (1970), *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*, Ann. Math Statistics, vol. 41, pp. 164-171.
- [3] Calliope, (1989), *La parole et son traitement automatique*, Paris : Masson.
- [4] Coifman, R. R., & Wickerhauser, M. V., (1992), *Entropy-based algorithms for best basis selection*, *IEEE Transactions on Information Theory*, 38(2 pt II), pp. 713-718.
- [5] Daubechies, I., (1992), *Ten lectures on wavelets*, Philadelphia, Pa.: Society for Industrial and Applied Mathematics.
- [6] Daubechies, I., (1988), *Orthonormal bases of compactly supported wavelets*, *Commun. Pure Appl. Math.*, 91, pp. 909-996.
- [7] Farooq, O., & Datta, S., (2001a), *Mel filter-like admissible wavelet packet structure for speech recognition*, *IEEE Signal Processing Letters*, 8(7), pp. 196-198.
- [8] Gish, H., & Schimdt, M., (1994), *Text-Independent Speaker recognition*, *IEEE Signal Processing Magazine*, pp. 18-32.
- [9] Higging A.L. et. al., (1991), *Speaker Verification Using Randomized Phrase Prompting*, *Digital Signal Processing*, Vol. 1, pp. 89-106.
- [10] Long, C. J., & Datta, S. (1996), *Wavelet based feature extraction for phoneme recognition*, *ICSLP 96. Proceedings., Fourth International Conference on Spoken language Processing. 1*, pp. 264-267.
- [11] Mallat S., (1989), *A theory for multiresolution signal decomposition : The wavelet representation*, *IEEE Trans. Patt. Anal Machine Intell.*, 11(7), pp. 674-693.
- [12] Mallat, S., (1989), *Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$* , *Trans. Amer. Math. Soc* ; 315, pp. 69-87.

- [13] Mallat, S., (1996), *Wavelet Signal Processing*, Academic Press.
- [14] Mallat, S. (1998), *A wavelet tour of signal processing*. New York: Academic Press.
- [15] Matsui T. and Furui S., (1994), *A New Similarity Normalization Method for Speaker Verification Based on a Posteriori Probability*, ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 59-62.
- [16] Meyer, Y., (1986), *Ondelettes, fonctions splines et analyses graduées*, lectures données à l'Université de Torino, Italie.
- [17] Meyer, Y., (1990), *Ondelettes et Opérateurs*, Ed. Hermann, France.
- [18] Papoulis A., (1984), *Probability, Random Variables and Stochastic Process*, McGraw-Hill, 2nd ed.
- [19] Rabiner L. R., (1989), *A tutorial on hidden markov models and selected applications in speech recognition*, Proceeding of IEEE, vol. 77, no 2.
- [20] Rioul, O., Vetterli, M., (1998), *Wavelets and signal, Processing*, IEEE Trans. Signal Processing Mag., pp. 14-38.
- [21] Saito, N. & Coifman, R.R., (1994), *Local discriminant basis*, Proc. Of Mathematical imaging signal and image processing, pp. 20-23.
- [22] Sarkar, T. K., et al. (1998). *Tutorial on wavelets from an electrical engineering perspective, part 1: discrete wavelet techniques*. IEEE Antennas and Propagation, vol. 40, No. 5.
- [23] Smith, M.J.T. & Barnwell, T.P., (1986), *Exact reconstruction techniques for tree-structured subband coders*, IEEE Trans. ASSP 34, pp. 434-441.
- [24] Truchelet, F., (1998), *Ondelettes pour le signal numérique*, HERMES.
- [25] Young S.J., Woodland P.C. , and Byrne W.J., (1993), *HTK: Hidden Markov Model Toolkit V1.5*, Cambridge University Engineering Department Speech Group and Entropic Research Laboratories Inc.