UQAR
Université du Québec
à Rimouski

ÉTUDE DES TECHNIQUES DE DÉTECTION DES SIBILANTS DANS LES SONS RESPIRATOIRES EN VUE D'UN TRAITEMENT TEMPS-RÉEL SUR FPGA

MÉMOIRE PRÉSENTÉ

dans le cadre du programme de maîtrise en ingénierie

en vue de l'obtention du grade de maître ès sciences appliquées (M. Sc. A.)

PAR

©ONS BOUJELBEN

Février 2017

Composition du jury :

Adrian Ilinca (Ph.D.), président du jury, Université du Québec à Rimouski

Mohammed Bahoura (Ph.D.), directeur de recherche, Université du Québec à Rimouski

Hassan Ezzaidi (Ph.D.), examinateur externe, Université du Québec à Chicoutimi

Dépôt initial le [15-12-2016]   Dépôt final le [22-02-2017]

À ma mère

À mon père

À ma sœur

À mon frère

À mes grands parents

REMERCIEMENTS

Je tiens à adresser ma sincère reconnaissance à mon directeur de recherche, Mohammed Bahoura, professeur au département de Mathématiques, d'Informatiques et de Génie de l'Université du Québec à Rimouski, pour m'avoir accompagné durant mes travaux de recherche. J'ai particulièrement apprécié de travailler à ses côtés. Il m'a donné de nombreux informations pertinentes, se montrant disponible dès que j'avais besoin de ses conseils.

J'exprime toute ma gratitude au professeur Adrian Ilinca d'avoir accepté d'être le président du jury pour l'évaluation de mon mémoire.

J'adresse également mes plus vifs remerciements au professseur Hassan Ezzaidi d'avoir accepté d'examiner mon travail.

Cette recherche a été rendue possible grâce au support financier du Conseil de Recherches en Sciences Naturelles et en Génie (CRSNG).

RÉSUMÉ

L'identification des bruits pulmonaires normaux et anormaux est une opération importante pour le diagnostic médical des poumons. De nos jours, le stéthoscope est l'outil le plus utilisé pour l'auscultation pulmonaire ; il permet aux spécialistes d'écouter les sons respiratoires du patient pour un usage complémentaire. En dépit de ses avantages, l'interprétation des sons fournis par le stéthoscope repose sur la perception auditive et l'expertise du médecin. L'asthme est une maladie respiratoire caractérisé par la présence d'un son musical (sibilant) superposé aux sons respiratoires normaux.

Dans la première étape du projet, nous proposons une étude comparative des techniques de classification les plus pertinentes : le *k*-plus proches voisins (*k*-NN), la machine à vecteurs de support (SVM) et le perceptron multicouche (MLP). Nous utilisons pour l'extraction des caractéristiques des sons respiratoires : les coefficients cepstraux à l'échelle de Mel (MFCC) et la transformée par paquets d'ondelettes (WPT). Des étapes de prétraitement ont été appliquées aux signaux respiratoires qui ont été échantillonnés à la fréquence de 6000 Hz et segmentés en utilisant des fenêtres de Hamming de 1024 échantillons.

Dans la deuxième étape, nous proposons d'implémenter sur le circuit de réseau de portes logiques programmables (FPGA) un détecteur automatique des sibilants permettant aux spécialistes de disposer d'une source d'information fiable qui peut les aider à établir un diagnostic pertinent de la maladie d'asthme. L'architecture matérielle proposée, basée sur la combinaison MFCC-SVM, a été implémentée en utilisant l'outil de programmation haut-niveau générateur système de XILINX (XSG) et le kit de développement ML605 construit autour du circuit FPGA Virtex-6 XC6VLX240T. La phase d'apprentissage du classificateur SVM est faite sur le logiciel MATLAB alors que la phase de test est réalisée avec XSG.

Les résultats de classification des sons respiratoires fournis par l'outil XSG sont similaires à ceux produits par le logiciel MATLAB. Concernant l'étude comparative de techniques de classifications, la combinaison MFCC-MLP a présenté le meilleur résultat de classification avec un taux de reconnaissance de 86.2 %. L'évaluation des différentes combinaisons est réalisée avec les paramètres de spécificité et de sensitivité issues de la matrice de confusion.

Mots clés : Sons respiratoires, MFCC, SVM, XSG, Classifications, Sibilants, FPGA, *k*-NN, MLP, WPT.

ABSTRACT

Identification of normal and abnormal lung sounds is an important operation for pulmonary medical diagnosis. Nowadays, stethoscope is the most used tool for pulmonary auscultation; it allows experts to hear the patient's respiratory sounds for complementary use. Despite its advantages, the interpretation of sounds provided by the stethoscope is based on the sense of hearing and the expertise of the doctor. Asthma is a respiratory disease characterized by the presence of a musical sound (wheezing) superimposed on normal respiratory sounds.

First, we propose a comparative study between the most relevant classification techniques : *k*-Nearest Neighbor (*k*-NN), the Support Vector Machine (SVM) and the Multi-layer perceptron (MLP). The feature extraction techniques used are : Mel-Frequency Cepstrum Coefficients (MFCC) and the Wavelet Packet Transform (WPT). Preprocessing steps have been applied to the respiratory sounds that have been sampled at 6000 Hz and segmented using Hamming window of 1024 samples.

In a second step, we propose to implement on the FPGA (Field Programmable Gate Array) circuit an automatic wheezes detector, allowing specialists to have a reliable source of information, which can help them to establish an accurate diagnosis of the asthma disease. The proposed hardware architecture, based on MFCC-SVM combination, was implemented using the high-level programming tool XSG (Xilinx System Generator) and the ML605 development kit build around the Virtex-6 XC6VLX240T FPGA chip. The learning phase of the SVM classifier is made on the MATLAB software while the test phase is carried out using XSG.

Classification results of the respiratory sounds provided by XSG are similar to those produced by the MATLAB software. Regarding the comparative study of the classification techniques, the combination MFCC-MLP presented the best classification result with a recognition rate of 86.2 %. The evaluation of different combinations is carried out with the specificity and sensitivity parameters, which present the outcome of confusion matrix.

Keywords : Respiratory sounds, MFCC, SVM, XSG, Classifiers, Wheezing, FPGA, *k*-NN, MLP, WPT.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

ANN *Artificial Neural Network*.

Réseau de neurones artificiels.

FFT *Fast Fourier Transform*.

Transformée de Fourier rapide.

FPGA *Field Programmable Gate Array*.

Réseau de portes logiques programmables.

GMM *Gaussian Mixture Model*.

Modèle de Mélange de Gaussiennes.

***k*-NN** *k-Nearest Neighbor*.

k-plus proches voisins.

MFCC *Mel-Frequency Cepstrum Coefficients*.

Coefficients cepstraux à l'échelle de Mel.

MLP *Multi Layer Perceptron*.

Perceptron multi-couches.

RBF *Radial Basis Function*.

Fonction radiale de base.

STFT *Short Time Fourier Transform*.

Transformée de Fourier à court terme.

SVM *Support Vector Machine*.

Machine à vecteurs de support.

VHDL *Very high speed integrated circuit Hardware Description Language*.

Langage de description de matériel pour circuits à très haute vitesse d'intégration.

WPT *Wavelet Packet Transform*.

Transformée par paquets d'ondelettes.

XSG *Xilinx System Generator*.

Générateur Système de XILINX.

INTRODUCTION GÉNÉRALE

0.1 État de l'art

Les maladies respiratoires sont parmi les causes les plus fréquentes de morbidité et de mortalité à travers le monde (Billionnet, 2012). D'après les estimations de l'OMS (Organisation Mondiale de Santé), il y a plus 235 millions d'asthmatiques dans le monde (Boulet et al., 2014). Cette maladie chronique se manifeste par des crises sévères accompagnées de sensations de suffocation, d'essoufflement et peut engendrer dans certains cas une perte de contrôle. Ainsi, des travaux de recherche ont été consacrés afin d'améliorer le diagnostic et la surveillance de cette maladie par le développement des techniques de l'auscultation pulmonaire.

Au cours des dernières années, les techniques de traitement de signal ont évolué de façon très rapide, que ce soit dans les domaines de reconnaissance de la parole, de la reconnaissance de formes ou dans le domaine biomédical comme l'analyse des sons respiratoires et de l'électrocardiographie (ECG). La combinaison entre l'auscultation pulmonaire, les techniques de communication récentes et les outils avancées pour le traitement du signal fournissent aux médecins une source d'information supplémentaire à celle du stéthoscope traditionnel.

De nombreuses méthodes ont été utilisées par les chercheurs au cours des trois dernières décennies pour traiter les sons respiratoires afin d'identifier les sons pathologiques (Shaharum et al., 2012). Nous pouvons citer quelques combinaisons de techniques d'extraction de caractéristiques et de classificateurs qui ont été documentées dans la littérature : les coefficients cepstraux à l'echelle de Mel (MFCC) combinés avec la machine à vecteurs de support (SVM), le *k*-plus proches voisins (*k*-NN) (Palaniappan et al., 2014) et le modèle de mélange de gaussiennes (GMM) (Bahoura and Pelletier, 2004). La transformée par paquets d'ondelettes (WPT) a été utilisée avec le réseau de neurones artificiels (ANN) (Kandaswamy et al., 2004; Bahoura, 2009), ainsi que d'autres combinaisons ont été présentées dans la littérature (Bahoura, 2009; Palaniappan et al., 2013). Parmi ces techniques, la combinaison MFCC-SVM a été appliquée pour détecter les sibilants chez les patients asthmatiques ; elle a démontré une précision supérieure à 95 % (Mazic et al., 2015).

Le développement d'outils de reconnaissance ou de classification des sons respiratoires conduit à des techniques de traitement de plus en plus complexes. Le fonctionnement en temps-réel de ces algorithmes nécessite une implémentation matérielle sur circuits programmables de types DSP (Digital Signal Processor) ou FPGA (Field Programmable Gate Array). Malgré que plusieurs achitectures matérielles a été proposées dans la littérature pour les algorithmes de reconnaissance/classification (Wang et al., 2002; Amudha and Venkataramani, 2009; EhKan et al., 2011; Mahmoodi et al., 2011; Manikandan and Venkataramani, 2011; Ramos-Lara et al., 2013; Pan and Lan, 2014), nous trouvons un seul système à base de circuit DSP (Alsmadi and Kahya, 2008) a été proposé pour la classification des sons pulmonaire et deux autres à base de circuit FPGA pour la caractérisation (Bahoura and Ezzaidi, 2013) et la détection des sibilants (Lin and Yen, 2014).

0.2 Nomenclature et classification des sons respiratoires

Les sons respiratoires sont divisés en deux classes : sons respiratoires normaux et sons respiratoires pathologiques. Les sons respiratoires anormaux (adventices) sont répartis en deux sous-classes de sons : continus et discontinus.

0.2.1 Sons respiratoires normaux

Deux sons respiratoires symbolisent le passage normal de l'air ambiant inspiré à travers le conduit respiratoire : les sons trachéo-bronchiques, entendus dans la trachée du larynx ou sur les grandes voies aériennes, et les sons vésiculaires, obtenus sur la surface thorique (Pelletier, 2006). Le premier son respiratoire normal est un son râpeux de grande intensité et continu, entendu à la fois lors de l'inspiration et de l'expiration, le second son respiratoire normal est un murmure continu, moelleux et de faible intensité, entendu durant toute l'inspiration et seulement au début de l'expiration (Laennec, 1819). La figure 0.1 est une représentation du son respiratoire normal dans le domaine temps et sous forme de spectrogramme.

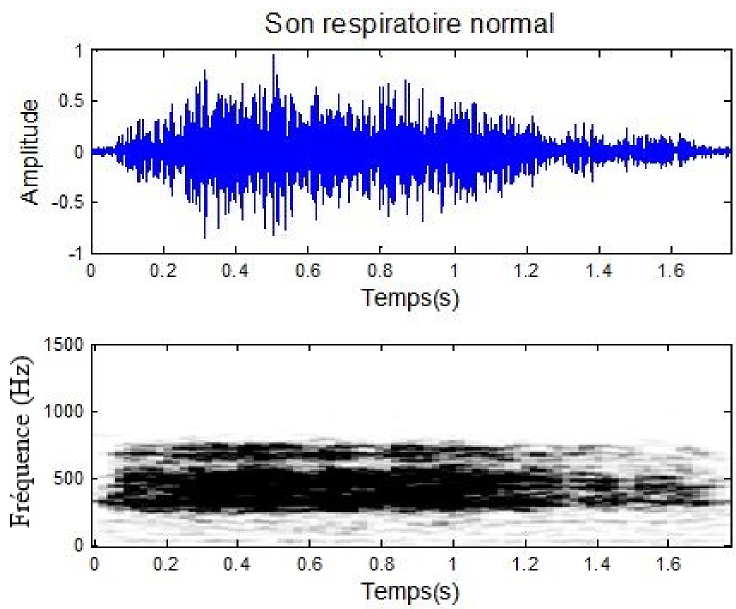

Figure 0.1: Représentation dans le domaine temps (haut) et sous forme de spectrogramme (bas) d'un son respiratoire normal.

0.2.2 Sons respiratoires adventices (pathologiques)

Les sons respiratoires adventices (pathologiques) sont des sons surajoutés aux sons respiratoires normaux qui marquent un dysfonctionnement du système respiratoire ou des inflammations. Ils sont subdivisés en deux classes selon leurs formes d'ondes : sons pathologiques continus et sons pathologiques discontinus. Nous nous intéressons dans ce projet à la classe des sons respiratoires pathologiques continus qui contient les sibilants et les ronchus. Il s'agit de sons à caractère musical, généralement de forte amplitude, produits lors d'une obstruction sévère provoquant des contacts entre les parois bronchiques. Selon CORSA (Computerized Respiratory Sound Analysis), la durée de ces sons est supérieure à 250 ms. Ils ont une plage de fréquence variable, la fréquence dominante du sibilant est supérieure à 400 Hz alors que celle du ronchus ne dépasse pas 200 Hz (Pelletier, 2006).

La figure 0.2 est une représentation dans le domaine temps et sous forme de spectrogramme de sons adventices continus (sibilants, ronchus).

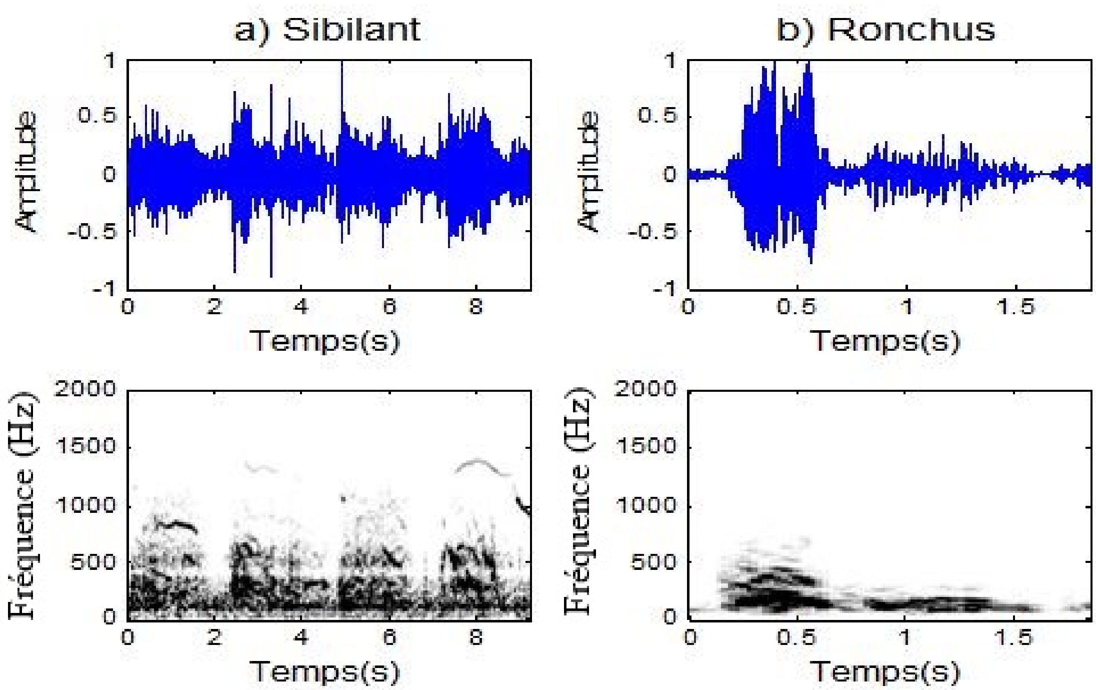

Figure 0.2: Représentation dans le domaine temps et sous forme de spectrogramme de sons respiratoires adventices continus : a) sibilant et b) ronchus.

0.3 Problématique de recherche

En dépit de ses avantages, l'analyse automatique des sons respiratoires n'a pas encore évolué pour atteindre à son utilisation clinique. La complexité des sons respiratoires, la diversification des systèmes d'acquisition (sites de prise de son, type de capteur, fréquences de filtrage, etc.), et l'absence ou la non disponiblité de base de données de réference rend difficile la comparaison entre les méthodes proposées par les différents chercheurs.

D'autre part, l'implémentation matérielle des techniques de reconnaissance/classification de sons respiratoires en vue d'un traitement temps-réel constitue encore un grand défit pour les chercheurs. La transcription des codes développés sous MATLAB en une architecture implémentable sur un circuit FPGA nécessite généralement un réaménagement (réadaptation) de ces codes.

La problématique du recherche consiste à concevoir un outil d'identification en temps-réel des sibilants dans les sons respiratoires des patients asthmatiques. Ce système peut aussi servir comme un instrument à usage domestique pour aider à l'évaluation et le suivi de l'état des patients à moindres frais.

0.4 Objectifs

Ce projet de recherche a deux principaux objectifs :

1. Élaborer une étude comparative des techniques de reconnaissance/classification les plus utilisées pour la classification des sons respiratoires, afin d'identifier la méthode la mieux adaptée à la détection des sibliants.
2. Concevoir un système de classification capable de discriminer en temps-réel les sibilants des sons respiratoires normaux.

0.5 Hypothèses

Pour atteindre les deux objectifs, notre démarche se base sur les hypothèses suivants :

1. Parmi les techniques de reconnaissance/classification, il faut se limiter à celles qui ont fait leur preuve dans le domaine de la reconnaissance automatique de la parole. En fait, les sons respiratoires ressemblent à un certain niveau aux signaux de parole.
2. Comme solution matérielle, choisir les circuits FPGA qui combinent les performances (parallèlisme) des circuits ASIC (Application Specific Integrated Circuit) et la flexibilité de programmation des circuits DSP (Digital Signal Processor).

0.6 Méthodologie

Afin d'atteindre les objectifs de cette recherche, nous utilisons des systèmes de classification appropriés pour l'études des sons respiratoires. Ces systèmes fonctionnent en deux phases : phase d'apprentissage et phase de test, comme le montre la figure 0.3. À partir d'une base de données contenant des sons respiratoires normaux et pathologiques et en utilisant une technique d'extraction des caractéristiques, chaque segment temporel des signaux à analyser sera représenté par quelques paramètres dans l'espace des caractéristiques. Pendant la phase d'apprentissage, ces données permettent d'obtenir un modèle pour chaque classe de sons. Lors de la phase de test et après extraction du vecteur de caractéristique du segment de test, le classificateur prend la décision d'appartenance en se basant sur la ressemblance entre le modèle de la classe établi au cours de l'apprentissage et le vecteur caractéristique du segment testé.

0.6.1 Étude des différentes méthodes d'identification des sibilants

L'extraction des caractéristiques d'un signal est un processus qui permet de réduire la dimension du signal tout en capturant l'information pertinente. Il peut être considéré comme

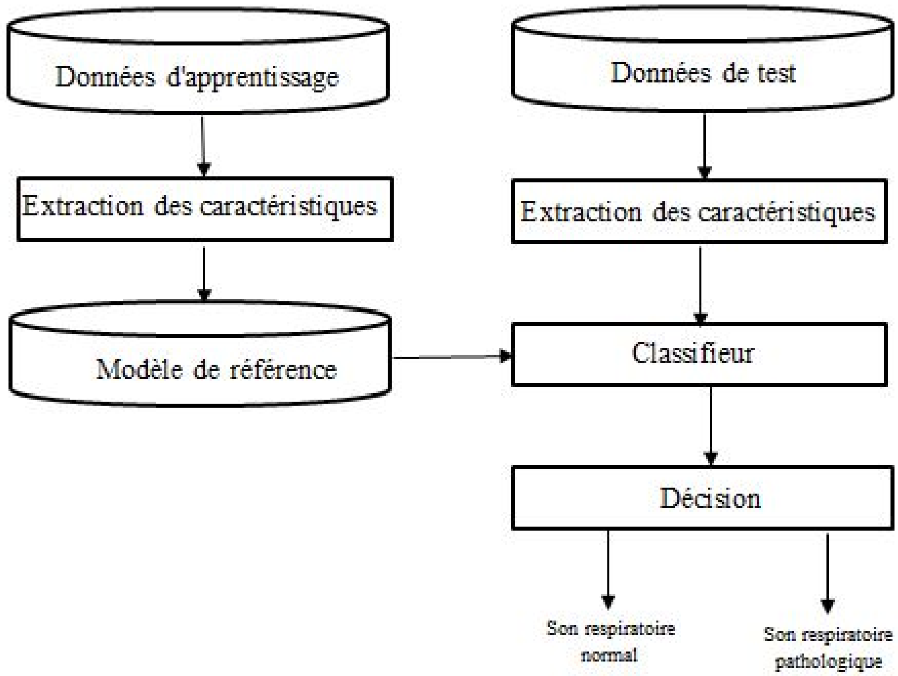

Figure 0.3: Principe de classification des sons respiratoires.

une projection dans l'espace de données :

$$\varphi : R^N \rightarrow R^K \qquad (N >> K) \tag{0.1}$$

où N et K représentent respectivement les dimensions de l'espace de données avant et aprés l'extraction des caractéristiques. Cette transformation se base sur deux étapes. D'abord, le signal est divisé en segments temporels contenant un nombre d'échantillons correspondant à largeur du fenêtre choisie pour le traitement du son. Ensuite, en appliquant la technique d'extraction des caractéristiques, chaque segment temporel sera caractérisé par un vecteur de caractéristiques contenant l'information essentielle du signal.

Dans ce projet de recherche, nous proposons d'utiliser deux techniques pour l'extraction des caractéristiques :

— Coefficients cepstraux à l'échelle de Mel (MFCC) ;

— Transformée par paquets d'ondelettes (WPT).

Pour la phase de calssification, nous faisons appel aux techniques liées à la problématique de la reconnaissance des sons respiratoires. Parmi les algorithmes de classification les plus pertinents, nous proposons de tester :

— Perceptron multicouche (MLP) ;
— Machine à vecteurs de support (SVM) ;
— *k*-plus proches voisins (*k*-NN).

0.6.2 Implémentation matèrielle d'un détecteur de sibilants

Les circuits FPGA se programment à la base, à l'aide d'un langage HDL (hardware description language), tels que VHDL ou Verilog. Cepeandant, il existe des langages de programmation de haut-niveau d'abstraction qui permettent de sauver beaucoup de temps dans la phase de développement. Dans ce projet, nous avons choisi d'utiliser l'outil de programmation XSG (Xilinx System Generator) qui nous permet de tirer profil de l'environement de simulation de MATLAB/SIMULINK. L'outil XSG dispose d'une librairie SIMULINK modélisant, au bit et au cycle près, les fonctions arithmétiques et logiques, les mémoires et des fonctions de traitement de signal. Il inclut aussi un générateur de code HDL automatique pour les circuits FPGA de XILINX.

L'architecture matérielle proposée est une combinaison de la technique d'extraction des caractéristiques par MFCC et de la technique de classification par SVM. Elle se base sur les architectures matérielles proposées pour le calcul des coefficients MFCC (Bahoura and Ezzaidi, 2013) et l'utilisation de la machine SVM (Mahmoodi et al., 2011).

0.7 Contributions

La première contribution de ce mémoire est l'étude compartive de plusieurs méthodes de classification des sons respiratoires en optimisant les paramètres de chaque technique. Les résultats (Article 1) seront soumis au moins à une conférence.

La seconde contribution est l'architecture matérielle MFCC-SVM, implémentée sur circuit FPGA. Une description détaillée (Article 2) sera très prochainement soumise à un journal spécialisé. Une version abrégée de son fonctionnement a été publiée dans la 2e conférence ATSIP :

— O. Boujelben and M. Bahoura, "FPGA implementation of an automatic wheezes detector based on MFCC and SVM," 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Monastir, 2016, pp. 647-650.

Ce système peut mener à un instrument embarqué de télé-surveillance de la fonction respiratoire chez les malades asthmatiques, permettant ainsi l'évaluation et le suivi de l'état des patients en temps-réel.

0.8 Organisation du mémoire

Ce mémoire en format d'articles comprend trois chapitres. Le premier chapitre présente une étude comparative des méthodes de classification des sons respiratoires en utilisant différentes combinaisons. Le deuxième chapitre concerne l'implémentation matérielle sur circuit FPGA d'un système de détection des sibilants basé sur la combinaison MFCC-SVM. Le dernier chapitre présente la conclusion générale ainsi que les perspectives de recherche.

ARTICLE 1

COMPARATIVE STUDY OF RESPIRATORY SOUNDS CLASSIFICATION USING DIFFERENT LEARNING MACHINES

1.1 Résumé en français du premier article

Les machines à apprentissage sont des outils performants de classification des formes, particulièrement dans le domaine de traitement de signal. Dans ce premier article, nous nous intéressons à l'étude et à la comparaison de plusieurs machines à apprentissage. Ces machines sont obtenues par des combinaisons de méthodes de caractérisation et de classification afin de reconnaître les sibilants dans les sons respiratoires.

Pour extraire les caractéristiques des sons respiratoires, nous proposons d'utiliser les coefficients cepstraux à l'échelle de Mel (MFCC) et la transformée par paquets d'ondelettes (WPT). Nous utilisons la machine à vecteurs de support (SVM), les *k*-plus proches voisins (*k*-NN) et le perceptron multicouche (MLP) comme classificateurs. Les résultats des tests révèlent que le meilleur taux de reconnaissance de 86.2 % est obtenu par la combinaison MFCC-MLP.

Ce premier article, intitulé "*Étude comparative de la classification des sons respiratoires utilisant différentes machines à apprentissage*", fut corédigé par moi-même ainsi que par le professeur Mohammed Bahoura. En tant que premier auteur, ma contribution à ce travail fut l'essentiel de la recherche sur l'état de l'art, le développement de la méthode, l'exécution des tests de performance et la rédaction de l'article. Le professeur Mohammed Bahoura, second auteur, a fourni l'idée originale. Il a aidé à la recherche sur l'état de l'art, au développement de la méthode ainsi qu'à la révision de l'article.

1.2 Abstract

A comparative study of different learning machines to classify respiratory sounds in two categories (normal and wheezing) is presented. The lung sounds are described by two feature extraction techniques: Mel-frequency cepstral coefficients (MFCC) and wavelet packet transform (WPT). As classifier, we use the k-nearest neighbor (k-NN), support vector machine (SVM) and multi-layer perceptron (MLP). In order to evaluate the performance of these combinations, we use a database composed of 24 respiratory sounds records: 12 are obtained from normal subjects and 12 records are obtained from asthmatic subjects. The test results reveal that the highest recognition rate is obtained by the combination MLP-MFCC with an accuracy of 86.2 %.

Key words: Respiratory Sounds, Classification, WPT, MFCC, Learning machine, k-NN, MLP.

1.3 Introduction

Pulmonary auscultation is an inexpensive and noninvasive medical examination technique that allows physicians to diagnose various respiratory disorders. The acoustic stethoscope has been proposed in 1816 to listen to lung and heart sounds, where the first attempt was realized by Dr. *René Laennec* who formed a notebook roller and apply one end on the patient's chest (Pasterkamp et al., 1997).

Lung sounds are divided into normal and adventitious classes. Adventitious (or pathological) respiratory sounds are even divided into continuous and discontinuous. Normal sounds present normal progression of the ambient air breathed across the respiratory tract. Pathological respiratory sounds are added to the normal lung sounds, they usually mark a dysfunction of the respiratory system or inflammation. These different respiratory sounds are classified according to their temporel and spectral characteristics. As shown in Fig. 1.1, normal sounds are characterized by a dominant frequency ranging from 37.5 Hz to 1000 Hz (Palaniappan et al., 2014). Wheezing sounds belong with continuous adventitious respiratory sounds, their duration is more than 250 ms and their frequency range is greater than 400 Hz (Sovijarvi et al., 2000). These sounds with musical character usually have high amplitude.

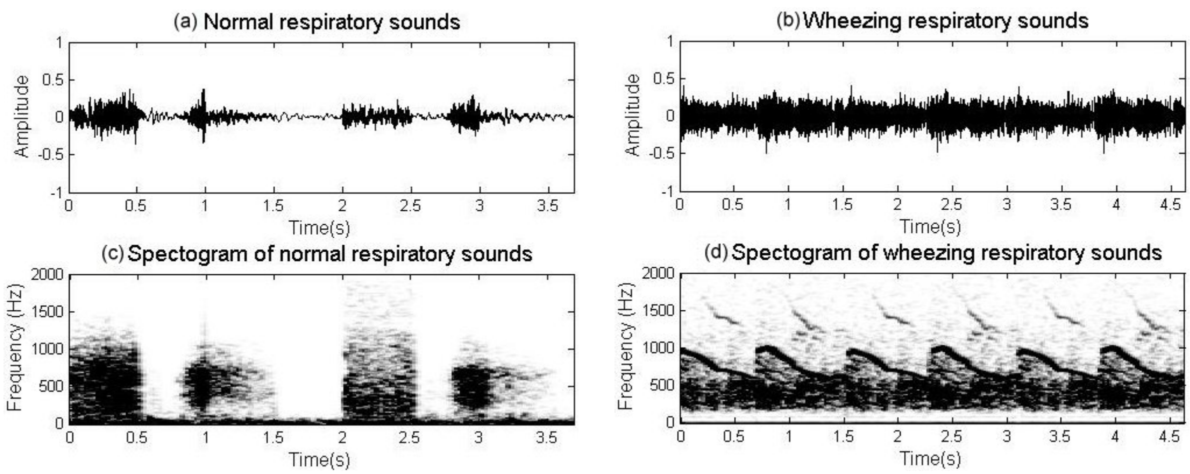

Figure 1.1: Normal (a) and wheezing (b) respiratory sounds and their associated spectrograms (c) and (d), respectively.

Nowadays, researchers emphasize a great importance to classify respiratory sounds using new signal processing techniques. In fact, some scientists consider that the stethoscope is an unreliable acoustic instrument, because it can amplify or attenuate sounds in the spectrum range of scientific concern (Pasterkamp et al., 1997). Moreover, pulmonary auscultation is a subjective process, it is based on the physician's expertise, personal experience as well as on his own hearing to distinguish the variety of sounds.

In pattern recognition field, the machine learning algorithms are widely used to classify different patterns database. Different feature extraction techniques and classifiers are used in the literature. Linear predictive coding (LPC) (Sankur et al., 1994), Fourier transform (FT) (Bahoura, 2009; Tocchetto et al., 2014), wavelet transform (WT) (Bahoura, 2009; Tocchetto et al., 2014), and Mel-frequency cepstral coefficients (MFCC) (Mazic et al., 2015; Palaniappan et al., 2014) techniques have been used for wheezing feature extraction, whereas

k-nearest neighbor (k-NN) (Chen et al., 2015; Palaniappan et al., 2014), artificial neural networks (ANN) (Bahoura, 2009; Tocchetto et al., 2014) and gaussian mixture models (GMM) (Pelletier, 2006) have been used for respiratory sounds classification.

In this paper, we use two features extraction techniques: Mel-frequency cepstral coefficients (MFCC) and the technique wavelet packet transform (WPT) using statistical features for classification. As classifier, we use the k-nearest neighbor (k-NN), support vector machine (SVM) and multi-layer perceptron (MLP). Finally, we propose a comparative study of different combinations of feature extraction techniques and machine learning classifiers.

1.4 Feature extraction

The feature extraction technique allows the selection of essential characteristic information from the analysed sound, which also reduces its vectors' dimension. In this research, we use two feature extraction techniques to characterize respiratory sounds: Mel-frequency cepstral coefficients (MFCC) and the wavelet packet transform (WPT) using statistical features.

1.4.1 Mel-Frequency Cepstral Coefficients (MFCC)

Mel-frequency cepstral coefficients (MFCC) present the cepstral coefficients across the Mel-scale. This feature extraction tool is extensively used in audio pattern recognition systems. The first step to obtain the MFCC coefficients is to segment the input signal $s(n)$ into successive frames of N samples using the following equation:

$$s_i(n) = s(n)w_i(n) \tag{1.1}$$

where w_i is a window function, i is the frame index, and n is the time index within the frame.

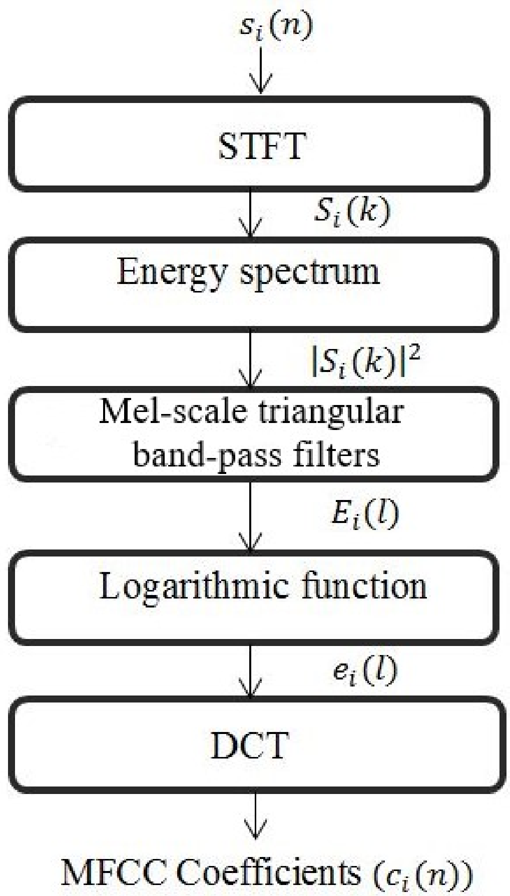

Figure 1.2: Block diagram for Mel-frequency cepstral coefficient (MFCC) feature extraction.

As shown in Fig. 1.2, the short-time Fourier transform (STFT) of the segmented signal $s_i(n)$ is calculated, then the energy spectrum $|S_i(k)|^2$ is computed. The energy spectrum is filtered by a Mel-scaled triangular band-pass filters. The Mel-scale can be approximated to the following equation:

$$Mel(f) = 2595 \log_{10}(1 + \frac{f}{700}) \tag{1.2}$$

A logarithm function is applied to the filter outputs $E_i(l)$ in order to compress their dynamic range. The output of the logarithmic function $e_i(l)$ is defined by Eq. 1.3.

$$e_i(l) = \log(E_i(l)) \tag{1.3}$$

The obtained MFCC coefficients are obtained by back-transformation in time domaine using discrete cosine transform (DCT)

$$c_i(n) = \sum_{l=1}^{P} e_i(l) cos(n(l - 0.5)\pi/P) \tag{1.4}$$

where ($n = 0, 1, ..., M - 1$) is the index of the cepstral coefficient, M is the number of desired MFCC coefficients and l present the index of the triangular band-pass filter ($1 \leq l \leq P$). In this research we use 14 triangular band-pass filter ($P = 14$). Since the coefficients $c_i(0)$ and $c_i(1)$ are often ignored because they represents the mean value of the input signal (Bahoura and Ezzaidi, 2013), the feature extraction vector is composed by 12 coefficients.
For a given frame m, the feature vector constructed from the MFCC coefficients (Eq. 1.4) is given by:

$$x_m = [c(2), c(3), ..., c(13)]^T \tag{1.5}$$

1.4.2 Wavelet Packet Transform (WPT)

The wavelet transform (WT) is widely used in signal analysis, because it provides a simultaneous time and frequency representation of signals. This decomposition conserves the important characteristics of signal (Tocchetto et al., 2014). The wavelet packet transform (WPT) provides a multichannel filtering analysis, where the number of filters and their frequency-bands depends on the level tree (Bahoura, 2009). It can be seen as an extension of the wavelet transform (WT). Fig. 1.3 presents a 2-level WPT decomposition tree for a signal $x(n)$ of length N.

The wavelet packet coefficients are defined by $w_k^i(n)$, where i is the level of decomposition, k represents the frequency subband index and $n = 0, ..., \frac{N}{2^i} - 1$ is the time index. The wavelet packet coefficients for even and odd subband k are given by equations (1.6) and (1.7), respectively.

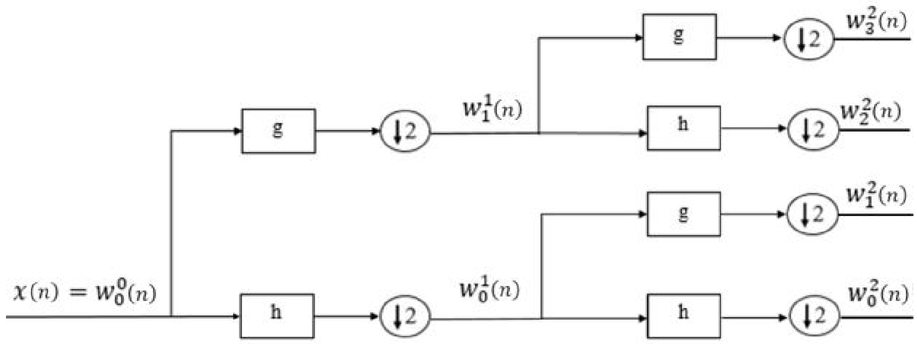

Figure 1.3: Wavelet Packet Transform for a 2-level decomposition tree.

$$w_{2p}^{i+1}(n) = \sum_m h(m-2n)w_p^i(m) \tag{1.6}$$

$$w_{2p+1}^{i+1}(n) = \sum_m g(m-2n)w_p^i(m) \tag{1.7}$$

where $p = 0, ..., 2^i - 1$ represents the subband of level i, h and g are the low-pass and high-pass filters, respectively.

In this study, the respiratory signals is uniformly divided into overlapped (50 %) segments, 1024 samples each. For each segment, the signal will be decomposed down to the sixth level ($i = 6$) of the wavelet packet decomposition tree, which leads to 64 packets. For a sampling frequency of 6000 Hz, the bandwidth of each node of the wavelet packet decomposition tree is 46.875 Hz. Since the lung sounds frequency spectrum ranges from 50 to 1000 Hz (Kandaswamy et al., 2004), we select for the feature extraction vector from the 2^{nd} node to the 22^{th} node of the wavelet decomposition, which correspond to the band 46.875-1031.25 Hz.

Different feature extraction functions have been proposed to reduce the dimensionality of the transformed signal as mean, standard deviation (Tocchetto et al., 2014) and variance (Gabbanini et al., 2004). In this study, the three statistical functions are tested and the variance function is chosen as feature extraction technique. In this research, we propose to use the variance feature extraction function based on the wavelet packet transform from the 2^{nd} node to the 22^{th} node so that the feature extraction vector is composed by 21 coefficients ($M = 21$).

For a given frame m the feature extraction vector is defined by:

$$x_m = [var(2), var(3), ..., var(22)]^T \tag{1.8}$$

1.5 Learning Machines

The learning machine belongs with the field of artificial intelligence (AI), it gives computers the possibility to elaborate analytical model without being explicitly programmed. This model can automatically learn to perform and change according to new data. Different classifiers were proposed in the literature for respiratory sounds classification: support vector machine (SVM) (Mazic et al., 2015; Palaniappan et al., 2014), k-nearest neighbor (k-NN) (Chen et al., 2015; Palaniappan et al., 2014), artificial neural networks (ANN) (Bahoura, 2009; Tocchetto et al., 2014) and gaussian mixture models (GMM) (Pelletier, 2006). In this research, we propose the use of three types of learning machine: k-nearest neighbor (k-NN), support vector machine (SVM) and the multi-layer perceptron (MLP).

1.5.1 *k*-Nearest Neighbor (*k*-NN)

The k-nearest neighbor (k-NN) is a supervised learning machine that allows to map a new observation x from the D-dimensional feature vectors space to its desired class from $[w_1, ..., w_K]$, where K is the number of classes (Kozak et al., 2006). The classification problem in a supervised classifier is defined by a labeled training set of n observations as described in

Eq. 1.9:

$$O_n = \{(x_1, w_1), (x_2, w_2), ..., (x_n, w_n)\} \tag{1.9}$$

where x_i are the feature vectors and w_i the associated scalar labels. For a given query instance x, the k-NN algorithm place a cell volume V around x to captures k prototypes. We denote by k_j the number of samples labeled w_j captured by the cell, so the captured k prototypes can be defined by the following equation:

$$k = \sum_{j=1}^{K} k_j \tag{1.10}$$

The joint probability can be calculated as (Bahoura and Simard, 2012) :

$$p_n(x, w_j) = \frac{k_j/V}{k} \tag{1.11}$$

The Eq. 1.11 is used to provide a reasonable estimate of the posterior probability

$$p_n(w_j|x) = \frac{p_n(x, w_j)}{\sum_{j=1}^{K} p_n(x, w_j)} = \frac{k_j}{k} \tag{1.12}$$

To estimate the class w for a new feature vector x, the k-NN uses the majority of vote among the k_j objects neighboring the new data:

$$w = arg \max_{1 \leq j \leq K} \{k_j\} \tag{1.13}$$

In this study, the analyzed respiratory sound is segmented uniformly to M overlapped frames so that the corresponding feature vectors sequence $X = [x_1, x_2, ..., x_M]$ is predicted into the class w by the following equation:

$$w = arg \max_{1 \leq j \leq K} \{\bar{k_j}\} \tag{1.14}$$

where the mean values $\bar{k_j}$ are computed over the M frames.

$$\bar{k_j} = \frac{1}{M} \sum_{i=1}^{M} k_{i,j} \tag{1.15}$$

In this research, the classification of unknown sound is made segment-by-segment ($M = 1$).

1.5.2 Support Vector Machine (SVM)

Since the 1900s, the support vector machine (SVM) technique is used to solve classification and regression problems (Vapnik, 1998). This learning algorithm is adopted for both binary and multiclass data. The SVM technique separates new data based on a predicted model which is generated during the training phase. Consider SVM for binary classification, a labeled training set of n observations as mentioned in Eq. 1.16:

$$O_n = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\} \tag{1.16}$$

where x_i are the feature vectors and $y_i \in \{1, -1\}$ the associated scalar labels. The SVM classifier computes an hyperplane that separates the training data in two sets corresponding to the desired classes. The optimal hyperplane is defined such that: all points labeled -1 are on one side of the hyperplane and all points labeled 1 are on the other side and the distance of the nearest vector of the hyperplane (both classes) is maximum.

This classifier is called support vector machine since the solution only depends on the support vectors. As shown in Fig. 1.4, the instances of the training data the most closest to canonical hyperplanes (H_1, H_2) are called support vectors (Ertekin, 2009). Where the

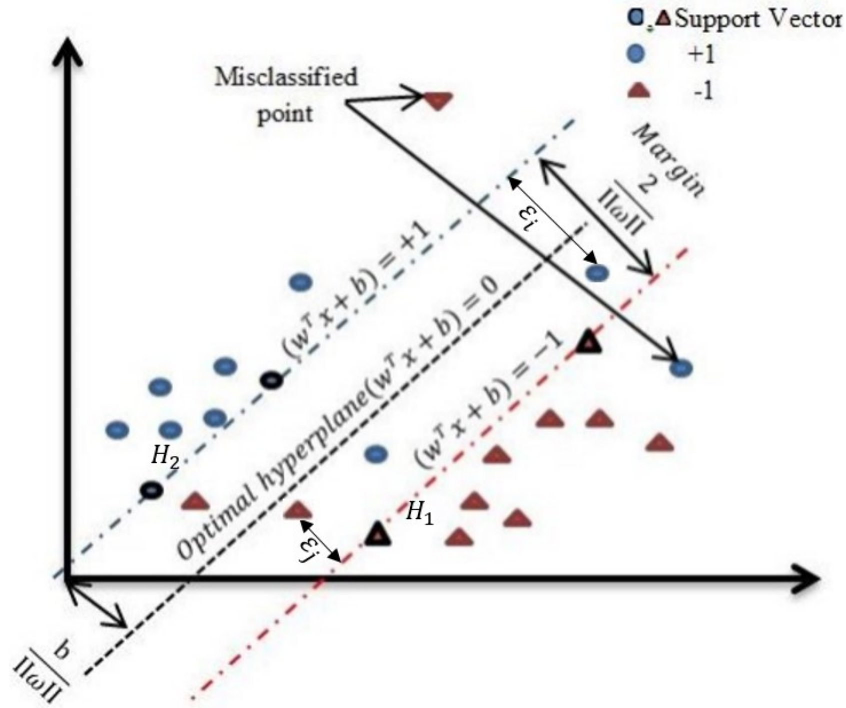

Figure 1.4: Optimal separating hyperplane and support vectors.

parameters w is the normal vector to the hyperplane, b is the bias, $\| w \|$ is the euclidean norm of w and ξ is slack variables representing the data that fall into the margin (Mahmoodi et al., 2011).

The maximum margin separating hyperplane can be constructed by solving the primal optimization problem in Eq. 1.17 introduced by (Vapnik, 1998).

$$Minimize \ \{\tau(w, \xi)\} = \frac{1}{2} \| w \|_2^2 + C \sum_{i=1}^{n} \xi_i \tag{1.17}$$

subject to :

$$y_i(w^T x_i + b) \geqslant 1 - \xi_i \\ \xi_i \geqslant 0 \tag{1.18}$$

where w is a n-dimensional vector, ξ_i is a measure of distance between the misclassified point and the hyperplane and C the misclassification penalty parameter dealing between maximization the margin and minimization the error. The first term is minimized to control the margin, the aim of the second term is to keep under control the number of misclassified points (Chapelle, 2004).

To classify an unknown data x, the decision for the linear SVM classifier is presented by Eq. 1.19.

$$d(x) = sign(w^T x + b) \tag{1.19}$$

To determinate the parameters w and b, we should first resolve the following dual Lagrange problem. Note that the penalty function related to the slack variables is linear, which disappears in the transformation into the dual formulation (Ertekin, 2009).

$$Maximize\ \{L_d(\alpha)\} = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{j,i=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{1.20}$$

subject to :

$$0 \leq \alpha_i \leq C \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{1.21}$$

where α_i are Lagrange multipliers, n is the number of samples. The Lagrange multipliers α_i are calculated by resolving the Lagrange equation 1.20. The parameters w and b can be determined using Eqs. 1.22 and 1.23 respectively:

$$w = \sum_{i=1}^{S} \alpha_i y_i x_i \tag{1.22}$$

$$b = y_i - w^T x_i \tag{1.23}$$

where x_i represents the support vectors parameter with $i = 1, ..., S$ and S is the number of the support vectors. The number of the support vectors S presents the number of the training instance that satisfy the primal optimization problem given by Eq.1.17. Geometrically, the support vectors parameter are the closest to the optimal hyperplane H_1 and H_2 as shown in Fig. 1.4.

The SVM technique is a kernel-based learning algorithm. In fact, the use of kernel facilitates the classification of complex data since the concept is based on the research of the similarity between linear and non-linear data according to linear separated data.

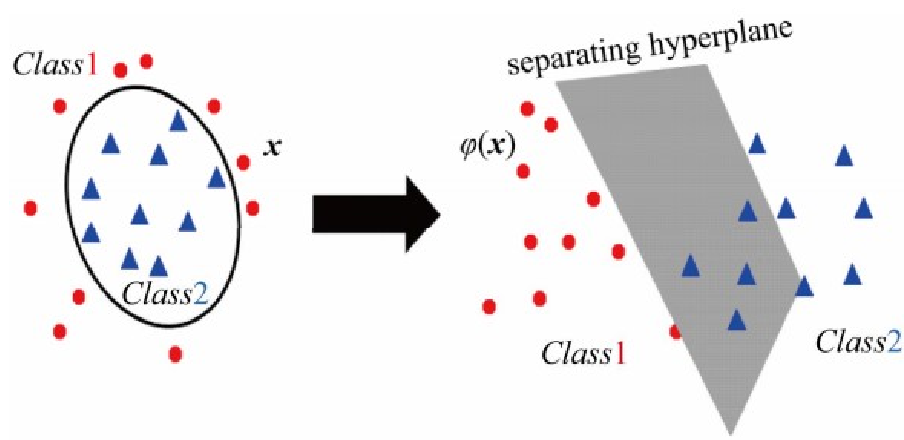

Figure 1.5: Kernel transform for two classes (Mahmoodi et al., 2011).

As shown in Fig. 1.5, the SVM technique uses kernel to maps input vectors into a richer feature space containing non linear features and constructs the proper hyperplane for data separation. In this case the vector x is transformed into $\varphi(x)$ and the kernel function is defined by the following inner product:

$$k(x_i, x) = \varphi(x_i)^T \times \varphi(x) \tag{1.24}$$

The decision function in the case of non linear data is defined by Eq. 1.25:

$$d(x) = sign(\sum_{i=1}^{S} \alpha_i y_i K(x_i, x) + b) \tag{1.25}$$

where the parameter b is given by:

$$b = y_i - \sum_{i=1}^{S} \alpha_i y_i K(x_i, x) \tag{1.26}$$

To ensure that a kernel function actually corresponds to some feature space, it must be symmetric (Núñez et al., 2002):

$$K(x_i, x) = \varphi(x_i)^T \times \varphi(x) = \varphi(x) \times \varphi(x_i)^T = K(x, x_i) \tag{1.27}$$

The kernel function constructs a different nonlinear decision hypersurface in an input space (Huang et al., 2006). The most commonly used kernel functions are (Zanaty, 2012):

— Linear kernel:

$$K(x_i, x) = x_i^T x \tag{1.28}$$

— Polynomial kernel:

$$K(x_i, x) = (1 + x x_i^T)^d \tag{1.29}$$

where d is the polynomial degree.

— Radial basis function (RBF) kernel:

$$K(x_i, x) = exp(-\gamma \parallel x - x_i \parallel^2) \tag{1.30}$$

where γ is a positive parameter controlling the radius.

In this paper, we propose to test the performance of respiratory sounds classification of linear, RBF and polynomial kernels. The accuracy of the classification for the RBF kernel depends on the choice of two parameters (C and γ). The experiment results reveals that the highest accuracy is obtained for $C = 1$ and $\gamma = 1$. For the polynomial kernel, the order of the polynomial kernel is fixed to $d = 4$ to get maximum accuracy, and for the linear kernel we fix the parameter $C = 1$. To classify a new observations x_n that is a D-dimensional feature vector. The SVM solves the Eq. 1.25 and defines the decision of classification d_n.

In this research, we classify respiratory sounds in two classes ($K = 2$), the respiratory sound is segmented uniformly to M overlapped frame so that the corresponding sequence is $X = [x_1, x_2, ..., x_M]$, after solving the Eq. 1.25 the actual output sequence $d = [d_1, d_2, ..., d_M]$ is obtained. For each output $k \in [normal, wheezing]$, the SVM gives a set of M values $[d_{j,k}]$, $j = 1, ..., M$. The sound class is identified as the reference sound with the largest value of the decision function:

$$\bar{k} = arg \max_{1 \le k \le K} \{\bar{d}_k\} \tag{1.31}$$

where the mean values $\bar{d}_k$ are computed over the M frames.

$$\bar{d}_k = \frac{1}{M} \sum_{j=1}^{M} d_{j,k} \;\; k = 1, ..., K \tag{1.32}$$

In this research, the classification of unknown sound is made segment-by-segment ($M = 1$).

1.5.3 Multi-layer perception (MLP)

Multi-layer perception (MLP) neural network is inspired from the biological neuron (Tocchetto et al., 2014). It is the most used type of feed-forward artificial neural network (ANN) (Singhal and Wu, 1988). Fig. 1.6 presents an example of MLP network characterized by D inputs, one hidden of N nodes, and $K = 2$ outputs.

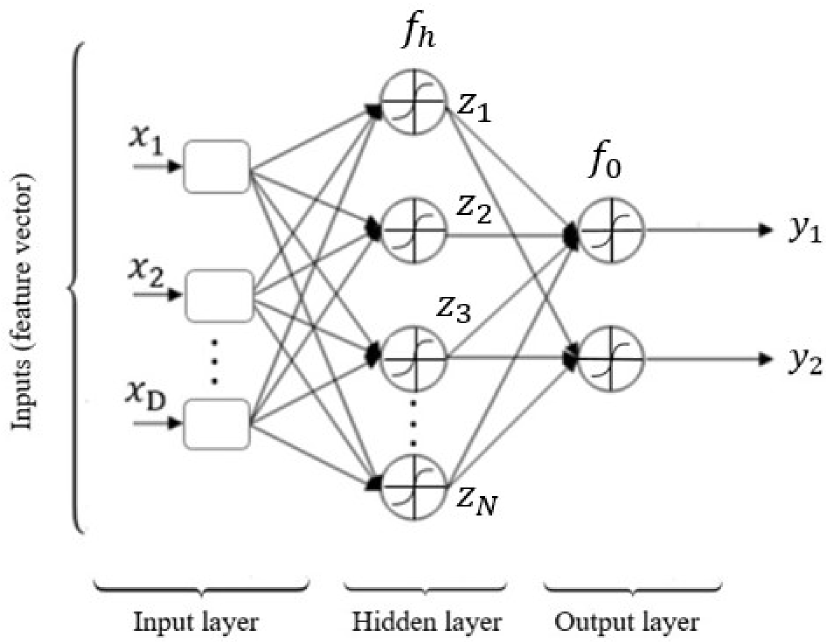

Figure 1.6: Multi-Layer Perception network architecture.

Each node j, in the hidden layer, receives the output of each node i from the input layer through a connection of weight $w^h_{j,i}$. Eq.1.33 presents the output of the node j in the hidden layer.

$$z_j = f_h(\sum_{i=0}^{D} w^h_{j,i} x_i), \quad j = 1, ..., N \tag{1.33}$$

where $f_h(.)$ is is the activation function, $x_0 = 1$, and $w^h_{j,0}$ is the bias of the j^{th} hidden neuron.

The output produced by the node j, in the output layer, is given by:

$$y_j = f_o(\sum_{i=0}^{N} w^o_{j,i} z_i), \quad j = 1, ..., K \tag{1.34}$$

where $f_o(.)$ is the transfer function, $w^o_{j,i}$ is the connection weight, $z_0 = 1$ and $w^o_{j,0}$ is the bias of the j^{th} output neuron.

During the training phase, the connection weights $w = [w^h_{j,i}, w^o_{j,i}]$ are calculated using a set of inputs for which the desired outputs are known $O_n = [(x_1, d_1), (x_2, d_2), ..., (x_n, d_n)]$. The desired outputs d_n is presented by K components which represent the number of the reference classes. For each desired vector d_i, only one component, corresponding to the presented input pattern x_i, is set to 1 and the others are set to 0 (Bahoura, 2016). The training task is done by using the backpropagation (BP) algorithm (Haykins, 1999). The process should be repeated until an acceptable error rate is obtained or a certain number of iterations (N_i) are completed using training examples (Bahoura, 2009).

In this research, the learning rate, the average squared error and the number of iterations (epochs) are fixed to $\eta = 0.01$, $E_{av} = 0.001$ and $N_i = 5000$, respectively. As the connections weights $w_{j,i}$ are initialized with random values, the learning and testing process is repeated 50 times and the average value is taken (Bahoura, 2009). Also, the feature vector is segmented uniformly to M overlapped frame so that the corresponding $X = [x_1, x_2, ..., x_M]$.

To classify an unknown respiratory sounds, the feature vectors set X is presented to the MLP neural network and produce the output sequence $Y = [y_1, y_2, ..., y_M]$. For each output $k \in [normal, wheezing]$, the network provides a set of M values $y_{i,k}$, $i = 1, ..., M$. An unknown respiratory sounds is associated to the reference corresponds to the largest mean value of the outputs:

$$\widehat{k} = arg \max_{1 \leq k \leq K} \{\bar{y_k}\} \tag{1.35}$$

where the mean values $\bar{y_k}$ are computed over the M frames.

$$\bar{y_k} = \frac{1}{M} \sum_{i=1}^{M} y_{i,k}, \; k = 1, .., K \tag{1.36}$$

In this paper, the classification of unknown sound is made segment-by-segment ($M =$ 1).

1.6 Methodology

As shown in Fig. 1.7, the process of classification is divided in two phases, training and testing. During the training phase, a model is generated from the feature characteristics vectors of the training data. In order to classify a new test set, the classifier uses both the extracted feature vector and the trained models. Each classifier uses its proper method to predict the class that data set belongs to.

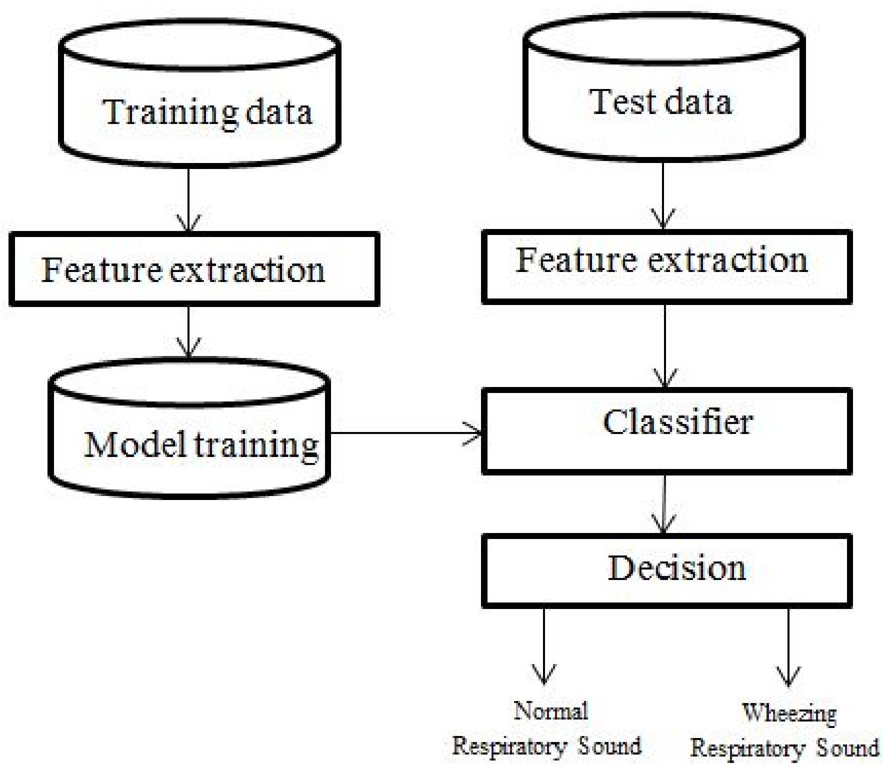

Figure 1.7: Respiratory sounds classification method.

In this study, we suggest to compare possible combination of feature extraction techniques (MFCC, WPT) and the selected machine learning classifiers (k-NN, SVM, MLP).

1.7 Results and discussion

In this section, we detail the experimentation protocol and the database used in this research. The classification accuracy of the proposed combinations are also presented.

1.7.1 Experimentation Protocol

The confusion matrix is used to evaluate classification performance. We define the total accuracy (TA) measurement, which can be calculated from the outcome of the confusion matrix:

$$TA = \frac{TN + TP}{TN + FP + TP + FN} \tag{1.37}$$

where TP (true positives), TN (true negatives), FP (false positives), FN (false negatives) are the outcome of confusion matrix. We use also the evaluation parameters *sensitivity* and *specificity* which represent the rate of identification of wheezing and normal sounds, respectively. This two quantities are defined by Eqs. 1.38 and 1.39, respectively:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1.38}$$

$$Specificity = \frac{TN}{TN + FP} \tag{1.39}$$

In this study, we use the method ”leave-one-out” method, it consists of testing all data sets by using $n - 1$ records for training and the n^{th} record for testing. This process is repeated to all database. For example, when sounds N02-N12 and W02-W12 are used for training, the combination N01-W01 is used for test.

1.7.2 Database

In order to test the classifiers, we use a database composed of 24 records: 12 are obtained from healthy subjects and 12 are obtained from asthmatic subjects. All respiratory sounds are sampled at 6000 Hz, normalized in amplitude, accentuated and manually labeled. The recording lung sounds are obtained from RALE database-CD, ASTRA database-CD and some websites (Bahoura, 2009). Sounds are uniformly divided into overlapped (50 %) segments, 1024 samples each.

Table 1.1: Database characteristics for normal and wheezing sounds.

Normal respiratory sounds			Wheezing respiratory sounds		
File name	Duration (s)	Number of segments	File name	Duration (s)	Number of segments
N01	15.68	183	W01	6.73	77
N02	17.14	199	W02	4.62	53
N03	32.39	378	W03	9.63	111
N04	10.10	117	W04	2.76	31
N05	16.84	196	W05	2.71	30
N06	17.77	207	W06	17.53	204
N07	7.68	88	W07	4.21	48
N08	8.22	94	W08	12.36	143
N09	6.84	179	W09	6.21	71
N10	7.24	183	W10	8.07	93
N11	9.16	106	W11	4.14	47
N12	7.91	91	W12	6.72	77
Total Normal	157.01	1822	Total Wheezes	85,71	985

1.7.3 Results and discussion

In this study, we test the influence of some parameters such as the kernel type for the SVM classifier, the k values for k-NN classifier and the number of hidden neurons (HN) for the MLP classifiers.

Table 1.2, Table 1.3 and Table 1.4 present the confusion matrix of the k-NN classifier with different k values, the SVM classifier with different kernel types and the MLP classifier with different numbers of hidden neurons (HN), respectively. It can be noted that for MFCC-MLP and the WPT-MLP methods, the learning and the testing process is repeated 50 times to take the mean value for each tested record.

Table 1.2: Confusion matrix of k-NN classifier with different k values.

Assigned CLass	MFCC-kNN						WPT-kNN					
	k = 1		k = 5		k = 9		k = 1		k = 5		k = 9	
	N	W	N	W	N	W	N	W	N	W	N	W
True Class												
N	1660	162	1687	135	1704	118	1637	185	1697	125	1711	111
W	241	744	277	708	292	693	315	670	336	649	362	623

Table 1.3: Confusion matrix of SVM classifier with different kernel types.

Assigned CLass	MFCC-SVM						WPT-SVM					
	Linear		RBF		Polynomial		Linear		RBF		Polynomial	
	N	W	N	W	N	W	N	W	N	W	N	W
True Class												
N	1602	220	1414	408	1570	252	1710	112	1742	80	1712	110
W	260	725	113	872	230	755	610	375	755	230	588	397

Table 1.4: Confusion matrix of MLP classifier with different numbers of hidden neurons (HN).

Assigned CLass	MFCC-MLP						WPT-MLP					
	HN = 13		HN = 30		HN = 50		HN = 13		HN = 30		HN = 50	
	N	W	N	W	N	W	N	W	N	W	N	W
True Class												
N	1630.4	191.6	1625.2	196.8	1630.6	191.4	1636.02	185.98	1626.94	195.06	1644.32	177.68
W	207.8	777.2	212.1	772.9	196.5	788.5	536.54	448.46	500.12	484.88	506.5	478.5

Figure 1.8 shows the performances of classifiers in term of the sensitivity (SE), where the highest performance was provided by the technique MFCC-SVM using the RBF kernel (SE = 88.50%). For the WPT extraction technique, the highest sensitivity (SE = 68.0%) is given when combined to the k-NN with (k = 1).

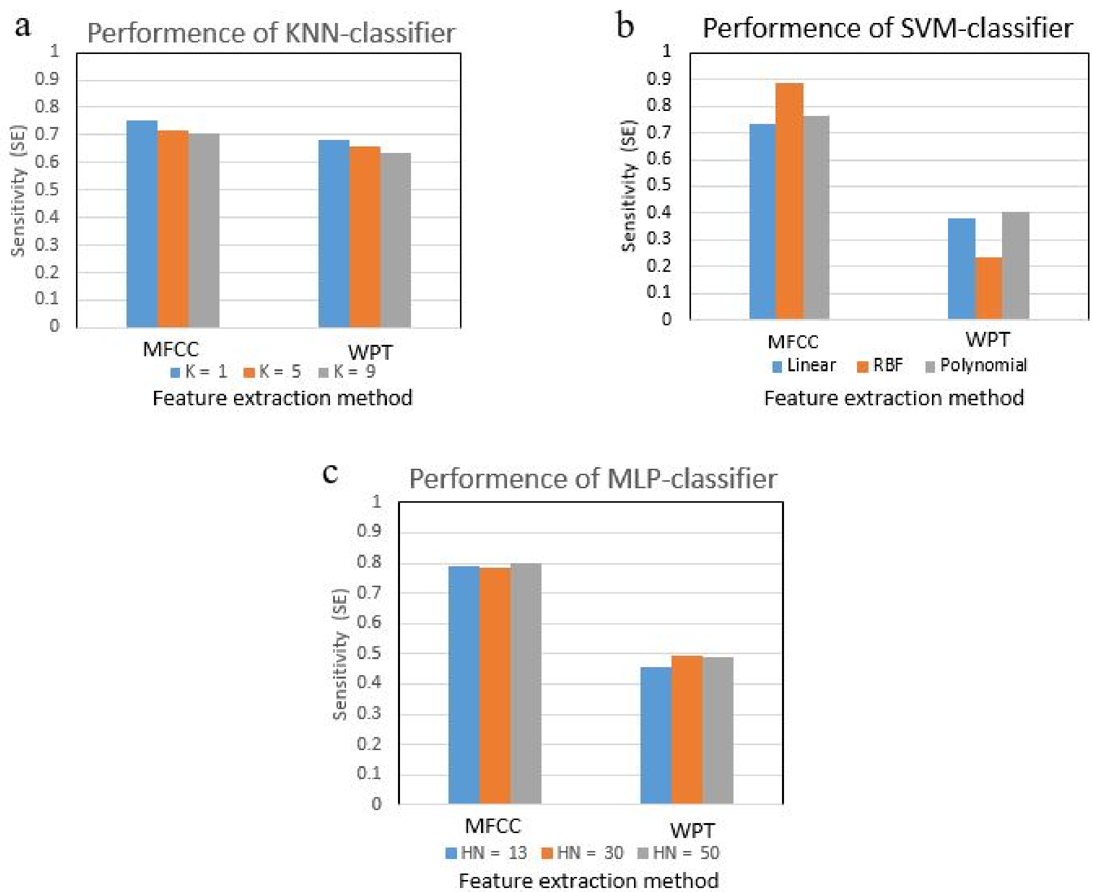

Figure 1.8: Sensitivity (SE) obtained with (a) k-Nearest Neighbor (k-NN), (b) Support Vector Machine (SVM) and (c) multi-layer perceptron (MLP) based classification for proposed feature extraction methods.

The specificity results of the different combinations is shown in Fig. 1.9. For the feature extraction technique MFCC, the highest specificity (SP = 93.50%) is given with the classifier k-NN with (k = 9). For the WPT technique, the combination WPT-SVM give the highest specificity (SP = 95.60%) with the RBF kernel.

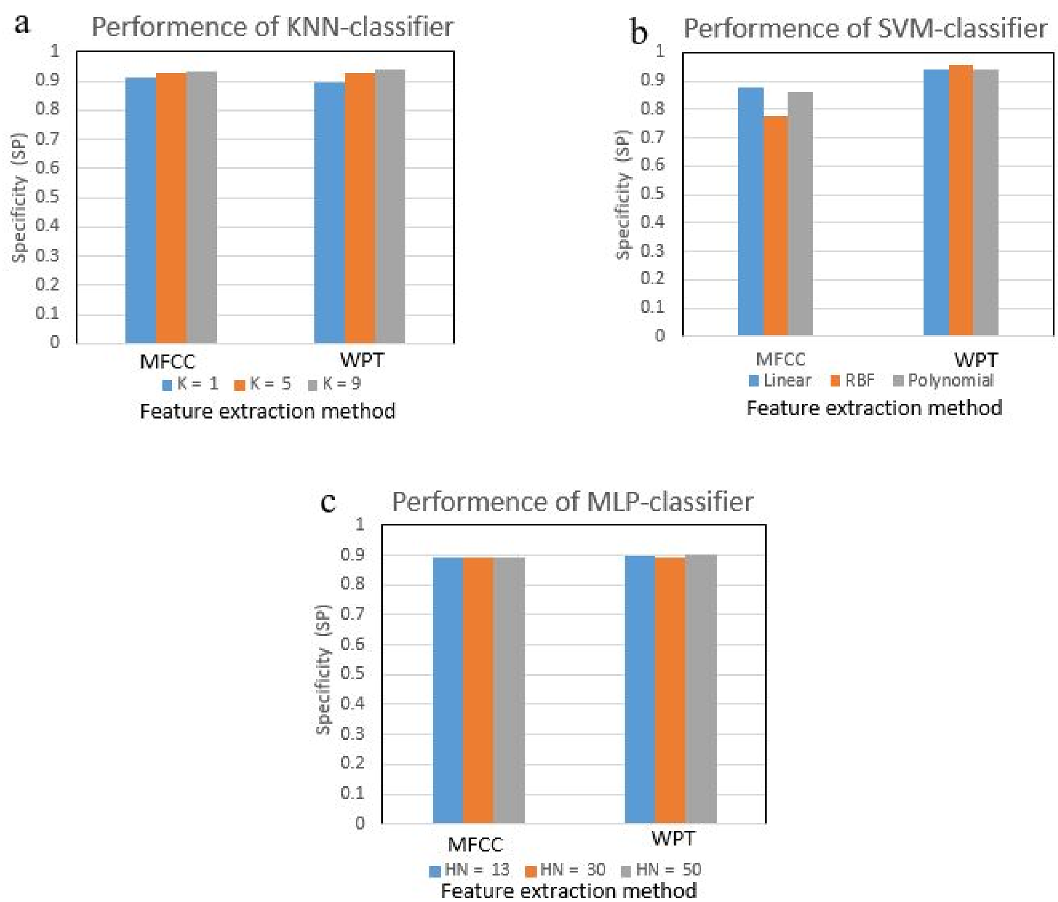

Figure 1.9: Specificity (SP) obtained with (a) k-Nearest Neighbor (k-NN), (b) Support Vector Machine (SVM) and (c) multi-layer perceptron (MLP) based classification for proposed feature extraction methods.

As shown in Fig. 1.10, the MFCC feature extraction technique offers the highest performance for the three classifiers. The combination MFCC-MLP gives the highest performance with an accuracy of (TA=86.2%) using 50 neurons. The MFCC-kNN provide an accuracy of (TA=85.6%) with one neighbor (k = 1). The combination MFCC-SVM provide an accuracy of (TA=82.9%) with the polynomial kernel. For the WPT feature extraction the highest accuracy is giving with the classifier k-NN for (k = 5) and gives an accuracy of (83.6%). The combination WPT-ANN gives an accuracy of (TA=75.6%) using 50 neurons.

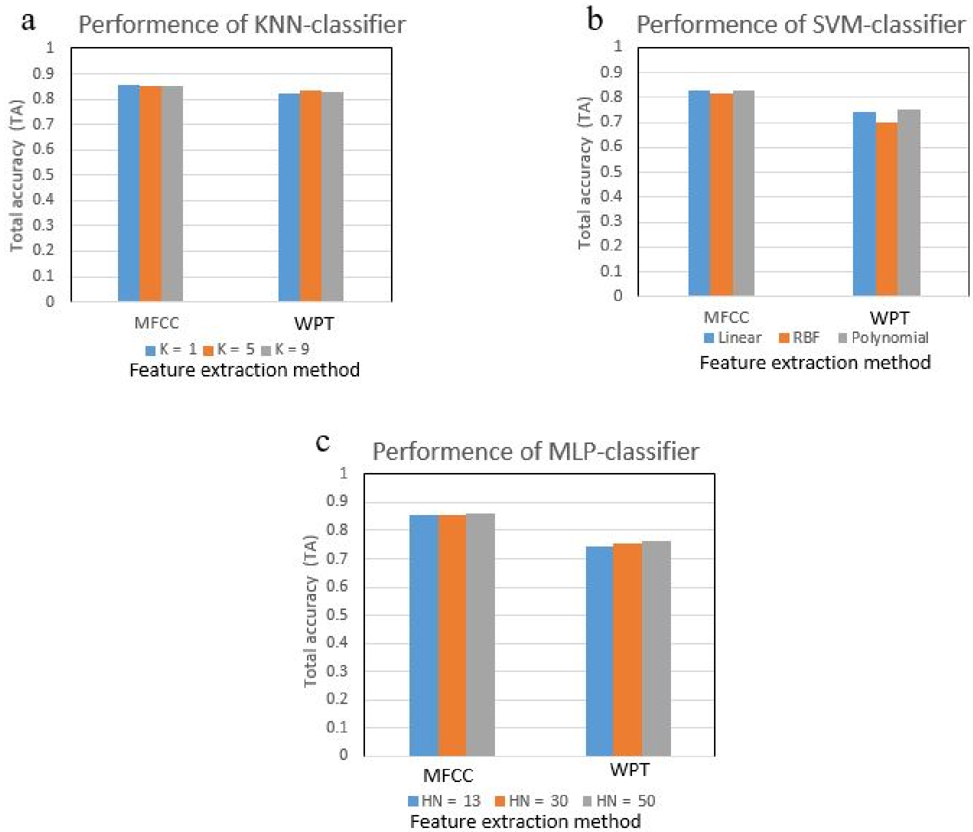

Figure 1.10: Total accuracy (TA) obtained with (a) k-Nearest Neighbor (k-NN), (b) Support Vector Machine (SVM) and (c) multi-layer perceptron (MLP) based classification for proposed feature extraction methods.

1.8 conclusion

In this paper, we propose a comparative study of different learning machines for respiratory sounds classification. The highest accuracy of 86.2% is obtained by the combination MFCC-MLP. It can be noted that the MFCC feature extraction technique improves the performance with the three classifiers.

As future work, we propose to test other combinations and increase the number of classes of respiratory sounds.

Acknowledgement

This research is financially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

ARTICLE 2

FPGA IMPLEMENTATION OF AN AUTOMATIC WHEEZES DETECTOR BASED ON THE COMBINATION OF MFCC AND SVM TECHNIQUES

2.1 Résumé en français du deuxième article

Dans cet article, nous proposons une implémentation matérielle d'un détecteur automatique des sibilants dans les sons respiratoires. Nous avons sélectionné un système bâti sur la combinaison de la technique d'extraction des caractéristiques basée sur les coefficients cepstraux à l'échelle de Mel (MFCC) et le classificateur à base de machine à vecteurs de support (SVM). L'architecture proposée est implémentée sur circuit FPGA en utilisant l'outil de programmation XSG dans l'environnement MATLAB/SIMULINK et la carte de développement ML605 à base du circuit FPGA Virtex-6 XC6VLX240T. Nous proposons d'utiliser la librairie LIBSVM avec le logiciel MATLAB pour extraire les paramètres de SVM pendant la phase d'apprentissage, tandis que la technique d'extraction des caractéristiques et la phase de test sont effectuées en temps-réel sur FPGA. Pour la validation de l'architecture conçue, nous utilisons une base de données composée de 24 sons respiratoires dont 12 sons respiratoires normaux et 12 sons respiratoires contenant des sibilants. L'architecture proposée est présentée en détails dans ce document. Nous présentons également l'utilisation des ressources et la fréquence de fonctionnement maximale pour le circuit FPGA Virtex-6 XC6VLX240T. Les performances de classification obtenues avec l'implémentation à virgule fixe de XSG/FPGA et l'implémentation la virgule flottante de MATLAB sont présentées et comparées.

Ce deuxième article, intitulé "*Implémentation sur FPGA d'un système de détection des sibilants basé sur la combinaison des techniques MFCC et SVM*", fut corédigé par moi-même ainsi que par le professeur Mohammed Bahoura. En tant que premier auteur, ma contribution à ce travail fut l'essentiel de la recherche sur l'état de l'art, le développement de la méthode, l'exécution des tests de performance et la rédaction de l'article. Le professeur Mohammed Bahoura, second auteur, a fourni l'idée originale. Il a aidé à la recherche sur l'état de l'art, au développement de la méthode ainsi qu'à la révision de l'article. Une version abrégée de cet article a été présentée à la conférence *"Conference on Advanced Technologies for Signal and Image Processing, Monastir, (Tunisia) du 21 - 23 March 2016"*.

2.2 Abstract

In this paper, we propose a hardware implementation of an automatic wheezes detector in respiratory sounds. We have chosen a system build on the combination of the feature extraction technique based on Mel-Frequency cepstral coefficients (MFCC) and the support vector machine (SVM) classifier. The proposed architecture is implemented on field programmable gate array (FPGA) using and Xilinx System Generator (XSG) and ML605 development board based on Virtex-6 XC6VLX240T FPGA chip. We propose to use the LIBSVM library in MATLAB environment to extract SVM parameters during the training phase, while the feature extraction technique and the testing phase are performed in real-time on FPGA chip. For the validation of the designed architecture, we use a database composed by 24 records including 12 normal respiratory sounds and 12 respiratory sounds containing wheezes. The implemented architecture is presented in details in this paper. We present also the resource utilization and the maximum operating frequency of the Virtex-6 XC6VLX240T FPGA Chip. The classification performances obtained the fixed-point XSG/FPGA and the floating-point MATLAB implementations are presented and compared.

KEY WORDS: Respiratory sounds, Wheezes, Classification, FPGA, SVM, MFCC.

2.3 Introduction

Asthma is a chronic obstructive pulmonary disease (COPD), for which the number of affected people is constantly increasing. This disease is characterized by the presence of wheezing sounds in patient's respiration. Wheezing sounds are superimposed to normal respiratory sounds and characterized by a duration over 250 ms and a frequency range greater than 400 Hz (Sovijarvi et al., 2000). These sounds with musical aspects, have high amplitude.

Computerized lung sound analysis (CLSA) provides objective evidence serving in the diagnosis of the respiratory illnesses. Significant consideration to lung sounds recognition problems is thoroughly studied by researchers, so that many techniques of signal processing are developed in order to classify different lung sounds. Since 1980, scientists have tried to automatically identify the presence of wheezing (Mazic et al., 2015). To classify respiratory sounds, different combinations of feature extraction and classifier techniques have been documented in the literature: Mel-frequency cepstral coefficients (MFCC) combined with Support vector machine (SVM) (Mazic et al., 2015), k-nearest neighbor (k-NN) (Palaniappan et al., 2014) and Gaussian mixture models (GMM) (Bahoura and Pelletier, 2004). The wavelet transform was used with artificial neural networks (ANN) (Kandaswamy et al., 2004; Bahoura, 2009), and other combinations can be found in (Bahoura, 2009; Palaniappan et al., 2013). Among these techniques, the combination MFCC-SVM based algorithms has been effectively applied to detect wheezing episodes, it can achieve an accuracy for classifying respiratory sounds higher than 95 % (Mazic et al., 2015).

In the last decades, researches have focused on the elaboration of new health care equipment. An effective health care system should be portable, performing in real-time and adaptable to both clinical and domesticated applications. Despite its advantages, the automatic respiratory sounds analysis cannot yet reach a level that can be used as a tool for clinical environment. The elaboration of a real-time sound analysis system is a great challenge for future investigation approaches. The field-programmable gate array (FPGA) is an integrated circuit programmed by the user after fabrication. The hardware description language (HDL) is used to configure FPGAs. The recent progress of these devices enables them to perform different ASICs applications. FPGAs contains DSP slices that can ensure an additional flexibility when programming these devices.

The literature review illustrates a significant use of FPGA approaches in signal processing field, feature extraction technique (Staworko and Rawski, 2010) and classifiers (EhKan et al., 2011; Gulzar et al., 2014). It can be noted that MFCC-based feature extraction technique (Schmidt et al., 2009; Bahoura and Ezzaidi, 2013) has been implemented on FPGA, while the SVM classifier was implemented on FPGA for Persian handwritten digits recognition (Mahmoodi et al., 2011).

In this paper, we propose an FPGA-based implementation of a real-time system to detect wheezing episodes in respiratory sounds of asthmatic patients using Xilinx system generator (XSG). The proposed system is based on the combination of MFCC and SVM techniques for feature extraction technique and classification tasks, respectively. The hardware design is generated and verified in the MATLAB/SIMULINK environment.

This article is organized as following: Section 2 and 3 describes mathematical equations for both the MFCC-based feature extraction technique and the SVM-based classifier, respectively. Section 4 presents the FPGA architecture design and discusses the details of the different blocks. The experimental results are described in Section 5. Finally, conclusion and potential for future works are provided in Section 6.

2.4 Feature Extraction

In this study, we propose to use the MFCC-based feature extraction technique, which approximate the responses of human auditory system. This firmly describes the sound that can be heard over the stethoscope (Mazic et al., 2015). The signal content owing to glottal speech stimulation $s(n)$ will be separated from the one owing to the vocal tract response $h(n)$ (Bahoura and Pelletier, 2004).

$$y(n) = s(n) * h(n) \tag{2.1}$$

As shown in Fig. 2.1, the computation of MFCC for a lung waveform input is composed of different phases. Every state is described by mathematical operations, which will be detailed in this section.

Figure 2.1: Algorithm of the feature extraction technique MFCC.

2.4.1 Signal windowing

The lung sound sampled at 6000 Hz, is first segmented into frames of N samples, and then multiplied by a Hamming window.

$$s(m,n) = s(n) * w(n - mL) \tag{2.2}$$

where m refers to the frame index, n represents the sample time index for the analyzed frame and L is the shift-time step in samples (Bahoura and Ezzaidi, 2013).

2.4.2 Fast Fourier Transform

The spectrum $X(m,k)$ of the windowed waveform is computed using the discrete Fourier transform (DFT).

$$X(m,k) = \sum_{n=0}^{N-1} s(m,n) e^{-j2\pi nk/N} \tag{2.3}$$

where N represents the number of discrete frequencies, $j = \sqrt{-1}$, and k is the frequency index ($k = 0, ..., N-1$).

2.4.3 Mel-Frequency Spectrum

In this step, the Mel-scale filter is applied to the energy spectrum. Fig. 2.2 presents Mel-scale filter bank, which is composed of successive triangular band-pass filters.

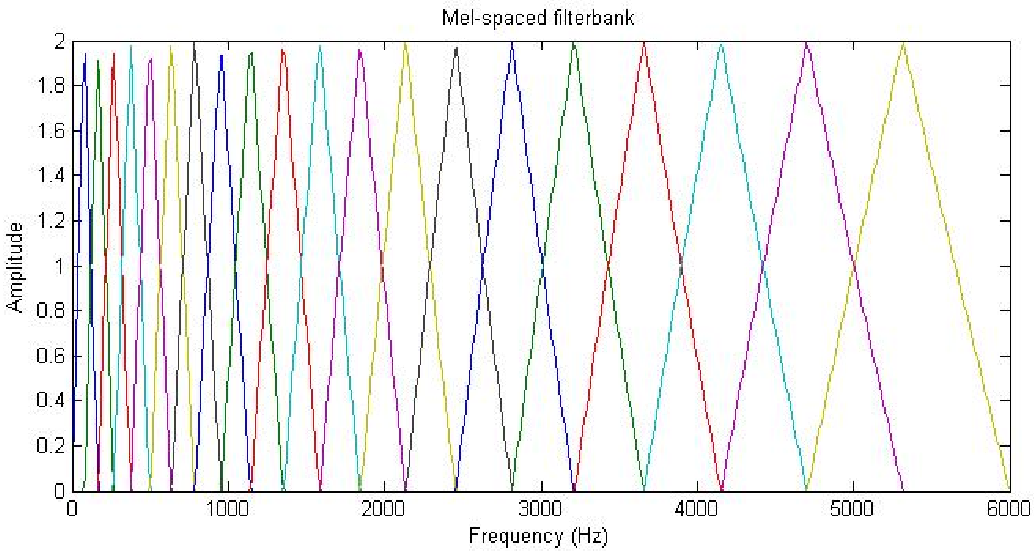

Figure 2.2: A bank of 24 triangular band-pass filters with Mel-scale distribution.

The Mel-scale is linear for the frequencies below 1000 Hz and logarithmic above 1000 Hz (Ganchev et al., 2005). The Mel-filtered energy spectrum is defined by the following equation

$$E(m, l) = \sum_{k=0}^{N-1} |X(m, k)|^2 H_l(k) \tag{2.4}$$

where $H_l(k)$ is the transfer function of the given filter ($l = 1, ..., M$) and $|X(m, k)|^2$ presents the energy spectrum.

2.4.4 Logarithmic energy spectrum

The logarithmic energy output of the l^{th} filter for the current frame m is defined as

$$e(m, l) = log(E(m, l)) \tag{2.5}$$

2.4.5 Discret cosine transform

The MFCC coefficients are obtained by the discrete cosine transform (DCT)

$$c(m, n) = \sum_{l=1}^{M} e(m, l)cos(n(l - 0.5)\pi/M) \tag{2.6}$$

where $(n = 0, ..., P - 1)$ is the index of the cepstral coefficient and $(P \leq M)$ is the needed number of the MFCC. In this case, 15 MFCC coefficient was used: $c_m(2), c_m(3), ..., c_m(16)$. The feature vector is constructed from the MFCC coefficients Eq. 2.6:

$$X_m = [c_m(2), c_m(3), ..., c_m(16)] \tag{2.7}$$

All equations and functions are designed using XSG blocks that are detailed in the section 3.

2.5 Classifier

The support vector machine (SVM) technique has been applied in classification and regression problems. It is a kernel-based learning algorithm that classifies binary or multiclass data. The SVM operates in training and testing phases. During the training phase, the SVM builds a pattern model from the training data and their corresponding class label values, then it uses this model to classify the test set.

Consider SVM for binary classification, a labeled training set of n observations as mentioned in Eq. 2.8:

$$O_n = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\} \tag{2.8}$$

where x_i are the feature vectors and $y_i \in \{1, -1\}$ the associated scalar labels.

As shown in Fig. 2.3, the main purpose of SVM is to define a hyperplane such that: the class labels of data $\{\pm 1\}$ are located on each side of the hyperplane and the distance of the nearest vector of the hyperplane (both classes) is maximum.

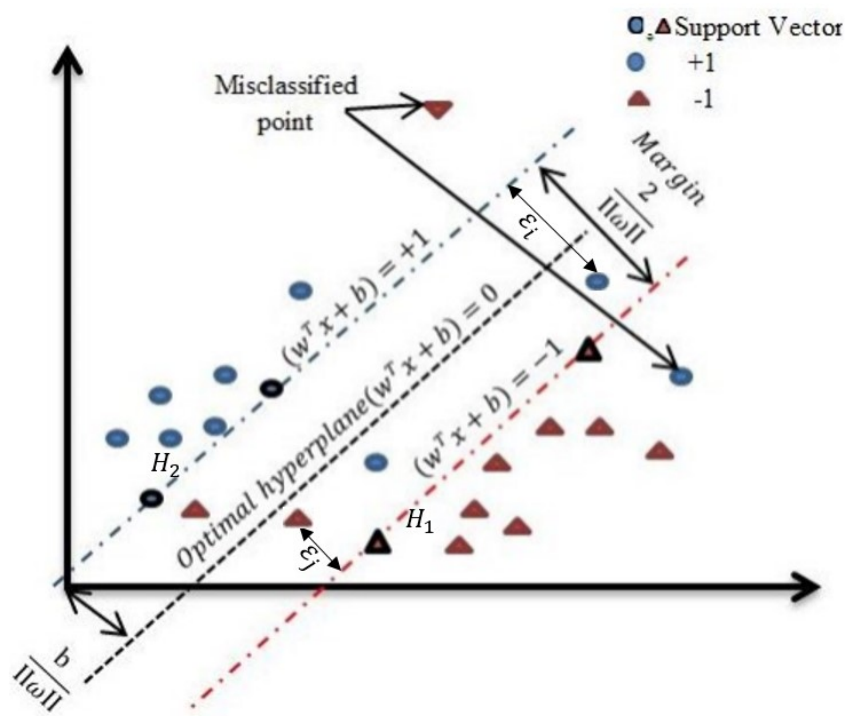

Figure 2.3: Maximum margin hyperplane for an SVM trained with samples from two classes.

where w is a n-dimensional vector, b is the bias, $\| w \|$ is the euclidean norm of w and ξ is slack variables which represent the data that fall into the margin (Mahmoodi et al., 2011).

To obtain the maximum margin separating hyperplane, (Vapnik, 1998) propose to solve the primal optimization problem given by Eq. 2.9.

$$Minimize \ \ \{\tau(w,\xi)\} = \frac{1}{2} \parallel w \parallel_2^2 + C\sum_{i=1}^{n} \xi_i \tag{2.9}$$

subject to :

$$\begin{aligned} y_i(w^T x_i + b) \geqslant 1 - \xi_i \\ \xi_i \geqslant 0 \end{aligned} \tag{2.10}$$

where the parameter C is the misclassification penalty, which is a tradeoff between maximization of the margin and minimization of the error. The primal optimization problem given by Eq. 2.9 concerns the minimization of two quantities. The first term permits to control the margin and the second term limits the number of misclassified points (Chapelle, 2004).

The decision of classification for the linear SVM classifier is presented by Eq. 2.11.

$$d(x) = sign(w^T x + b) \tag{2.11}$$

To determinate the parameters w and b, we should first compute the Lagrange multipliers α_i by solving the following dual Lagrange problem:

$$Maximize \ \ (L_d(\alpha)) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{j,i=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{2.12}$$

subject to:

$$\begin{aligned} 0 \leq \alpha_i \leq C \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{aligned} \tag{2.13}$$

where α_i are the Lagrange multipliers and n is the number of samples. The Lagrange multipliers α_i are calculated by resolving the Lagrange equation 2.12, so we can obtain the w and b coefficients by using the following equations:

$$w = \sum_{i=1}^{S} \alpha_i y_i x_i \tag{2.14}$$

$$b = y_i - w^T x_i \tag{2.15}$$

where x_i represents the support vectors parameter with $i = 1, ..., S$ and S is the number of the support vectors. The number of the support vectors S presents the number of the training instance that satisfy the primal optimization problem given by Eq. 2.9. Geometrically, the support vectors parameter are the closest to the optimal hyperplane H_1 and H_2 as shown in Fig. 2.3.

In the case of nonlinearly separated data, SVM maps data into a richer feature space (H) including nonlinear features, then constructs an hyperplane in that space. In this case the vector x is transformed into $\varphi(x)$.

$$\varphi : R^n \rightarrow H \tag{2.16}$$

$$x \rightarrow \varphi(x) \tag{2.17}$$

The Kernel function is defined by the following inner product:

$$k(x_i, x) = \varphi(x_i) \times \varphi(x) \tag{2.18}$$

The decision function in the context of non-linear data is defined by the following equation:

$$d(x) = sign(w^T \times \varphi(x) + b) \tag{2.19}$$

The software tests reveal that the linear SVM classifier gives the best classification accuracy of 92.78 %. In this study, we used the linear function, because it was demonstrated as quite efficient when classifying respiratory sounds. As mentioned previously, to classify a feature vector x the classifier SVM uses the sign of the following equation:

$$d(x) = sign(w^T x + b) \tag{2.20}$$

By replacing the terms of the Eq. 2.14 in the Eq. 2.20 to classify a new data x is equivalent to

$$d(x) = sign(\sum_{i=1}^{S} \alpha_i y_i x_i^T \times \; x + b) \tag{2.21}$$

In the following, the notation X_s is used to denote the support vectors parameter where $(1 < s < S)$ and S is the number of the support vectors. For an unknown feature vector extraction x_m, in our case we use 15 coefficients of MFCC, the decision of classification for each frame m is simplified to the sign of the following equation:

$$d(x_m) = sign\left(\begin{bmatrix} x_{m,1} & x_{m,2} & \cdots & x_{m,15} \end{bmatrix} \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,15} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,15} \\ \vdots & \vdots & \ddots & \vdots \\ X_{S,1} & X_{S,2} & \cdots & X_{S,15} \end{bmatrix}^T \begin{bmatrix} y\alpha_1 \\ y\alpha_2 \\ \vdots \\ y\alpha_S \end{bmatrix} + b \right) \tag{2.22}$$

where S is the number of the support vectors, $y\alpha_i$, X_s and b are the parameters obtained during the training phase using LIBSVM library in MATLAB environment.

In this study, the implementation of the SVM technique is based on three essential steps. We note that the multiplication of two matrix is based on the sum of products. At first, the support vectors (X_s) are stored in ROM blockset, then multiplicative and addition blocks are used to ensure the multiplication of the MFCC-based feature vector x_m and the support vectors matrix (X_s). The third block concerns the multiplication of the $y\alpha_i$ vector with the output of the additional block, the result is accumulated and added to the b value. The sign of the decision is identified with the threshold block, this process is then repeated for each frame.

2.6 FPGA Architecture Design

In this study, we use the Xilinx system generator (XSG) in MATLAB/SIMULINK environment to design the hardware wheezes detector system. The use of this high-level programming tool (XSG) ensures an easy and rapid prototyping of complex signal processing algorithms. This model serves directly for the hardware implementation through the compilation using the hardware co-simulation in XSG environment. In the current implementation, we use Hamming window with a length of 1024 samples, 24 triangular mel-filters, and 15 DCT coefficients.

Beforehand, we performed the training phase of the SVM classifier in MATLAB using LIBSVM library (Chang and Lin, 2011), which is available online. The extracted parameters are used during the test phase on the hardware chip. The LIBSVM generates a modeltrain that contains the needed parameters (X_S, $y\alpha$, b). Table 2.1 shows the training results of the SVM classifier for the combination of test (Normal01-Wheeze01). The test reveals that the highest accuracy is obtained with $C = 1$.

Table 2.2 shows the used hardware resources for the Virtex-6XC6VLX240T FPGA device and the maximum operating frequency of the implemented architecture, as reported by Xilinx ISE Design Suite 13.4.

Table 2.1: Computed SVM parameters as reported by LIBSVM with $C = 1$ for the combination of test (Normal01-Wheeze01).

	Class 1 & Class 2	
	Normal respiratory sounds	Wheezing respiratory sounds
S	36	36
$y\alpha$	$[y\alpha_1, y\alpha_2, ..., y\alpha_{72}]^T$	
b	5.372	

Table 2.2: Resource utilization and maximum operating frequency of the Virtex-6 XC6VLX240T Chip, as reported by Xilinx ISE Design Suite 13.4.

Resource utilization	
Flip Flops (301,440)	15,207 (5%)
LUTs (150,720)	17,945 (11%)
Bonded IOBs (600)	20 (3%)
RAMB18E1s (832)	4 (1%)
DSP48E1s (768)	122 (15,0%)
Slice (37,680)	5,373 (14%)
Maximum Operating Frequency	27.684 MHz

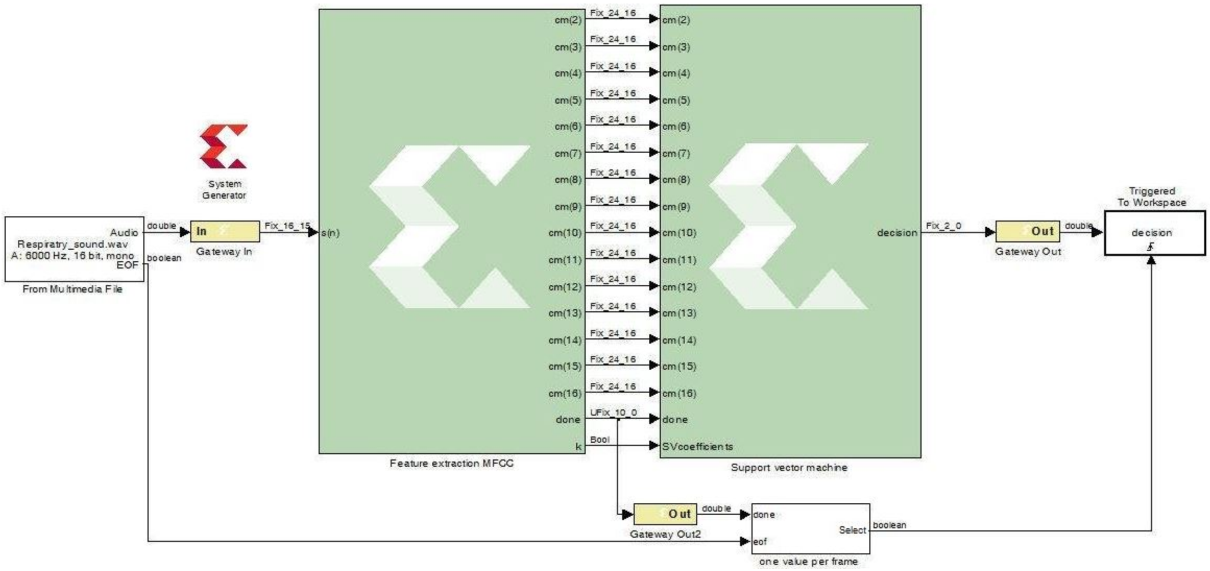

Figure 2.4: MFCC-SVM architecture based on Xilinx system generator (XSG) blockset for the automatic wheezes detector.

Figure 2.4 represents the top-level block diagram of the proposed automatic wheezes detector. The hardware architecture uses Xilinx System Generator (XSG) and the Virtex-6 FPGA ML605 evaluation kit.

Figure 2.5 represents the subsystem for feature extraction technique MFCC with block details. Fig. 2.6 shows the block details of the linear-kernel SVM design and an optional subsystem designed with SIMULINK blocks, which select one decision of classifications for every frame. More details on the FPGA implementation of MFCC feature extraction technique and the SVM classifier can be found in (Bahoura and Ezzaidi, 2013) and (Mahmoodi et al., 2011), respectively.

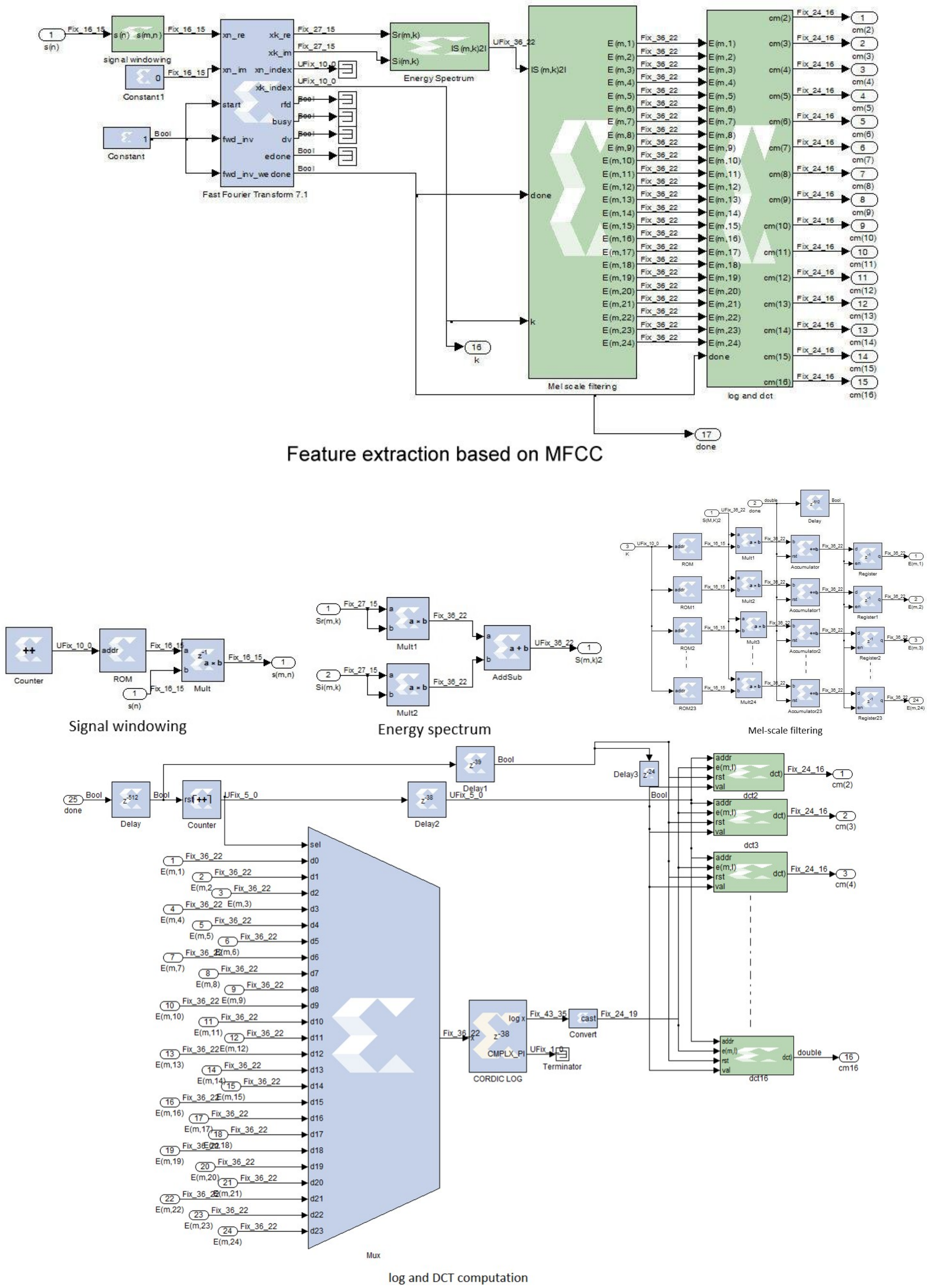

Figure 2.5: MFCC feature extraction technique architecture based on Xilinx system generator (XSG) blockset. The complete subsystem of the MFCC is given on the top followed by details of the different subsystems.

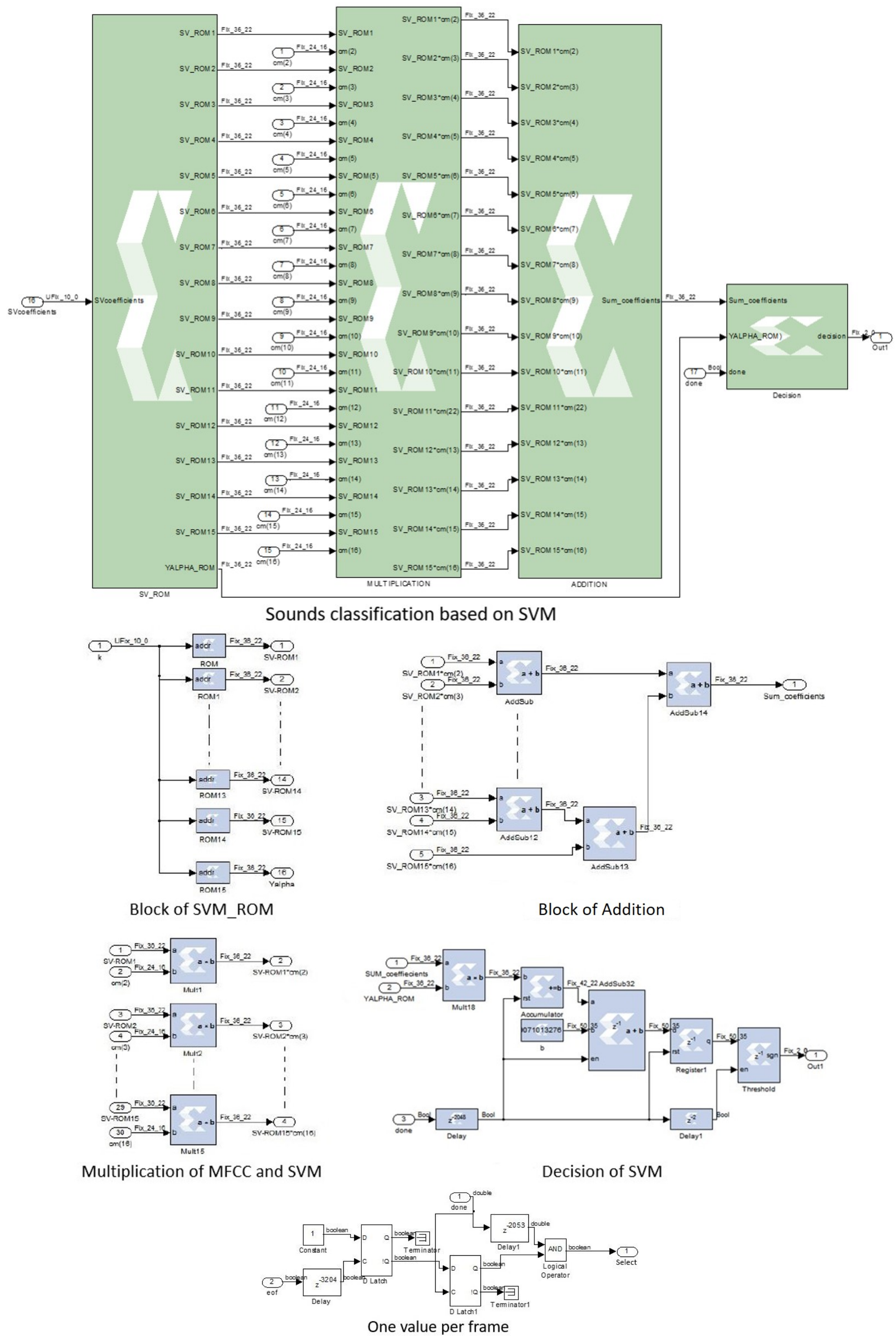

Figure 2.6: SVM classifier architecture based on Xilinx system generator (XSG) blockset for wheezes classification. The complete subsystem of the SVM is given on the top followed by details of the different subsystems.

2.7 Results and Discussion

2.7.1 Database

To evaluate the proposed architecture, two classes of respiratory sounds (normal and wheezing) are used for training and testing records. The recording lung sounds are obtained from the RALE database-CD, ASTRA database-CD and some online websites. 12 records from healty subjects and 12 others provided from asthmatic subjects, where some wheezing sounds include monophonic and polyphonic wheezes. Each respiratory sound is sampled at 6000 Hz. The $\{\pm 1\}$ labels indicate the type of class that the tested segment belongs to.

Table 2.3: Database characteristics for normal respiratory sounds and asthmatics.

Normal respiratory sounds			Wheezing respiratory sounds		
File name	Duration(s)	Number of segment	File name	Duration(s)	Number of segment
Normal01	5,24	30	Wheezes01	4.55	26
Normal02	3.68	21	Wheezes02	3.56	18
Normal03	3.85	22	Wheezes03	7.50	45
Normal04	6,89	40	Wheezes04	2.72	15
Normal05	7,67	44	Wheezes05	4.20	24
Normal06	6,83	40	Wheezes06	7.10	41
Normal07	6,66	39	Wheezes07	8.02	47
Normal08	3,75	22	Wheezes08	3,08	18
Normal09	4,5	26	Wheezes09	6.72	39
Normal10	4,72	27	Wheezes10	3.70	21
Normal11	9,15	53	Wheezes11	5.12	30
Normal12	7,90	46	Wheezes12	5.31	31
Total normal	70,84	410	Total wheezes	61,58	355

It can be noted that the wheezing database contains 6 mixed records (normal and wheezing) so the total of normal segment is 483 and the total of wheezes segment is 282 segments.

2.7.2 Protocol

Sampled at the frequency of 6000 Hz, the respiratory sounds used in this paper are manually labeled into their corresponding classes. We named class 1 with label $\{+1\}$ for normal data and class 2 with label $\{-1\}$ for wheezing data. As mentioned in the previous section, we perform the training phase of the SVM technique off-line. However, the feature extraction technique and the test phase are both performed on FPGA. For the training phase, we use ”leave-one-out” method, it consists of testing all data sets by using $n-1$ records of data for training and the n^{th} record for testing. For example, when sounds Normal02-Normal12 and Wheeze02-Wheeze12 are used for training, the combination Normal01-Wheeze01 is used for test.

2.7.3 Simulation of XSG blocks

As shown in Fig. 2.7, a Hamming window is applied to the input signal $s(n)$ to provide non-overlapping frames $s_m(n)$. We observe a delay time of 2180 samples for the freq.index(k) signal in the Fig. 2.7 (c). It represents the delay between the first sample $s_m(n)$ and the first sample $|S_m(k)|^2$ corresponding to the FFT block delay. The power spectrum $|S_m(k)|^2$ of the frequency components is computed and the third feature vector component is illustrated for each frame in Fig. 2.7 (d).

Figure 2.7 (g) shows an additional delay of 1024 at the output addition block of Fig. 2.6. So, the signal is delayed by 3204 = 2180 + 1024 caused by both MFCC and SVM classifier, the first decision of classification is made in the next frame with a delay of two samples caused by the decision block . Fig. 2.7 shows the simulation results of respiratory sounds containing normal and wheezing episodes. The Fig. 2.8 represents the computation performance of the feature extraction vectors based on MFCC technique using the fixed-point XSG and the floating-point MATLAB implementation. Fig. 2.8 (a,c) presents the result of the MFCC-based features using the floating-point MATLAB implementation for normal and wheezing

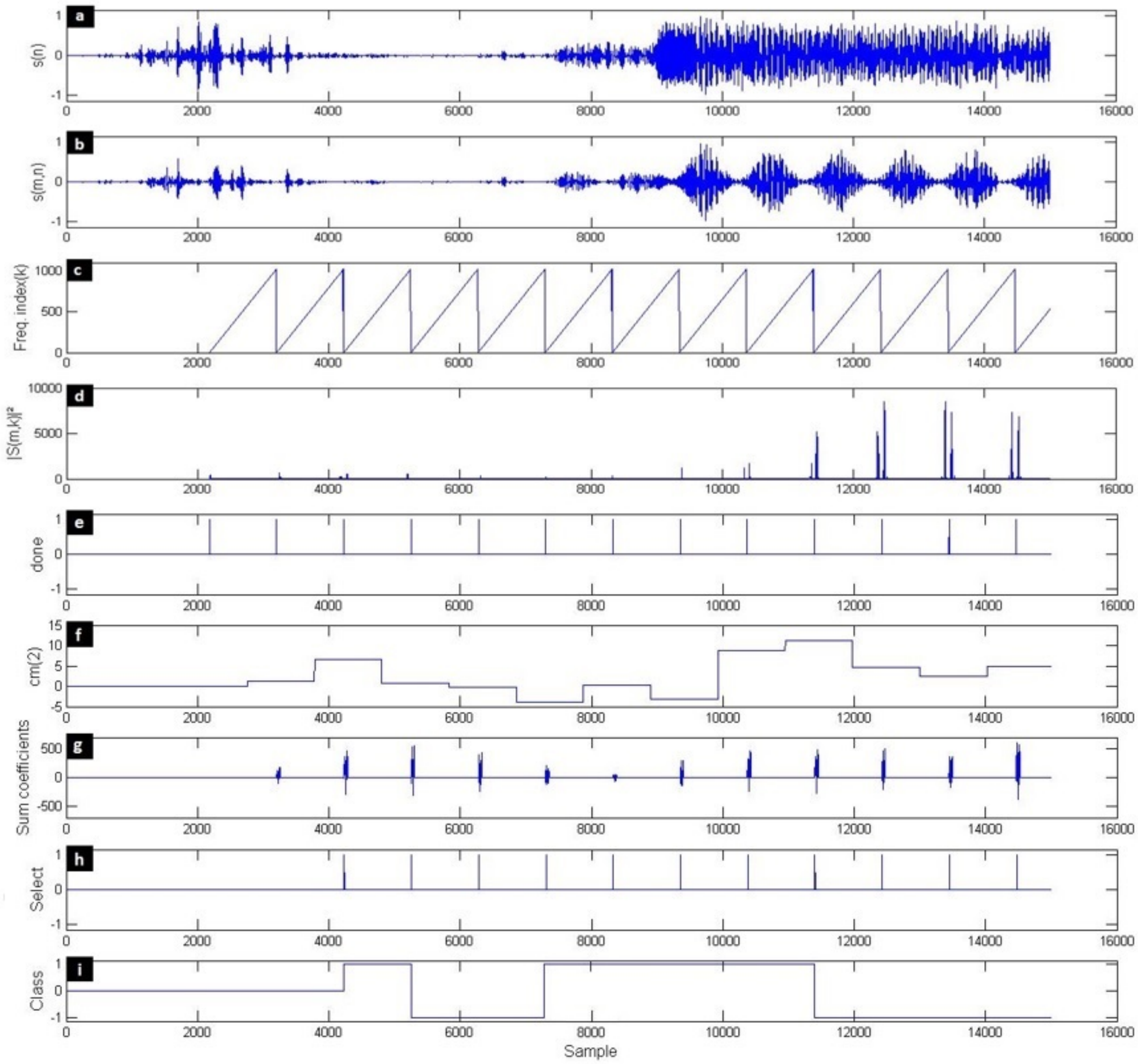

Figure 2.7: Response signals obtained during the characterization/classification of respiratory sounds. (a) input signal s(n); (b) windowed signal s(m,n); (c) frequency.index(k); (d) power spectrum $|S_m(k)|^2$; (e) done signal; (f) third component of the feature vector cm(2); (g) output of the addition block; (h) select; (i) recognized class.

respiratory sounds, respectively, and the Fig. 2.8 (b,d) shows the associated MFCC-based features using the fixed-point XSG for the same signals. The computation performance of the feature extraction vectors based on MFCC represents equivalent results for both the fixed-point XSG and the floating-point MATLAB implementation.

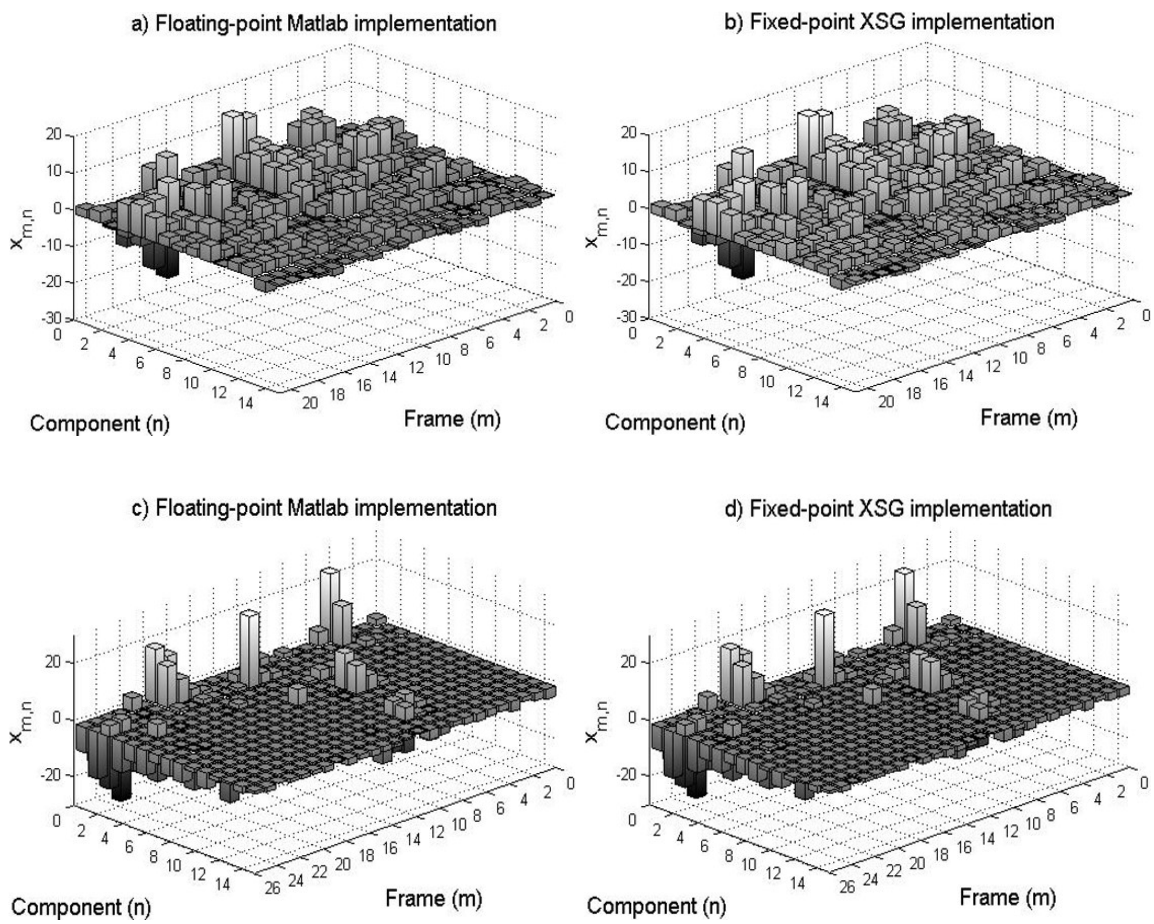

Figure 2.8: Feature extraction vectors based on MFCC technique. (a) MFCC-based features X_m obtained with MATLAB implementation for normal respiratory sound; (b) MFCC-based features X_m obtained with fixed-point XSG implementation for normal respiratory sound; (c) MFCC-based features X_m obtained with MATLAB implementation for wheezing respiratory sound; (d) MFCC-based features X_m obtained with fixed-point XSG implementation for wheezing respiratory sound.

2.7.4 Hardware Co-Simulation

After the simulation test of the proposed design in the SIMULINK/XSG environment, we are interested in a second step to generate the hardware model. In our study, we configure the ML605 evaluation board in XSG to succeed the hardware co-simulation. In fact, the maximum operating frequency of 27.684 MHz is inferior to the minimum clock frequency target design in the ML605 evaluation kit. Fig. 2.9 presents the hardware co-simulation block generated by the hardware co-simulation XSG compilation mode. As shown in this figure, the input wave is provided from the multimedia file of the SIMULINK environment and sent to Virtex-6 chip via the JTAG connection. The designed architecture is performed in FPGA device and the cable ensures the recovery of classifications result on the other hand.

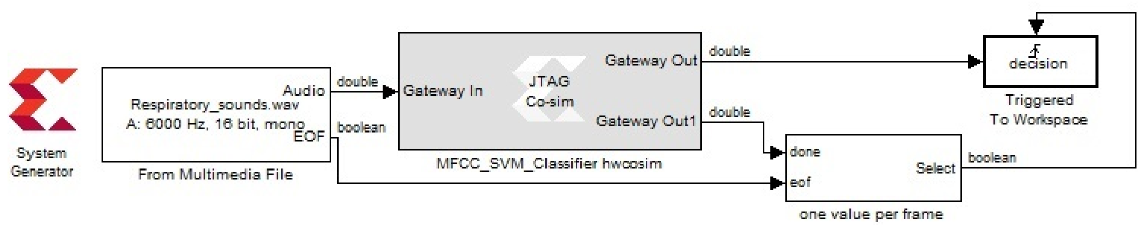

Figure 2.9: The hardware co-simulation of the MFCC-SVM classifier.

2.7.5 Classification Accuracy

In order to compare the performance provided by the floating-point MATLAB and the fixed-point XSG implementations, we use confusion matrix to evaluate classification performance. We define the total accuracy (TA) measurement, which can be calculated from the outcome of the confusion matrix as such:

$$TA = \frac{TN + TP}{TN + FP + TP + FN} \tag{2.23}$$

where TP (True Positives), TN (True Nositives), FP (False Positives), FN (False Negatives) are the outcome of confusion matrix.

2.7.6 Simulation results using XSG blockset and MATLAB

In this section, we presents the simulation results of the designed XSG-based architecture, compared with those provided by MATLAB floating-point implementation.

The classification performances for the proposed architecture are presented in Table 2.4. The fixed-point XSG implementation gives equivalent performance as the floating-point MATLAB using the described database. Table 2.5 present the confusion matrix of the well classified sounds against the class recognizer provided with both the fixed-point XSG and the floating-point MATLAB implementations of the MFCC/SVM based classifier.

The total accuracy of the floating-point MATLAB is 92.78 % whereas the accuracy provided by XSG-based architecture is 93,24 %. The difference can be justified by the quantization errors in XSG-based SVM classifier (Mahmoodi et al., 2011).

Table 2.4: Performances obtained with XSG and MATLAB based implementations.

Respiratory sounds	Total Accuracy %	
	XSG	MATLAB
Normal	96.06	95.85
Wheezes	90.42	89.71
Total	93.24	92.78

As shown in Fig. 2.10, the designed architecture implemented with fixed-point XSG provides equivalent performances than the floating-point MATLAB for both feature extraction technique MFCC and the classifier SVM. The results of feature extraction technique give the same performance using MATLAB and XSG, so that he slight difference shown in Ta-

Table 2.5: Confusion matrix of XSG and MATLAB based implementations.

True class	Assigned class (XSG)		Assigned class (MATLAB)	
	Normal	Wheezes	Normal	Wheezes
Normal	464	19	463	20
Wheezes	27	255	29	253

ble 2.4 and Table 2.5 is caused by the block of the classifier SVM, this difference is referred to the quantization errors in XSG (Mahmoodi et al., 2011).

The classification results of normal and pure wheezing respiratory sounds are presented in Fig. 2.10 (a,b) for both XSG-based architecture and MATLAB software. The respiratory sound record presented in Fig. 2.10 (c) contains normal and wheezes sounds. In this case, both architectures (XSG and MATLAB) can distinguish between the frame containing normal lung sounds from those that containing wheezes. Finally, the designed architecture implemented with the fixed-point XSG gives equivalent accuracy than those obtained with the floating-point MATLAB.

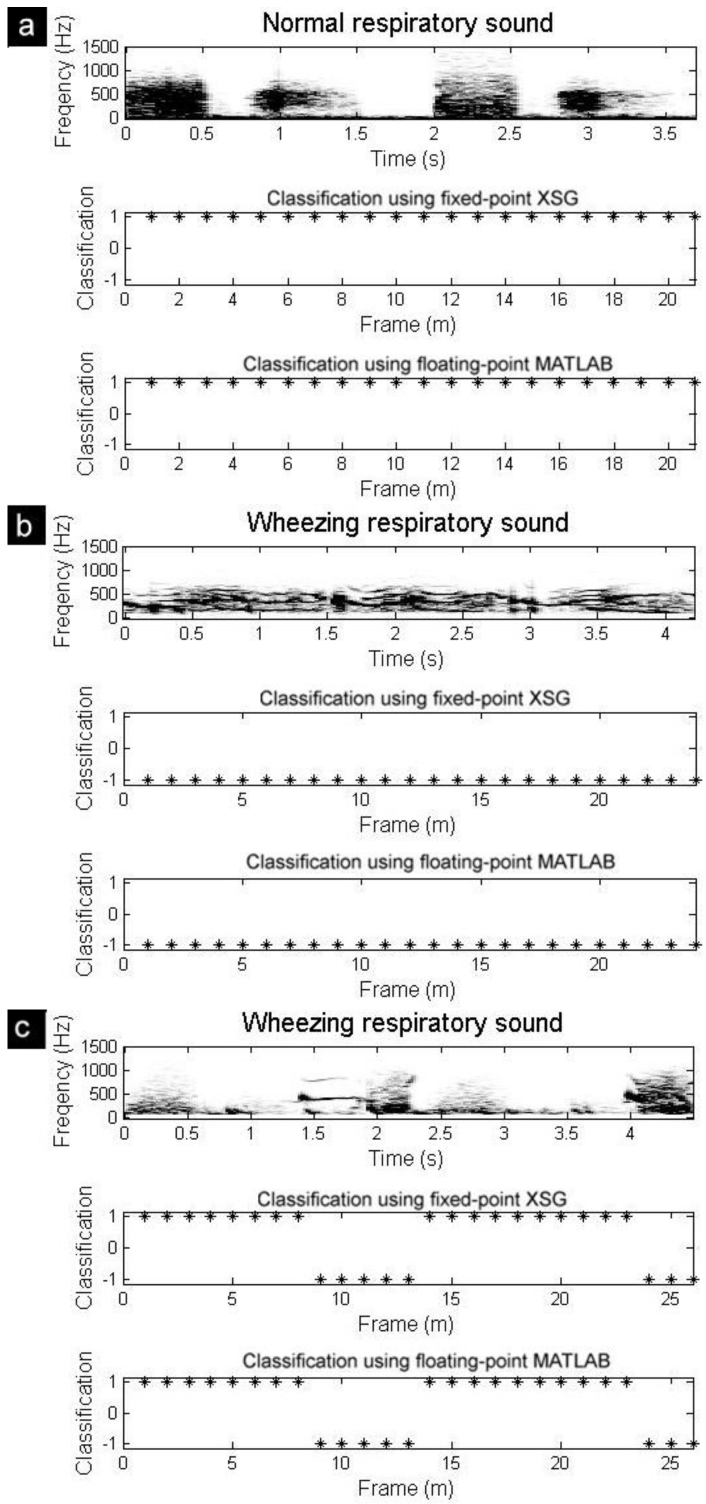

Figure 2.10: Classification of normal (a) and wheezing (b and c) respiratory sounds into normal $\{+1\}$ and wheezing $\{-1\}$ frames. Each subfigure includes the spectrogram of the tested sound (top) and the classification results using fixed-point XSG (middle) and floating-point MATLAB (bottom).

2.8 Conclusion

In this paper, FPGA-based architecture of an automatic wheezes detector using MFCC and SVM has been suggested. Based on the tested respiratory sound records, the classification performances obtained with hardware fixed-point architecture are compared to those obtained with the floating-point MATLAB implementation. The designed architecture can be applied to other respiratory sound classes.

In future works, the implementation of other feature extraction techniques is recommended to improve the identification accuracy.

Acknowledgment

This research is financially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

CONCLUSION GÉNÉRALE

Le travail présenté dans ce mémoire s'inscrit dans le domaine du traitement des sons respiratoires. L'objectif de cette étude se résume dans l'étude et l'implémentation des différents systèmes de détection des sibilants dans les sons respiratoires enregistrés sur des patients asthmatiques dans le but d'un traitement en temps-réel.

Les systèmes de reconnaissance utilisés dans cette recherche fonctionnent en deux phases : apprentissage et test. La phase d'apprentissage consiste à créer un modèle prédictif à partir des vecteurs de caractéristiques issus de la technique d'extraction des caractéristiques de la base de données. Pendant la phase de test, le classificateur utilise le modèle conçu ainsi que des opérations et des méthodes propres pour chaque classificateur pour définir la décision de l'élément du test.

Dans la première partie de ce projet, nous avons proposé une étude comparative des machines d'apprentissage les plus utilisées pour la classification des sons respiratoires. Nous avons choisi de tester trois classificateurs : k-NN, SVM et MLP. Nous proposons d'utiliser les coefficients cepstraux à l'échelle de Mel (MFCC) et la transformée par paquets d'ondelettes (WPT). Le taux de précision maximale de 86.2 % est obtenu par la combinaison MFCC-MLP. Nous mentionnons que la technique d'extraction des caractéristiques MFCC donne de bonnes performances avec les différents classificateurs, à savoir avec un taux de reconnaissance supérieur à 80 %. Concernant la technique WPT le meilleur taux de reconnaissance est 83.6 % et il est obtenu avec le classificateur k-NN.

La deuxième partie du projet propose de concevoir un système de classification en temps-réel des sons respiratoires en deux catégories : normaux et sibilants. La combinaison des techniques d'extraction des caractéristiques et de classification est implémentée sur l'outil Xilinx system generator (XSG) opérant dans l'environnement MATLAB/SIMULINK. Les résultats des tests révèlent que les performances de la classification obtenues avec l'implémentation matérielle sont semblables à ceux obtenus avec le logiciel MATLAB. L'architecture conçue peut être généralisée à d'autres classes de sons respiratoires.

Comme futur travaux, nous proposons de tester d'autres combinaisons de techniques d'extraction des caractéristiques ainsi que de classficateurs, nous proposons aussi d'augmenter le nombre de classes de sons respiratoires comme les ronchus et les crépitants.

RÉFÉRENCES

Alsmadi, S., Kahya, Y. P., 2008. Design of a DSP-based instrument for real-time classification of pulmonary sounds. Computers in biology and medicine 38 (1), 53–61.

Amudha, V., Venkataramani, B., 2009. System on programmable chip implementation of neural network-based isolated digit recognition system. International Journal of Electronics 96 (2), 153–163.

Bahoura, M., 2009. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Computers in biology and medicine 39 (9), 824–843.

Bahoura, M., 2016. FPGA implementation of blue whale calls classifier using high-level programming tool. Electronics 5 (1), 8.

Bahoura, M., Ezzaidi, H., 2013. Hardware implementation of MFCC feature extraction for respiratory sounds analysis. In: 8th Workshop on Systems, Signal Processing and their Applications. Algiers, Algeria, May 12-15, 2013, pp. 226–229.

Bahoura, M., Pelletier, C., 2004. Respiratory sounds classification using cepstral analysis and gaussian mixture models. In: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS). Vol. 1. San Francisco, Canada, September 1–5, 2004, pp. 9–12.

Bahoura, M., Simard, Y., 2012. Serial combination of multiple classifiers for automatic blue whale calls recognition. Expert Systems with Applications 39 (11), 9986–9993.

Billionnet, C., 2012. Pollution de l'air intérieur et santé respiratoire: prise en compte de la multi-pollution. Ph.D. thesis, Université Pierre et Marie Curie-Paris VI, Paris, France.

Boulet, L.-P., Côté, P., Bourbeau, J., 2014. Le réseau québécois de l'asthme et de la maladie pulmonaire obstructive chronique (RQAM): un modèle d'intégration de l'éducation thérapeutique dans les soins. Education Thérapeutique du Patient-Therapeutic Patient Education 6 (1), 10301.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3), 27.

Chapelle, O., 2004. Support vector machines: principes d'induction, réglage automatique et connaissances a priori. Ph.D. thesis, Université Pierre et Marie Curie-Paris VI, Paris, France.

Chen, C.-H., Huang, W.-T., Tan, T.-H., Chang, C.-C., Chang, Y.-J., 2015. Using k-nearest neighbor classification to diagnose abnormal lung sounds. Sensors 15 (6), 13132–13158.

EhKan, P., Allen, T., Quigley, S. F., 2011. FPGA implementation for GMM-based speaker identification. International Journal of Reconfigurable Computing. Vol. 2011, pp. 1–8.

Ertekin, S., 2009. Learning in extreme conditions: Online and active learning with massive, imbalanced and noisy data. Ph.D. thesis, The Pennsylvania State University, Pennsylvania, USA.

Gabbanini, F., Vannucci, M., Bartoli, G., Moro, A., 2004. Wavelet packet methods for the analysis of variance of time series with application to crack widths on the brunelleschi dome. Journal of Computational and Graphical Statistics 13 (3), 639–658.

Ganchev, T., Fakotakis, N., Kokkinakis, G., 2005. Comparative evaluation of various MFCC implementations on the speaker verification task. In: 10th International Conference Speech and Computer (SPECOM). Vol. 1. Patras, Greece, 17-19 October, 2005, pp. 191–194.

Gulzar, T., Singh, A., Rajoriya, D. K., Farooq, N., 2014. A systematic analysis of automatic speech recognition: an overview. International Journal of Current Engineering and Technology 4 (3), 1664–1675.

Haykins, S., 1999. A comprehensive foundation on neural networks, 2nd ed. Prentice Hall Upper Saddle River, NJ, USA.

Huang, T.-M., Kecman, V., Kopriva, I., 2006. Kernel based algorithms for mining huge data sets. Springer, Heidelberg.

Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., Jayaraman, S., Malmurugan, N., 2004. Neural classification of lung sounds using wavelet coefficients. Computers in Biology and Medicine 34 (6), 523–537.

Kozak, K., Kozak, M., Stapor, K., 2006. Weighted k-nearest-neighbor techniques for high throughput screening data. International Journal of Biomedical Sciences 1 (3), 155–160.

Laennec, R. T., 1819. De l'auscultation médiate ou traité du diagnostic des maladies des poumons et du coeur. Paris: Brosson and Chaudé, pp. 181–210.

Lin, B.-S., Yen, T.-S., 2014. An FPGA-based rapid wheezing detection system. International Journal of Environmental Research and Public Health 11 (2), 1573–1593.

Mahmoodi, D., Soleimani, A., Khosravi, H., Taghizadeh, M., 2011. FPGA simulation of linear and nonlinear support vector machine. Journal of Software Engineering and Applications 4 (05), 320–328.

Manikandan, J., Venkataramani, B., 2011. Design of a real-time automatic speech recognition system using Modified One Against All SVM classifier. Microprocessors and Microsystems 35 (6), 568–578.

Mazic, I., Bonkovic, M., Dzaja, B., 2015. Two-level coarse-to-fine classification algorithm for asthma wheezing recognition in children's respiratory sounds. Biomedical Signal Processing and Control 21, 105–118.

Núñez, H., Angulo, C., Català, A., 2002. Rule extraction from support vector machines. In: European Symposium on Artificial Neural Networks. Bruges, Belgium, April 24-26, 2002, pp. 107–112.

Palaniappan, R., Sundaraj, K., Ahamed, N. U., 2013. Machine learning in lung sound analysis: a systematic review. Biocybernetics and Biomedical Engineering 33 (3), 129–135.

Palaniappan, R., Sundaraj, K., Sundaraj, S., 2014. A comparative study of the SVM and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. Biomedicalcentral bioinformatics 15 (1), 1–8.

Pan, S.-T., Lan, M.-L., 2014. An efficient hybrid learning algorithm for neural network-based speech recognition systems on FPGA chip. Neural Computing and Applications 24 (7-8), 1879–1885.

Pasterkamp, H., Kraman, S. S., Wodicka, G. R., 1997. Respiratory sounds: advances beyond the stethoscope. American journal of respiratory and critical care medicine 156 (3), 974–987.

Pelletier, C., 2006. Classification des sons respiratoires en vue d'une détection automatique des sibilants. Université du Québec à Rimouski, Québec, Canada.

Ramos-Lara, R., López-García, M., Cantó-Navarro, E., Puente-Rodriguez, L., 2013. Real-time speaker verification system implemented on reconfigurable hardware. Journal of Signal Processing Systems 71 (2), 89–103.

Sankur, B., Kahya, Y. P., Güler, E. Ç., Engin, T., 1994. Comparison of ar-based algorithms for respiratory sounds classification. Computers in Biology and Medicine 24 (1), 67–76.

Schmidt, E. M., West, K., Kim, Y. E., 2009. Efficient acoustic feature extraction for music information retrieval using programmable gate arrays. In: 10th International Society for Music Information Retrieval Conference (ISMIR). Kobe, Japan, October 26-30, 2009, pp. 273–278.

Shaharum, S. M., Sundaraj, K., Palaniappan, R., 2012. A survey on automated wheeze detection systems for asthmatic patients. Bosnian Journal of Basic Medical Sciences 12 (4), 249–255.

Singhal, S., Wu, L., 1988. Training multilayer perceptrons with the extende kalman algorithm. In: Advances in Neural Information Processing Systems. pp. 133–140.

Sovijarvi, A., Malmberg, L., Charbonneau, G., Vanderschoot, J., Dalmasso, F., Sacco, C., Rossi, M., Earis, J., 2000. Characteristics of breath sounds and adventitious respiratory sounds. European Respiratory Review 10 (77), 591–596.

Staworko, M., Rawski, M., 2010. FPGA implementation of feature extraction algorithm for speaker verification. In: Proceeding of the 17th International Conference "Mixed Design of Integrated Circuits and Systems", MIXDES 2010. Wroclaw, Poland, June 24-26, 2010, pp. 557–561.

Tocchetto, M. A., Bazanella, A. S., Guimaraes, L., Fragoso, J., Parraga, A., 2014. An embedded classifier of lung sounds based on the wavelet packet transform and ANN. In: IFAC Proceedings Volumes 47 (3), 2975–2980.

Vapnik, N. V., 1998. Statistical learning theory. Wiley New York.

Wang, J., Wang, J.-F., Weng, Y., 2002. Chip design of MFCC extraction for speech recognition. The VLSI Journal Integration. Vol. 32, pp. 111–131.

Zanaty, E., 2012. Support vector machines (SVMs) versus multilayer perception (MLP) in data classification. Egyptian Informatics Journal 13 (3), 177–183.