

وزارة التعليم العالي والبحث العلمي

BADJIMOKHTAR-ANNABAUNIVERSITY

UNIVERSITE BADJI MOKHTAR ANNABA



جامعة باجي مختار - عنابة

Année : 2019

Faculté: Sciences de l'Ingéniorat

Département: Electronique

MEMOIRE

Présenté en vue l'obtention du diplôme de :MASTER

Intitulé :

Evaluation des techniques de reconnaissance des sons des oiseaux

Domaine : Sciences et Technologies

Filière : Télécommunications

Spécialité: Système de Télécommunications

Par : DJEMIL SARRA

DEVANT Le JURY

Président : K. Boukari

MCA UBM Annaba

Directeur de mémoire: A. Boulmaiz

MCB UBM Annaba

Examineur1 : D. Messadeg

Prof UBM Annaba

Examineur2 : R. Hamdi

Prof UBM Annaba

Remerciement

Je tiens à remercier l'ensemble des professeurs du département et en particulier à leurs têtes mon encadreur professeur consultant Mme BOULMAIZ Amira qui m'ont porté beaucoup de soutien durant ma scolarité et aussi pour le choix du sujet et leurs apports bénéfiques et laborieux pour l'élaboration de mon thème.

Je remercie également l'encadrement pour le choix du sujet qui m'a élargie mes connaissances dans ce domaine.

Aux membres de jury : Je remercie pareillement tous les membres du jury pour leurs critiques que je considère comme bénéfique pour l'enrichissement, et l'argumentation de tous les points de mon sujet.

A ma familles : Je remercie infiniment mes parents et surtout l'aide et le soutien moral continuellement lors de mes études, un salut exceptionnel a mon père HAMDI, ma mère NDIWA, tous mes frères et mes sœurs SEYF EDDINE, AMINE, KELTHOUM, HALIMA, RIHANA, et YAKOUT pour leurs apports et aides et qui m'ont pas délaissé.

Amis et camarades de classe : Sans oublier mes chers amis et camarades de classe, pour leurs compréhensions et soutien durable lors de mes études et préparations de mon projet de diplôme; tels que Benaida Nour El Houda, Bireche chaima, Hamida Chahinez, Labed Nour El Islem et ceux et elles qui je n'ai pas pu les citer.

Vive la scolarité, les études de tous les cycles où j'ai connus de humbles et nobles personnes en dehors de ma famille tels que mes professeurs et tout le staff administratif pour leurs aides minutieuses.

Avec de fortes émotions et affections particulières que je ressens au fond de moi-même, je salut vivement l'ensemble des personnes de loin ou de près qui m'ont apporté un plus à mes connaissances qui me serviront dans ma vie pratique.

À ma famille.

À tous ceux que j'aime.

À tous ceux qui ont participé de près ou de loin à ma réussite...

À ma chère amie Benaida NOOR

À mon collègue de classe Boulama B. Hadji

Sarah...

Table de matière

Liste des tableaux

Liste des figures

Liste des abréviations

Résumé (Arabe, Français, Anglais)

Introduction générale

Chapitre 1 : Etat de L'art

1.1. Introduction.....	3
1.2. Mécanisme de production sonore chez les oiseaux	3
1.3. Relation entre la vocalisation des oiseaux et la parole humaine	5
1.4. Travaux connexes.....	6
1.4.1. McIlraith et al (1995)	6
1.4.2. Fagerlund (1997).....	7
1.4.4. Cai et al (2010).....	8
1.4.5. Ventura et al (2015)	9
1.5. Limites et complications	12
1.5.1. La collecte des enregistrements des sons d'oiseaux	12
1.5.2. Le bruit d'environnement	12
1.4.3. La largeur de bande de fréquence	13
1.6. Conclusion.....	14

Chapitre 2 : Système de reconnaissance automatique des sons d'oiseaux

2.1. Introduction	15
2.2. Mise en situation	15
2.2.1. L'habitat des oiseaux	15
2.2.2. Le bruit d'environnement :	16
b. Débruitage audio temps-fréquence	18
c. Estimation non diagonal	18
d. Seuillage par blocs temps-fréquence	19
e. Algorithme de seuillage par bloc	19
f. Seuillage de bloc adaptatif.....	19
g. Choix de bloc.....	20
h. Choix du niveau de seuillage λ	20
2.3.2. Extraction des caractéristiques.....	21
a. Caractérisation par MFCC	21
b. Caractéristiques HFCC (Human Factor Cepstral Coefficients)	23
2.3.3. Classification.....	26
a. Déformation temporelle dynamique (DTW)	26
2.4. Conclusion.....	27

Chapitre 3 : Simulations et Résultats

3.1. Introduction.....	29
3.2. Base de données et oiseaux étudiés	29
3.3. Résultats et discussion.....	30
3.3.1. Performance de la méthode de segmentation et de débruitage.....	30
3.3.2. Extraction des caractéristiques	34
3.3.2.1. Caractéristiques MFCC.....	34
3.3.2.2. Caractéristiques HFCC	34

3.3.3. L'impact de la méthode de segmentation sur le taux de reconnaissance	35
3.3.4. L'impact du bruit additif blanc gaussien sur le système proposé.....	36
3.3.5. L'impact du bruit d'environnement sur le système proposé	37
3.4. Conclusion.....	39

Conclusion Générale

Annexe

Références Bibliographiques

Liste des tableaux

Tableau	Titre	Page
Tableau 1.1	Travaux précédents et les méthodes utilisées pour la reconnaissance des sons d'oiseaux.	10/11
Tableau 3.1	Désignation des catégories d'espèces d'oiseaux étudiés. Les colonnes donnent les catégories des espèces étudiés, du nom scientifique, du nom commun et du nombre d'enregistrements utilisés.	29
Tableau 3.2	Paramètres de configuration pour les stratégies de caractérisation considérées.	30
Tableau 3.3	Evolution du taux de reconnaissance des méthodes étudiées à différents niveaux de SNR du bruit additif blanc gaussien	36

Liste des figures

Figure	Titre	Page
Figure 1.1	Une vue schématique du mécanisme de production sonore chez les oiseaux.	4
Figure 1.2	Spectrogramme du chant du pinson des arbres [2].	6
Figure 1.3	Les effets (a) de plusieurs oiseaux, (b) de la pluie et (c) du vent sur un enregistrement de son d'oiseau	13
Figure 2.1	Étapes de segmentation et débruitage audio par seuillage de bloc temps-fréquence	17
Figure 2.2	Partition de blocs macro	20
Figure 2.3	La structure de l'extraction des caractéristiques MFCC	21
Figure 2.4	Banc de filtre (Mel) des coefficients MFCC	22
Figure 2.5	Le banc de filtres (ERB) des coefficients HFCC	23
Figure 2.6	La structure de l'extraction des caractéristiques HFCC	24
Figure 3.4	Représentation du signal audio de Fuligule morillon perturbé par le bruit continu (vent) à un SNR = 0 dB (a), à un SNR = 5 dB (c), à un SNR = 15 dB (e). (b), (d), (f) après l'application de la méthode de débruitage temps-fréquence par seuillage de bloc.	31
Figure 3.2	Représentation du signal audio de Fuligule morillon perturbé par le bruit intermittent (accident de voiture) à un SNR = 0 dB (a), à un SNR = 5 dB (c), à un SNR = 15 dB (e). (b), (d), (f) après l'application de la méthode de débruitage temps-fréquence par seuillage de bloc.	32
Figure 3.3	Représentation du signal audio de Fuligule morillon perturbé par le bruit impulsif (explosion d'une bombe) à un SNR = 0 dB (a), à un SNR = 5 dB (c), à un SNR = 15 dB (e). (b), (d), (f) après l'application de la méthode de débruitage temps-fréquence par seuillage de bloc.	33
Figure 3.4	Vecteurs de caractéristiques MFCC.	34

Figure 3.5	Vecteur de caractéristiques HFCC	34
Figure 3.6	Evolution du taux de reconnaissance des méthodes étudiées à différents niveaux de SNR du bruit d'environnement, (a) avant l'application de la méthode utilisée pour le débruitage et la segmentation du son d'oiseau, (b) après l'application de la méthode utilisée.	35
Figure 3.7	Evolution du taux de reconnaissance des méthodes étudiées à différents niveaux de SNR d'un bruit continu.	37
Figure 3.8	Evolution du taux de reconnaissance des méthodes étudiées à différents niveaux de SNR d'un bruit intermittent.	38
Figure 3.9	Evolution du taux de reconnaissance des méthodes étudiées à différents niveaux de SNR d'un bruit impulsif.	38

Listes d'abréviations

Abréviations	Forme longue
<i>ANS</i>	Asymmetric Noise Suppression
<i>ANN</i>	Artificial Neural Network
<i>DCT</i>	Discret Cosinus Transform
<i>DTW</i>	Dynamic Time Warping
<i>ERB</i>	Equivalent Rectangular Bandwidth
<i>FFT</i>	Fast Fourier Transform
<i>GMM</i>	Gaussian Mixture Model
<i>HFCC</i>	Human Factor Cepstral Coefficients
<i>IFFT</i>	Inverse Fast Fourier Transform
<i>KDE</i>	Kernel Density Estimation
<i>LP</i>	Linear Prediction
<i>LPC</i>	Linear Prediction Coefficient
<i>LPCC</i>	Linear Prediction Cepstral Coefficients
<i>MFCC</i>	Mel-Frequency Cepstral Coefficient
<i>MMSE</i>	Minimum Mean Square Error
<i>PMVDR</i>	Perceptual Minimum Variance Distortionless Response
<i>PNCC</i>	Power-Normalized Cepstral Coefficients
<i>SNR</i>	Signal to Noise Ratio
<i>STFT</i>	Short Time Fourier Transform
<i>SVM</i>	Support Vector Machine
<i>VAD</i>	Voice Activity Detector

Résumé

L'Algérie abrite une grande diversité des zones humides, qui sont des sites importants pour l'hivernage et la nidification des oiseaux d'eau migrateurs du Paléarctique. La population d'oiseaux est l'une des espèces animales les plus menacées de l'écosystème aquatique. Ceci rend les zones humides d'Algérie d'une grande importance et nécessitent une attention particulière.

Dans ce mémoire, nous proposons un système de reconnaissance de différentes espèces d'oiseaux à travers des caractéristiques acoustiques HFCC extraites de leurs sons. Une méthode de segmentation et de débruitage audio par seuillage en bloc temps-fréquence est utilisée pour la segmentation. Enfin, pour la classification, l'algorithme DTW est appliqué.

Les résultats expérimentaux pour l'identification de 10 espèces d'oiseaux, démontrent que la méthodologie adoptée pour reconnaître les sons d'oiseaux est efficace et fonctionne de manière satisfaisante par rapport aux méthodes populaires considérées dans notre étude à savoir (MFCC, PNCC, PMVDR). De plus, les performances de reconnaissance ainsi que l'immunité au bruit d'environnement sont considérablement améliorées après l'application de la méthode de segmentation, ce qui indique que c'est une approche appropriée pour la reconnaissance acoustique des oiseaux dans des environnements complexes.

Mots clé : Reconnaissance des sons d'oiseaux, HFCC, Seuillage en bloc temps-fréquence, DTW, bruit d'environnement.

Abstract

Algeria is home to a wide variety of wetlands, which are important sites for wintering and nesting Palearctic migratory waterbirds. The bird population is one of the most endangered animal species in the aquatic ecosystem. These make the wetlands of Algeria of great importance and require special attention.

In this thesis, we propose a system for recognizing different bird species through acoustic HFCC characteristics extracted from their sounds. An audio segmentation and denoising method using time-frequency blocking is used for segmentation. Finally, for classification, the DTW algorithm is applied.

Experimental results for the identification of 10 bird species, demonstrate that the methodology adopted to recognize bird sounds is effective and works satisfactorily compared to the popular methods considered in our study namely (MFCC, PNCC, PMVDR). In addition, recognition performance and environmental noise immunity are significantly improved after the application of the segmentation method, indicating that it is an appropriate approach for acoustic recognition of birds in complex environments.

Key words: Recognition of bird sounds, HFCC, Time-frequency blocking, DTW, environment noise.

ملخص

الجزائر هي موطن لمجموعة واسعة من الأراضي الرطبة، والتي تعد من المواقع المهمة لفصل الشتاء وتعشيش الطيور المائية المهاجرة في منطقة البحر القديم. تعد الطيور من أكثر الأنواع الحيوانية المهددة بالانقراض في النظام البيئي المائي. وهذا يجعل الأراضي الرطبة في الجزائر ذات أهمية كبيرة وتتطلب عناية خاصة.

في هذه المذكرة، نقترح نظامًا للتعرف على أنواع الطيور المختلفة من خلال خصائص HFCC الصوتية المستخرجة من أصواتها. يتم استخدام طريقة تجزئة الصوت وتقليل الضوضاء باستخدام طريقة للتجزئة تعتمد على التردد و الزمن. أخيرًا، للتصنيف، يتم تطبيق خوارزمية DTW .

أظهرت النتائج التجريبية لتحديد 10 أنواع من الطيور أن المنهجية المعتمدة للتعرف على أصوات الطيور فعالة وتعمل بشكل مرضٍ مقارنة بالطرق الشائعة التي تم دراستها في دراستنا وهي (MFCC, PNCC, PMVDR) بالإضافة إلى ذلك، تم تحسين أداء الاعتراف والحصانة من الضوضاء البيئية بشكل كبير بعد تطبيق طريقة التجزئة، مشيرًا إلى أنه هو النهج المناسب للاعتراف الصوتي للطيور في البيئات المعقد.

الكلمات المفتاحية: التعرف على أصوات الطيور ، HFCC ، طريقة التجزئة معتمدة على التردد و الزمن، DTW ، ضوضاء البيئة.

Introduction générale

Introduction générale

L'analyse acoustique des sons d'oiseaux est un aspect important et en augmentation dynamique de la recherche ornithologique. L'identification des espèces d'oiseaux peut servir d'exemple. Elle est souvent utilisée à des fins éducatives et pédagogiques mais peut également s'appliquer à la recherche et aux études biologiques et agricoles. L'application spécifique de l'identification des espèces d'oiseaux est la détection des espèces d'oiseaux dans la zone surveillée. Une telle détection permettrait de réaliser des économies, en particulier pour les activités de surveillance des oiseaux qui prennent un temps et une présence régulière sur le terrain. Dans les actions de suivi et de dénombrement des espèces d'oiseaux, généralement organisées par les institutions nationales (parcs nationaux, autres) et différentes organisations (nationales ou privées), on compte généralement jusqu'à plusieurs centaines de personnes (nécessaires pour être efficaces). Elles sont entreprises dans des zones difficilement accessibles et nécessitent un travail intensif de spécialistes formés et qualifiés pour reconnaître les oiseaux uniquement par vocalisation.

0.1. Problématique

Les spectres de fréquence des voix d'oiseaux, en plus des sons situés dans la plage de fréquence humaine, incluent également certaines composantes de fréquence sous infra ou ultrasonores, qui ne sont pas perçues par l'homme [5, 6]. L'analyse de fréquence effectuée doit pouvoir traiter des signaux très rapides et une modulation de signal large. Certains oiseaux mélangent parfois des signaux provenant des deux cordes vocales (une construction anatomique unique pour les oiseaux) lors du chant. Ces conditions rendent l'analyse de la voix des oiseaux difficile dans une certaine mesure dans l'application à la reconnaissance des espèces d'oiseaux.

Le premier facteur affectant le bon fonctionnement de tout système de reconnaissance audio est le bruit sous ses différentes formes. Le bruit quelle que soit sa nature, rend difficile, voire impossible la mesure et l'acquisition des sons d'oiseaux dans leurs habitats naturels. En effet, les sons d'oiseaux observés sont souvent noyés dans ce bruit de l'environnement (tels que le vent, la pluie, les sons d'autres animaux, etc...).

L'identification acoustique d'espèces d'oiseaux est généralement basée sur les méthodes utilisées pour la reconnaissance de la parole chez l'homme [1, 2, 3]. Comme vecteurs caractéristiques, on utilise généralement des coefficients Cepstral Mel-Fréquence (MFCC) [1, 3], tandis que les modèles de Markov cachés (HMM) et « Dynamic Time Warping » (DTW) sont les classificateurs les plus courants [2, 3]. Récemment, des méthodes d'extraction de caractéristiques basées sur le suivi des pics spectraux [2] et des histogrammes de paires de syllabes [4] ont été introduites.

0.2.Organisation du mémoire

Dans ce mémoire, un système de reconnaissance audio des espèces d'oiseaux est proposé. Ce système sera basé sur trois principales étapes : la première étape consiste en la segmentation et le débruitage des sons des oiseaux. La méthode adoptée se base sur le seuillage adaptatif de bloc temps-fréquence. Pour l'étape de caractérisation des segments débruités, les coefficients de facteur humain (HFCC ou Human Factor Cepstral Coefficients), proposés dans le contexte de la reconnaissance de la parole humaine [4], ont été utilisés. Enfin, l'algorithme DTW est exploité pour l'étape de classification. Les expériences ont été effectuées sur des une base de données d'oiseaux d'eau sédentaires/migrateurs du nord-est de l'Algérie.

Dans le chapitre 1, la reconnaissance automatique des sons des oiseaux : un état de l'art. La première partie de ce chapitre présente le mécanisme de production sonore chez les oiseaux. La deuxième partie présente la relation entre la vocalisation des oiseaux et la parole humaine, et la partie suivante sur les travaux précédents utilisés. La dernière partie présente les limites et complications du système de reconnaissance.

Dans le chapitre 2, Système de reconnaissance automatique des sons des oiseaux pour la surveillance d'habitat. Le deuxième chapitre présente les méthodes de segmentation, de caractérisation et de classification adoptées dans ce mémoire.

Dans le chapitre 3, Expérimentations et résultats. Ce chapitre est consacré aux résultats obtenus pour notre système de reconnaissance des sons d'oiseaux.

Chapitre 1
Etat de l'Art

1.1. Introduction

Les oiseaux représentent un rôle essentiel au sein de la biodiversité. Un grand nombre d'études biologiques obtiennent des enregistrements de la vocalisation d'oiseaux collectés dans leurs habitats pour la recherche technique. Un grand nombre de ces enregistrements sont analysés à l'aide d'une représentation des sons d'oiseaux sous forme de spectrogramme, qui est utilisée pour reconnaître les espèces à partir de leurs sons. Généralement, les recherches de ce type sont faites dans des environnements naturels des oiseaux, ce qui engendre des difficultés techniques en ce qui concerne la localisation et la surveillance de ces oiseaux.

Les vocalisations humaines sont constituées de sous-unités organisées, mots et phrases hiérarchisés, et cela s'applique également aux oiseaux, avec leurs éléments, syllabes et phrases. Jusqu'ici, relativement peu de travail a été accompli pour développer des logiciels capables d'appliquer ce type de reconnaissance aux vocalisations animales [6].

Dans la première partie de ce chapitre, nous présentons une description du mécanisme de production sonore chez les oiseaux.

Dans la deuxième partie, nous comparons la vocalisation des oiseaux et la parole humaine qui nous permettra plus tard de modéliser et synthétiser leurs sons de la même manière que la parole humaine.

Dans la troisième partie, nous citerons quelques travaux précédents et les méthodes utilisées pour la reconnaissance des sons d'oiseaux.

Enfin, nous exposerons les limites et les complications auxquels sont confrontées les techniques de reconnaissance audio automatique lors de l'application sur les sons des oiseaux.

1.2. Mécanisme de production sonore chez les oiseaux

Les oiseaux utilisent comme les humains leur système respiratoire pour produire et émettre des sons. Le système respiratoire des oiseaux est assez complexe. Il se compose de deux poumons attachés à des petites expansions bulbeuses ou des sacs aériens qui se diffèrent en nombre suivant les espèces, ces organes sont comblés d'air et vidés grâce à l'action des muscles pectoraux. Les sacs aériens jouent un rôle majeur lors de la respiration, ils aident à refroidir le corps grâce à leur surface importante d'évaporation, dont ils permettent d'aérer les poumons en permanence en fonctionnant comme des soufflets.

Les narines et le bec des oiseaux sont les moyens d'absorption d'air. Après, l'air passe par la trachée qui se distribue aux deux bronches reliée chacune à un poumon et aussi aux sacs pulmonaires.

Dans tout ce système exactement à la jonction des deux bronches existe le syrinx qui joue un rôle essentiel de création de son. Le syrinx joue le rôle des variations du passage de l'air et cela grâce à sa complexité et aussi à la présence d'une musculature sophistiquée utilisée pour modifier le rythme d'évacuation de l'air en dehors des poumons et des sacs aériens ce qui explique le phénomène de vibration de membranes.

Les bronches des oiseaux sont en fait constituées d'anneaux cartilages assemblé par un tissu qui laisse leur rigidité. Ce tissu devient très souple au niveau du syrinx pour s'adhérer à la forme de membranes souples. La figure 1.1 présente une vue schématique du mécanisme. Selon les espèces d'oiseaux les dimensions et le mécanisme varient, mais l'organisation est pareille.

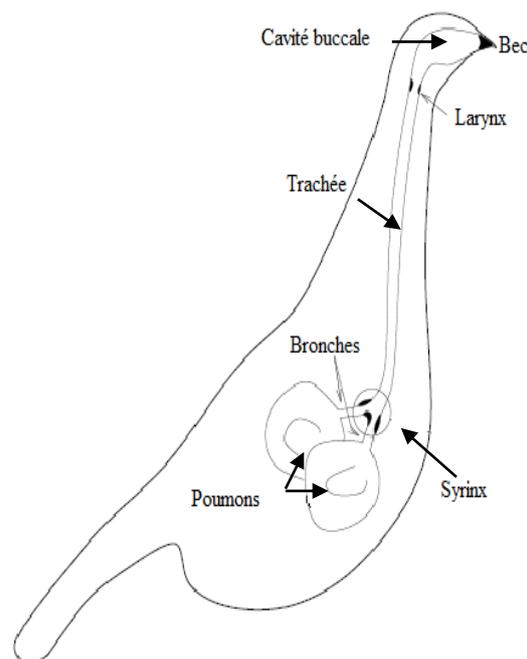


Figure 1.1 : Une vue schématique du mécanisme de production sonore chez les oiseaux.

1.3. Relation entre la vocalisation des oiseaux et la parole humaine

Les oiseaux et les humains produisent des sons de signaux acoustiques complexes. Chez l'homme, cela se produit lorsque l'air traverse l'appareil vocal lors de l'expiration, ce qui produit un son. Pendant l'expiration, des ondes se forment dans les cordes vocales. Les composants des ondes sont modifiés par le reste du canal vocal, à savoir la bouche, le nez, la langue et les dents.

La production de son chez l'être humain est similaire à celle des oiseaux. Dont les oiseaux produisent du son dans l'ensemble vocal par le biais du syrinx en accord avec l'appareil vocal. Ce dernier est considéré comme un tube résonateur qui convertit le signal provenant de syrinx

Malgré la différence observée dans la composition du système vocal, les humains et les oiseaux sont capables de produire des voix organisées et de modifier la vitesse élevée. Ce qui nécessite un contrôle neurologique sophistiqué et une coordination du système vocal.

Pour déterminer le chant d'oiseau, la méthode d'analyse est utilisée selon un spectrogramme donnant une représentation visuelle montrant la variation du signal. La densité spectrale du chant d'oiseaux peut être examinée en fonction du temps utilisé pour le traitement du signal temps-fréquence.

Le spectrogramme affiche des petites sections continues, ces sections représentant le niveau de base du chant, appelé élément. Cela ressemble aux éléments de base de la parole chez l'homme. Ces éléments du chant sont acoustiquement distincts, comme des notes de musique, et ils constituent une séquence caractéristique.

La présence de plusieurs éléments consécutifs ou parfois séparés avec des moments de silence forme des syllabes. La figure ci dessous illustre le chant d'un pinson des arbres et les sous-unités de son chant, avec des éléments, des syllabes et une phrase tous visibles.

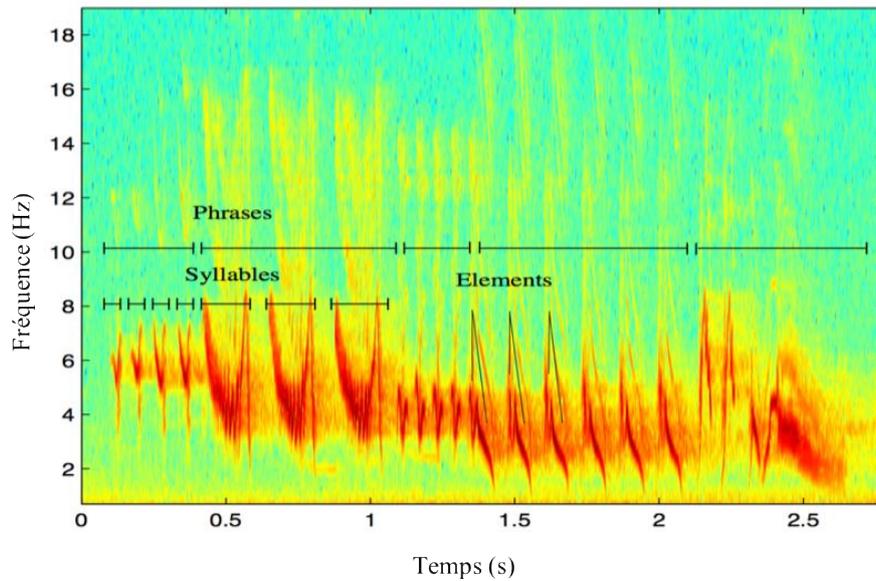


Figure 1.2 : Spectrogramme du chant du pinson des arbres [7].

Les jeunes oiseaux chanteurs, cependant, suivent le même processus d'apprentissage vocal à l'instar des êtres humains. Ils écoutent le chant émis par leurs aînés, après ils essayent à imiter leurs chants. Puis, ils utilisent la réaction de leur auditoire pour parfaire leurs capacités motrices vocales, créent des motifs sonores structurés et apprennent à communiquer. Ceci est très similaire à ce qui a été observé chez l'être humain.

Et de ce principe, chez l'homme comme chez les oiseaux, le mécanisme sonore est semblable. Donc, il est raisonnable que l'utilisation des méthodes relatives à la modélisation de parole et des algorithmes de reconnaissance vocale automatisés, utilisés par l'analyse de la parole, puisse être utilisée pour la reconnaissance du chant des oiseaux.

1.4. Travaux connexes

1.4.1. McIlraith et al (1995)

Une enquête sur la reconnaissance des espèces d'oiseaux a été réalisée à l'aide d'enregistrements des chants d'oiseaux de six espèces différents, en essayant de créer un réseau de neurones artificiel.

Les méthodes de prétraitement incluaient un codage prédictif linéaire et des transformations de Fourier rapide. Le classificateur peut identifier proprement environ 82 % à 93 % des échantillons qui ont été évalués.

Il est difficile de déterminer le succès relatif de ce travail, en raison du petit ensemble de données et de l'absence de modèle de classification comparative. De plus, l'utilisation de

la longueur des chansons utilisées pour aider à déterminer le type d'oiseaux n'est pas pratique à l'heure actuelle et ne donne pas de succès fructueux en termes de détermination du succès relatif, ce qui rend les résultats artificiels dans une certaine mesure.

1.4.2. Fagerlund (1997)

Ce travail a pour but de développer un système entièrement automatique pour la reconnaissance des espèces d'oiseaux à partir de leurs vocalisations produites sur leurs habitats. Ceci est fait en utilisant deux représentations paramétriques: les coefficients cepstrale mel-fréquence et un ensemble de paramètres de signal descriptifs.

La base de données est représentée par deux groupes d'espèces d'oiseaux : un avec 6 espèces, et l'autre avec 8 espèces.

Pour effectuer l'étape de segmentation, un algorithme itératif de domaine temporel a été appliqué pour déterminer les syllabes.

La méthode de caractérisation MFCC et l'ensemble de paramètres de signal descriptif sont utilisés pour les caractéristiques des syllabes segmentées. Les deux ensembles de données ont été segmentés manuellement en syllabes.

Pour l'étape de classification, l'identification est effectuée dans un arbre de décision avec un algorithme de classification SVM (machine à vecteurs de support). Pour obtenir les résultats de la reconnaissance, une combinaison de la fonction descriptive et des fonctions MFCC a été utilisée dans les deux ensembles de données.

D'après ces travaux, les résultats ont montré que le classificateur SVM fonctionne bien avec les MFCC. L'avantage d'un classificateur SVM est leur précision. Le MFCC est l'algorithme optimal pour le traitement du signal.

1.4.3. Brown et al (2009)

Brown et al. ont comparé la précision des DTW, des GMM et des SVM ainsi que du logiciel Song Scope¹ lors de la reconnaissance de 608 échantillons sur un total de 5 espèces différentes. Les enregistrements ont été traités manuellement pour supprimer tout bruit de fond.

¹ Est un logiciel qui permet de visualiser les appels sur un spectrogramme et de créer des "dispositifs de reconnaissance" permettant de rechercher automatiquement des enregistrements pour des vocalisations spécifiques.

Ensuite, ces données ont été partagées en deux ensembles égaux de données. L'extraction des caractéristiques a été effectuée sur les deux ensembles.

Les caractéristiques utilisées ont été calculées en divisant les données dans des cadres ou en 256 échantillons, puis en calculant une FFT et en calculant la transformation directe en cosinus de ces caractéristiques.

Les résultats de l'expérience ont montré que, dans l'ensemble, l'approche GMM atteignait la précision la plus élevée (55,2%), suivie du logiciel Song Scope (32%). DTW a eu le pire rendement avec 10%, et le SVM avait une précision de 23,8%.

Les recherches menées dans le cadre de ce projet ont montré que le classificateur GMM est capable de surpasser celui du classificateur HMM de Song Scope. Cependant, les tests n'ont jamais testé l'exactitude au niveau syllabique du SVM ou du GMM. Bien que ce niveau d'extraction de syllabes ait été réalisé par Song Scope en interne ainsi que par le modèle DTW.

1.4.4. Cai et al (2010)

Dans ce travail, Cai et al. étudient les performances de la reconnaissance des espèces d'oiseaux à l'aide de réseaux neurones artificiels (ANN). Les auteurs ont créé leur propre algorithme de réduction du bruit et l'ont appliqué efficacement au prétraitement des espèces d'oiseaux. L'ensemble de données comprend 14 espèces.

Cet algorithme permet d'estimer le bruit provenant de n'importe quelle trame avec ou sans signaux et ne nécessite pas un VAD (détecteur d'activité vocale). L'algorithme créé n'a pas besoin d'une période de silence pour obtenir l'estimation de fond et fournit une estimation à partir de n'importe quelle trame, cet algorithme est comparé à un filtre de réduction de bruit à erreur quadratique moyenne (MMSE).

Les performances de l'architecture du réseau neuronal du contexte ont été comparées avec des coefficients cepstraux Mel-fréquence (MFCC). Les résultats atteignent le taux de reconnaissance de 86,8%. Les auteurs analysent les performances de leur système et avaient constaté que les interférences provenant des cris d'autres oiseaux ou d'animaux constituaient toujours le principal obstacle à la reconnaissance des espèces d'oiseaux.

1.4.5. Ventura et al (2015)

Ce travail à pour but de développer un système entièrement automatique pour la reconnaissance des espèces d'oiseaux à partir de leurs vocalisations produites sur leurs habitats. La base de données est représentée par quarante espèces d'oiseaux. L'étape de segmentation et de la réduction du bruit sont faite à l'aide d'un filtre morphologique.

La méthode de caractérisation MFCC (Mel Frequency Cepstral Coefficients) est utilisée pour les caractéristiques des syllabes segmentées. Une précision de classification de 72% à été obtenue après l'utilisation du classificateur HMM.

Nous citons dans le tableau ci-dessous quelques travaux précédents et les méthodes utilisées pour la reconnaissance des sons d'oiseaux et la performance des méthodes:

Étude	Prétraitement, segmentation, estimation du bruit	Caractéristiques utilisés	Classification	A/M ²	Nombre d'espèces	Performances	Commentaires
Brown et al (2009)	Non mentionné	L'extraction de syllabes est réalisée par Song Scope.	Déformation temporelle dynamique (DTW), SVM, Modèle de mélange gaussien (GMM), sont comparés avec le logiciel Song Scope.	M	5	Précision de 55.2% pour le GMM, et 32 % pour le logiciel Song Scope, le SVM avec précision de 23.8 % et 10% pour le DTW	Le classificateur GMM est capable de surpasser celui du classificateur HMM de Song Scope
Cai et al (2010)	Estimation du bruit de chaque trame de bruit sans utiliser des détecteurs d'activité vocale (VAD)	Coefficient Cepstral Mel-Fréquence (MFCC) Coefficients de fréquence linéaire	Un réseau de neurones de perception multicouches (MLP NN)	M	14	Une précision globale de 86,8% pour les MFCC	Un seul obstacle au processus semble être l'interférence d'autres oiseaux et animaux sur les enregistrements.

² Automatique/Manuelle

Étude	Prétraitement, segmentation, estimation du bruit	Caractéristiques utilisés	Classification	A/M ³	Nombre d'espèces	Performances	Commentaires
Zhang et Li (2015)	Détection d'énergie adaptative.	Transformées en ondelettes discrètes	Machine à Vecteur de support (SVM)	M	30	Précision de classification de 85%	Enregistrements propres utilisés, manuellement par du bruit.
Ventura et Al. (2015)	Réduction du bruit et sélection de trame à l'aide d'un filtre morphologique du spectrogramme.	Coefficient Cepstral Mel-Fréquence (MFCC)	Modèle de Markov caché (HMM). Modèle de Mélange Gaussien (GMM)	M	40	La précision (72%) était supérieure à celle de MFCC après la détection du cadre avec le détecteur d'énergie GMM (70 %), segmentation des syllabes (65 %), détecteur de région d'intérêt (48%).	Cette méthode conduit à une précision relativement plus élevée par rapport aux méthodes traditionnelles à base de GMM et à base de syllabes.
Oliveira et al(2015)	Ouverture morphologique (érosion et dilatation)	Non mentionné	Modèle de Markov caché (HMM).	A	1	56% de rappel.	Requis de grandes quantités de données de formation.

Tableau 1.1 : Travaux précédents et les méthodes utilisées pour la reconnaissance des sons d'oiseaux

³ Automatique/Manuelle

1.5. Limites et complications

L'utilisation de la surveillance bioacoustique est un outil fonctionnel d'évaluation de la population d'oiseaux. Néanmoins, il faut tenir compte des circonstances atténuantes pour l'usage des techniques de reconnaissances automatiques des paroles lors de l'application de la distinction des sons des oiseaux.

1.5.1. La collecte des enregistrements des sons d'oiseaux

Par exemple, en considère de rassembler des échantillons de données des humains est plus facile que de reconnaître des échantillons des oiseaux. Car le chercheur peut rassembler dans un milieu contrôlé, ce qui lui permet d'avoir un enregistrement bien défini dépourvus de temps de silence et de produire un échantillon avec un signal élevé.

Comparons ceci avec des études sur les voix des oiseaux, collectant les échantillons d'oiseaux selon leurs milieux d'origines. L'appareil d'enregistrement peut être à plusieurs distances des lieux d'oiseaux. D'où peut avoir des perturbations et obstacles d'enregistrement en provoquant des réverbérations sur les fichiers audio. Parmi ces perturbations qui apparaissent le bruitage des feuilles d'arbres.

1.5.2. Le bruit d'environnement

D'autres complications surviennent à cause de la quantité importante de bruit de fond affectant l'enregistrement, avec le bruit associé à d'autres animaux, d'autres sons d'oiseaux, la présence humaine, y compris les avions, les trains, moteurs des véhicules et des événements naturels tels que, le son des eaux, le soufflement des vents et les pluies. Tous ces derniers peuvent nous donner un signal avec SNR faible, et l'appareil de reconnaissances trouve des difficultés à identifier l'espèce d'oiseau.

Des exigences supplémentaires en matière de traitement du signal sont donc nécessaires avant de pouvoir utiliser le fichier audio brut, tels que l'insertion du signal dans des filtres passe bande et la normalisation.

La figure 1.3 montre un exemple de trois enregistrements différents. Il est facile de voir les effets du bruit. Il est essentiel de supprimer la majorité du bruit avant que la chanson ne soit reconnue avec succès. Sans suppression correcte de cette interférence, une précision inférieure à celle attendue.

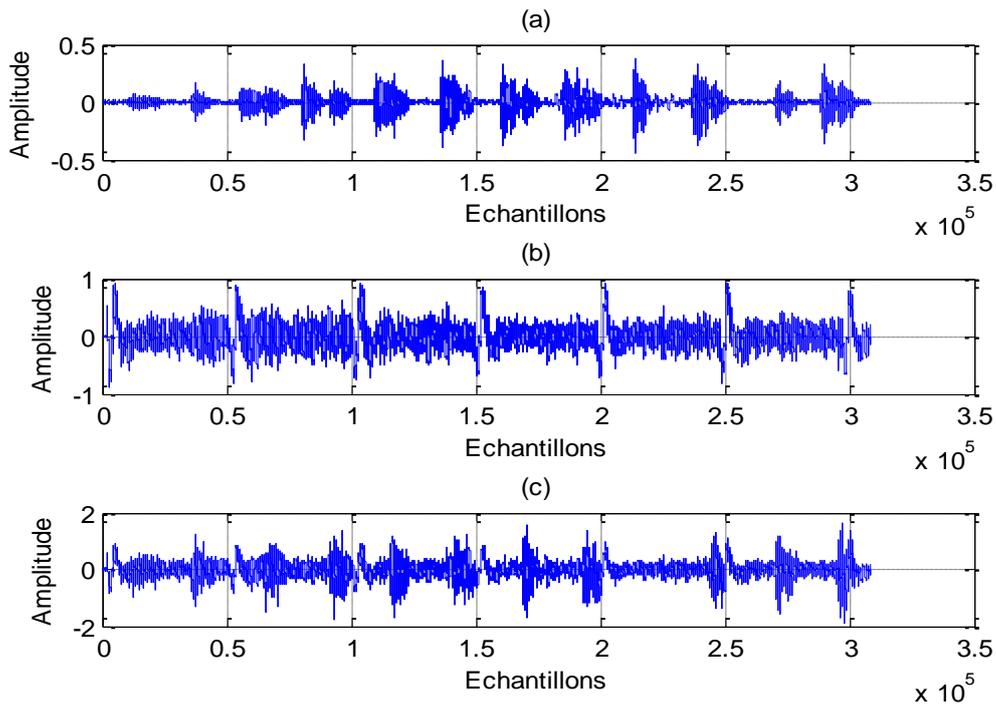


Figure 1.3 : Les effets (a) de plusieurs oiseaux, (b) de la pluie et (c) du vent sur un enregistrement de son d'oiseau.

1.4.3. La largeur de bande de fréquence

Une autre difficulté trouvée, est la largeur de bande de fréquence. La largeur de bande de la parole de l'être humain est unique, elle est concentrée dans une largeur de bande d'environ 4 kHz. Mais avec les oiseaux les sons peuvent varier considérablement, sur une large gamme de fréquences allant de 10Hz à 10 kHz. Cela permet à différentes espèces d'oiseaux de produire des sons très différents. La variation du chant rend la détection difficile.

Certains appels sont courts et à bandes étroites aux propriétés spectrales distinctes. Tandis que d'autres chants peuvent être longs avec des différences spectrales complexes. A cause de cette différence, il est difficile de créer un algorithme capable de détecter avec succès un spectre aussi large.

En outre, la grande quantité de données de formation disponibles pour la reconnaissance de la parole humaine facilite la modélisation des variations individuelles grâce aux données rassemblées de milliers d'individus. Une grande partie des recherches menées sur la vocalisation des oiseaux se limite à un nombre beaucoup moins important de données d'essais, ce qui rend difficile la modélisation des grandes variations chez de nombreuses espèces [6].

En utilisant les informations précédentes, on peut sans aucun problème comprendre les complications de la détection des oiseaux, par rapport à la reconnaissance vocale automatisée, et les complications de l'application d'un algorithme. Pour appliquer un algorithme il est indispensable d'utiliser une procédure de prétraitement agressive, afin d'utiliser l'enregistrement pour un classificateur qui a une capacité de déterminer une espèce à partir d'un chant d'oiseau.

1.6. Conclusion

Dans ce premier chapitre, nous présentons un état de l'art sur la reconnaissance automatique des sons des oiseaux. Ainsi, la relation entre la vocalisation des oiseaux et la parole humaine, ou nous avons démontrés que les vocalisations humaines sont constituées de sous-unités organisées, mots et phrases hiérarchisés, et cela s'applique également aux oiseaux, avec leurs éléments, syllabes et phrases, et c'est ce qui a incité pour développer des logiciels capables d'appliquer ce type de reconnaissance aux vocalisations animales.

Les travaux précédents évoqués dans la troisième partie montrent les résultats et leurs performances.

En fin de ce chapitre, nous avons cités les limités et les circonstances atténuantes engendrés lors de l'utilisation des techniques de reconnaissances automatiques, tels que le son d'autres animaux, la présence humaine, et les événements naturels. Tous ces inconvénients ont poussés à trouver des solutions et les moyens de les éliminer tels que les algorithmes de prétraitement.

Dans le chapitre suivant, on expose ce qui suit : l'habitat des oiseaux, bruits d'environnement, limites énergétiques des capteurs sans fils. De plus, nous allons présenter les différentes étapes du système de reconnaissance automatique des sons des oiseaux pour la surveillance d'habitat qui consolident la segmentation, l'extraction des caractéristiques, et la classification, et les techniques utilisées pour chaque étape.

Chapitre 2
Systeme de
reconnaissance
automatique des sons
d'oiseaux

2.1. Introduction

Le système de reconnaissance automatique des sons d'oiseaux basé sur trois étapes essentiels. Premièrement, l'étape de la segmentation qui a pour but de détecter les segments significatifs des sons d'oiseaux. Deuxièmement, l'extraction de caractéristiques, les signaux audio sont représentés avec un nombre faible d'attributs (comparés aux données d'origines) qui caractérisent des événements sonores particuliers. L'extraction de caractéristiques réduit les dimensions en supprimant les informations redondantes des données d'origine. Une fois les caractéristiques sont extraites l'étape de classification des caractéristiques est effectuée.

Dans ce chapitre, nous présentons les méthodologies adoptées dans notre étude pour l'étape de segmentation, d'extraction de caractéristiques, et la classification pour la conception de système de reconnaissance de sons d'oiseaux.

La surveillance de l'écosystème en ce qui concerne la population aviaire est menée dans des environnements naturels, ce qui entraîne des difficultés techniques à la localisation et la surveillance de ces oiseaux. Parmi ses difficultés nous pouvons citer : le bruit d'environnement, la consommation énergétique des nœuds-capteurs sans fil, que nous allons détailler dans ce chapitre.

2.2. Mise en situation

2.2.1. L'habitat des oiseaux

L'habitat est le milieu où vit une espèce naturellement et qui lui offre nourriture et abri à l'un des stades de son cycle biologique. Les oiseaux ont aussi besoin de sites pour la nidification.

Afin de préserver la qualité de l'habitat, il est important de garder des forêts bien aménagées avec des peuplements diversifiés et hétérogènes, des bandes de végétation en bordure des cours d'eau, de maintenir l'intégrité des milieux humides et de favoriser la flore indigène.

Il y a environ 10000 espèces d'oiseaux dans le monde, les différences climatiques, topographiques et géologiques expliquent dans une large mesure cette biodiversité.

Des pôles aux forêts équatoriales, des déserts aux centres des océans et des plus hautes montagnes au cœur de nos villes, partout les oiseaux sont parmi les formes les plus visibles de la vie animale. Les oiseaux ont un rôle majeur dans la biodiversité. La consommation d'insectes

nuisibles à l'agriculture, la dispersion des graines pour le reboisement et la pollinisation des plantes, voici quelques faits essentiels démontrant l'utilité des oiseaux [7].

2.2.2. Le bruit d'environnement :

Lors de l'enregistrement des sons d'oiseaux vivant dans un milieu bruyant, il est difficile d'obtenir des sons clairs sans aucun bruit interférent.

Dans un milieu naturel, les enregistrements des sons d'oiseaux sont perturbés par plusieurs sources de bruit, tels que la présence humaine, le son interférent produit par d'autres animaux, et aussi le son des phénomènes naturels comme par exemple (le soufflement des vents, la pluie, le tonnerre) et les mouvements de la végétation provoqués par les phénomènes naturels.

Une étude sur l'impact du bruit de fond environnemental sur l'interprétation de chants territoriaux a été examinée, dont une analyse des niveaux de pression acoustique a révélé que, dans les endroits moins bruyants, les mâles chantaient avec des niveaux sonores supérieurs à ceux des oiseaux dans les territoires moins affectés par les bruits de fond. Le bruit de l'environnement, quelle que soit sa nature, rend difficile, voire impossible la mesure et l'acquisition des sons d'oiseaux dans leurs habitats naturels [7].

A fin de diminuer les impacts du bruit sur la vocalisation des oiseaux pour obtenir des bons résultats, il est essentiel d'utiliser des appareils spéciaux, tels que des microphones directionnels, pour assurer un bon SNR.

On peut diviser les bruits d'environnement selon leur évolution temporelle en trois catégories : Le bruit continu, le bruit intermittent et le bruit impulsif [7].

2.3. Système de reconnaissance automatique des sons des oiseaux

Dans cette partie, nous présentons les trois étapes du système de reconnaissance automatique des sons des oiseaux qui sont : la segmentation, l'extraction des caractéristiques et la classification.

2.3.1. Segmentation

La segmentation audio est une étape indispensable pour identifier les segments significatifs pour les évènements audio distincts existant dans le signal d'entrée. Elle consiste à diviser les données brutes en des segments, appelés scènes auditives.

Une segmentation précise est également une étape cruciale dans le système de reconnaissance parce qu'elle conditionne les performances du classificateur. Si la segmentation est imprécise elle peut accroître le bruit dans les caractéristiques et conduire à des erreurs de classification.

Il est plus facile de construire des systèmes d'analyse et de classification des objets segmentés que pour les données brutes. Cette étape réduit l'espace de données et les calculs en phase d'analyse (caractérisation) si elle est faite après segmentation [7].

a. Méthode adoptée : segmentation et débruitage audio par seuillage de bloc temps-fréquence

La suppression du bruit des signaux audio nécessite un traitement non diagonal des coefficients temps-fréquence afin d'éviter la production du bruit. Les algorithmes de l'état de la technique effectuent un filtrage paramétré des coefficients du spectrogramme avec des paramètres fixés de manière empirique. Une procédure d'estimation de seuil par bloc est introduite, qui ajuste tous les paramètres de manière adaptative à la propriété du signal en minimisant une estimation du risque.

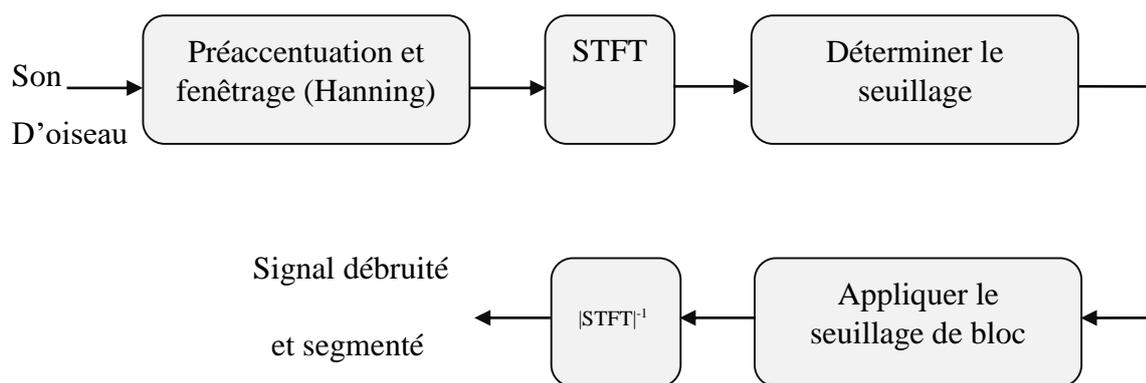


Figure 2.1: Étapes de segmentation et débruitage audio par seuillage de bloc temps-fréquence

b. Débruitage audio temps-fréquence

Les procédures de réduction de bruit audio temps-fréquence calculent une transformation de Fourier à court terme (STFT), et traitent les coefficients résultants afin d'atténuer le bruit. Le signal audio est contaminé par un bruit qui est souvent modélisé comme un processus gaussien à moyenne nulle, indépendamment de:

$$y[n] = f[n] + e[n], \quad n = 0, 1, \dots, N - 1 \tag{2.1}$$

Le STFT décompose le signal audio sur une famille d'atomes temps-fréquence

$$g_{l,k}[n] = w[n - lu] \exp\left(\frac{i2\pi kn}{K}\right) \tag{2.2}$$

Où l et k sont respectivement les indices de temps et fréquence, et $w[n]$ est une fenêtre temporelle du support de la taille K , c'est le racine carrée de la fenêtre de *Hanning* qui est décalée d'un pas : $u \leq K$.

Un algorithme de débruitage modifie les coefficients temps-fréquence en multipliant chacun d'eux par un facteur d'atténuation $a[l, k]$ afin d'atténuer la composante de bruit.

Les algorithmes de débruitage temps-fréquence diffèrent par le calcul des facteurs d'atténuation $a[l, k]$. La variance du coefficient de bruit est définie comme :

$$\sigma^2[l, k] = E\{|e, g_{l,k}\rangle|^2\} \tag{2.3}$$

c. Estimation non diagonal

Afin de réduire le bruit ainsi que le risque d'estimation, l'estimation du SNR a priori avec une régularisation temps-fréquence du SNR a posteriori a été proposé. Les facteurs d'atténuation résultants dépendent des valeurs de données dans un voisinage entier et l'estimateur résultant est dit non diagonal. A est le facteur de redondance.

$$f'[n] = \frac{1}{A} \sum_{l,k} a[l, k] Y[l, k] g_{l,k}[n] \tag{2.4}$$

Matz et Hlawatsch [3] ont également proposé d'estimer le SNR avec un filtre temps-fréquence rectangulaire et de l'utiliser avec estimateur empirique de *Wiener* pour montrer un gain de performance notable par rapport à une estimation de SNR en diagonale.

La régularisation de l'estimation du SNR réduit le bruit musical ainsi que le risque d'estimation est défini comme suit :

$$r = E\{\|\hat{f} - f\|^2\} \tag{2.5}$$

d. Seuillage par blocs temps-fréquence

Pour le débruitage du signal audio, nous décrivons un estimateur non diagonal à seuillage de blocs adaptatif qui ajuste automatiquement tous les paramètres. Il repose sur la possibilité de calculer une estimation du risque, sans modèle de signal audio stochastique préalable, ce qui rend cette approche particulièrement robuste.

e. Algorithme de seuillage par bloc

Un estimateur de seuillage par blocs temps-fréquence permet de régulariser l'estimateur de soustraction de puissance en calculant un seul facteur d'atténuation sur des blocs temps-fréquence. Le plan temps-fréquence $\{l, k\}$ est segmenté en I blocs B_i dont la forme peut être choisie arbitrairement. L'estimateur de signal \hat{f} est calculé à partir des données de bruit y avec un facteur d'atténuation constant a_i sur chaque bloc B_i .

$$\hat{f}[n] = \sum_{i=1}^I \sum_{(l,k) \in B_i} a_i Y[l, k] g_{l, k}[n] \tag{2.6}$$

Un estimateur par seuillage de bloc peut donc être interprété comme un estimateur non diagonal dérivé d'estimations SNR moyennes sur des blocs. Chaque facteur d'atténuation est calculé à partir de tous les coefficients de chaque bloc, ce qui régularise l'estimation du coefficient temps-fréquence.

f. Seuillage de bloc adaptatif

Un seuillage de bloc segmente le plan temps-fréquence en blocs rectangulaires disjoints de longueur L_i en temps et de largeur W_i en fréquence. Dans ce qui suit, «taille de bloc» signifie un choix de formes et de tailles de bloc parmi un ensemble de possibilités. Le seuillage de bloc adaptatif choisit les tailles en minimisant une estimation du risque.

Le risque $E\{\|\hat{f} - f\|^2\}$ ne peut pas être calculé car f est inconnu, mais il peut être estimé. Les meilleures tailles de blocs sont calculées en minimisant ce risque estimé.

Pour régulariser la segmentation adaptative en blocs, le plan temps-fréquence est d'abord décomposé en macro-blocs, comme illustré à la Figure 2.2. Chaque macro-bloc est segmenté en blocs de même taille, ce qui signifie qu'il est constant sur un macro-bloc. Plusieurs segmentations de ce type sont possibles et nous voulons choisir celle qui conduit à la plus petite estimation du risque.

$Prob \{ \bar{e}^2 > \lambda \sigma^2 \}$ est la probabilité de conserver un bruit résiduel. Ajuster λ et la taille des blocs B peut être interprété comme une optimisation entre le biais et la variance de notre estimateur de seuillage de blocs. Le paramètre λ est défini en fonction de B en ajustant la probabilité de bruit résiduel.

$$Prob \{ \bar{e}^2 > \lambda \sigma^2 \} = \sigma^2 \tag{2.7}$$

La probabilité σ est un paramètre perceptuel. Certaines des spécifications sur le choix des paramètres sont discutées ci-dessous.

g. Choix de bloc

On regroupe les coefficients de Fourier à court terme contigus temps-fréquence dans des blocs rectangulaires disjoints. La taille du bloc est $B_i = L_i * W_i$, où L_i et W_i sont respectivement la longueur du bloc en temps et la largeur du bloc en fréquence.

h. Choix du niveau de seuillage λ

Soit un choix de taille de bloc et de niveau de probabilité de bruit résiduel σ que l'on tolère, le niveau de seuillage λ . Pour chaque largeur et longueur de bloc, λ est estimé.

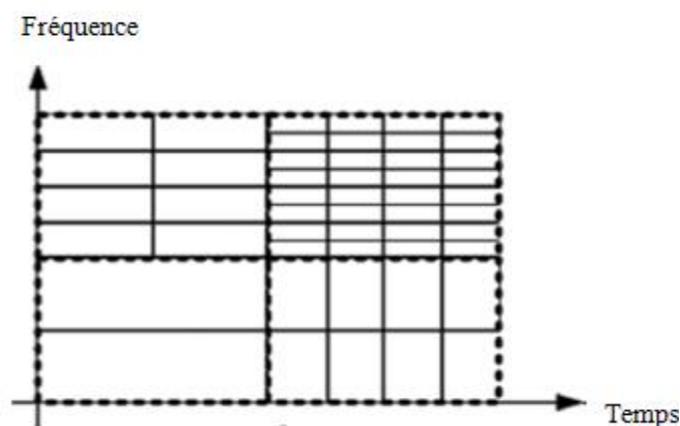


Figure 2.2 : Partition de blocs macro

2.3.2. Extraction des caractéristiques

L'extraction de caractéristiques discriminantes est une phase fondamentale du système de reconnaissance audio, préalable à la classification. Elle crée des vecteurs caractérisant les signaux audio segmentés. Depuis plus de deux décennies, les coefficients cepstraux de fréquence à échelle Mel (MFCC) sont les caractéristiques acoustiques les plus largement utilisées dans les systèmes de reconnaissance automatique de la parole et de son d'oiseaux, à cause de leurs bonnes performances de précision. Ces coefficients nous permettent de réduire la quantité d'information nécessaire pour décrire un signal.

a. Caractérisation par MFCC

La caractérisation par les coefficients (MFCC) est une caractéristique largement utilisée dans l'analyse de signaux audio. Bien qu'elle soit principalement conçue pour la reconnaissance automatique de la parole.

Dans le traitement du son, Les MFCC sont les coefficients qui constituent collectivement le Cepstrum de fréquence de Mel (MFC), représentation du spectre de puissance à court terme du son basé sur la transformation linéaire en cosinus d'un spectre de puissance logarithmique sur une échelle de fréquence logarithmique Mel non linéaire. Les MFCC sont linéairement espacés sur l'échelle de fréquence Mel qui se rapproche beaucoup de la réponse du système auditif humain. Cette représentation du signal sonore extrait des caractéristiques discriminantes qui permettent d'atteindre une classification du son avec une bonne précision.

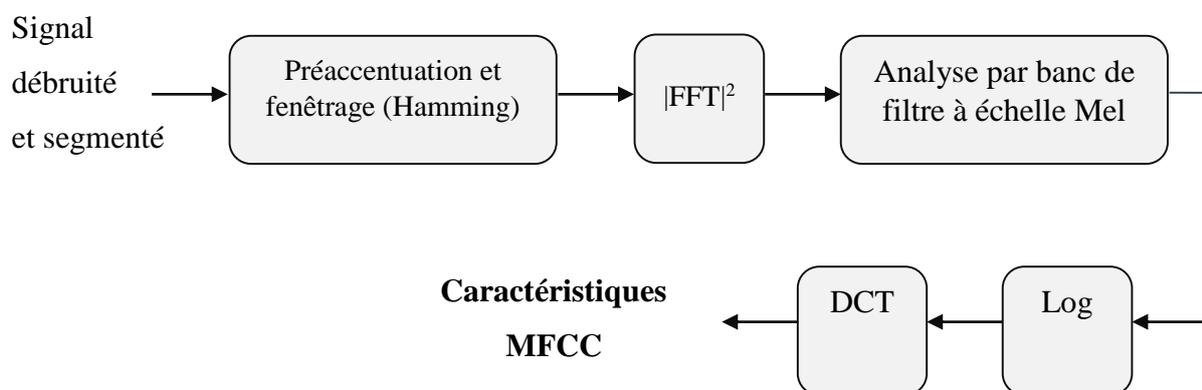


Figure 2.3 : La structure de l'extraction des caractéristiques MFCC

L'échelle de Mel est définie comme suit:

$$f_{mel} = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right) \quad (2.8)$$

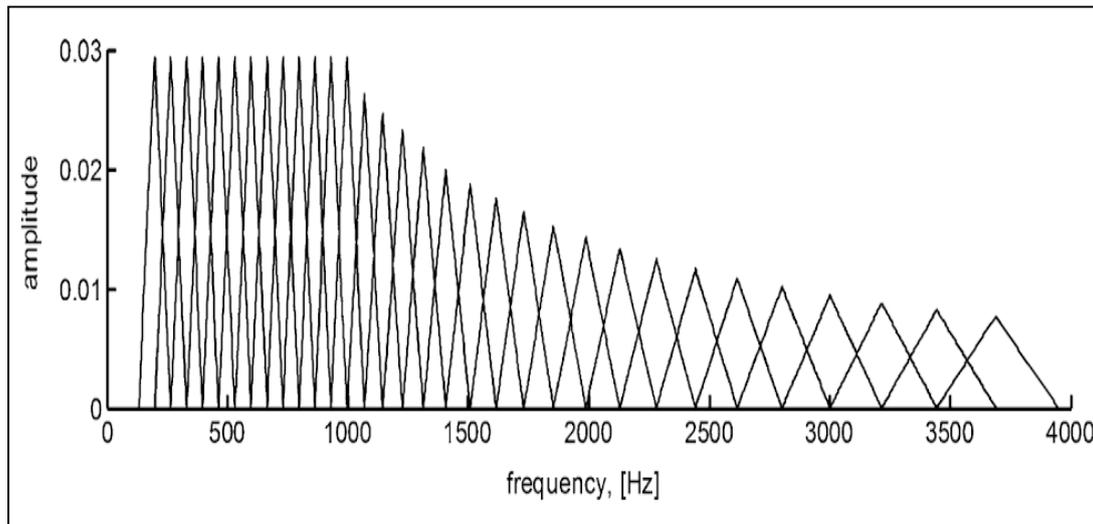


Figure 2.4 : Banc de filtre (Mel) des coefficients MFCC

Les caractéristiques sont extraites au cours de plusieurs étapes successives, comme illustré dans la figure 2.3:

- Dans la première étape, un filtre passe-haut de préaccentuation de la forme $H(z) = 1.97z^{-1}$ est appliqué. L'analyse par transformée de Fourier à court terme (STFT) est effectuée à l'aide de fenêtres de Hamming d'une durée de 20 ms, espacées de 10 ms entre elles. Le spectre de puissance basé sur le périodogramme est estimé pour chaque trame segmentée.
- Dans l'étape suivante, un ensemble de bancs de filtres à l'échelle de Mel est calculé et appliqué au spectre de puissance de chaque cadre. Pour calculer le spectre d'énergie du banc de filtre, chaque filtre est multiplié par le spectre de puissance et les coefficients sont additionnés. La sortie de bancs de filtres est transformée en une échelle logarithmique.
- Dans la dernière étape, la transformée en cosinus discrète (DCT) est appliquée sur les amplitudes logarithmique du banc de filtre pour obtenir les vecteurs caractéristiques MFCC.

b. Caractéristiques HFCC (Human Factor Cepstral Coefficients)

Les coefficients Cepstral de facteur humain (HFCC) introduits en 2004 par Skowronski et Harris [9] constituent la dernière mise à jour du banc de filtres MFCC afin d'améliorer les performances en termes de résistance au bruit. La différence la plus significative entre le HFCC et le MFCC est que, dans le HFCC, la largeur de bande du filtre est découplée de l'espacement du filtre. Plus précisément, la bande passante en HFCC est un paramètre de conception indépendant de l'espacement des filtres. Pour déterminer la largeur de bande, HFCC utilise l'approximation de la largeur de bande critique de mesurée en largeur de bande rectangulaire équivalente ERB.

$$ERB = 6.23 \cdot f^2 + 93.39 \cdot f + 28.52 \text{ Hz} \tag{2.9}$$

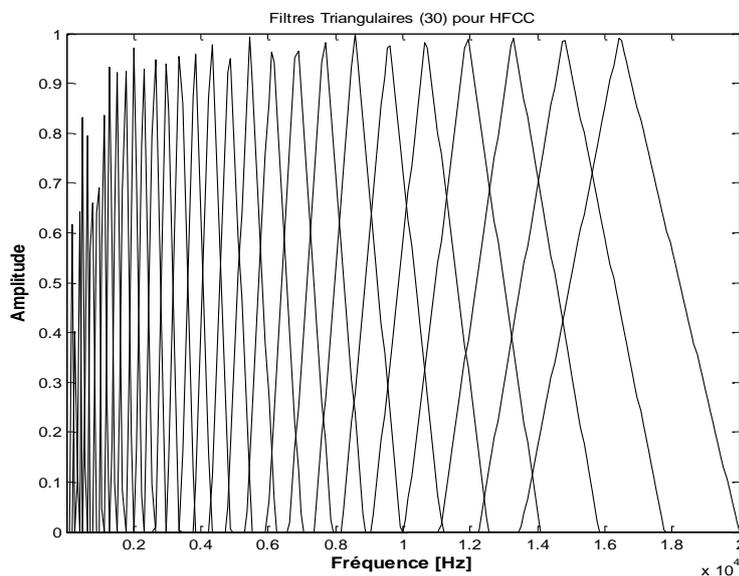


Figure 2.5 : Le banc de filtres (ERB) des coefficients HFCC

Où la fréquence centrale f est en kHz avec l'ajustement de la courbe valide est comprise entre 0,1 et 6,5 kHz. L'ERB d'un filtre passe-bande est la largeur d'un rectangle dont la hauteur est égale au maximum de la réponse d'amplitude du filtre et dont l'aire est identique à la réponse du filtre. L'ERB est utilisé comme alternative aux points à 3 dB pour décrire la largeur de bande du filtre. Les fréquences centrales du filtre sont également espacées de manière égale dans la fréquence mel en reliant la fréquence mel f' à la fréquence linéaire f .

$$f' = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \tag{2.10}$$

Les étapes d'extraction des coefficients HFCC sont présentées comme suit, figure 2.4

- Premièrement, une application de préaccentuation est appliquée au signal audio avant l'analyse spectrale à court terme pour augmenter l'amplitude des hautes fréquences. Un filtre passe-haut de préaccentuation de la forme $H(z) = 1.95z^{-1}$ est appliqué.
- Ensuite, l'analyse par transformée de Fourier à court terme (STFT) est effectuée à l'aide de fenêtres de Hamming d'une durée de 20 ms, espacées de 10 ms entre elles. Le spectre d'amplitude est calculé et mis à l'échelle par le banc de filtres triangulaire.
- Dans l'étape suivante, un ensemble de bancs de 40 filtres de facteur humain est calculé et appliqué pour filtrer le signal de son d'oiseau. La somme des échantillons mis à l'échelle par chaque filtre du banc, la sortie de bancs de filtres est transformée en une échelle logarithmique.
- Après, une application de la transformée en cosinus discrète (DCT) est effectuée pour obtenir les coefficients cepstraux HFCC.
- Finalement, les treize premiers coefficients ont été choisis et le premier coefficient a été remplacé par une énergie logarithmique de chaque trame. Le vecteur de caractéristiques HFCC comprend également douze coefficients.

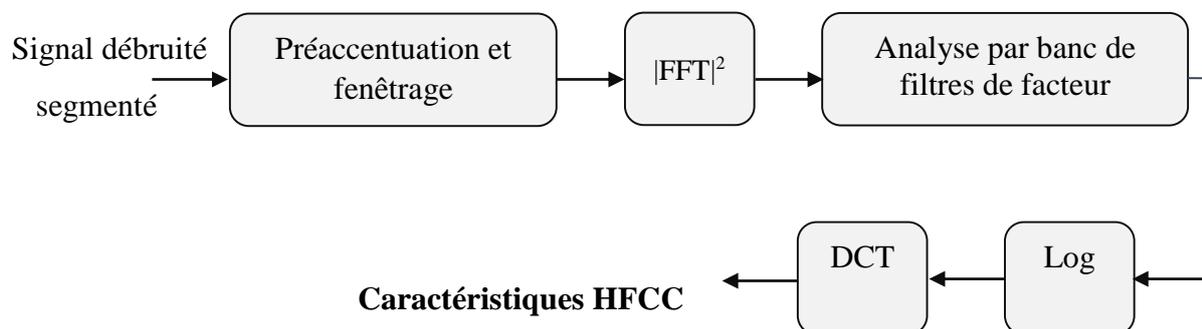


Figure 2.6 : La structure de l'extraction des caractéristiques HFCC

La conception du banc de filtres HFCC comprend les étapes suivantes : On choisit tout d'abord les limites basses f_{\min} , et hautes f_{\max} de l'ensemble du banc de filtres, et le nombre M de filtres est choisi.

Les fréquences centrales f_{c1} et f_{cM} du premier et du dernier des filtres, respectivement, sont calculées comme suit:

$$f_{ci} = \frac{-\bar{b} \pm \sqrt{\bar{b}^2 - 4\bar{c}}}{2} \tag{2.10}$$

Où l'indice i est 1 ou M et, \bar{b} , \bar{c} définis comme suit:

$$\bar{b} = \frac{b-\hat{b}}{a-\hat{a}} \text{ Et } \bar{c} = \frac{c-\hat{c}}{a-\hat{a}} \quad (2.11)$$

Recevoir des valeurs différentes pour les deux cas. Les valeurs a, b, c sont celles de (2.9), 6.23, 93.39, 28.52, respectivement. Pour le premier filtre, les valeurs des coefficients $\hat{a}, \hat{b}, \hat{c}$, sont calculées comme suit:

$$\hat{a} = \frac{1}{2} \frac{1}{700+fmin}, \hat{b} = \frac{700}{700+fmin}, \hat{c} = \frac{fmin}{2} \left(1 + \frac{700}{700+fmin}\right) \quad (2.12)$$

Pour le dernier filtre ce sont:

$$\hat{a} = \frac{1}{2} \frac{1}{700+fmax}, \hat{b} = \frac{700}{700+fmax}, \hat{c} = \frac{fmax}{2} \left(1 + \frac{700}{700+fmax}\right) \quad (2.13)$$

Une fois que les fréquences centrales des premier et dernier filtres sont calculées, les fréquences centrales des filtres situés entre eux sont faciles à calculer car elles sont équidistantes sur l'échelle de Mel. Le pas $\Delta\hat{f}$ entre les fréquences centrales des filtres adjacents est calculé comme suit:

$$\Delta\hat{f} = (\hat{f}_{cM} - \hat{f}_{c1}) / (M - 1) \quad (2.14)$$

où toutes les fréquences sont en mel. Les conversions $f_{c1} \rightarrow \hat{f}_{c1}$ et $f_{cM} \rightarrow \hat{f}_{cM}$ sont données par (2.9). Ayant $\Delta\hat{f}$, les fréquences centrales \hat{f}_{ci} sont calculées comme suit:

$$\hat{f}_{ci} = \hat{f}_{c1} + (i - 1) \cdot \Delta\hat{f}, \text{ pour } i = 2, \dots, M - 1 \quad (2.15)$$

Ensuite, à travers(2.16), la transformation inverse $\hat{f}_{ci} \rightarrow f_{ci}$ est effectuée,

$$f_{ci} = 700 \cdot (10^{\hat{f}_{ci}/2595} - 1) \quad (2.16)$$

et à travers (2.9) la transformation ERB_i est effectuée pour chaque f_{ci} . Enfin, les basses et hautes fréquences $fmin_i$ et $fmax_i$ respectivement du i filtre sont déduites de:

$$fmin_i = -(700 + ERB_i) + \sqrt{(700 + ERB_i)^2 + fci(fci + 1400)} \quad (2.17)$$

$$fmax_i = fmin_i + 2 \cdot ERB_i \quad (2.18)$$

Avec tous les paramètres calculés par (2.10) ÷ (2.18), la conception du banc de filtres HFCC est terminée.

2.3.3. Classification

Le classificateur utilise les coefficients acoustiques pour décider la classe à laquelle appartient le modèle de test par rapport à un ensemble de formation. Ceci est réalisé en comparant la meilleure correspondance entre un échantillon de test et un modèle de classe ou un modèle cible.

Dans les tâches de classification, l'espace vectoriel est divisé en régions correspondant à différentes classes. Lors de la classification des syllabes d'oiseaux, chaque classe (représentant une espèce d'oiseau) peut comprendre différents types de représentations de caractéristiques, chacune correspondant à un type de syllabes différent pour chaque espèce. Dans ce cas, certains modèles peuvent avoir la même étiquette que la classe.

Plusieurs méthodes de classification sont utilisées pour la reconnaissance des sons des oiseaux, y compris la déformation temporelle dynamique DTW (Dynamic Time Warping).

a. Déformation temporelle dynamique (DTW)

La déformation temporelle dynamique est un algorithme permettant de prendre la décision de reconnaissance des données qui peuvent varier au cours du temps.

Le principe de base de cette technique est de trouver l'alignement global optimal entre l'échantillon de test et l'échantillon de référence, tout en créant une base de données contenant des échantillons de référence marqués, lors de l'identification de l'échantillon de test, on essaye d'associer l'échantillon de la référence la plus proche à l'échantillon de test en minimisant le coût d'association.

Ceci peut être visualisé par la construction d'une matrice, avec le délai temporel d'un échantillon sur l'axe des abscisses « x » et de l'autre échantillon sur l'axe des ordonnées « y ».

Le but est de trouver le chemin le plus court, commençant en bas à gauche et se terminant en haut à droite [2]. Le coût d'une association correspond à la distance entre les deux éléments (allant vers le haut, à la droite ou en diagonale: vers la droite et en haut) a un coût qui lui est associé. Ce principe est résumé en ces deux étapes :

Premièrement :

Supposons qu'on a deux séquences *I* et *J*

On créer une matrice : $g = I * J$ (2.19)

Où *I* et *J* sont la taille de la séquence 1 et de la séquence 2 respectivement sur l'axe du temps.

Où la distance entre les éléments de la matrice est calculée comme suit :

$$D(i, j) = dist(i, j) + \min \begin{cases} D(i - 1, j) \\ D(i, j - 1) \\ D(i - 1, j - 1) \end{cases} \quad (2.20)$$

Où $D(i, j)$ est la distance de déformation optimale entre 2 séquences de taille *I* et *J*.

Deuxièmement :

Construire un chemin ou un alignement basé sur la matrice *g*.

L'algorithme DTW est soumis à trois conditions qui leur assurent une convergence rapide:

1. monotonie - le chemin ne revient jamais, cela signifie que les deux indices *I* et *J* utilisés pour le passage en séquence ne diminuent jamais.
2. continuité - le chemin avance progressivement, pas à pas; les indices *i* et *j* augmentent d'un maximum de 1 unité sur une marche.
3. Limite - le chemin commence dans le coin inférieur gauche et se termine dans le coin supérieur droit. De (1,1) à (*i, j*).

Le déplacement en diagonale est généralement le moins coûteux. Pour chaque point dans le treillis, la distance entre les deux vecteurs de caractéristiques associés aux trames temporelles correspondantes aux échantillons d'entrée est calculée et ajouté au coût. De cette façon, un compromis est trouvé entre un chemin simple (autant de diagonales que possible) et correspondant aux données [2].

2.4. Conclusion

Dans ce chapitre, nous avons présenté un système de reconnaissance audio des oiseaux constitué des étapes suivantes :

- Une méthode de segmentation des sons des oiseaux, basés sur le débruitage audio par seuillage de bloc temps-fréquence.

- Dans la phase d'extraction de caractéristiques nous nous sommes basé sur la méthode HFCC (Human Factor Cepstral Coefficients) utilisant un banc de filtre de facteur humain.
- Pour la classification, nous avons opté pour le classificateur DTW

Pour pouvoir évaluer les performances de la méthode utilisée HFCC, nous allons comparer cet algorithme avec trois autres méthodes de caractérisation utilisées dans la reconnaissance de la parole et des sons d'oiseaux qui sont MFCC, PNCC et PMVDR.

Chapitre 3
Simulations et
Résultats

3.1. Introduction

Dans ce dernier chapitre, nous allons présenter notre base de données et les oiseaux étudiés dans ce travail. Ensuite, nous allons discuter les résultats atteints et les performances de chaque étape du système de la reconnaissance des sons d’oiseaux proposé.

3.2. Base de données et oiseaux étudiés

Dix espèces d'oiseaux faisaient partie de la base de données dans ce travail. Ces espèces d'oiseaux sont courantes dans les zones humides du Nord-Est de l’Algérie.

	<i>Nom scientifique</i>	<i>Nom commun</i>	<i>Enregistrements</i>	
			<i>Test</i>	<i>Apprentissage</i>
1	<i>Anas acuta</i>	Canard pilet	3	9
2	<i>Anas crecca</i>	Sarcelle d’hiver	5	10
3	<i>Aythya fuligula</i>	Fuligule morillon	5	9
4	<i>Aythya nyroca</i>	Fuligule nyroca	2	4
5	<i>Fulica atra</i>	Foulque macroule	4	9
6	<i>Ixobrychus minutus</i>	Blongios nain	3	7
7	<i>Mareca penelope</i>	Canard siffleur	3	7
8	<i>Oxyura leucocephala</i>	l’Erismature à tête blanche	1	3
9	<i>Phalacrocorax carbo</i>	Grand cormoran	3	6
10	<i>Plegadis falcinellus</i>	Ibis falcinelle	3	7

Tableau 3.1 : Désignation des catégories d’espèces d’oiseaux étudiés. Les colonnes donnent les catégories des espèces étudiées, leur nom scientifique, leur nom commun et le nombre d'enregistrements utilisés.

Des enregistrements audio de ces espèces ont été pris de l’archive Xeno-canto⁴. La fréquence d’échantillonnage pour les formes d'onde du son d'oiseau est 44, 1 KHz avec 16 bits pour chaque échantillon et sont enregistrés sur leurs environnements naturels, et donc ils contiennent des syllabes similaires (des cris et des chants de diverses espèces d’oiseaux et des bruits causés par la nature tels que le vent, la pluie). Tous ces fichiers audio sont convertis en format « wav » monophoniques.

⁴. Un projet de science citoyenne dans laquelle les bénévoles enregistrent, téléchargent et annotent des enregistrements de chants et cris d’oiseaux. Tous les enregistrements sont publiés sous licences, dont certaines sont ouvertes. Le site officiel (www.xeno-canto.org) est soutenu par un certain nombre d’institutions académiques et ornithologiques dans le monde entier, avec son support principal étant aux Pays-Bas.

Les éléments de base de données sont générés de manière déterministe et fournissent donc un moyen de déterminer la sensibilité de la procédure de mise en correspondance. Les désignations des catégories d'espèces d'oiseaux sont présentées au tableau 3.1. Dans cette étude, nous avons d'abord appliqué à la base de données une méthode de segmentation et débruitage temps-fréquence par seuillage de bloc, et quatre méthodes pour extraire les caractéristiques (HFCC, MFCC, PNCC et PMVDR) utilisées pour la reconnaissance automatique du son des oiseaux, puis avec la méthode de DTW pour la classification des caractéristiques obtenus.

Cent trois enregistrements audio étaient disponibles pour chacune des 10 espèces. Nous avons utilisé 32 enregistrements pour la création de modèles et de 71 enregistrements d'apprentissage.

Après l'étape de segmentation, quatre méthodes d'extraction de caractéristiques décrites dans le tableau 3.2 sont utilisées.

Paramètres	HFCC	MFCC	PMVDR	PNCC
Prétraitement	Oui	oui	oui	oui
Fenêtrage	Hamming	Hamming	Hamming	Hamming
Taille trame	25.6 ms	25.6 ms	25.6 ms	25.6 ms
Espacement	10 ms	10 ms	10 ms	10 ms
Banc de filtres	Triangulaire	Triangulaire	/	Gammatone
Echelle Fréq.	facteur humain	Mel	Mel	ERB
No. Bandes	40	24	/	24
Compression	log	log	/	Non-linéaire
Coefficients	13 + Δ + $\Delta\Delta$			

Tableau 3.2 : Paramètres de configuration pour les stratégies de caractérisation considérées

3.3. Résultats et discussion

3.3.1. Performance de la méthode de segmentation et de débruitage

Les figures 3.1 et 3.2 et 3.3 présentent l'enregistrement du son de Fuligule morillon d'une durée de 7 seconds et un taux d'échantillonnage de 44100 Hz perturbé par trois bruits d'environnement : le sifflement du vent comme un bruit continu, accident de voiture comme bruit intermittent, et l'explosion d'une bombe comme bruit impulsif respectivement. Avec un SNR variant entre « 0 dB à 15 dB ».

Un bruit continu (le sifflement du vent) a été rajouté au modèle de test avec un SNR variant entre « 0 et 15 dB », le signal audio du son bruité est représenté sur les figures 3.1.1, 3.1.2, 3.1.3 respectivement.

En appliquant la méthode proposée pour la segmentation des sons des oiseaux basée sur le débruitage audio temps-fréquence par seuillage par blocs après l'ajout du bruit continu nous obtenons un signal débruité constitué de segments contenant les informations. Les figures représente le signal audio du son débruité segmenté de l'oiseau après l'application de la méthode utilisée.

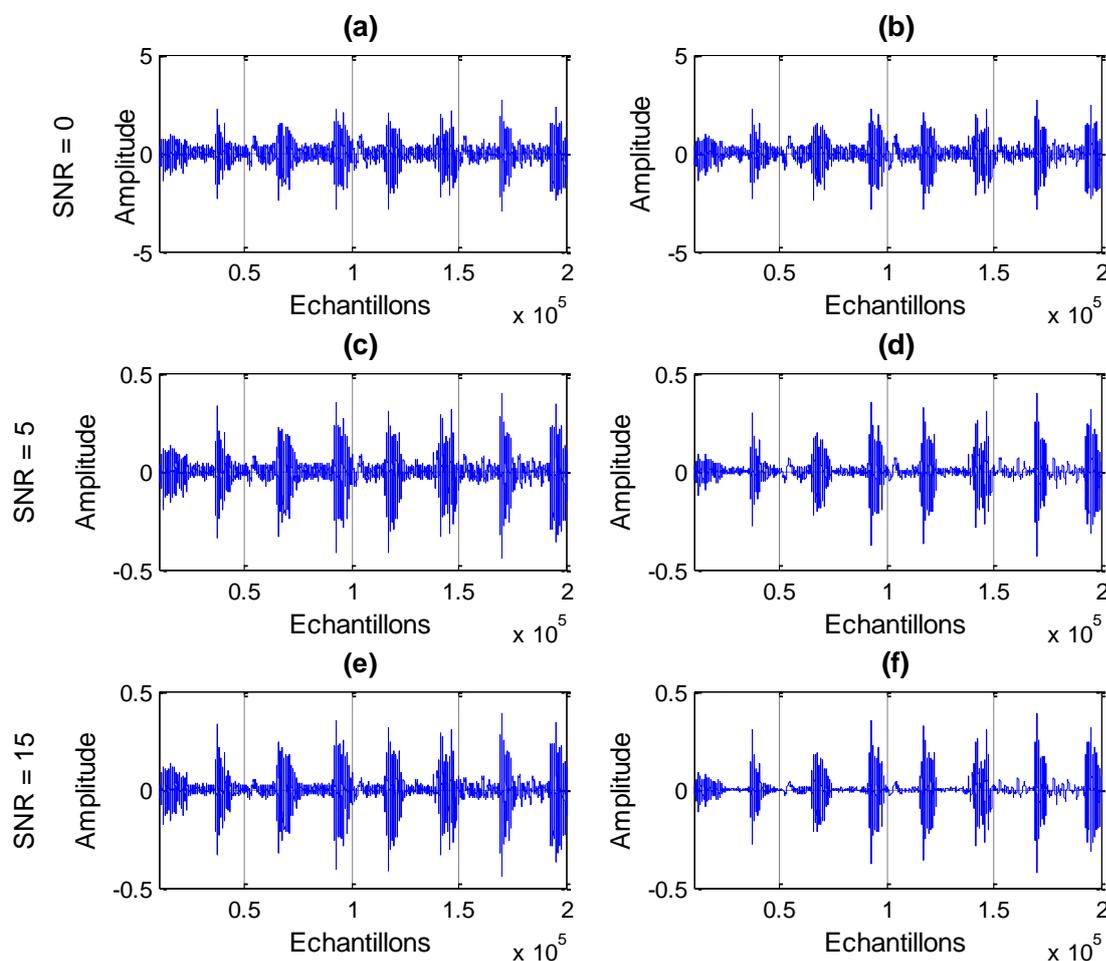


Figure 3.1 : Représentation du signal audio de Fuligule morillon perturbé par le bruit continu (vent) à un SNR = 0 dB (a), à un SNR = 5 dB (c), à un SNR = 15 dB (e). (b), (d), (f) après l'application de la méthode de débruitage temps-fréquence par seuillage de bloc.

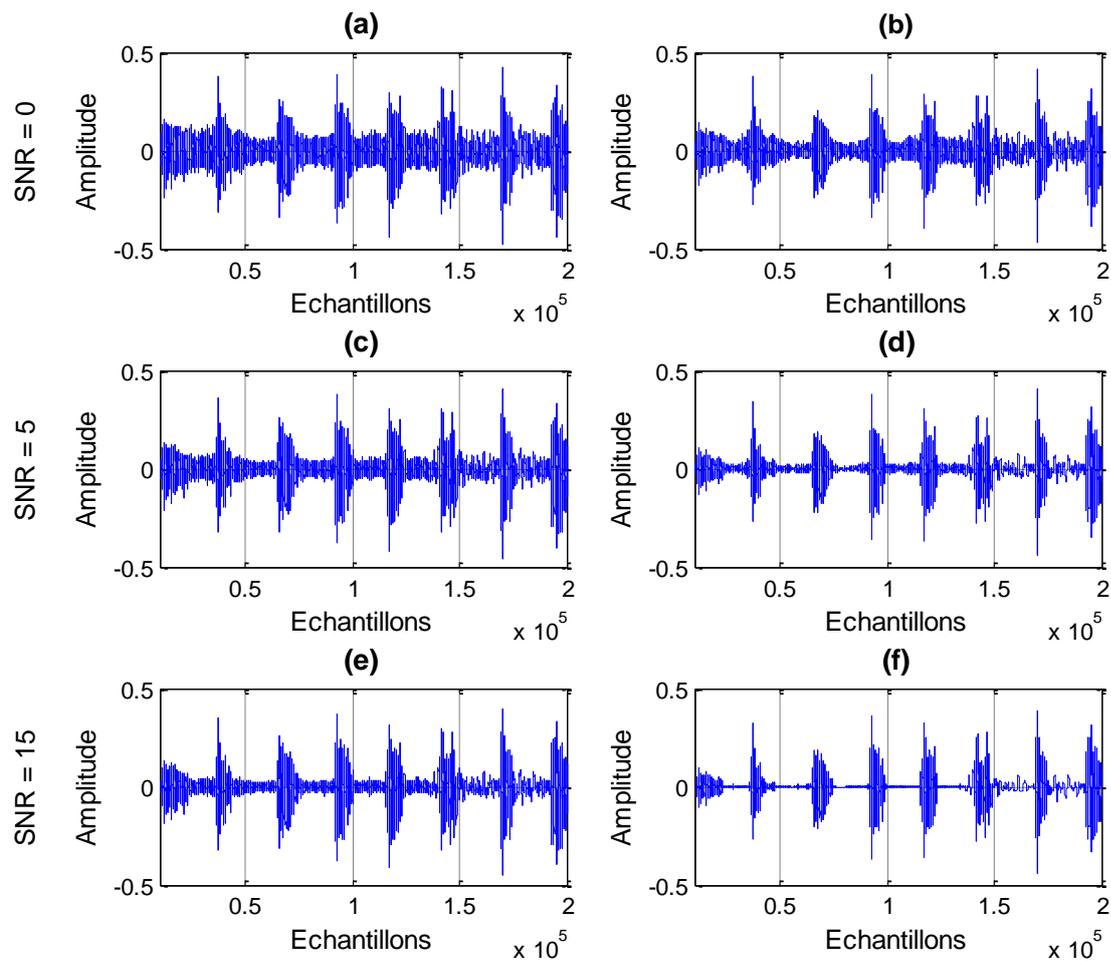


Figure 3.2 : Représentation du signal audio de Fuligule morillon perturbé par le bruit intermittent (accident de voiture) à un SNR = 0 dB (a), à un SNR = 5 dB (c), à un SNR = 15 dB (e). (b), (d), (f) après l’application de la méthode de débruitage temps-fréquence par seuillage de bloc.

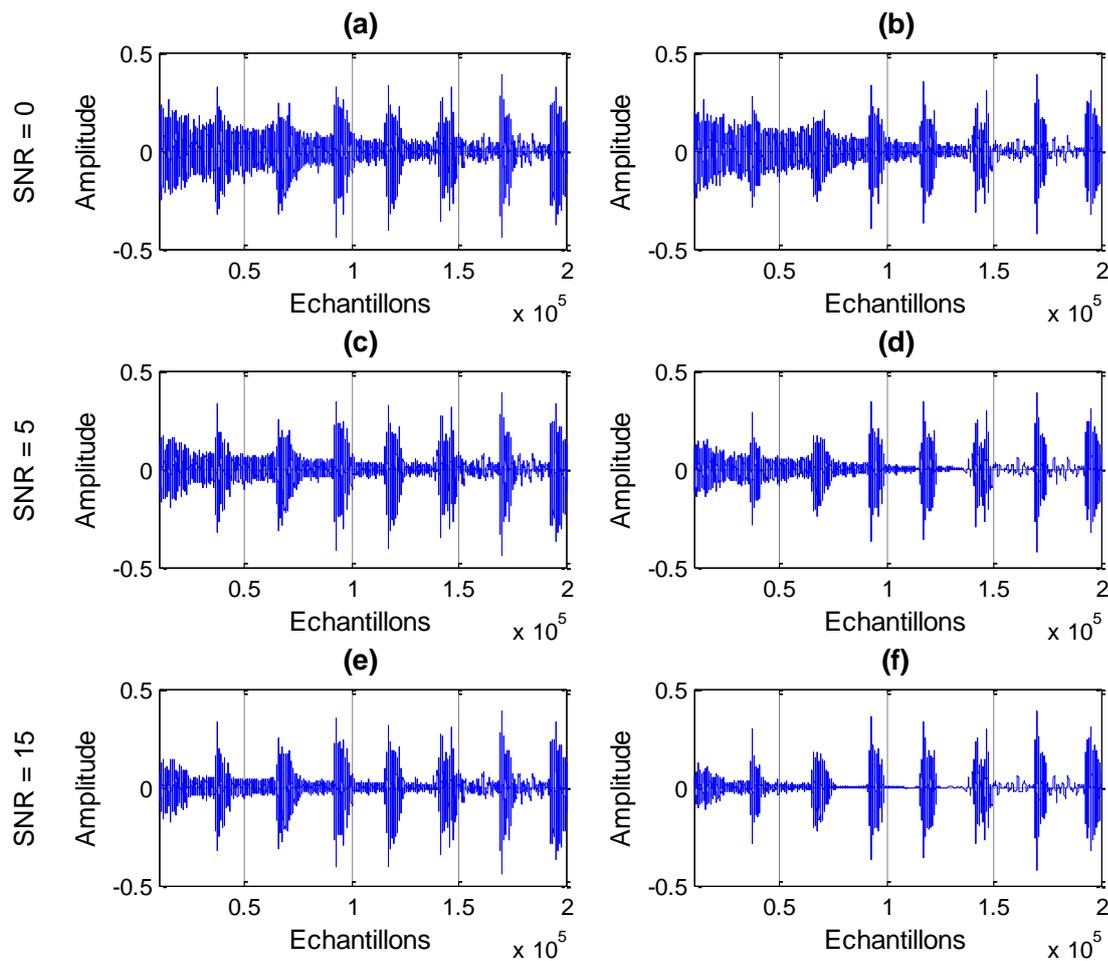


Figure 3.3 : Représentation du signal audio de Fuligule morillon perturbé par le bruit impulsif (explosion d’une bombe) à un SNR = 0 dB (a), à un SNR = 5 dB (c), à un SNR = 15 dB (e). (b), (d), (f) après l’application de la méthode de débruitage temps-fréquence par seuillage de bloc.

À chaque fois que le SNR diminue, l’effet du bruit de l’environnement sur le signal audio du son d’oiseau augmente et le signal audio sera plus perturbé sous le bruit de l’environnement.

3.3.2. Extraction des caractéristiques

3.3.2.1. Caractéristiques MFCC

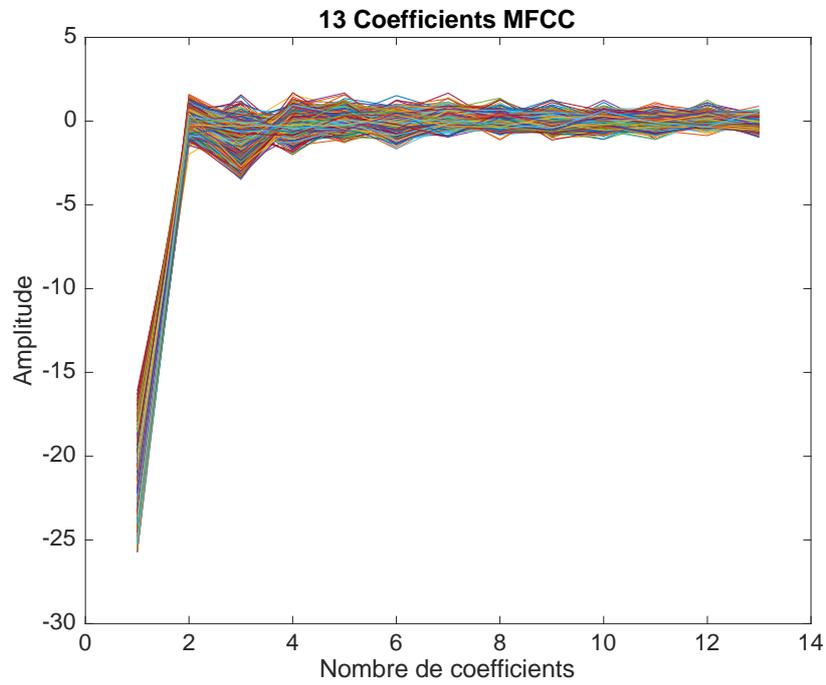


Figure 3.4 : Vecteurs de caractéristiques MFCC.

3.3.2.2. Caractéristiques HFCC

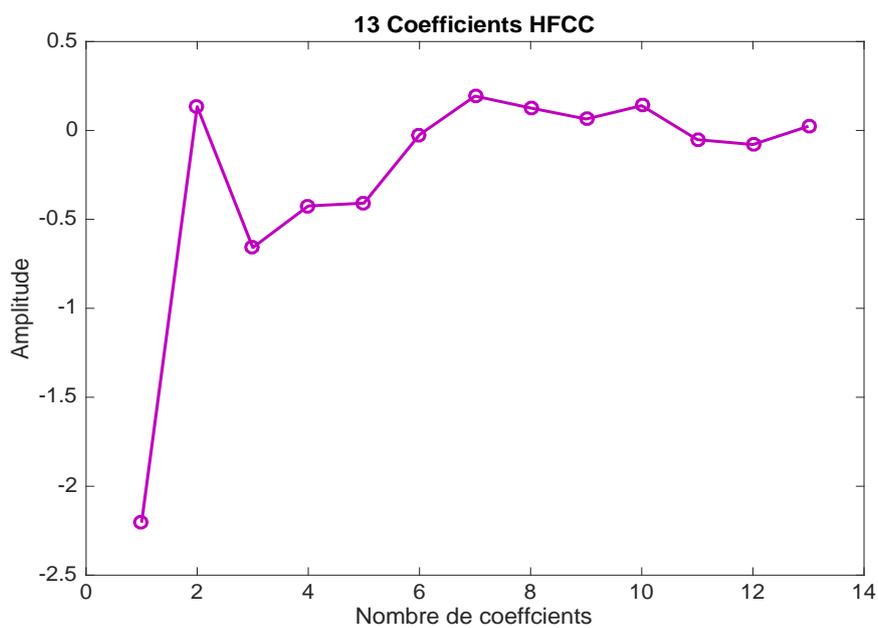


Figure 3.5 : Vecteur de caractéristiques HFCC

3.3.3. L'impact de la méthode de segmentation sur le taux de reconnaissance

Dans cette partie nous allons étudier l'impact de la méthode de segmentation sur notre système. Le taux de reconnaissance pour les méthodes de d'extraction de caractéristiques est illustré dans la figure 3.6.

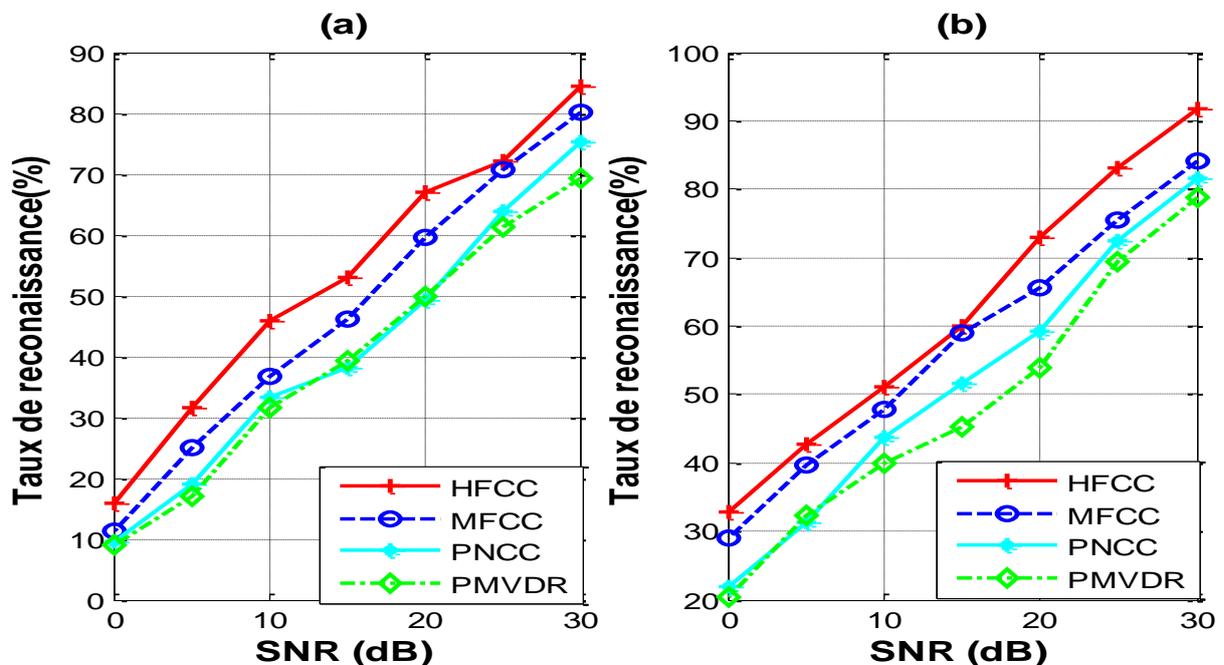


Figure 3.6: Evolution du taux de reconnaissance des méthodes étudiées à différents niveaux de SNR du bruit d'environnement, (a) avant l'application de la méthode utilisée pour le débruitage et la segmentation du son d'oiseau, (b) après l'application de la méthode utilisée.

D'après la figure 3.6, à chaque fois que le SNR diminue, les exactitudes globales d'identification moyenne des quatre caractéristiques (HFCC, MFCC, PNCC et PMVDR) diminuent sous le bruit de l'environnement, mais la tendance à la chute est évidemment différente.

La tendance de taux de précision avant le débruitage et la segmentation (Figure 3.6 (a)) diminue rapidement, alors que la tendance après l'application de la méthode utilisée (Figure 3.6 (b)) diminue lentement. Tous les taux de reconnaissance atteignent au plus haut, une précision de reconnaissance élevée de 80.45 % pour l'utilisation des Caractéristiques HFCC, et 80.19 % pour les MFCC, suivi par les PNCC avec une précision de 75.17 et à la fin avec 69.36 % pour les caractéristiques PMVDR à un SNR élevé (30 dB), la précision de reconnaissance moyenne le plus faible est de 8.99 % quand le SNR est faible (0 dB).

Après l'application de la méthode de débruitage et segmentation (Figure 3.6 (b)), la précision de reconnaissance augmente et le taux de reconnaissance le plus élevé atteint jusqu'à 91.73 % pour les HFCC.

D'après les résultats obtenus, la méthode adoptée de segmentation et débruitage audio par seuillage de bloc temps-fréquence a une bonne résistance au bruit, en particulier une meilleure robustesse aux faibles SNR ce qui la rend adapter pour l'identification des sons des oiseaux dans des environnements complexes.

3.3.4. L'impact du bruit additif blanc gaussien sur le système proposé

Dans cette partie nous allons étudier l'impact du bruit additif blanc gaussien sur notre système. Le taux de reconnaissance pour les méthodes de d'extraction de caractéristiques (HFCC, MFCC, PNCC et PMVDR), sont illustrées dans la table 3.3.

		SNR (dB)/ Taux de reconnaissance (%)						
		0	5	10	15	20	25	30
Avant segmentation	HFCC	20.63	31.65	44.68	53.20	66.80	72.65	82.90
	MFCC	18.30	28.62	40.21	48.55	53.14	61.74	77.99
	PNCC	16.04	26.95	37.09	42.34	48.61	54.35	69.15
	PMVDR	11.52	21.21	30.36	38.97	42.64	49.54	60.87
Après segmentation	HFCC	32.54	41.32	53.24	59.91	72.35	82.16	90.35
	MFCC	26.24	39.60	47.80	57.70	64.27	74.24	83.67
	P NCC	22.02	30.25	42.47	52.14	60.89	72.65	79.65
	PMVDR	19.51	28.36	39.21	45.24	53.60	68.47	77.60

Tableau 3.3 : Evolution du taux de reconnaissance des méthodes étudiées à différents niveaux de SNR du bruit additif blanc gaussien.

D'après la table 3.3, on peut remarquer que la méthode de caractérisation HFCC est la meilleure et plus efficace en termes de reconnaissance, avant et après l'application de la e méthode de segmentation et débruitage en présence du bruit aditif blanc gaussien. HFCC donne un taux de reconnaissance égal à 82,90% avant la segmentation et une précision de 90.35% après l'application de la segmentation qui est le plus haut taux par rapport aux autres méthodes

étudiées, suivie par MFCC avec un taux de reconnaissance égal à 83.67 %, PNCC avec un taux de reconnaissance égal à 79.65 % et enfin PMVDR avec un taux de reconnaissance égal à 77.60 % après l'application de la segmentation.

La HFCC est plus efficace pour réduire le bruit aditif blanc gaussien en particulier quand le SNR égal à 30 dB, et la PMVDR est la plus faible en termes de reconnaissance.

3.3.5. L'impact du bruit d'environnement sur le système proposé

Dans cette partie nous allons étudier l'impact du bruit d'environnement sur notre système. Le taux de reconnaissance pour les méthodes de d'extraction de caractéristiques est illustré dans les figures (3.7), (3.8), (3.9).

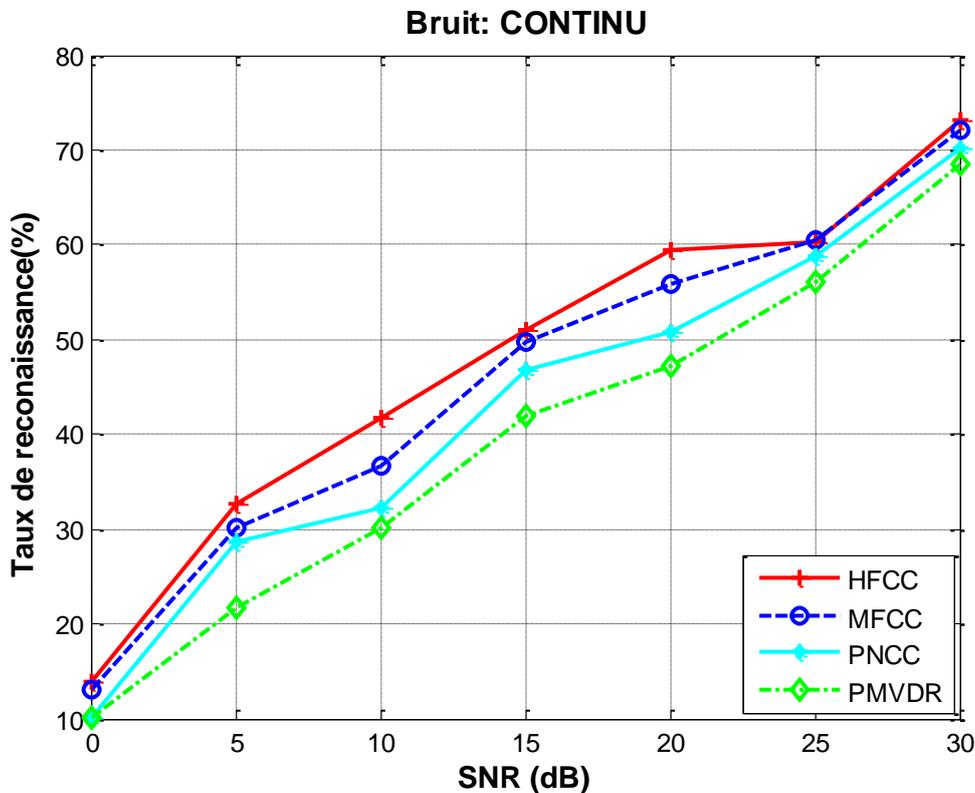


Figure 3.7 : Evolution du taux de reconnaissance des méthodes étudiées (HFCC, MFCC, PNCC et PMVDR) à différents niveaux de SNR d'un bruit continu.

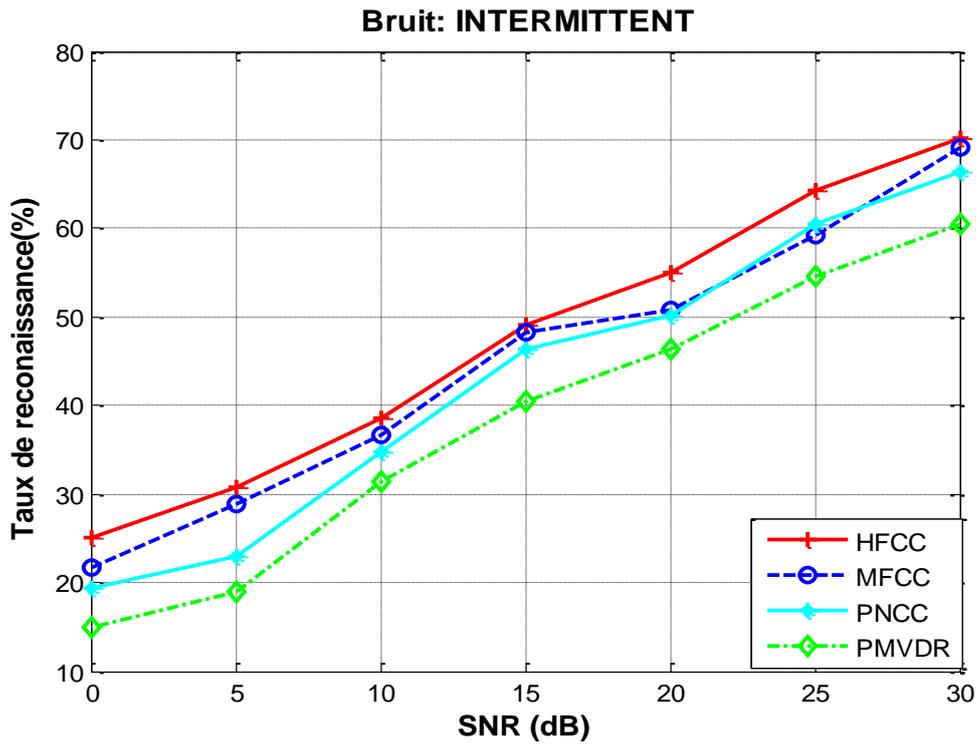


Figure 3.8 : Evolution du taux de reconnaissance des méthodes étudiées (HFCC, MFCC, PNCC et PMVDR) à différents niveaux de SNR d'un bruit intermittent.

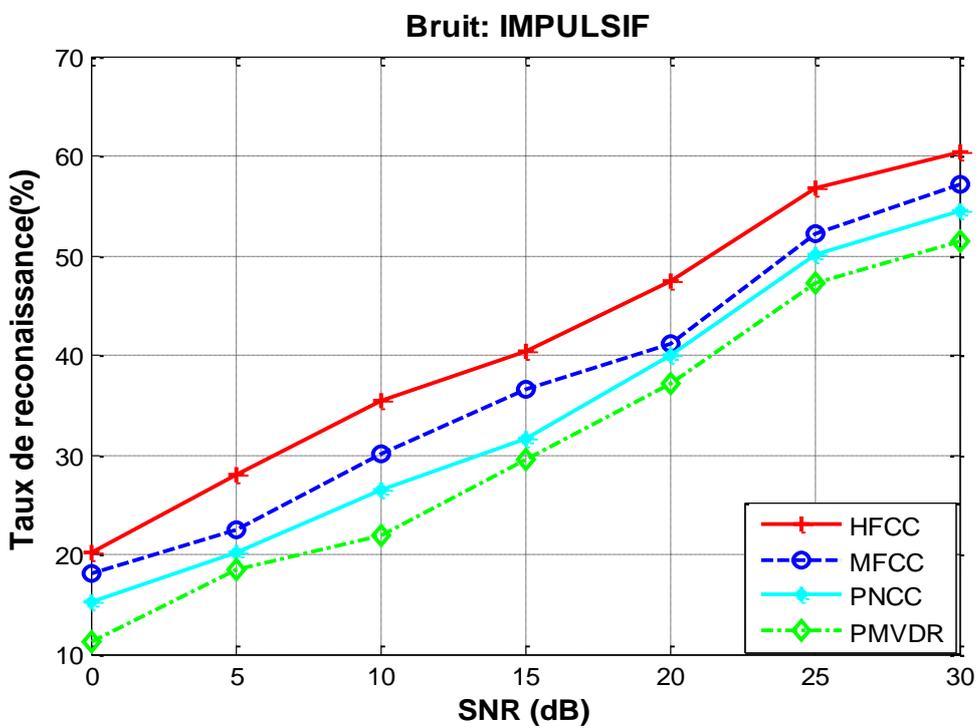


Figure 3.9 : Evolution du taux de reconnaissance des méthodes étudiées (HFCC, MFCC, PNCC et PMVDR) à différents niveaux de SNR d'un bruit impulsif.

Les figures (3.7), (3.8), (3.9), illustrent la robustesse des méthodes de caractérisation étudiées (HFCC, MFCC, PNCC et PMVDR) aux trois bruits d'environnement : le sifflement du vent comme bruit continu, le son d'accident de voiture comme bruit intermittent et une explosion d'une bombe comme un bruit impulsif avec un SNR variant entre 0 et 30 dB avec un pas de 5 dB.

D'après les résultats obtenus, on observe que le taux de reconnaissance est élevé en présence du bruit continu (le sifflement du vent) (Figure 3.7) avec une précision de 73.15 % pour le HFCC à un SNR égal 30 dB. Lors de la présence du bruit intermittent (le son d'un accident de voiture) (Figure 3.8), on observe que le taux de reconnaissance est un peu plus atténué avec une précision variante entre 70.23 % et 60.52 % pour les quatre méthodes. D'un autre côté, le bruit impulsif (explosion d'une bombe) est le bruit qui n'a pas pu être correctement estimé par les quatre méthodes de caractérisation avec un taux de reconnaissance de 60.33%, 57.14%, 54.57% et 51.49% respectivement avec un SNR égal 30 dB.

Le taux de reconnaissance de ces méthodes est affaibli selon le type du bruit d'environnement.

3.4. Conclusion

Dans ce dernier chapitre nous avons étudié les performances de la méthode de segmentation et de débruitage audio par seuillage de bloc temps fréquence, ainsi les performances de la méthode d'extraction de caractéristiques HFCC en comparant avec le MFCC, PNCC et PMVDR en terme de taux de reconnaissance.

Cela a permis une meilleure compréhension du processus d'extraction de caractéristiques étudié « segmentation et débruitage audio par seuillage de bloc temps fréquence- HFCC-DTW »

Il a été démontré que la segmentation et débruitage audio par seuillage de bloc temps-fréquence a une robustesse remarquable au bruit en particulier pour de faibles SNR. L'efficacité de l'algorithme HFCC a été prouvée, dont la reconnaissance des espèces d'oiseaux basée sur HFCC donne de meilleurs résultats en comparaison avec les méthodes MFCC, PNCC et PMVDR, en particulier dans les conditions bruyantes et avec des SNR faibles.

Conclusion Générale

Conclusion générale

La reconnaissance des sons des oiseaux n'est pas une tâche facile, et ceci est essentiellement dû à la mauvaise qualité des enregistrements des sons fortement perturbés par les bruits d'environnement.

Dans ce mémoire, nous avons présenté un système de reconnaissance robuste des sons d'oiseaux constitué de trois étapes suivantes : une méthode de segmentation et de débruitage basée sur un seuillage adaptatif en bloc temps-fréquence, une méthode de caractérisation basée sur les coefficients HFCC et une méthode de classification utilisant l'algorithme DTW.

Ce travail c'est basé sur trois principaux chapitres :

Dans le premier chapitre, nous avons présenté en premier lieu le mécanisme de production sonore chez les oiseaux. Ensuite, nous avons étudié la relation entre la vocalisation des oiseaux et la parole humaine. Après, nous avons fait un état de l'art sur les travaux connexes à la reconnaissance des sons d'oiseaux trouvés dans la littérature. Enfin, les limites et complications de la surveillance bioacoustique ont été présentés.

Dans le deuxième chapitre, nous avons expliqué en détails chaque étape du système de reconnaissance des sons d'oiseaux proposé.

Dans le troisième chapitre, nous avons évalué les performances du système de reconnaissance audio proposé. Les résultats obtenus par simulation en utilisant le logiciel MATLAB, ont inclus les éléments suivants :

- Impact de la méthode de segmentation et de débruitage par seuillage adaptatif en bloc temps-fréquence sur le taux de reconnaissance
- Impact du bruit blanc gaussien sur le taux de reconnaissance
- Impact des différents types de bruit d'environnement sur le taux de reconnaissance
- Comparaison de la méthode de caractérisation HFCC à trois autres méthodes populaires à savoir MFCC, PNCC et PMVDR.

Les résultats obtenus ont démontré l'efficacité du système proposé même dans le cas d'enregistrements fortement bruités.

Annexe

Annexe

1. segmentation et débruitage audio par seuillage de bloc temps-fréquence

```
clc, close all, clear all
f = audioread('./Duck.wav');
fn=audioread('./vent7.wav');
sigma_noise = 0.047;
% fréquence d'échantillonnage(Hz)
f_sampling =41000;
[noisy,noise] = addnoise( fn, f, 0.1); figure (1);
tmp = sprintf("The SNR of the noisy signal is %.2f", get_SNR(f,noisy));
disp(tmp) %display:pour afficher le snr
subplot (3,2,1); plot(noisy)
title('(a)'), xlabel('Echantillons'), ylabel('Amplitude')
hold on; grid on
time_win = 50;
[f_rec, AttenFactorMap, flag_depth] =BlockThresholding(noisy, time_win,
f_sampling, sigma_noise);
audiowrite('./Duck-denoised_snr0.wav', f_rec, f_sampling);
tmp0 = sprintf("The SNR of the denoised signal is %.2f", get_SNR(f,f_rec));
disp(tmp0)
Subplot (3,2,2); plot(f_rec)
title('(b)'), xlabel('Echantillons'), ylabel('Amplitude')
grid on; hold on
disp('le signal débruité est a été écrit dans le fichier ./signal_debruite.wav');
```

2. Extraction des caractéristiques

Calcul des MFCCs

```
[ceps,fb,mfccDCTMatrix] = mfcc(S,fs);
figure
plot(ceps)
xlabel ('Fréquence (Hz)'); ylabel('Amplitude'); grid on ; title ('13 Coefficients
MFCC');
```

Annexe

Calcul des HFCCs

```
st=1;
fftSize=512;
Nhfcc=30;
numCoeffs=14;
fRange=[133 20000];
ERBscaleFactor=1;
deltaSize=4;
[fbHfcc,dctHfcc,fcHfcc] = fbInit(fs,fftSize,Nhfcc,numCoeffs,fRange,ERBscaleFactor);
[Y,YTemp,ceps1,logEnergy]= getFeatures(S,st,fbHfcc,dctHfcc,fftSize,deltaSize);
figure
plot(ceps1)
xlabel ('Fréquence (Hz)'); ylabel('Amplitude'); grid on ; title ('13 Coefficients HFCC');
```

- **Calcul du banc de filtre des HFCC**

```
[fbHfcc,dctHfcc,fcHfcc] = fbInit(fs,fftSize,Nhfcc,numCoeffs,fRange,ERBscaleFactor);
```

- **Extraire les caractéristiques HFCC**

```
[Y,YTemp,ceps1,logEnergy]= getFeatures(S,st,fbHfcc,dctHfcc,fftSize,deltaSize);
```

3. Classification par DTW

```
clc, close all, clear all,
[inputA,fe] = audioread('./White-headed Duck - Oxyura leucocephala1-denoised.wav');
[cepsA,fb,mfccDCTMatrix] = mfcc(inputA,fe);
[inputC,fe] = audioread('./White-headed Duck - Oxyura leucocephala3-denoised3.wav');
[cepsC,fb,mfccDCTMatrix] = mfcc(inputC,fe);
figure
sptr1= cepsA;
sptr2 = cepsC;
[D]=distdtw(sptr1,sptr2)
plot(D)
```

Références Bibliographiques

- [1]. S. Fagerlund, A. Harma, "Parametrization Of Inharmonic Bird Sounds For Automatic Recognition", in *Proc. EUSIPCO 2005*, Antalya, Turkey, September 4-8 2005.
- [2]. Z. Chen, R. C. Maher, "Semi-automatic classification of bird vocalizations using spectral peak tracs", *J. Acoust. Soc. Am.*, Vol. 120, No. 5, November 2006
- [3]. J. A. Kogan, D. Margoliash, "Automated bird song recognition elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study", *J. Acoust. Soc. Am.* 103 (4), April 1998
- [4]. P. Somervuo, A. Härmä, "Bird Song Recognition Based on Syllable Pair Histograms", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)* Montreal, Canada, May 17-21, 2004.
- [5]. M.D. Skowronski, J.G. Harris "Exploiting independent filter band-width of human factor cepstral coefficients in automatic speech recognition", *J. Acoust. Soc. Am.*, 116(3): 1774-1780, 2004.
- [6]. Shaw, S. (2011). "An evaluation of birdsong recognition techniques". *University of Sheffield, Department of Computer Science, Tech. Rep.*
- [7]. BOULMAIZ Amira (2017). *Traitement et Reconnaissance Automatique des Cris et Chants des Oiseaux dans les Zones Humides Annaba-El Taref.*
- [8]. Matz, G., & Hlawatsch, F. (2000). Minimax robust nonstationary signal estimation based on a p-point uncertainty model. *Journal of the Franklin Institute*, 337(4), 403-419.
- [9]. Skowronski, M. D., & Harris, J. G. (2004). Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition. *The Journal of the Acoustical Society of America*, 116(3), 1774-1780.
- [10]. Fagerlund, S. (2007), "Bird species recognition using support vector machines. *EURASIP Journal on Applied Signal Processing*, 2007(1), 64-64.
- [11]. Wolff, D. (2008). Detecting bird sounds via periodic structures: a robust pattern recognition approach to unsupervised animal monitoring. *To be found at* < <http://www->

mmdb. iai. uni-bonn. de/download/Diplomarbeiten/Diplomarbeit_Daniel_Wolff.pdf>[quoted 01.06. 2011].

- [12]. Cai, J., Ee, D., Pham, B., Roe, P., & Zhang, J. (2007, December). Sensor network for the monitoring of ecosystem: Bird species recognition. In *2007 3rd international conference on intelligent sensors, sensor networks and information* (pp. 293-298). IEEE.
- [13]. Kwan, C., Mei, G., Zhao, X., Ren, Z., Xu, R., Stanford, V., ... & Ho, K. C. (2004, May). Bird classification algorithms: Theory and experimental results. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 5, pp. V-289). IEEE.
- [14]. McIlraith, A. L., & Card, H. C. (1997, June). A comparison of backpropagation and statistical classifiers for bird identification. In *Proceedings of International Conference on Neural Networks (ICNN'97)* (Vol. 1, pp. 100-104). IEEE.
- [15]. Ross, D. J. (2006). Bird call recognition with artificial neural networks, support vector machines, and kernel density estimation.
- [16]. Fagerlund, S. (2014). Studies on bird vocalization detection and classification of species.
- [17]. Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology*, 49(5), jav-01447.
- [18]. Yu, G., Mallat, S., & Bacry, E. (2008). Audio denoising by time-frequency block thresholding. *IEEE Transactions on Signal processing*, 56(5), 1830-1839.
- [19]. Rahali, H., Hajaiej, Z., & Ellouze, N. (2014, July). Feature extraction method human factor cepstral coefficients in automatic speech recognition. In *2014 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP)* (pp. 266-270). IEEE.
- [20]. Ganchev, T., Fakotakis, N., & Kokkinakis, G. (2005, October). Comparative evaluation of various MFCC implementations on the speaker verification task. In *Proceedings of the SPECOM*(Vol. 1, No. 2005, pp. 191-194).
- [21]. Sharma, K. K., Kapoor, P., Chakraborty, P., & Nandi, G. C. (2011). Dynamic Spectrum Derived MFCC and HFCC Parameters and Human Robot Speech Interaction. In *International Conference on Advances in Computer Engineering–ACE*.