Faculty of Engineering of the University of Porto

FEUP

Heart Sound Analysis for Cardiac Pathology Identification: Detection and Characterization of Heart Murmurs

João Manuel Patrício Pedrosa

FINAL VERSION

Dissertation conducted under the
Integrated Master in Bioengineering
Biomedical Engineering Branch

Advisor: Dr. Tiago T. V. Vinhoza
Co-advisor: Dr. Ana Castro

31st July 2013

João Manuel Patrício Pedrosa, Author

Tiago Travassos Vieira Vinhoza, Advisor

Resumo

A auscultação cardíaca é um meio de diagnóstico inegável, no entanto, recentemente tem vindo a perder a sua importância devido ao crescimento de novas tecnologias como o ecocardiograma. Isto tem vindo a tornar a análise digital de sons cardíacos um domínio de investigação em rápida evolução à medida que são feitas tentativas para criar sistemas de suporte à decisão que sejam capazes de diminuir os gastos hospitalares ajudando os médicos da primeira linha a chegar ao diagnóstico através de algoritmos capazes de segmentar um fonocardiograma nos seus ciclos cardíacos e detectar e caracterizar a presença de sopros. A análise de sons cardíacos já foi abordada de várias maneiras incluindo análise temporal, tempo-frequência, análise não-linear ou análise baseada nos elementos caóticos de um sinal, para além de combinações destes campos. A separação de um som nas suas várias componentes também tem surgido como um campo promissor nesta área.

A partir de uma base de dados construída num ambiente clínico real, a base de dados DigiScope, o trabalho proposto consiste no desenvolvimento de novos algoritmos para a segmentação de sons cardíacos em ciclos cardíacos assim como a extração de características que permitam a deteção de sopros e a sua classificação. Para cumprir este propósito, um algoritmo baseado na função autocorrelação (ACF) foi desenvolvido de modo a estimar a frequência cardíaca média, excluir regiões corrompidas por ruído externo e realizar a segmentação do sinal. A classificação em S_1 ou S_2 foi feita através da duração da sístole e da diástole ou, nos casos em que tal não era possível, através de um modelo de Markov oculto (HMM) baseado nas características tempo-frequência de cada som. Uma grande variedade de características, 250, foi extraída de modo a descrever completamente cada segmento. Um classificador k-médias foi utilizado para detectar os sopros.

O algoritmo de segmentação foi testado na base de dados do desafio "PASCAL Classifying Heart Sounds" *sendo obtida uma sensibilidade e valor preditivo positivo de 89,2% e 98,6% respectivamente. O desvio médio entre o valor dos tempos anotados na base e o valor estimado pelo algoritmo foi de 9,8ms. A classificação feita pela HMM foi avaliada em ambas as bases de dados tendo sido obtidos os valores de erro de 11,88% para a base de dados Pascal e 13,57% para a base de dados DigiScope. A deteção de sopros foi avaliada na base de dados DigiScope em duas situações diferentes: uma com divisão aleatória dos segmentos em teste e treino e a outra com a mesma divisão feita de acordo com os pacientes. A primeira situação originou uma sensibilidade de 98,42% e especificidade 97,21%. A segunda situação teve um desempenho inferior com um erro mínimo de 33,65%. O ponto de operação foi no entanto alterado para uma sensibilidade de 69,67% e uma especificidade de 46,91% obtendo um erro total de 38,90%. Isto foi feito variando a percentagem de segmentos classificados como sopro necessários para o sinal ser considerado como tal.*

Abstract

Cardiac auscultation has been an undeniable bedside diagnostic modality but is recently losing its importance due to the rise of new technologies such as the echocardiogram. This has turned the digital analysis of heart sounds an evolving field of study as attempts are made to create decision support systems capable of diminishing hospital costs and helping physicians in the first screening through algorithms capable of segmenting a phonocardiogram into its cardiac cycles and detecting and characterizing murmurs. Heart sound analysis has been approached in several ways namely time domain analysis, time-frequency domain analysis, nonlinear and chaos based analysis, perceptual analysis and combinations between these. Blind source separation has also emerged as a promising field of study.

From a database acquired in a realistic clinical environment, the DigiScope database, the work proposed consist in the development of novel algorithms for the segmentation of the heart sounds into heart cycles as well as feature extraction and murmur detection and classification. For this, an autocorrelation function (ACF) based algorithm was developed to estimate the average heart rate, exclude regions corrupted by noise and perform the signal's segmentation. The classification into S_1 or S_2 of each sound was conducted according to the length of systole and diastole or, in dubious cases, by a time-frequency based hidden Markov model (HMM). A wide amount of features, 250, were extracted to provide a complete description of each segment. A k-means classifier was used to detect the murmurs.

The segmentation algorithm was tested in the "PASCAL Classifying Heart Sounds" challenge database and a sensitivity and PPV of 89,2% and 98,6% were obtained, respectively. The average deviation between the time value annotated in the database and the value returned by the segmentation algorithm was computed obtaining the value of 9,8ms. The HMM classification was evaluated in both databases both obtaining similar values of 11,88% error for the Pascal database and 13,57% for the DigiScope database. The murmur detection was evaluated in the DigiScope database in two different situations, with a random division between train and test set and a division according to patients. The first returned sensitivity and specificity of 98,42% and 97,21% respectively. The second division had a far worse performance with a minimum error of 33,65%. The operating point was chosen at sensitivity 69,67% and a specificity 46,91% for a total error of 38,90% by varying the percentage of segments classified as murmurs needed for a signal to be classified as presenting murmur.

Agradecimentos

A realização deste trabalho não teria sido possível sem a contribuição de variadas pessoas às quais dirijo o meu mais sincero obrigado. Aos orientadores que me acolheram, Tiago Vinhoza e Ana Castro pelos seus muitos conselhos e acompanhamento e também pela liberdade que me deram de perseguir os meus próprios objetivos e a todo o grupo DigiScope que permitiu que este trabalho fosse alguma vez pensado. Outros me ajudaram no caminho percorrido neste semestre e devo agradecer principalmente à Dra. Filipa Pedrosa por todo o apoio em questões fisiológicas e médicas e ao Prof. João Monteiro pelas questões computacionais e algorítmicas.

Além das referidas acima, muitas outras pessoas me ajudaram percorrer o caminho até aqui, mais do que aquelas que poderei referir. Aos meus pais e à minha família que estiveram sempre do meu lado e me carregaram ao colo pelo menos na primeira metade do caminho. À Helena Monte por toda a coragem que me conseguiu transmitir um dia atrás do outro nestes últimos cinco anos sem nunca deixar de acreditar em mim. Aos meus amigos, que sempre tomaram conta de mim e aos quais peço que o continuem a fazer por anos vindouros. A Metal&Bio e à Praxe que sejam sempre o Porto seguro que para mim o foram ao longo deste curso.

"A man who dares to waste one hour of time
has not discovered the value of life."

x

Contents

List of Figures

List of Tables

Abbreviations and Symbols

List of abbreviations (ordered alphabetically)

ACF	Autocorrelation Function
CWT	Continuous Wavelet Transform
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
HMM	Hidden Markov Model
MFCC	Mel-Frequency Cepstrum Coefficient
PCG	Phonocardiogram
PPV	Positive Predictive Value
ROC	Receiver Operating Characteristic
S_1	First Heart Sound
S_2	Second Heart Sound
S_3	Third Heart Sound
S_4	Fourth Heart Sound
SFFS	Sequential Forward Feature Selection
TFR	Time-Frequency Representation
VFD	Variance Fractal Dimension

List of symbols

ð	Average Temporal Deviation of Heart Sound Detection

Chapter 1

Introduction

1.1 - Overview and Motivation

In the continually developing clinical environment, occasionally there are opportunities for certain diagnosis modalities to emerge and become a part of the clinical practice. For some time now and for a number of historical and technical reasons there has been an opportunity for computer-aided auscultation. Computer-aided auscultation consists in the digital processing of a PCG signal to aid physicians in the task of correctly interpreting it and achieving a diagnosis. The importance of auscultation has been proved through extensive clinical use throughout the last centuries and, consequently, any tool that aids physicians in the difficult process of assessing an auscultation would be more than welcome. One of the main areas within this field of opportunity is the detection of murmurs to assess important cardiopathies. Murmurs, symptomatic of some cardiopathies can be recorded in a PCG and even though cardiologists can easily perceive them, general practitioners may have more difficulty and thus, a system that would be able to detect a murmur, characterize it and diagnose or aid in the diagnosis of a patient would be of major importance. A system of computer-aided auscultation must however fulfill many requirements to be reliable, not only in terms of algorithmy, but also in the methods of database acquisition and classification [1].

1.2 - Goals

The goal of this dissertation was then to design a robust algorithm capable of segmenting a PCG signal into heart cycles and heart sounds and also able to detect murmurs when they are present and pinpointing their temporal location within the heart cycle (systole/diastole). The designed algorithm should be independent of other sources of information such as ECG or echocardiography. Furthermore, the designed algorithm should be able to cope with the amount of noise and

variability that exists in a real clinical environment. The performance evaluation of the developed algorithms is also a priority, as it will allow its validation.

1.3 - Contributions

Three main contributions arose from the development of this project. First of all, a novel algorithm for heart sound segmentation was designed to tackle the problems of noise and variability caused by PCG acquisition in a real clinical environment. Secondly, a non-duration based HMM was designed to classify the heart sounds in a sequence as S_1 or S_2. This HMM, based mainly in observations obtained from the spectral shape of the heart sounds, is especially important in the heart sound segmentation of children and adult with higher heart rates. Finally, a new and extensive set of features for murmur detection in systolic segments was experimented with composed of features from different analysis domains.

1.4 - Structure of the Dissertation

Besides the introduction, this dissertation is divided into 5 more chapters. Chapter 2 addresses the historical evolution of auscultation and of the heart, namely its anatomy, physiology and the sounds produced by it, pathological or not. Chapter 3 provides an insight into the usual methods used on computer-aided auscultation, focusing on the segmentation of heart sounds and detection of murmurs, together with some of the results obtained so far and their limitations. In Chapter 4 the proposed algorithm is presented together with all the scientific background needed to apply it. Chapter 5 exposes the results obtained and discusses their significance. Finally, Chapter 6 serves as a conclusion to this dissertation providing final remarks and future improvements for the algorithm.

Chapter 2

Auscultation and the Heart

As one of the vital organs, the heart was long identified as the centre of the entire body. Its significance was however converted from philosophical to completely physiological as scientific knowledge and medicine evolved.

2.1 - Historical Overview

The practice of auscultation was first registered during the Hippocratic period, from 460 to 370 BC. This was done by applying one's ear to the chest or abdomen to listen to sounds from within the body. This approach is called immediate auscultation because it uses no apparatus to transport the sound from the body to the physician. Hippocrates described a number of different internal sounds from which different diagnosis could be formulated such as *"You shall know by this that the chest contains water but not pus, if in applying the ear during a certain time on the side, you perceive a noise like that of boiling vinegar."* [2].

A formal description of the heart sounds however, was only completed in 1628 in William Harvey's *De Motu Cordis* in which he concludes that the main function of the heart is to pump the blood to the extremities of the body and the heart sounds are characterised as *"two clacks of a water bellows to raise water"*. Nevertheless, immediate auscultation presented a number of challenges. For one, it was far from what can be called an efficient method. The sounds that could be heard were of low volume and only its main components could be heard. This problem was even further aggravated in the case of overweight patients. Not only this, but the technique was far from suitable to be used by male physicians, the majority at the time, to examine female patients due to the fact that the physician needed to press his ear against the patients' breasts and because the very technique became unpractical when the patient had larger breasts. In summary, auscultation remained a technique used by a select few being palpation and percussion the main examination techniques at the time [2].

The turning point for this situation was in 1816 when Réné Théophile Hyacinthe Laennec

invented mediate auscultation. Faced with the problems of immediate auscultation, Laennec used a quire of paper rolled into a cylinder to hear the heart sounds of a patient. Later on, he developed a proper apparatus consisting of a perforated wooden cylinder, a funnel-shaped plug and a stopper. He named this device the Cylinder or Stethoscope (from Greek *stēthos* 'breast' + *skopein* 'look at'). About this he wrote "*I could perceive the action of the heart in a manner much more clear and distinct than I had ever been able to do by the immediate application of the ear...*" [2, 3].

In little time, Laennec's stethoscope gained popularity and despite the rather lack of scientific exactness of his interpretations of the heart sounds, by the 1830s it was an undeniable bedside tool for examination of chest problems and something expected by patients seeing a doctor. The use of the stethoscope grew throughout the 19th and 20th centuries to become a medical tool of excellence that could obtain remarkable results depending on a meticulous physician training for a correct and trained use. It was however this same need for a thorough clinical formation that made the stethoscope so useful that generated its current near-demise [2].

More recent stethoscopes were given the ability to record sound and transmit them to a computer not only for a later review and patient follow-up but also for computer processing and analysis [2].

Auscultation is, in spite of its advantages, being replaced by evolving technologies, such as echocardiography and other diagnostic modalities, which provide larger amounts of information in a considerably easier way. This, combined with the ever more compressed time that limits opportunities for clinical trainees to gain mastery through practice and repetition has caused the stethoscope to lose its importance. For the final diagnosis, physicians depend more and more on other technologies. These new modalities bring greater costs for hospitals and a time delay in the patient's final diagnosis, which results in greater distress as they await the diagnostic and run sequential tests. In consequence, attention has been recently given to decision support systems to help first screening general clinicians make faster and cheaper decisions using a tool that they are certain to carry along with them at all times, the stethoscope [4].

Figure 2.1 - Stethoscope evolution; on the left Laennec's cylinder, on the right a digital stethoscope (Littmann® 3200).

2.2 - The Heart and Heart Sounds

The heart is the center of the cardiovascular system. It serves as a pump with the purpose of transporting the blood from the tissues to the lungs and vice versa. To perform this action, the heart, throughout our existence, goes through a sequence of events in an organized, timely and most precise manner. Each sequence of events of the heart, from the beginning of one heartbeat to the beginning of the next is called cardiac cycle [5].

A cardiac cycle is divided into systole and diastole. During diastole, first the semilunar (aortic and pulmonary) valves close, the atrioventricular (mitral and tricuspid) valves are open, and the whole heart is relaxed. After this, the atrium contracts, and blood flows from the atrium to the ventricles. During the systole, the atrioventricular valves close, the ventricles begin to contract and there is no change in volume. After this, the ventricles become empty and continue contracting, and the semi-lunar valves open. Finally, the pressure decreases, no blood enters the ventricles, the ventricles stop contracting and begin to relax and the semilunar valves close due to the pressure of blood in the aorta to renew the cycle [5, 6].

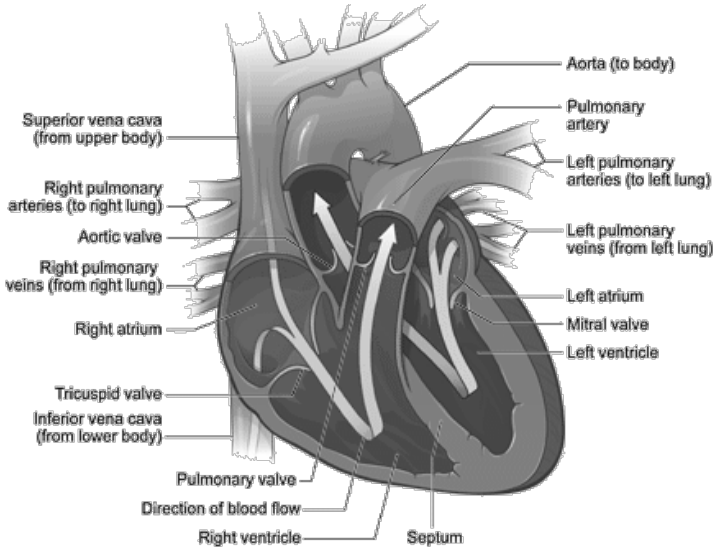

Figure 2.2 - A human heart showing inner chambers, valves, blood flow and general anatomy.

The heart sounds are generated by the beating heart and the resultant blood flow within, particularly the turbulent blood flow that, whenever it occur,s causes vibrations that may be heard by the human ear. Laminar blood flow, which occurs throughout almost the whole cardiovascular system, produces no sound. For very long it was thought that the very vibrations caused by the closure of the valves produced the heart sounds, but this theory was proven wrong. A healthy adult heart produces two sounds (S_1 and S_2), often described as *lub-dub*. Additional sounds such as murmurs and gallop rhythms (S_3 and S_4) may also be present and may or may not be a sign of pathology. Because different sounds irradiate to different areas on the chest depending on their origin, normally auscultation is done in what are called the auscultation areas or spots: the aortic, pulmonic, tricuspid and mitral. The location of these spots is shown on **Figure 2.3**. An audio signal of a collection of heart sounds is called a PCG and its total frequency range extends within 20-1000Hz [6-8].

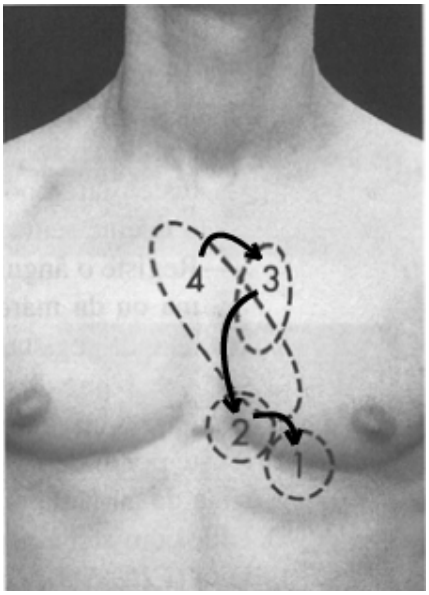

Figure 2.3 - Chest with the location of the auscultation spots: 1-mitral spot; 2-tricuspid spot; 3-pulmonar spot; 4-aortic spot.

2.2.1- First and Second Heart Sound

The first and second heart sounds are completely non-pathological and correspond to the turbulence of blood caused by the closure of the heart valves. S_1 is caused by the closure of the atrioventricular valves and S_2 is caused by the closure of the semilunar valves. These two sounds can then be used as boundaries between systole and diastole. Systole occurs between S_1 and S_2 and diastole from S_2 to S_1. S_1 and S_2 are normally the highest amplitude events in a PCG and have frequencies between 20-200Hz. The normal duration of these sounds varies from 70ms to 140ms. Their frequency spectrum is very similar but it has been shown that S_2s have larger amplitudes above 150Hz than S_1s [5, 9, 10].

2.2.2 - Gallop Rhythms

Besides S_1 and S_2, gallop rhythms S_3 and/or S_4 may be present. The term gallop rhythms comes from the fact that instead of the usual *lub-dub* sound sequence, an additional sound is heard causing a sound similar to a gallop, *lub-dub-ta* or *ta-lub-dub*. Both of these sounds occur within the diastole but S_3 comes right after S_2 whereas S_4 comes right before S_1. The origin of S_3 is discussed but is thought to be caused by the tension of the chordae as the blood distends the left ventricle. It occurs normally in children and young adults but is, in other subjects, usually pathologically indicating heart failure. S_4 is known to be caused by the forcing of the blood from the atrium to the left ventricle when it is still noncompliant. S_4 marks atrial contraction and is always pathological. S_3 and S_4 have low amplitudes and frequencies between 15-65Hz. The duration of these sounds varies from 40ms to 60ms [5, 9].

2.2.3 - Murmurs

Heart murmurs are distinguishable from heart sounds due to their longer duration. They may either be innocent or pathological. Most murmurs are caused by the turbulent blood flow that results, for example, from the narrowing (stenosis) or leaking (regurgitation) of the heart valves or due to abnormal blood passages in the heart. According to the physiological situation leading to the murmur, different sounds are generated. Murmurs can be systolic, diastolic or continuous according to its temporal location within the heart cycle. The intensity variation of the murmur is also important and murmurs are characterised as *crescendo*, *decrescendo*, *crescendo-decrescendo* or *plateau*. Murmurs usually have higher frequencies than the rest of the components of a PCG (200-700Hz) and its amplitude may vary according to the severity of the condition being, at times, even higher than the first and second heart sounds. Examples of some of the most common murmurs are shown on **Figure 2.4** [5, 6, 9].

2.2.4 - Additional Sounds

Apart from the aforementioned, other sounds may be heard such as the early systolic ejection sound or the mitral opening snaps, which are pathological sounds. Adventitious sounds which are

sounds not originated within the heart such as respiratory sounds may also be present [5].

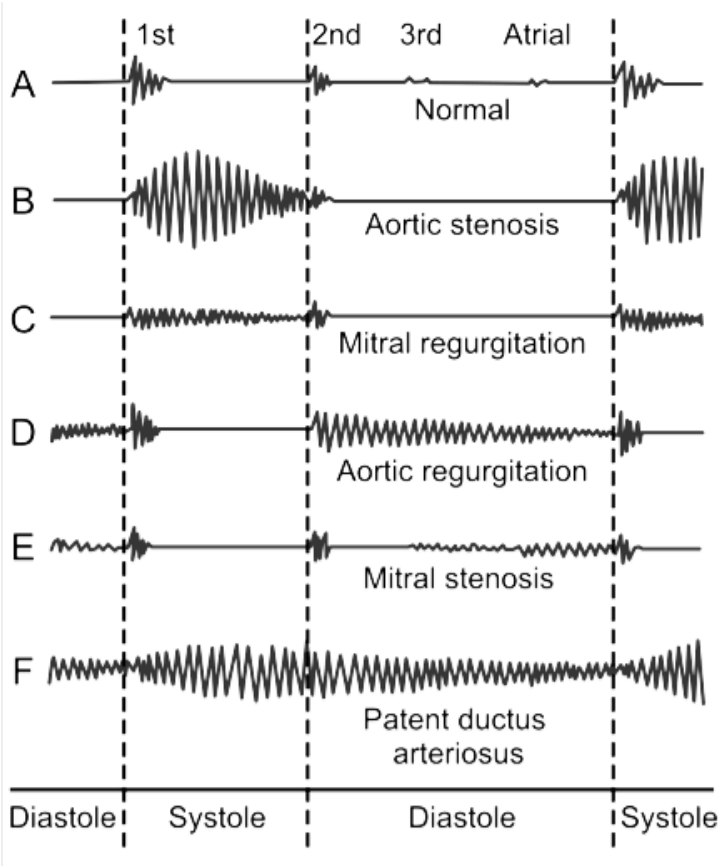

Figure 2.4 - Examples of systolic, diastolic and continuous murmurs and correspondent pathologies.

Chapter 3

State of the Art

In this chapter, the current methodologies of computer-aided auscultation are exposed. Section 3.1 presents the usual domains of analysis used in the digital auscultation field. Section 3.2 explains the different approaches to PCG acquisitions and their implication. Sections 3.3 and 3.4 give more information about specific applications concerning the problems of heart sound segmentation and murmur detection and results achieved.

3.1 - Overview

Computer-aided auscultation has been a subject of research for some time and many different methods have been applied to solve this problematic not only in terms of the proper algorithm and classification but also in the very way the data is acquired and treated. PCG signal processing can be crudely divided into two main research areas. One is focused in the detection of events such as S_1 and S_2 in order to perform a segmentation of the PCG into heart cycles. The other is the detection of murmurs and consequently of cardiac pathologies. However, because both objectives are interconnected and are accomplished from the same base signal, the PCG, both objectives share the same basic processing tools. These processing tools range, nevertheless, a large amount of techniques from the most simple to some utterly complex [1].

The simplest techniques are comprised in the time domain analysis, such as the low-order statistics. Envelopes are also fairly used to simply represent the change in amplitude throughout the PCG. Slightly more complicated envelopes such as energy envelopes, Shannon's Energy or Teager Filter's are also of very common use. **Figure 3.1** shows examples of the amplitude and energy envelope [7, 11-13].

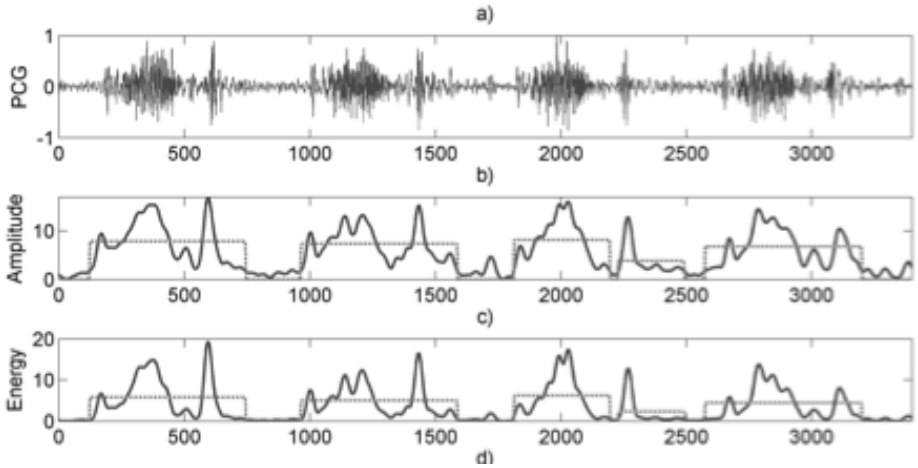

Figure 3.1 - PCG plot and respective instantaneous amplitude and energy envelopes. Adapted from [11].

The PCG is a sound wave composed of different sources with different frequency signatures each, the time-frequency domain analysis is a very important and close to undeniable tool for PCG analysis. As pointed out by the very name, these techniques allow observing the evolution of the frequency components of the signal over time. Techniques used in this domain range from the simplest frequency envelope that represents the main frequency at a given moment to more complex methods that allow the TFR of the signal and thus the observation of its different frequency components. Such techniques can be for example the Short-Time Fourier Transform, Wigner-Ville Distributions, the Gabor Transforms and the Wavelet Transform. The difference between these techniques in terms of application is the trade-off between time and frequency resolution. Because these techniques originate TFR matrices, tools such as singular value decomposition are often used to retrieve features from them. **Figure 3.2** shows an example of a TFR matrix obtained by Short-Time Fourier Transform [7, 11, 12, 14-16].

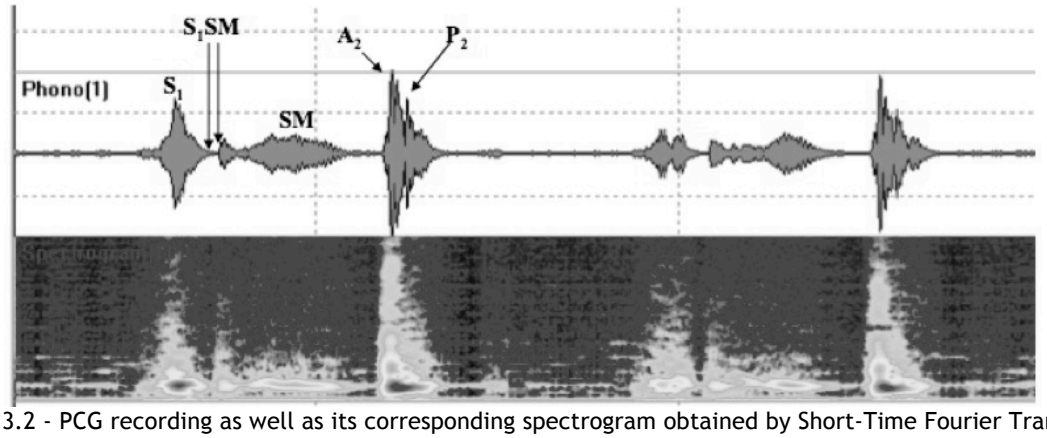

Figure 3.2 - PCG recording as well as its corresponding spectrogram obtained by Short-Time Fourier Transform within the frequency range 0- 1000Hz. Adapted from [16].

A somewhat more reduced field of PCG processing is the perceptual analysis. Similarly to many other PCG processing tools, the perceptual analysis was imported from speech recognition software. Unlike digital processing that treats frequency in a linear dimension, the human ear treats frequency in a logarithmic scale, the Mel scale. Perceptual analysis is then any technique

dependent of the conversion of the signal or signal spectrum to Mel scale. One of the most used tools within this domain is the MFCC extraction that allows studying the spectrum shape of a signal in the Mel scale [7, 17].

The nonlinear and chaotic nature has also been a subject of much analysis. It has been stated that in a PCG, murmurs are the most chaotic component, followed by the heart sounds and then noise, which has no structure at all. This makes it a refined tool especially in more complex algorithms. Higher-order statistics such as the bispectrum are seldomly used to measure the non-linearity or non-Gaussianity of a signal. The reconstruction of the state space is an important tool that gives access to a wide range of nonlinear and chaos based analysis tools. These tools measure characteristics of the state space trajectory which by themselves measure the nonlinearity or chaoticity of the signal. These may be the correlation dimensions, Lyapunov exponents, the simplicity or a simple characterization of the state space through Gaussian Mixture Models. Recurrence Quantification Analysis has also been used to characterize the recurrence of the trajectory in the state space. The VFD is used to measure the structure's variance in the fractal dimension, which is a measure of the signal's complexity [7, 12, 18-23].

A recent effort is also being made in the separation of the different components of the PCG. Blind source separation methodologies include, for example, principal component analysis and singular value decomposition. However, for such an approach to be possible, there is a need for a number of signals equal or larger to the number of sources. Nevertheless, the results are quite promising as shown on **Figure 3.3** [24, 25].

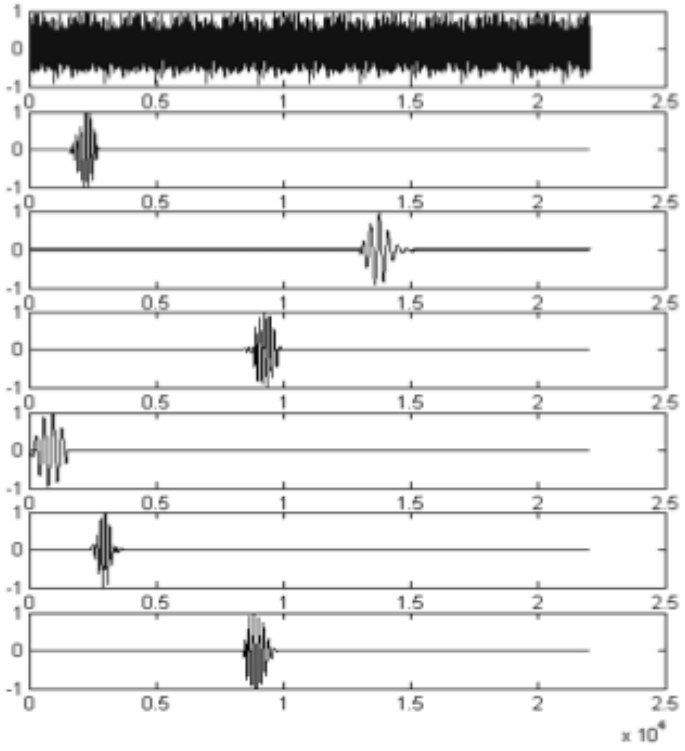

Figure 3.3 - The separate components of a normal PCG signal. From top to bottom: background noise, mitral component of S1, S3, aortic component of S2, S4, tricuspid component of S1 and pulmonary component of S2. Adapted from [25].

3.2 - PCG Acquisition

The method of PCG signal acquisition is just as important as the processing algorithm implemented and has implications for the whole algorithm and application of the results.

One of the approaches used is the acquisition of the PCG signal simultaneously with another biosignal. The most common is the ECG but other signals such as the respiratory rate or the echocardiogram have been reported. The trend of simultaneous acquisition of PCG and ECG is common due to the synchronization between both signals. As shown on **Figure 3.4**, S_1 corresponds in timing to the QRS complex in the ECG and S_2 follows the systolic pause in the normal cardiac cycle. This is advantageous for two reasons. Not only is the PCG heart sound detection a difficult task due to the complexity of the signal, but also the event detection of ECG signals is a consolidated field with already proven design methods. This approach has however a huge disadvantage. The simultaneous acquisition of PCG and ECG does not occur in normal clinical procedures. Furthermore, the preparation of the patient for this procedure is lengthy and the introduction of this method in clinical practice would be impossible; it is simply not practical [1, 26].

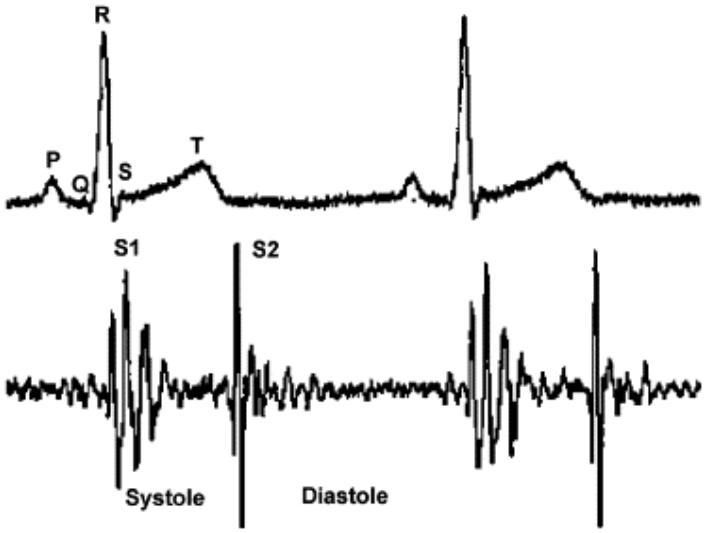

Figure 3.4 - Synchronized ECG (top) and PCG (bottom) showing the QRS complex-S1 and T wave-S2 time relations. Adapted from [26].

The alternative approach is the sole acquisition of the PCG with no other biosignal. This approach is much more consensual as any results obtained through this method could be extrapolated to an actual clinical practice where the physician would acquire the PCG with a digital stethoscope and obtain the results in real time. However, due to its complexity, the task of PCG heart sound detection is much more difficult and new algorithms are required. Promising results have been achieved with PCG databases acquired in controlled environments as will be shown in the following Sections (3.3 and 3.4). The challenge is then to apply algorithms to databases composed of PCG signals acquired in real clinical environment and in procedures similar to the normal auscultation methods used by physicians.

3.3 - Heart Sound Segmentation

The large variety of segmentation algorithms available in literature makes the complete description of the existing methods impractical and thus a select few were chosen and are described hereinafter.

M. El-Segaier developed a method based on ECG gating. This method uses the simultaneous acquisition of PCG and ECG signals. Using an envelope-based detection algorithm, the R-waves of the ECG were detected and the distance R-R computed. The T-waves were also computed. The Short-Time Fourier Transform was then used to obtain the spectrum of the PCG. Using the temporal relations between the PCG and ECG, intervals of search for S_1 and S_2 were defined and the maximum in the spectrum in each of those intervals was defined as S_1 or S_2. An additional tool was also developed to determine if the maximum obtain formed a well-defined peak in the time domain [16].

H. Liang et al. designed a segmentation algorithm dependent on the Shannon energy envelope. A threshold is set to select the peaks from the Shannon energy envelope. Time relationships between the obtained peaks are evaluated to reject extra peaks and to recover low-amplitude heart sounds that were not obtained due to the threshold used. The S_1s and S_2s are then separated by comparing the lengths of systoles and diastoles [13].

J. Martínez-Alajarín et R. Ruiz-Merino developed a segmentation method dependent mostly on time domain analysis. Using the amplitude, energy and frequency envelopes, the heart rate is found throught the ACF, a simple function specialized in finding the periodic elements of a signal. The amplitude envelope and a series of empirically defined rules is used to find the events of interest, the heart sounds [11].

H. Naseri et M. R. Homaeinezhad developed a method based both on time domain and time-frequency analysis. A specific function was designed to be sensitive to high amplitudes and the specific frequencies of the main heart sounds. This was done using the Fast Fourier Transform. The peaks of the envelope function obtained were then considered as candidates for heart sounds. Their shape and duration were also evaluated and, if validated according to these parameters, the events could be classified into S_1 or S_2. This process was done iteratively along the PCG signal [9].

D. Gill et al. developed an algorithm using self-organizing probabilistic maps. A homomorphic filter is used to obtain a smooth envelogram. This method handles split or serrated peaks using a scalable smoothness. The peaks of the envelogram are then used in a HMM using as observations the amplitude of the peak, the temporal distance between the adjacent peaks, their amplitudes and the second derivative of the peak. The Baum-Welch algorithm was used to provide the self-organization of the HMM with three states [27].

T. Oskiper et R. Watrous designed a time-delay neural network to accomplish the segmentation task. 40 wavelet scales encompassing the frequencies 10-299Hz were used to obtain a TFR matrix of the PCG in the frequencies of the main heart sounds. A simultaneously acquired ECG is used as a fiducial point and two different time-delay neural networks are trained. The first will determine

the locations of the S_1s and the second will determine the locations of the S_2s. These neural networks may then be applied without the aid of the ECG signal [28].

Table 3.1 shows the results obtained by each of the methods exposed above as well as some of their characteristics. Nevertheless, many other methods have been developed that were not mentioned here.

Table 3.1 – Performance evaluation of the reviewed segmentation algorithms (T- time domain; TF- time-frequency domain).

Authors	Analysis Used	Auxiliary Signals	Sensitivity	PPV
M. El-Segaier [16]	T;TF	ECG	100% (S1); 97% (S2)	-
H. Liang et al. [13]	T	-	94,11%	98,76%
J. Martínez-Alajarín et R. Ruiz-Merino [11]	T	-	-	-
H. Naseri et M. R. Homaeinezhad [9]	T;TF	-	99%	98,60%
D. Gill et al. [27]	T;TF	-	98,40%	96,70%
T. Oskiper et R. Watrous [28]	T;TF	ECG (training)	98,40%	97,80%

3.4 - Murmur Detection

In the heart murmur detection field, many articles have also been published, with different results. A common guideline is nevertheless present between these articles; a number of features are extracted from the interval of interest and they are introduced into a classifier for training and testing to find a performance value. There is however, a wide range of variations that can be made within this common protocol.

The first variation, and the starting point of the whole process, is the database. In this case, unlike the segmentation, this is not related the simultaneous acquisition of other signals. Even if other signals are used in the segmentation algorithm they are, to the author's knowledge, not used in murmur detection. The database issue is then related to the very way the acquisition is performed i.e. the conditions in which it is conducted; the subjects of the study, the environment and precautions taken and the amount of "true" information available about each signal. Some studies use inclusively computer simulated PCGs and this will of course have its implications on the true application of the final result. The size of the database is, as expected, a big issue too as a larger database will have a larger variation of signals and thus its results will be much more reliable [29].

The number and nature of the features is also an issue. The features used are included in the four analysis domains explained in Section 3.2 and are, of course, very different from article to article. Feature selection routine such as the SFFS are common and a number of articles has

conducted reviews of the performance of the several domains to find which are the optimal features for heart murmur detection [7, 12].

The classification process used will also be of major importance in the final results obtained in spite of the fact that this topic is usually not given much importance. The number and type of classes are also variable. Two main division methods exist. One aims to identify the exact pathology that is present and thus presents a class for each of the pathologies found in the database. The second aims at identifying whether or not there is a murmur present independent of the pathology that originated it. The first approach requires obviously a much more profound knowledge of the patient's clinical data.

Table 3.2 presents some of the algorithms reviewed in this Section, the features used and results obtained by the authors.

Table 3.2 - Performance evaluation of the reviewed murmur detection algorithms (T- time domain; TF- time-frequency domain; P- Perceptual analysis; NLC- nonlinear and chaos based analysis).

Authors	Analysis Used	Results
E. Delgado-Trejos et al. [7]	TF;P;NLC	TF: 95,28% accuracy P: 88,7% accuracy NLC: 97,17% accuracy all: 96,11% accuracy
C. Ahlstrom et al. [12]	T;TF	86% accuracy
D. Kumar et al. [18]	NLC	91,09% sensitivity and 95,25% specificity
D. Kumar et al. [22]	TF;NLC	89,1% sensitivity and 95,5% specificity

Chapter 4

Methodology

This chapter provides a description of the data used and the methods used to treat it. Section 4.1 contains information about the database used and Section 4.2 gives an overview of the entire algorithm designed while Sections 4.3 and 4.4 present the heart cycle segmentation algorithm and the feature extraction and classification routines respectively.

4.1 - Database

The database used was collected in the *Real Hospital Português* in *Recife*, *Brasil* using a *Littmann 3200* stethoscope and consists on a total of 72 signals. This stethoscope was used with the *DigiScope* prototype developed within the homonymous project to collect, transmit and record heart sounds without interfering with clinical routine. All the sound samples were collected in the clinical environment with lengths under one minute. The selected procedure was to sequentially auscultate all four auscultation spots in the order 4-3-2-1 as shown in **Figure 2.3**. Each physician was given the freedom to decide how much time to spend on each spot depending on whether there was something particular, as they would do if examining the patient in a normal situation. Physicians were also instructed to make no additional effort to find a quiet environment for signal acquisition. The patients auscultated were of ages comprised from six months to 17 years old. This database will, hereinafter, be referred to as the DigiScope database.

The information regarding the acquired signals is however limited to the presence of a murmur and its temporal location (systole/diastole). Because the heart sounds were unmarked, a different database was also used to test and validate the heart cycle segmentation: the *PASCAL CHSC 2011* database. Only a section of the database comprised of 111 signals of varying lengths between 1 and 30 seconds was used. This section is uniquely composed of signals without murmur and the signals have very little or no noise. The ages of the patients auscultated are unknown but it is known that both children and adults are present in this database [30].

The differences between the two databases used can be easily perceived in **Table 4.1**, which summarizes the characteristics of each database.

Table 4.1 – Database characteristics summary.

Database	Number of Signals	Duration of Signals	Labelled S_1/S_2	Labelled Murmurs
DigiScope	72	~1min	No	Yes
Pascal	111	1-30s	Yes	No

4.2 - Overview

To treat the PCG signals, perform the heart cycle segmentation and subsequent murmur detection, an algorithm was developed using MATLAB® and a number of its toolboxes available. A very brief schematic of the dataflow in the designed algorithm is shown on **Figure 4.1**. Starting from the minute-long PCG signals, the first part of the algorithm is used to separate the PCG into segments composed of either a systole or a diastole. Once all the possible segments of a PCG are obtained, each segment is subjected to the second part of the algorithm where a number of features are extracted from it to determine if it presents a murmur or not. The amount of segments said to present a murmur are then used to determine if the patient has a murmur or not.

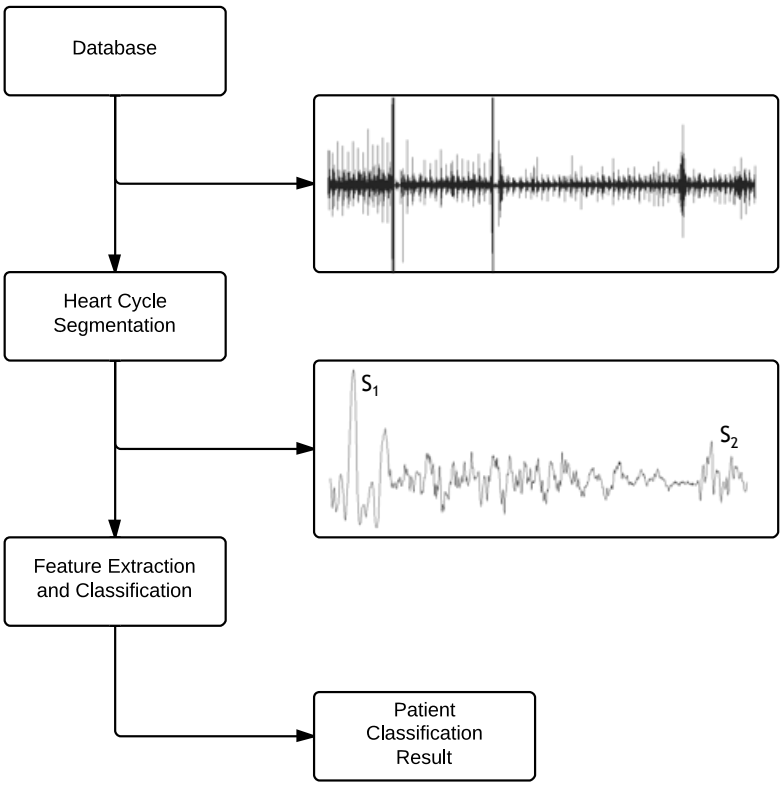

Figure 4.1 – Full schematic of the developed algorithm showing its several phases and the intermediate results namely the full PCG from the database and one of the systoles obtained from with the heart cycle segmentation.

4.3 - Heart Cycle Segmentation

The segmentation of the heart cycles and thus, the detection of the different heart sounds, namely S_1 and S_2, is of paramount importance. Not only will it allow the classification of each segment as presenting murmur or not, but will also allow for the temporal localization of the murmur, if detected. The ideal segmentation algorithm will perform the total detection of the heart sounds with no false positives and identify each sound as S_1 and S_2. As previously shown many different algorithms have been developed to solve this problematic. However, the signals of the DigiScope database have specific characteristics that the segmentation algorithm must adapt to. For one, the signals are extensive and have varying amounts of noise throughout the signal. Strategies have been developed to take advantage of the length of the signal and lessen the effect the noise may have on the segmentation results. Furthermore, the variability of the database implies that the algorithm must be applicable for a broad amount of patients from children to adults. This required additional strategies so that these different signals can be processed correctly. A novel algorithm was then developed and a simplified flowchart is shown on Figure 4.2. The whole algorithm can be divided into three main parts: pre-processing, systole length estimation and heart sound sequence analysis. Each of these sections will be described in detail hereinafter.

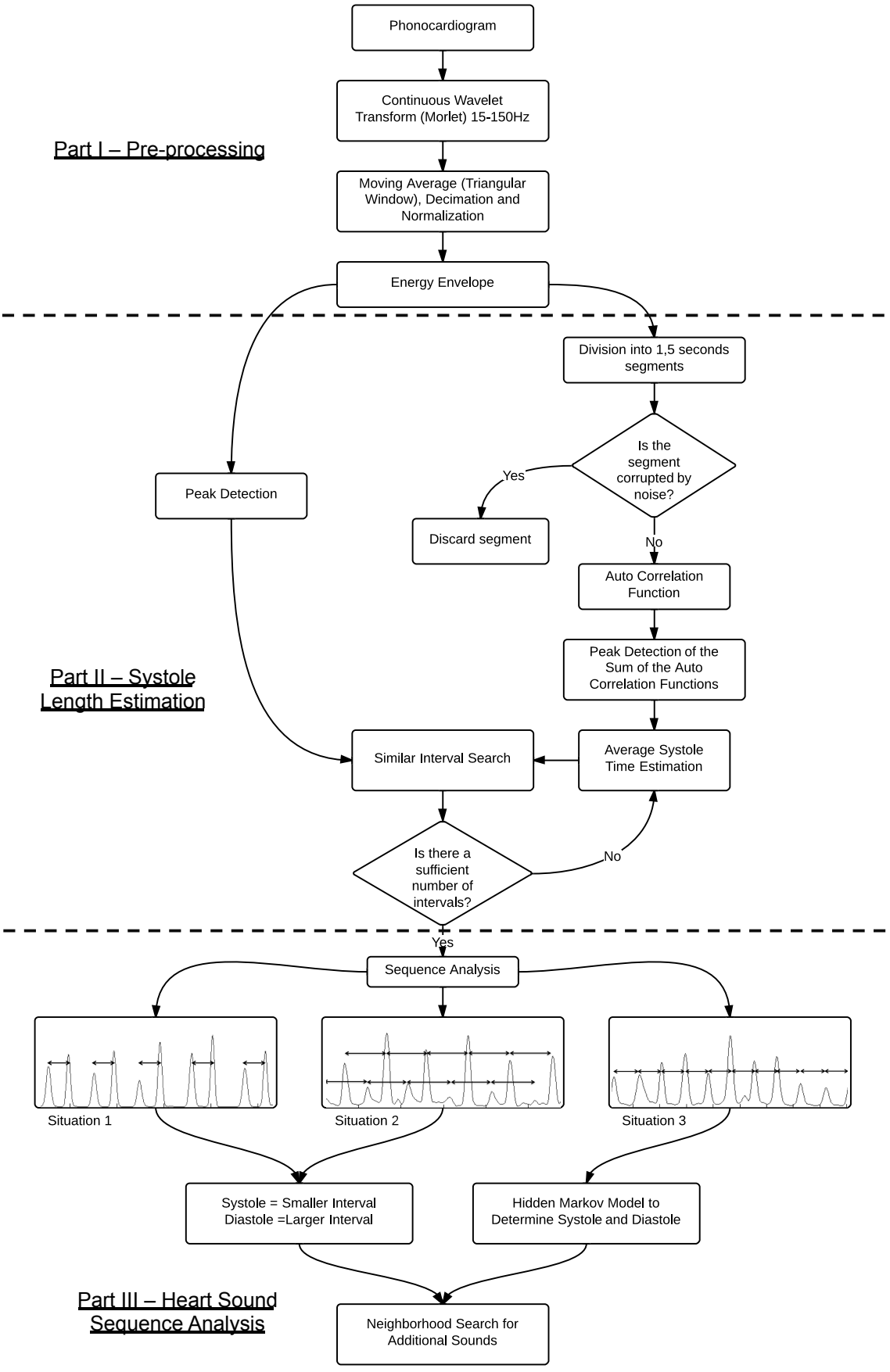

Figure 4.2 – Simplified Segmentation Algorithm Flowchart.

4.3.1 - Pre-processing

The first part of the algorithm consists in the pre-processing of the original PCG to amplify the desired traits. A Continuous Wavelet Transform (CWT) is used to filter out PCG components out of the frequency range of the heart sounds. The CWT is a signal decomposition method that uses a set of basis functions obtained by dilations, contractions and shifts of a unique small wave called the mother wavelet. The CWT computes the degree of correlation between the original signal and the wavelet. Because the wavelet is contracted and dilated it is possible to have both a large time and frequency resolution [14]. The CWT may be obtained using

$$CWT(m,k) = \frac{1}{\sqrt{|k|}} \sum_{n=1}^{N} s(n) w(\frac{n-m}{k}) \qquad \textbf{(4.1)}$$

where *w* denotes the mother wavelet function, *s* is the signal to be transformed, *m* and *k* are the dilation and translation parameters, respectively. *N* is the length of the signal *s* [14].

The mother wavelet used was the Morlet wavelet and the scaling function was selected to encompass frequencies from 15Hz to 150Hz [9]. The Morlet wavelet, shown on **Figure 4.3**, was chosen as it has been shown it is the most appropriate wavelet for the TFR of PCG signals obtained through CWT [14].

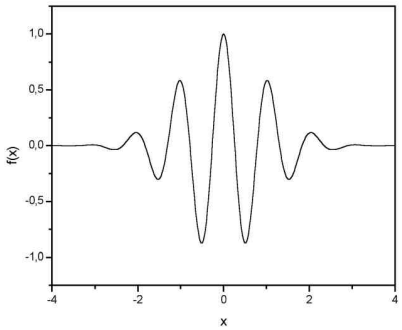

Figure 4.3 – The Morlet Wavelet.

The resulting signal is then submitted to smoothing operations to obtain a simple envelope. A moving average with a triangular window of 60ms long is used to smooth the signal without compromising the amplitude of the peaks and a downsampling in a factor of 12 allows not only a further rejection of short high frequency noise but also for a reduction in data size and therefore smaller computational effort. The signals are also normalized to avoid different ranges of values according to

$$\overline{x} = \frac{x - \mu_x}{\sqrt{\sigma_x}} \qquad \textbf{(4.2)}$$

where ***x*** is the signal to be normalized, μ_x is the mean of ***x*** and σ_x is its variance. An additional step, the subtraction of the minimum of ***x***, is also performed to obtain a minimum value of zero [11].

Finally the energy envelope of the signal is computed to further amplify the peaks of the signal. The energy envelope is computed by squaring each sample of the normalized signal.

4.3.2 - Systole Length Estimation

The second part of the algorithm focuses on the estimation of the average systole length throughout the signal. The main tool used to accomplish this task is the autocorrelation function (ACF). This function can be used, among other things, to find periodic events in a signal according to the lag between such events. It consists in performing the cross-correlation of a signal with itself, that is:

$$ACF_x(j) = \sum_{n=1}^{N} x_n x_{n-j} \qquad (4.3)$$

where x_n is the signal of interest, j the lag and N the length of the signal x_n, for a real signal x_n. [11].

However, as stated earlier most signals have a large amount of noise. Furthermore, there are high amplitude events when the stethoscope is changed from one auscultation spot to another as shown on **Figure 4.4**. These events would completely drown out the rest of the events if the whole signal were to be subjected to an ACF. To overcome this problem, rather than applying the ACF to the whole signal, the signal was divided into segments of 1,5 seconds to which the ACF can be applied. This allows for the segments with a large amount of noise to be discarded. This specific length, 1,5 seconds, was chosen empirically based on the fact that the shorter the interval, the more selective the process of rejecting corrupted segments became. However, a minimum length has to be maintained for at least one systole to be present in all segments. The length of 1,5 seconds was then chosen to accommodate both requirements allowing for a full heart cycle to be present at all times.

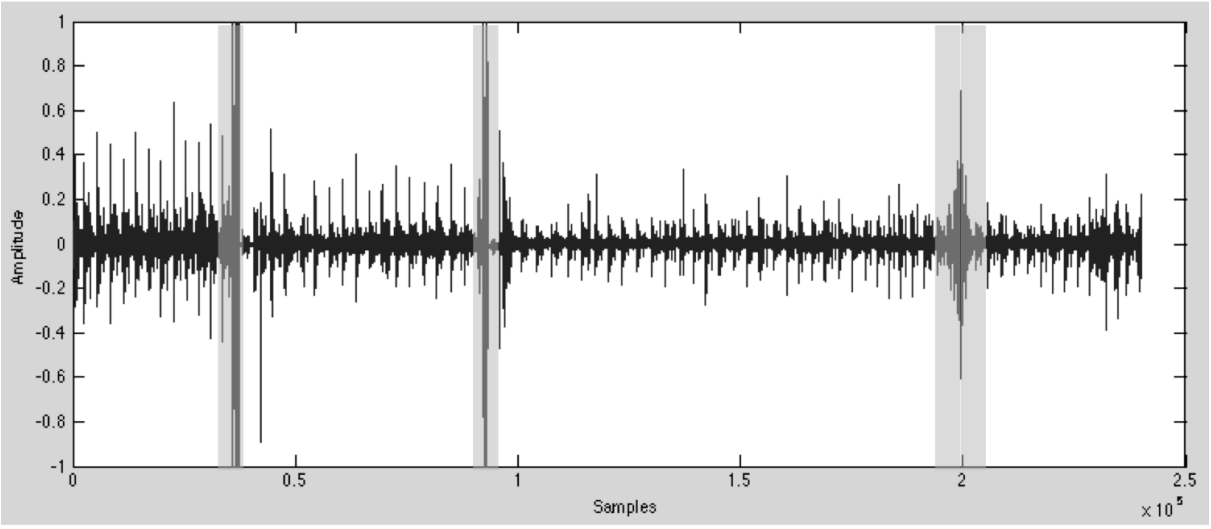

Figure 4.4 - Full PCG showing the three high amplitude events (shaded regions) correspondent to the switches between auscultations spots.

The criterion used to select or discard a segment is based in its mean amplitude. If the mean of a segment is larger than the mean of the signal plus the standard deviation of the mean amplitude of the segments, the segment is discarded. Only then would the ACF of each of the non-discarded segments was computed. **Figure 4.5** shows an example of a segment and its ACF. To merge the

information present in each ACF the sum of all the ACFs was computed and normalized to an amplitude range of zero to one.

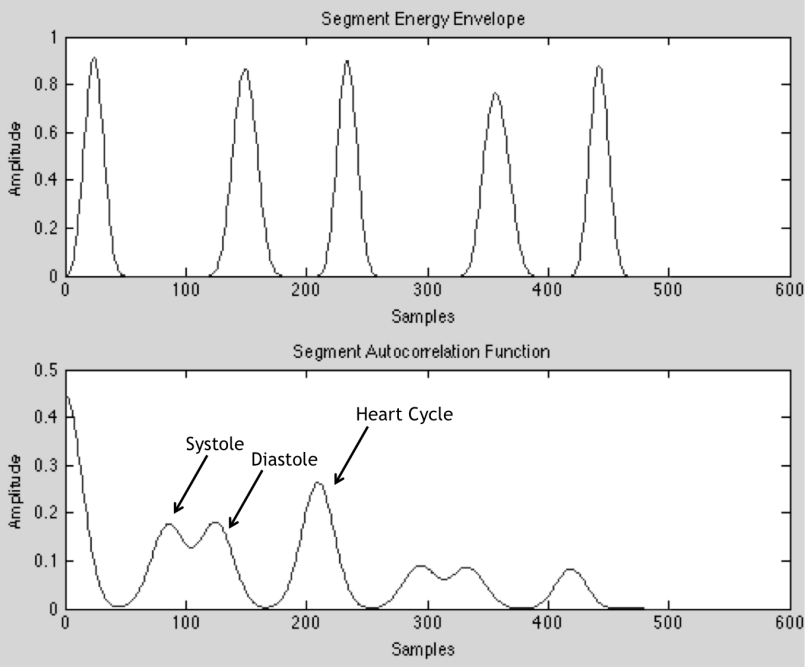

Figure 4.5 – Energy envelope segment and corresponding ACF with systole, diastole and heart cycle peaks pointed out.

The resulting ACF can then be used to estimate the periodic elements within the signal by finding its peaks. As shown on **Figure 4.5** the ACF peaks will represent each of the sections of the heart cycle (the first peak, at lag zero, is meaningless and is therefore rejected). The second peak is then selected as an estimation of the systole length. Because sometimes other elements such as a murmur can be represented in the ACF by a small peak, only peaks that were at least 1/8 larger than the surrounding data were used [31]. This value was chosen empirically and its exact value had little repercussions on the final result as the remaining algorithm was designed to not only check if the estimated value was a significant periodic element of the PCG but also to allow a multitude of situations to occur. Both these aspects will be explored in detail later on.

In parallel, the peaks of the PCG energy envelope are computed and a thorough search is conducted for intervals between peaks of similar length to the previous estimation. By doing this, one can see if the systole estimation is found throughout the signal. If it is not, the systole estimation is rejected and a new estimation is done based on the summed ACF by moving from the second to the third ACF peak and so on. The criterion used for the amount of intervals considered sufficient is 70% of the peaks of the energy envelope with amplitude larger than 0,1. This excludes smaller peaks with no significance.

4.3.3 - Heart Sound Sequence Analysis

The final part of the algorithm is necessary to determine which peaks are S_1 and which are S_2 and also to search for additional peaks. This procedure however is different according to the systole

estimation obtained and the signals' characteristics. Because of the intrinsic way the algorithm is designed through the ACF the estimation result may vary. Three different situations were identified and will be explained hereinafter.

Situation 1

The first situation is the original situation in which the estimation corresponds to the systole and is identified by the interval search. **Figure 4.6** shows an example of this situation. Because the systole estimation was correctly performed, the intervals are correctly identified and the remaining spaces are consequently identified as the diastoles. S_1 and S_2 are also easily identified as the first peak of an interval and the second peak, respectively. If the remaining spaces are found to be smaller than the intervals, the estimation identified the diastole rather than the systole and the order needs only to be switched.

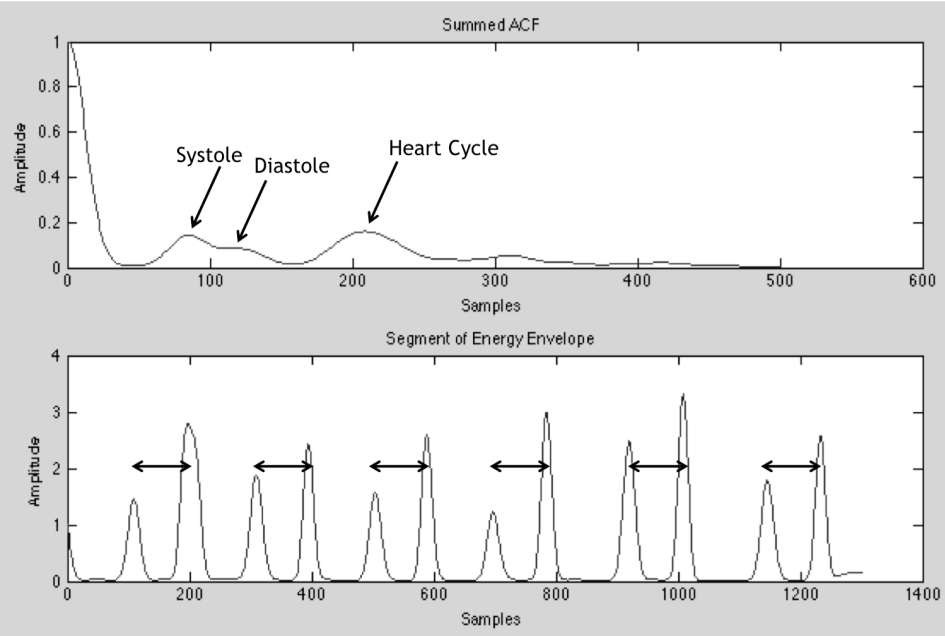

Figure 4.6 – Example of a summed ACF and segment with identified intervals for an estimation equal to the systole.

Situation 2

The second situation occurs when the estimation algorithm identifies the heart cycle length rather than the systole. This occurs mainly when one of the heart sounds is of very small amplitude compared to the other. The resulting summed ACF has a systole/diastole peak of such small amplitude that the algorithm discards it. **Figure 4.7** shows an example of this situation. Even though the systole/diastole peak is present it is not identified and the heart cycle peak is used. The interval search performed in this situation identifies the heart sounds but links them together in a different fashion than the one observable in **Figure 4.6**. The S_1s are linked with each other and the same happens with the S_2s. To identify which is which the lengths between peaks are analyzed. The largest corresponds to the diastole and the shortest to the systole.

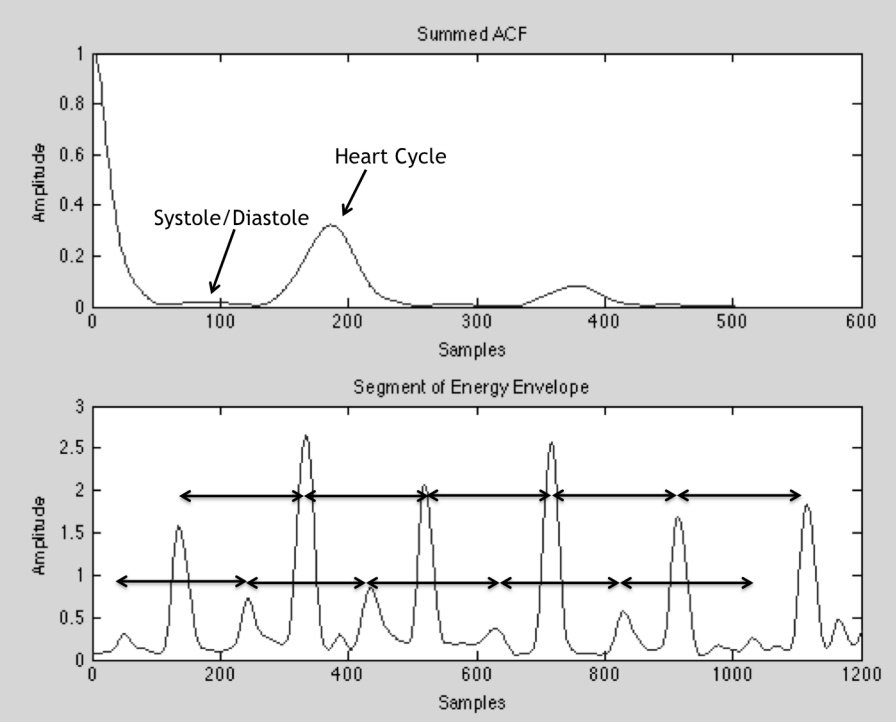

Figure 4.7 - Example of a summed ACF and segment with identified intervals for an estimation equal to the heart cycle.

Situation 3

The third and final situation is the most complicated. It occurs when the systole and diastole are of similar lengths. Even though the systole is almost always shorter than the diastole there are situations where this is not true. In children, and in adults with higher heart rates, it is normal for the diastole and systole to have similar lengths. **Figure 4.8** shows an example of this situation. The ACF peaks of systole and diastole are so similar that they are joined into a single systole/diastole peak. When this value is used in a search for similar length intervals, both the systole and diastole are identified as shown on **Figure 4.8**. To solve this problematic and distinguish S_1 and S_2 a Hidden Markov Model (HMM) was applied.

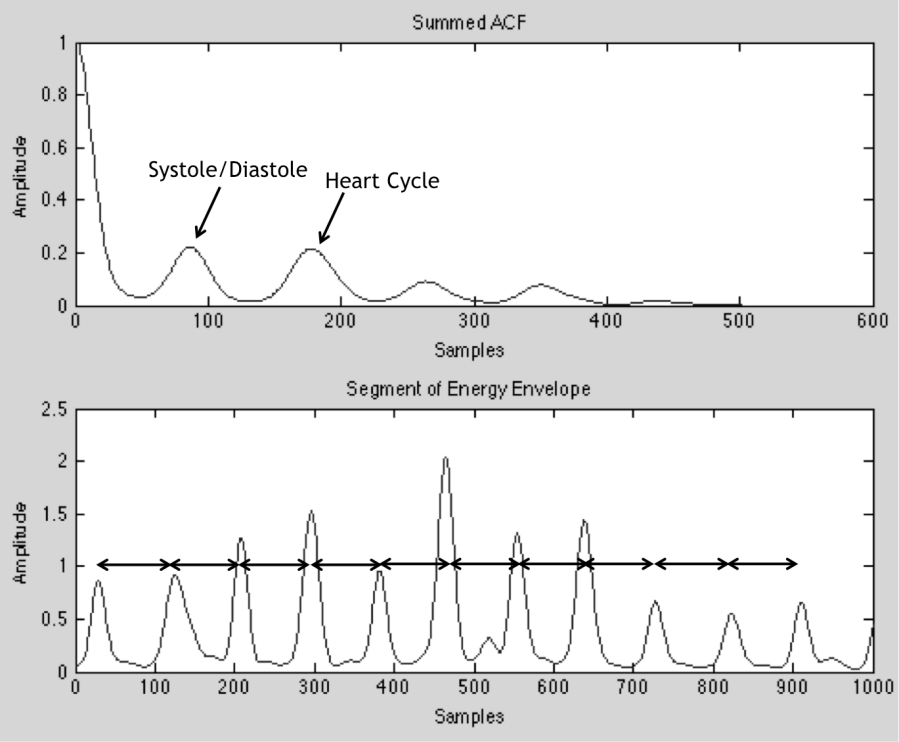

Figure 4.8 - Example of a summed ACF and segment with identified intervals for systoles and diastoles of equal length.

A HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with hidden states. It is useful, among other things, to find the most probable path of a system between states given a set of observations. This particular problem can be described by the schematic shown on **Figure 4.9**. Assuming that only the two main heart sounds are detected and that more complicated situations such as arrhythmias do not arise, the system is easily modeled. There are two possible states in this system, which correspond to the heart sounds, S_1 and S_2. These two states will change between them intermittently as a S_1 is always followed by a S_2 and vice-versa. This means that the transition probabilities p_{12} and p_{21} will be equal to 1. Another parameter needed is the prior probabilities. Given a sequence of heart sounds, the prior probability of S_1 is the probability of the first sound in that sequence being S_1. Because S_1 and S_2 exist in equal amounts both prior probabilities are 0,5. Finally, y_1 and y_2 are the emissions from each of the states. In this case the emissions will be continuous observations performed at each heart sound to retrieve characteristics from it. It is these characteristics that will give the information needed to determine the most probable path between the two states [32].

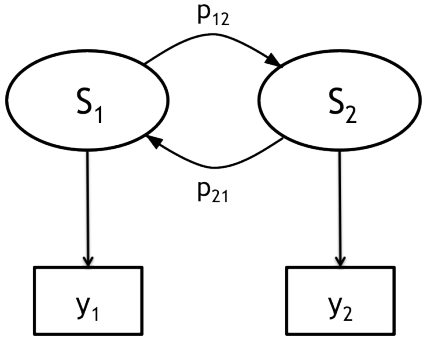

Figure 4.9 – Probabilistic parameters of the HMM designed.

Observations must then be retrieved for each of the states in known conditions such as the ones described in situations 1 and 2. The probability densities of the observations for each state must also be determined. These probability densities were parametrized by using a mixture of Gaussians with one component. No larger number of components was possible as the ratio of samples by features was too small. One can then determine the most probable path in the sequences of interest by retrieving the predefined characteristics for each heart sound. Due to the simplicity of the probabilistic parameters of the system, a simple brute force method is used. For any given sequence of heart sounds, only two possible paths exist, one starting with S_1 and the other with S_2. The most probable path is then given by calculating the maximum log-likelihood of each of the two possible paths by using

$$logP(Y \mid S, \Theta) = \sum_{n=1}^{N} logP(y_n \mid s_n, \Theta) \quad \textbf{(4.4)}$$

where *Y* is the set of observations of a sequence, *S* the proposed state sequence and *Θ* is the set of parameters of the system (prior, transition probabilities and emission probability densities). *N* is the number of heart sounds of the sequence, y_n the observation of heartsound *n* and s_n the proposed state for that observation. $P(y_n|s_n,\Theta)$ is obtained by the probability density function of observation *n* in the Gaussian mixture model designed for state s_n.[32].

Situation 3 – HMM Emissions

For the correct determination of the most probable path the emissions must be as selective as possible. A total of 46 features were chosen for this model as shown on **Table 4.2**. It has been shown that there are differences between S_1 and S_2 in what concerns their frequency range [33]. In consequence, 44 out of the 46 features considered are frequency based. The other two are the PCG amplitude at the heart sound and the standard deviation of the PCG on a 40ms neighbourhood around the heart sound. These are simple features, extracted easily, which represent the instantaneous amplitude of the PCG. The standard deviation is a measure of the sharpness of the peak.

Table 4.2 – Features extracted for the HMM observations.

Analysis Domain	Feature Name	Amount
Time Domain	Amplitude	1
	Standard Deviation	1
Time-Frequency Domain	CWT	8
	DWT	16
Perceptual	MFCC	20
	Total	46

The same neighbourhood was used to retrieve the Mel-Frequency Cepstrum Coefficients (MFCCs). The MFCCs are a perceptual signal analysis tool. It is a tool that perceives frequency in a logarithmic fashion, in a manner similar to the human ear, rather than in a linear fashion, the

common way of frequency analysis in digital processing. This logarithmic perception of frequency is achieved through the use of the Mel-scale. The MFCCs are the main components of a Mel-Frequency Cepstrum originated by a Discrete Cosine Transform (DCT). Starting from the frequency spectrum originated by Fourier Transform, the power of the spectrum is mapped to the Mel scale by using a Mel-scaled filter bank. The logarithm at each Mel-frequency is then obtained and by applying a DCT (particularly the DCT-II) the MFCCs are obtained, as the amplitudes of the resulting spectrum. The MFCCs are given by the following equations

$$X_{Mel}[m] = ln(\sum_{i=1}^{N-1} |X[i]|^2 H_{Mel}[i]) \quad \textbf{(4.5)}$$

$$c[p] = \sum_{m=0}^{M-1} X_{Mel}[m] cos\left(\frac{\pi p}{M}\left(m - \frac{1}{2}\right)\right), 0 \leq p < M \quad \textbf{(4.6)}$$

where *X[i]* is the spectrum of the signal of interest and $H_{Mel}[i]$ a Mel-scaled filter bank. $X_{Mel}[m]$ resultant from (4.5) is the Mel-scaled power spectrum to which the DCT is applied as shown on (4.6). A total of twenty-one MFCCs were obtained in this manner but the first MFCC was discarded as it represents the average of the spectrum rather than its spectral shape. **Figure 4.10** shows the Mel filter bank designed for this purpose [7, 17, 33, 34].

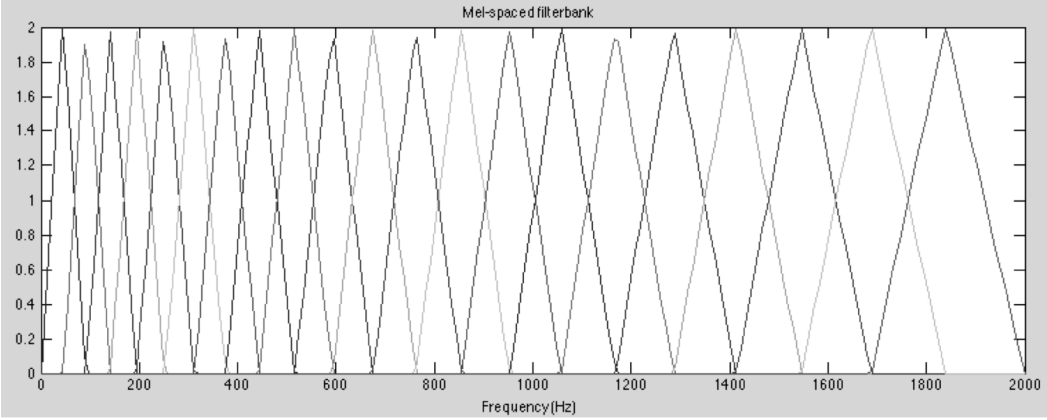

Figure 4.10 – Mel filter bank designed for the MFCC calculation.

Sixteen features were generated through the Discrete Wavelet Transform (DWT). Unlike the CWT which analyses the signals by using a set of functions related by scaling and translation operations, the DWT uses digital filtering techniques resulting in a much faster computation of the Wavelet Transform. As shown on **Figure 4.11** the signal is subjected to high and low pass filters H_0 and G_0 respectively. Each of these results is then decimated and generates the wavelet detail and approximation coefficients of level 1. The approximation coefficients can then be further separated into level 2 approximation and detail coefficients and so on [12, 35, 36].

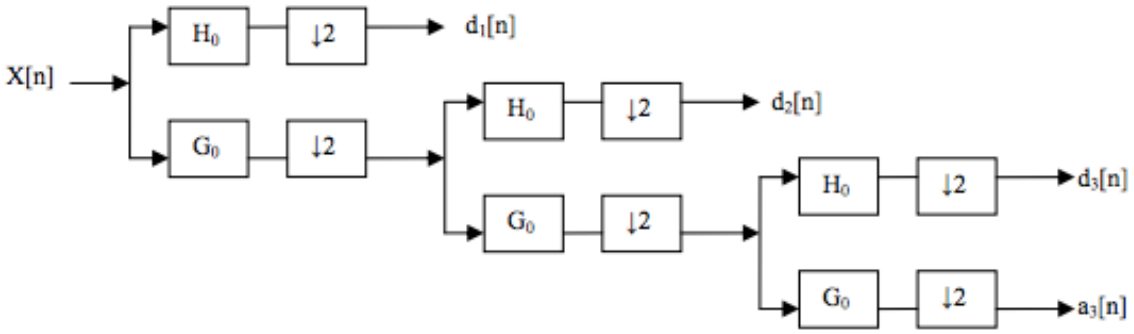

Figure 4.11 – Three-level wavelet decomposition tree. Adapted from [35].

The first five detail coefficients were extracted from the aforementioned neighbourhood using the Daubechies 1 mother wavelet and three features were extracted from each detail level. A. Castro et al. used a similar procedure to distinguish S_1 from S_2 with promising results. The two first features were the mean of the detail coefficients and the detail coefficient of the center of the segment, which corresponds to the position of the heart sound peak. The third feature was the Shannon Energy of the detail level coefficients given by

$$E_{ShannonE}(n) = \frac{-1}{N_{seg}} \sum_{n=1}^{N_{seg}} s^2(n) \cdot log(s^2(n)) \quad \textbf{(4.7)}$$

where *s(n)* is the input signal and N_{seg} its length. The Shannon Energy differs from other energy computations not only in the way that it performs a simultaneous decimation and energy calculation but also in the very way the energy is computed. The sixteenth feature was obtained by the quotient between the minimum Shannon energy and the maximum among the several detail levels [12, 13, 37].

The final eight features were obtained by using the CWT with the Morlet wavelet to encompass the frequencies from 25Hz to 275Hz in a neighbourhood of 12,5ms. The mean and median frequencies of the heart sound neighbourhood are simple measurements of the frequency distribution. They were then computed from the signal generated by the CWT so as to give maximum importance to the frequency range of S_1 and S_2. The mean frequency was computed according to

$$f_{mean} = \frac{\sum_{i=0}^{n} I_i \cdot f_i}{\sum_{i=0}^{n} I_i} \quad \textbf{(4.8)}$$

where I_i is the intensity of the frequency spectrum obtained by the CWT and f_i is the frequency for that particular intensity. To find the median frequency the whole spectrum intensity was computed. The median frequency is then the frequency with cumulative spectrum intensity equal to half the whole spectrum intensity. Five other features were obtained from the CWT spectrum by computing the spectrum intensity in intervals of 50Hz and a sixth feature was obtained through the ratio of the minimum CWT spectrum interval intensity and the maximum.

All the features were normalized according to (4.2) to have mean zero and standard deviation of one. Additionally, the extracted features were subjected to a SFFS algorithm to avoid irrelevant

features and improve the HMM performance. The criterion used was the 1-nearest neighbour error and the algorithm was programmed to return the optimal set of features for the best performance [38].

Neighbourhood Search for Additional Sounds

The final stage of the algorithm is the search for additional sounds starting in the borders of each sequence of sounds. In situations 1 and 2 an additional sub-stage is needed before this can be done. Each of the sequence of sounds is checked to see if there aren't any mistakes. Any eventual mistake is corrected or, otherwise, erased. For example, a sequence S_2-S_1-S_1-S_1-S_2 can be assumed to be S_2-S_1-S_2-S_1-S_2 whereas a sequence S_1-S_1-S_2-S_2-S_1 is a more dubious case and it is best to discard it. To separate these cases and make a decision, two different masks are applied to the sequence corresponding to the two possible sequences, one starting with S_1 and the other with S_2. The amount of differences between each of these sequences and the original is computed. If the amount of differences of one of the sequences is twice the other than the second is considered as the true sequence. Dubious cases are discarded but these rarely occur, nevertheless, due to the intricacies of the processes conducted in situations 1 and 2.

The neighbourhood search is conducted in equal circumstances for all three situations. By using the heart sound labeling already done to each sequence, and starting from either of the ends of the sequence, the average systole or diastole length (according to the case) is used to search for a peak in that region. Deviations from that length are allowed according to what phase of the heart cycle is being considered. A deviation of ±0,1*(Systole + Diastole) is allowed for the systole and a larger deviation of ±0,2*(Systole + Diastole) is allowed for the diastole as the diastole is more prone to deviate between heart cycles. This search increases the amount of heart sounds found and helps joining nearby sequences into a larger single one.

4.4 - Feature Extraction and Classification

Following the segmentation algorithm, a total of 250 features were extracted from each segment. These segments were either systoles or diastoles from one heart sound to the other. The extracted features are to be used as means for the classification of the signal as having murmur or not and will be presented thoroughly hereinafter.

4.4.1 - Feature Extraction

In the decision of which features to extract, an attempt was made to distribute the features among several analysis domains, namely time domain, time-frequency domain, perceptual and nonlinear and chaos based. **Table 4.3** shows the features extracted as well as its amount and analysis domain.

Table 4.3 – Features extracted for the murmur classification

Analysis Domain	Feature Name	Amount
Time Domain	Shannon Energy	7
Time-Frequency Domain	CWT	13
	DWT	55
	Singular Value Decomposition	48
Perceptual	MFCC	100
Nonlinear and Chaos Based	Bispectrum	16
	VFD	9
	Lyapunov Exponents	2
	Total	250

Time Domain Analysis

The time domain features extracted were obtained with the Shannon energy equation mentioned above in (4.7). The Shannon energy was computed for seven specific points of each PCG segment. The length of the segments considered for the Shannon energy computation was 40ms. A point was placed at each of the heart sounds (points 1 and 7) and two were placed 20ms into the signal, measured from the heart sounds (points 2 and 6). The final three points were obtained by dividing the segment into four equal parts and using the three border points needed to do so (points 3-5). An example of a PCG segment, its Shannon energy calculated in 40ms segments and points considered are shown on **Figure 4.12** [12].

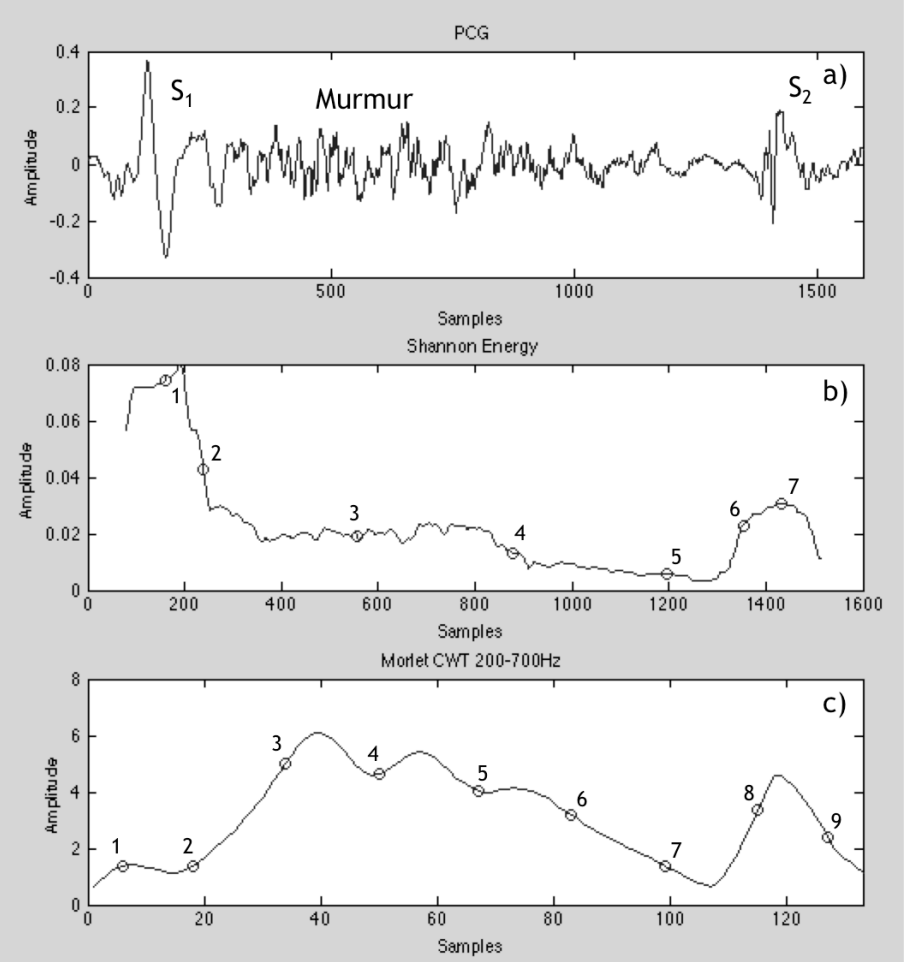

Figure 4.12 – a) PCG segment with heart sound components S_1, S_2 and systolic murmur marked; b) corresponding Shannon energy and the seven points considered as features; c) corresponding CWT obtained with scales encompassing 200Hz to 700Hz with points used for features extraction marked.

Time-Frequency Domain Analysis

Fifty-five features were extracted from each segment using the DWT. A similar routine to the one used to extract features for the HMM emissions was used. The first five level detail coefficients were obtained with the Daubechies 1 mother wavelet. The mean and Shannon energy of each segment was obtained. Finally, the ratio between the minimum Shannon energy and the maximum between levels returned the last feature. This was done for five different sections of the segment: the whole segment, the whole segment except for the S_1 and S_2 and each of the thirds of the segment after removing the S_1 and S_2.

Thirteen features were generated using the CWT. The Morlet mother wavelet was used to obtain the CWT of the PCG with scales encompassing the frequencies from 200Hz to 700Hz. This frequency range was chosen to remove any low or high frequency components and maintain the main murmur bandwidth. A moving average was performed to the CWT signal obtained with a triangular window of 60ms [14].

Eleven out of the thirteen features were defined as the sum of the CWT envelope within a predefined neighbourhood. Two features were obtained in a 40ms neighbourhood around the heart sounds (intervals 1-2 and 8-9). The remaining signal between these two neighbourhoods was divided

into six parts (intervals 2-3, 3-4, 4-5, 6-7 and 7-8). Each of them was used to generate a feature. Summing the adjacent intervals into larger ones two by two and calculating the sum of the CWT generated three other features (intervals 2-4, 4-6 and 6-8). The final two features were defined as the ratio between the minimum CWT sum of the peaks and the minimum CWT sum of the larger intervals and the ratio between the minimum CWT sum of the peaks and the maximum CWT sum of the larger intervals. **Figure 4.12** shows an example of a PCG and its corresponding CWT.

An alternative method for obtaining features from a TFR matrix such as the one obtained from the CWT is by using singular value decomposition. Singular value decomposition is an effective method for data reduction of matrices, with nonetheless a maintainability of its substructure. This technique allows the reduction of non-square non-symmetric matrices as the ones obtained from CWT by

$$M = U \Sigma V^T \tag{4.9}$$

where *M* is the TFR matrix, *U* and *V* its left and right eigenvectors respectively and *Σ* a diagonal matrix containing the eigenvalues of *M*. The left and right eigenvectors of a TFR matrix will be the eigentime and eigenfrequency vectors, respectively. These vectors can be interpreted as the main components of the TFR matrix [12, 39].

Singular value decomposition was then applied to each PCG segment. The eight largest eigenvalues were considered as features. The eigentimes and eigenfrequencies corresponding to the two largest eigenvalues were also used as features by applying a histogram (10 bins) to each of the eigenvectors probability density function. This function can be obtained by squaring the eigenvector elements due to its natural orthonomality. This returned ten features by eigenvector summing up to a total of 48 features generated by singular value decomposition from each segment [12, 39].

Perceptual Analysis

MFCCs were also used as features using the same filter bank used for the HMM emissions. A total of twenty coefficients per section returned a total of 100 MFCCs. The sections used for MFCC extraction were the same that were used for the DWT [12].

Nonlinear and Chaos Based Analysis

The bispectrum, a common higher order statistics, was used to extract features from the PCG regarding its nonlinear interactions. The bispectrum differs from the power spectrum normally obtained through the use of the Fourier transform due to the additional phase information it provides. Bispectral analysis detects phase relationships between different frequency components. This analysis measures the interdependency of the phase of such components. The higher the bispectrum, the higher the degree of interdependency is. The bispectrum can be mathematically defined by

$$B_x(f_1, f_2) = \sum \sum E\left\{x_t, x_{t+\tau_1}, x_{t+\tau_2}\right\} e^{-j2\pi(f_1\tau_1 + f_2\tau_2)} \tag{4.10}$$

where f_1 and f_2 are the frequencies being compared, E is the expectation operator and τ_1 and τ_2 are two lag variables. However, the bispectrum can be, and was, calculated in a much more efficient manner by using an FFT-based approach as follows

$$B_x(f_1, f_2) = X(f_1)X(f_2)X^*(f_1 + f_2) \quad \textbf{(4.11)}$$

where X is the Fourier transform of the signal and X^* its conjugate [12, 40].

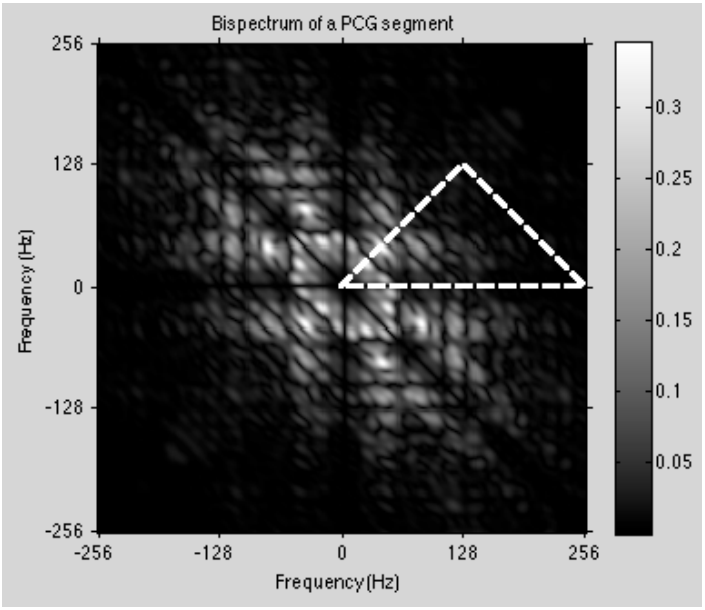

Figure 4.13 – Center of the bispectral analysis matrix obtained from the PCG signal shown on Figure 4.12 a). The inherent symmetry is easily observed and the dashed triangle delimits the first non-redundant region.

A bispectral analysis returns a matrix, which must be reduced into suitable features. **Figure 4.13** shows an example of such a matrix. It can be easily observed that there is an inherent symmetry to the bispectrum matrix. Thus, the first non-redundant region was divided into 16 smaller sections as shown on **Figure 4.14** and the mean amplitude of each section was used as a feature. The bispectral analysis was conducted from heart sound to heart sound [41].

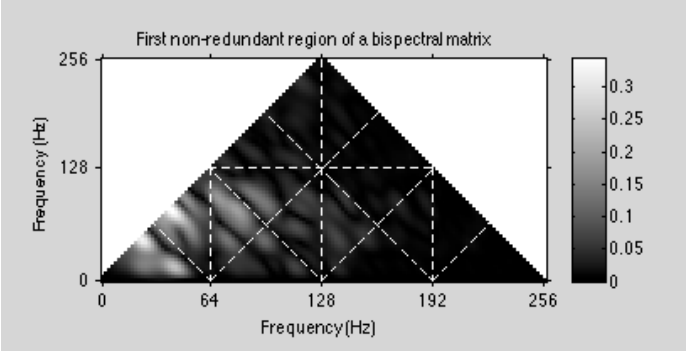

Figure 4.14 – First non-redundant region of the bispectral matrix center shown on Figure 4.13. The dashed lines separate the 16 different regions considered for feature extraction.

The VFD, an estimate of the fractal dimension was applied to several sections of the signal to obtain a total of eight features. The fractal dimension is a good measure of the signal's complexity

and can be computed through the Hurst exponent. Its use may then allow the detection of higher complexity elements such as murmurs. The Hurst exponent is mathematically defined as

$$H = \lim_{(n_2 - n_1) \to 0} \left(\frac{log\left[Var(s(n_2 - n_1))\right]}{2log(n_2 - n_1)} \right) \quad \textbf{(4.12)}$$

where *s* is the signal of interest. This approach is however computationally complex and thus an estimate of the Hurst exponent was used. This estimate is obtained by calculating the rescaled scale *R*/*s* as defined by Hurst where *R* is the range of the signal and *s* its standard deviation. Once this procedure is repeated *n* times, by averaging adjacent points in pairs one can create the log-log plot of the rescaled range versus the length of the time series considered. The slope of the plot created can be considered an approximation of the Hurst exponent. The VFD is then obtained by

$$VFD = D + 1 - H = 2 - H \quad \textbf{(4.13)}$$

where *H* is the Hurst exponent and *D* is the Euclidian dimension of the series considered, which is in this case equal to one [7, 12, 42, 43].

This method was used to obtain the VFD trajectory of each segment from heart sound to heart sound as shown on **Figure 4.15**. A neighbourhood of 40ms was used for the computation of the VFD at each point. To smooth the VFD trajectory, a moving average filter of 10ms wide was used. The VFD was obtained at each heart sound (points 1 and 7) and 20ms within the signal at each heart sound (points 2 and 6). The division of the segment into four parts returned three separating points, which were also used to retrieve the VFD features (points 3 to 5). Two additional features were obtained by the ratio between the minimum VFD of the heart sounds and the minimum VFD between the three center points and the ratio between the minimum VFD of the heart sounds and the maximum VFD between the three center points [12, 43].

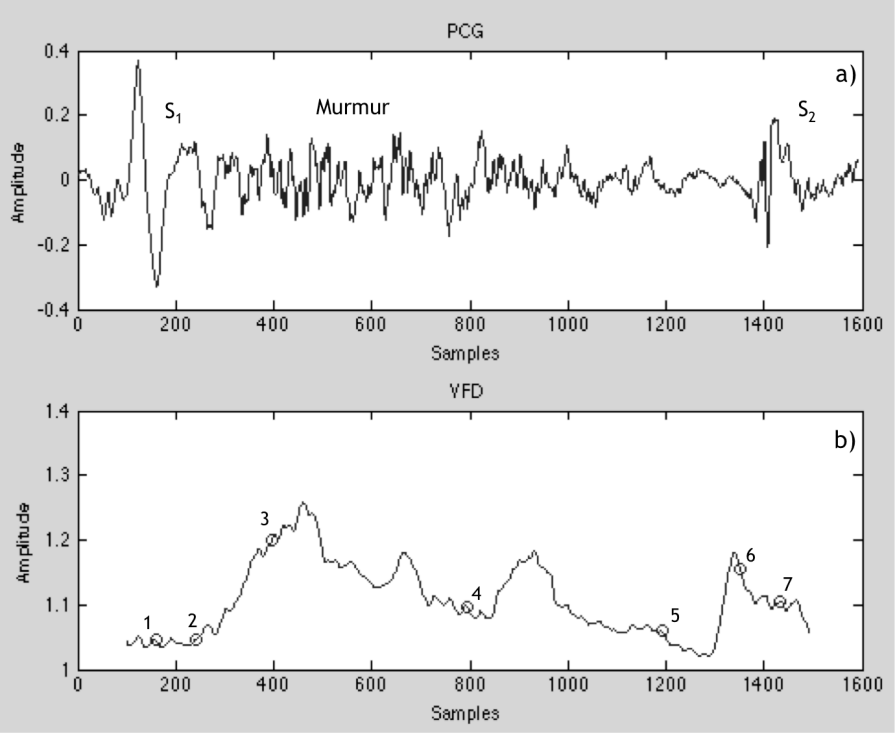

Figure 4.15 – a) PCG segment with heart sound components S_1, S_2 and systolic murmur marked; b) corresponding VFD trajectory showing points 1-7 used as features.

Finally, two features were extracted using the Lyapunov exponents. These are a measure of the chaoticity of a signal. It is known that any signal is a mere observation of the projection of a system, which may have any number of dimensions. This multivariate state space can be reconstructed, or at least its topological equivalent, according to Taken's delay embedding theorem

$$a(n) = \{s(n), s(n+\tau), s(n+2\tau), ..., s(n+(d-1)\tau)\} \quad \textbf{(4.14)}$$

where *s(n)* is the original signal, *a(n)* its reconstructed state space, τ its time delay and *d* the number of dimensions of *a(n)*. Both these embedding parameters, τ and *d*, must however be estimated previously before the state space can be reconstructed [7, 12, 19, 44].

The choice of the time delay is extremely important and will determine the distribution of the state space trajectories. If the time delay is too small there is almost no difference between the different elements of the delay vectors. This will cause a distribution along the bisectrix of the state space as shown on **Figure 4.16** a). If the time delay is however too large the coordinates may become uncorrelated which will complicate the state space trajectories (**Figure 4.16** c)) [19, 44].

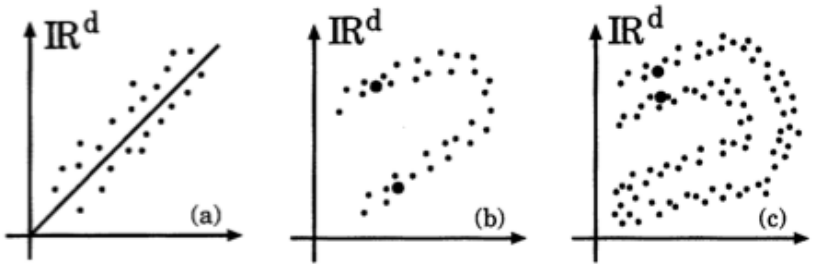

Figure 4.16 – Representation of three state space reconstructions with different time delays. (a) too small; (b) optimal; (c) too large. Adapted from [44].

These geometrical/visual observations of the state space are however limited and more specific techniques to determine the optimal time delay exist. The technique used is based on the average mutual information. The average mutual information is calculated for several time delays and can be described as the information about $s(n+\tau)$ that we possess just by knowing $s(n)$. The optimal time delay is then the first minimum of the average mutual information as it marks the time delay for which $s(n+\tau)$ adds maximal information to $s(n)$ and thus, the redundancy is minimal. **Figure 4.17** shows the distribution of the first minimum between the signals of the DigiScope database. The most common first minimum of the average mutual information is clearly 8 and was thus chosen as the optimal timelag [19, 44, 45].

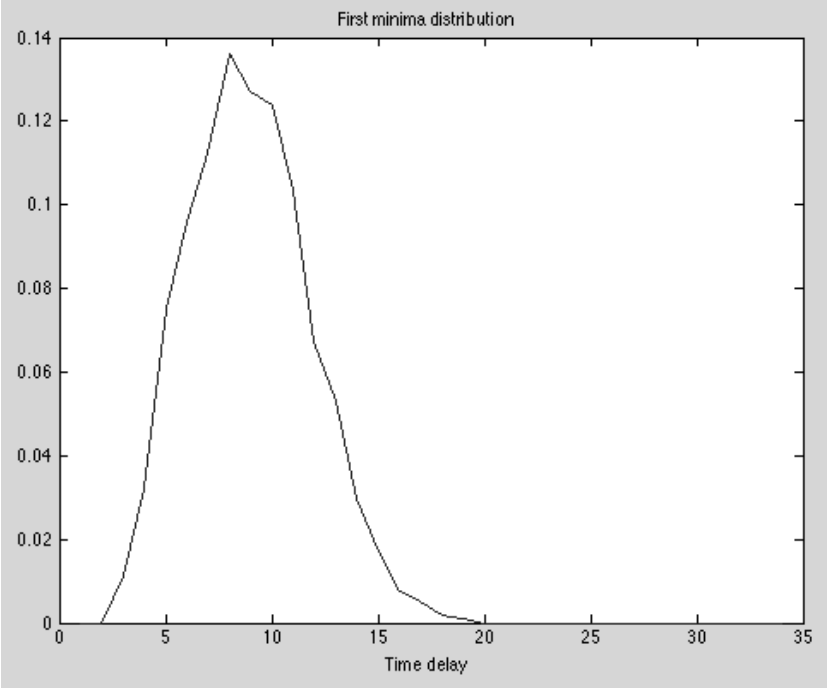

Figure 4.17 – Distribution of the first minimum of the average mutual information between the signals of the DigiScope database. The value 8 is the maximum of the curve and was thus chosen as the time delay.

Similarly to the time delay estimation, many different techniques exist to estimate the minimum embedding dimension. The method with wider acceptance is however Cao' method and thus was the one used. This method relies on the concept of false neighbours. These are neighbouring points in the embedding space that should, however, not be, as their future temporal evolution is too different. An optimal embedding dimension is one without false neighbours for the maximal unfolding of the state space reconstruction. Cao's method studies the evolution of the distance between neighbours throughout different embedding dimensions. To do so, the ratio $a(i,d)$ is defined as

$$a(i,d) = \frac{\left\| y_i(d+1) - y_{n(i,d)}(d+1) \right\|}{\left\| y_i(d) - y_{n(i,d)}(d) \right\|} \qquad \textbf{(4.15)}$$

where $y_i(d)$ is the i^{th} reconstructed state space vector of dimension d and $y_{n(i,d)}(d)$ is the nearest neighbour of $y_i(d)$. The operation $||.||$ is the maximum norm between the members of the state

space vector and its corresponding nearest neighbour. The amount of false neighbours of a dimension may then be estimated, to a certain extent, by the mean of all the ratios *a(i,d)*

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i,d) \qquad \textbf{(4.16)}$$

which depends only of the dimension considered and the time delay used. Finally, the minimum embedding dimension can be found by studying the evolution of *E(d)* from *d* to *d+1*. For this purpose a final ratio was defined as *E1(d)=E(d+1)/E(d)*. By plotting *E1(d)* one can see the evolution of the false neighbours ratio as the dimension increases. When *E1(d)* stops changing the minimum embedding dimension has been found. An additional parameter *E2(d)* is also defined in Cao's method to determine if the signal is deterministic or stochastic. This is however not a meaningful determination for this project as the phonocardiogram is known to be deterministic. **Figure 4.18** shows the average *E1* values obtained for the DigiScope database signals using a time delay of 8. Even though the *E1* values don't stop changing abruptly it is clear that in the higher dimensions it is stable. The value of 10 dimensions was consequently chosen for the reconstruction of the state space [46].

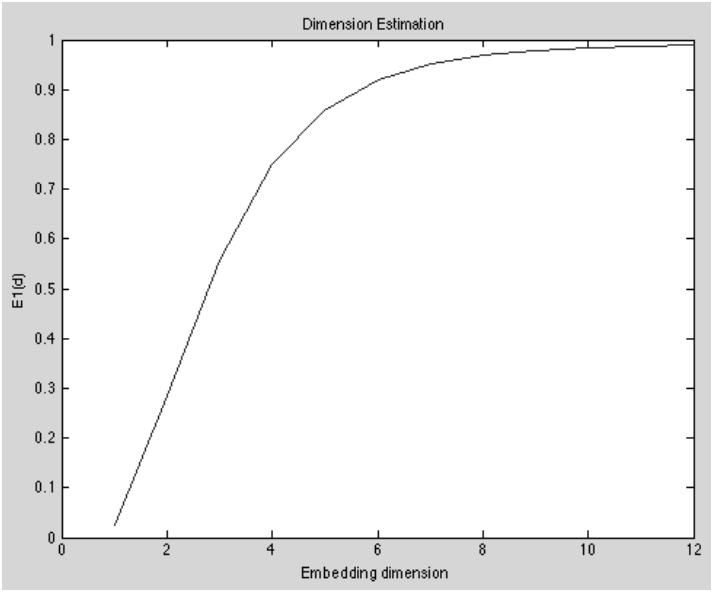

Figure 4.18 – Average *E1(d)* values obtained using Cao's method showing the stabilization of *E1(d)*.

Figure 4.19 shows the trajectory of the reconstructed state space in the first three dimensions of a segment of the PCG for the estimated embedding parameters.

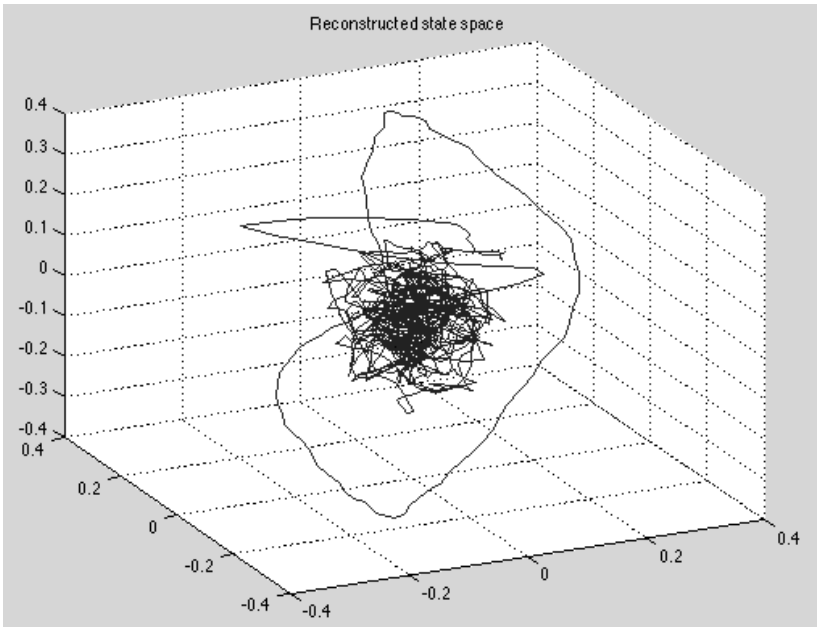

Figure 4.19 – Reconstructed state space of a PCG segment plot in the first three dimensions. The total ten dimensions would be needed to unfold the trajectory.

The Lyapunov exponents are a measure of the system's chaoticity as they effectively measure the rate of divergence/convergence of the trajectories of the state space. It is not alarming that trajectories within the state space diverge nevertheless; if this divergence is exponentially fast this is a sign of chaoticity. The more divergent a state space is, the more chaotic the system is. It is known that the murmurs are highly chaotic components within a PCG and thus would substantially increase the Lyapunov exponents if present. The number of Lyapunov exponents of a reconstructed state space is equal to the number of dimensions. However, the maximum exponent is often used as it represents the maximum divergence of the entire system. These exponents may be obtained by many different methods. Sato's method to obtain the maximum exponent was used as it is a direct method of simple execution. This algorithm studies the average exponential growth of the distance of neighbouring trajectories in a logarithmic scale by using the prediction error according to

$$p(k) = \frac{1}{Nt_s} \sum_{n=1}^{N} log_2 \frac{||y_{n+k} - y_{nn+k}||}{||y_n - y_{nn}||} \qquad \textbf{(4.17)}$$

where y_{nn} is the nearest neighbour of y_n and t_s is the time between two samples. The evolution of the prediction error along k encompasses three different phases. In Phase I the neighbouring orbit converges to the direction of the maximum Lyapunov exponent. Phase II is a linear slope whose value is equal to the maximum Lyapunov exponent. In Phase III the distance increases slower until it decreases again due to the folding of the state space trajectories. If Phase II has enough length the maximum Lyapunov exponent can be determined by the slope of the prediction error. This is shown on **Figure 4.20**. The maximum Lyapunov exponent was extracted for two segments: one including the heart sounds and the other excluding them [7, 18, 19, 44].

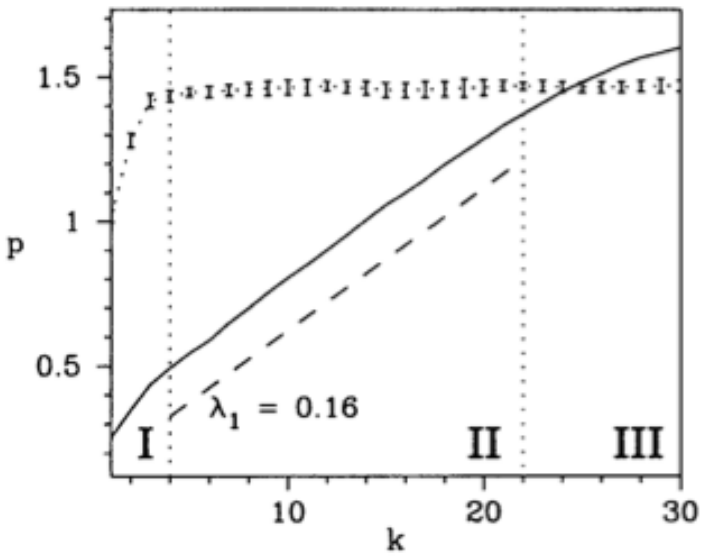

Figure 4.20 – Example of a prediction error plot. The slope of Phase II is equivalent to the maximum Lyapunov exponent λ_1. Adapted from [44].

4.4.2 - Feature Selection and Classification

In resemblance to what was done for the HMM emissions, the feature space was subjected to a SFFS algorithm to find the optimal feature set with a 1-nearest neighbour criterion. Finally, a k-means classifier was trained using the extracted features. In spite of the fact that the original goal was to develop two different classifiers, one for systolic segments and the other for diastolic segments, the latter was unable to be developed due to the fact that the database only contained one patient with a diastolic murmur. All the results presented regarding murmur detection are then solely regarding the systolic segments.

Chapter 5

Results

This chapter presents the results of the algorithms proposed earlier in Chapter 4. Similarly, the results are exposed in two sections, first the results of the segmentation algorithm in Section 5.1 and then the results of the murmur detection in Section 5.2.

5.1 - Heart Sound Segmentation

Due to the complexity of the algorithm itself and the fact that different situations are treated in different ways within the algorithm, two different analyses were made to assess the performance of the segmentation. The first analysis focuses on the precision of the segmentation itself by the amount of true heart sounds identified and also the temporal precision of each heart sound identified. The second analysis focuses on the precision of the HMM classification of sequences.

5.1.1 - Heart Sound Detection Performance

To assess the performance of the heart sound segmentation algorithm three different measurements were used. The first two, the sensitivity and the PPV, are widely used as was shown in Section 3.3. These two statistical measures will evaluate the ability of the algorithm to detect a heart sound. The sensitivity can be obtained by $\frac{TP}{TP+FN}$ where TP is the amount of true positives and FN the amount of false negatives. The PPV can be obtained by $\frac{TP}{TP+FP}$ where FP is the amount of false positives. The sensitivity will evaluate the proportion of heart sounds that are found whereas the PPV will evaluate the proportion of the events identified that are actually heart sounds [9].

An additional measurement was also used to determine the temporal precision of such detections. This measurement, δ, can be described as the average temporal deviation of the heart sound detections and can be obtained according to

$$\delta = \frac{1}{N_s}\sum_{k=1}^{N_s}\frac{\sum_{i=1}^{N_k}|t(RS_i) - t(ES_i)|}{N_k} \quad (5.1)$$

where $t(RS_i)$ is the real location of the i-th heart sound and $t(ES_i)$ the estimated location. N_k is the total amount of heart sounds in a signal and Ns the total amount of PCG signals considered during the analysis [30].

Because the DigiScope database is unlabelled in terms of heart sounds the Pascal database was used. However, the PCG signals of the Pascal database differ in length. Due to the fact that the designed algorithm relies on the length of the PCG to detect the heart rate, the signals with less than 15 heart sounds were excluded, as the algorithm would be more vulnerable to an incorrect estimation of the systole caused by noise. This originated a subset of the original Pascal database composed of 50 signals.

It is a common practice to calculate the sensibility and PPV of the S_1 and S_2 heart sounds in separate however this is impossible due to the use of the HMM. Either the heart sounds that were classified as Situation 3 were excluded which would diminish even further the dataset being used, or the results would become corrupted with the HMM classification error. The distinction between S_1 and S_2 detection was not performed and the sensitivity, PPV and δ were computed for both heart sounds in conjunction returning values of 89,2%, 98,6% and 9,8ms respectively.

In comparison to the results shown on **Table 3.1**, the performance of the developed algorithm is not superior. In fact it has a lower sensitivity than any other method. However, a few things must be taken into consideration. First of all, different databases were used in each study and consequently direct comparisons are always subjective. Secondly, the segmentation algorithm was designed for a different database than the one used. This would not be a big problem except for the size of the PCG signals that, as mentioned earlier, hinder the heart rate estimate process. Furthermore, the goal of the designed algorithm must be taken into account. Unlike the algorithms in **Table 3.1**, this algorithm's goal is not only to identify the heart sounds but also to discard noisy regions even if the algorithm is able to pinpoint the exact locations of the heart sounds in these regions. This design was implemented because the priority was to infer in the presence of murmur and not the perfect segmentation of the signal. To ensure that the classification process was conducted as smoothly as possible, the noisy regions are discarded here and thus will not interfere with the classification. This can be observed by the imbalance between the sensibility and the PPV. In a long signal such as the ones in the DigiScope database, it is not extremely important to identify every sound (lower sensitivity) but it is important to make sure the identified sounds are correct (high PPV).

In regard to the average temporal deviation value obtained of 9,8ms, no value in the literature was found to make a comparison. However, knowing that the average duration of the S_1 and S_2 is of approximately 100ms, these 9,8ms are a small deviation very much within the borders of the heart sound detected.

5.1.2 - HMM Classification Performance

As explained above, the HMM classification will determine the order of the heart sounds when this order is unknown as in Situation 3. To be able to perform this, one must first obtain the Gaussian mixture models of each of the heart sounds. Two different analyses were conducted depending on the database used.

First, the Pascal database was used. Because this is a labelled database the heart sound sequences considered could be used directly without depending on a segmentation algorithm. This ensures the absence of false heart sounds. The database was then randomly divided into test and train datasets in a 60-40 percentage. The heart sound sequences of the train dataset were used to obtain the Gaussian mixture models and these were used to classify the test dataset. This procedure was done repeatedly to ensure the randomness of the results.

The very same procedure was used with the original database. However, because the order of the heart sound sequences of Situation 3 was unknown these were excluded. Situations 1 and 2 were assumed to be 100% accurate, with no false positives. As mentioned above, the true PPV was of 98,6% however this difference was found negligible and thus, Situations 1 and 2 were used to test the HMM performance.

The performance of the HMM classification in both these databases was also tested with SFFS subsets. **Table 5.1** shows the features chosen by the SFFS algorithm and thus used for each of the SFFS subsets. The Pascal optimal subset is composed of 21 features whereas the DigiScope optimal subset is composed of 29 features. Nevertheless, it is clear that there is a consistency between the features chosen from the two databases for the optimal set. Most features are common to both subsets. The first four MFCCs, for example, are common to both as they represent the frequencies from 44-306Hz approximately. These frequencies encompass the core of the heart sound frequencies thus better representing their shape. The same is true for the CWTs that are all selected for both subsets. The DWT has a more scattered pattern but two things can be observed. First, the 4^{th} detail has the most features selected between all the detail levels, which was expected as it encompasses the frequencies 125-250Hz. This finding is corroborative with the publications of A. Castro et al.. Secondly, from the three methods of extracting information of the DWT, the Shannon Energy was selected for both the SFFS subsets in detail levels one to four which may mean it is the most efficient method for this purpose. The 5^{th} detail level (62,5-125Hz) had a single feature selected, which reveals a small importance in the classification procedure. However, the heart sounds' frequency spectrum is normally extended into these frequencies and the MFCC and CWT encompassing these frequencies were selected for both subsets. This leads into believing that the high decimation ratio achieved in the 5^{th} level, taking into account the reduced size of the segment considered, caused the information to be lost and thus rendered these features redundant [37].

Table 5.1 – Features chosen for the HMM emissions that were selected by the SFFS algorithms for each of the databases.

Feature	Pascal SFFS	DigiScope SFFS	Feature	Pascal SFFS	DigiScope SFFS
1. Amplitude		✔	24. Median Frequency	✔	✔
2. Standard Deviation	✔	✔	25. 1st Detail DWT Mean		
3. 1st MFCC	✔	✔	26. 1st Detail DWT Centre	✔	✔
4. 2nd MFCC	✔	✔	27. 1st Detail DWT Energy	✔	✔
5. 3rd MFCC	✔	✔	28. 2nd Detail DWT Mean		✔
6. 4th MFCC	✔	✔	29. 2nd Detail DWT Centre		
7. 5th MFCC		✔	30. 2nd Detail DWT Energy	✔	✔
8. 6th MFCC		✔	31. 3rd Detail DWT Mean		
9. 7th MFCC		✔	32. 3rd Detail DWT Centre	✔	✔
10. 8th MFCC		✔	33. 3rd Detail DWT Energy	✔	✔
11. 9th MFCC			34. 4th Detail DWT Mean	✔	✔
12. 10th MFCC	✔		35. 4th Detail DWT Centre		✔
13. 11th MFCC		✔	36. 4th Detail DWT Energy	✔	✔
14. 12th MFCC		✔	37. 5th Detail DWT Mean		✔
15. 13th MFCC			38. 5th Detail DWT Centre		
16. 14th MFCC			39. 5th Detail DWT Energy		
17. 15th MFCC	✔		40. DWT Energy Ratio		
18. 16th MFCC			41. CWT (25Hz-75Hz)	✔	✔
19. 17th MFCC			42. CWT (75Hz-125Hz)	✔	✔
20. 18th MFCC			43. CWT (125Hz-175Hz)	✔	✔
21. 19th MFCC			44. CWT (175Hz-225Hz)	✔	✔
22. 20th MFCC			45. CWT (225Hz-275Hz)	✔	✔
23. Mean Frequency	✔	✔	46. CWT Ratio		

Figure 5.1 shows the classification error in each of the situations considered. The error obtained for the DigiScope database was of 13,84% with the whole feature set and of 13,57% with the SFFS subset. For the Pascal database the whole feature set error was 16,72% and the SFFS subset error was of 11,88%. This level of error is clearly not ideal. An error in the HMM classification will imply that in that specific sequence every S_1 and S_2 will be switched. This will most likely cause the murmur classification to fail either because it gives a wrong classification result or if it does classify correctly, that very same classification will be temporally wrong, as a systolic murmur will appear as a diastolic murmur or vice versa. However, an HMM classification routine independent of the duration between peaks had not been developed, at least not to the author's knowledge. As a primary result under the difficult conditions imposed these are quite promising results.

The fact that the results are similar between the two databases allows inferring that the proposed segmentation algorithm is also working well for the DigiScope database otherwise the error for this database would be substantially larger.

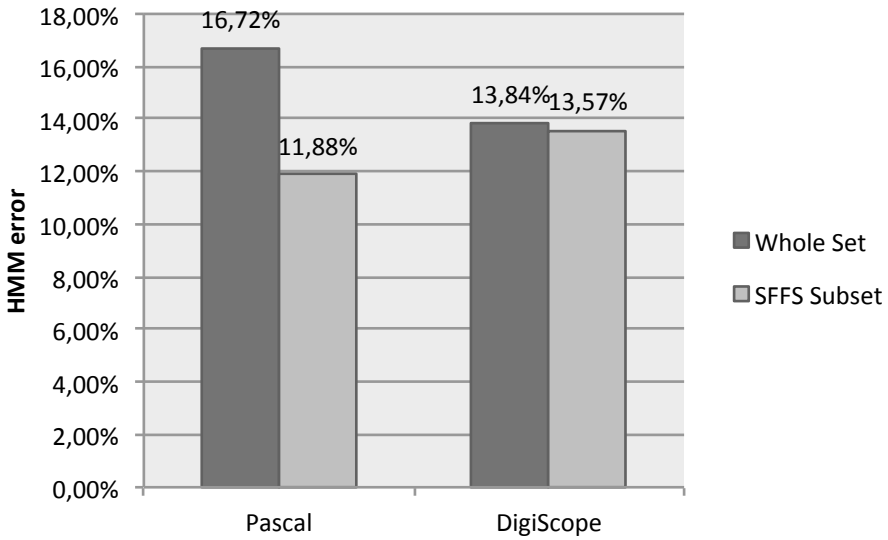

Figure 5.1 – HMM classification error in both databases for the whole feature set and correspondent SFFS subsets.

5.2 - Murmur Detection

To evaluate the murmur detection, different situations were tested to maximize the information obtained. Because the Pascal database is not labelled in terms of murmurs and their location, only the DigiScope database was used. From the DigiScope database the signals that were classified as Situation 3 were excluded (23 cases). If included, the classification error would be augmented by the HMM classification error and thus the true classification error would remain unknown.

Six different feature sets were used; the whole feature set, one for each analysis domain as described in **Table 4.3** and the SFFS subset.

5.2.1 - SFFS Algorithm

The SFFS subset was originated using Situations 1 and 2 and was composed of 167 features. Because the complete list of features selected was too extensive, their distribution between feature classes is displayed on **Figure 5.2** showing that all the feature classes except for the Singular Value Decomposition, the MFCC and the DWT had all their features maintained.

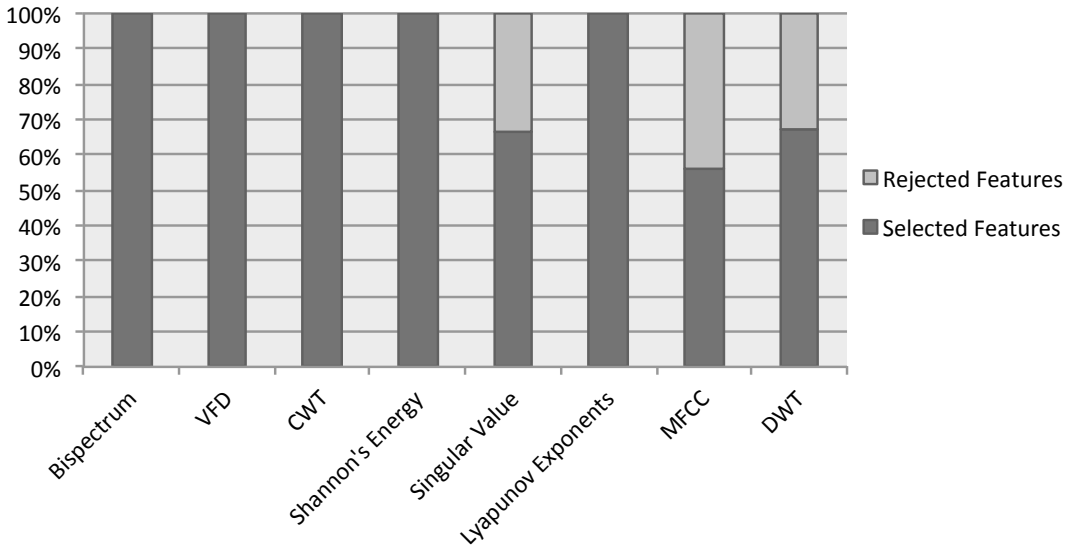

Figure 5.2 – Percentage of features selected/rejected by the SFFS algorithm discriminated by feature class.

These results must however be interpreted in this context as they do not necessarily mean that these three feature classes are the least significant. Upon closer inspection one sees that these three classes are the ones with the larger amount of features (**Table 4.3**). This could mean that among all the features extracted, some might not have a significant meaning for this problem. Furthermore, some of the features rejected could be correlated to others, which would not improve the performance of the classification and thus resulted in their rejection. The three feature classes that were partially rejected were inspected in detail.

The Singular Value Decomposition produces five different vectors: one with the largest eigenvalues, two eigentimes vectors and two eigenfrequencies vectors. The eigenvalues were all selected as they describe the larger components of the TFR matrices. The eigentimes and eigenfrequencies both had features rejected. However, more than half of the eigentimes features were rejected whereas only a quarter of the eigenfrequencies features suffered that same fate. As expected, in both cases there were more frequencies rejected from the second eigenvector than from the first. This is consistent with the findings of Ahlstrom et al. in which the importance of eigenfrequencies was shown through a similar SFFS algorithm [12].

In the MFCC feature selection, several observations can be made. In terms of preference between the five PCG segments considered, it is clear that there is a larger preference for the larger segment for which 16 MFCCs out of 20 were selected. The smaller segments (of a third of the entire PCG segment) portrayed an average of 8 MFCCs selected with no significant preference for any of the periods of the segment. The prevalence of the selection of each MFCC was also studied resulting in the chart shown on **Figure 5.3**. As expected, the higher MFCCs were proven of low significance, especially higher than the 13th MFCC – above 1067Hz approximately. The low significance of the higher MFCCs is easily explained by remembering the normal frequency range of the PCG - 20 to 1000Hz. Harder to explain is the gap of occurrences between the 5th and the 9th MFCCs. The frequency band of these MFCCs is in the frequency band of most murmurs and thus no possible explanation was found for this distribution of occurrences.

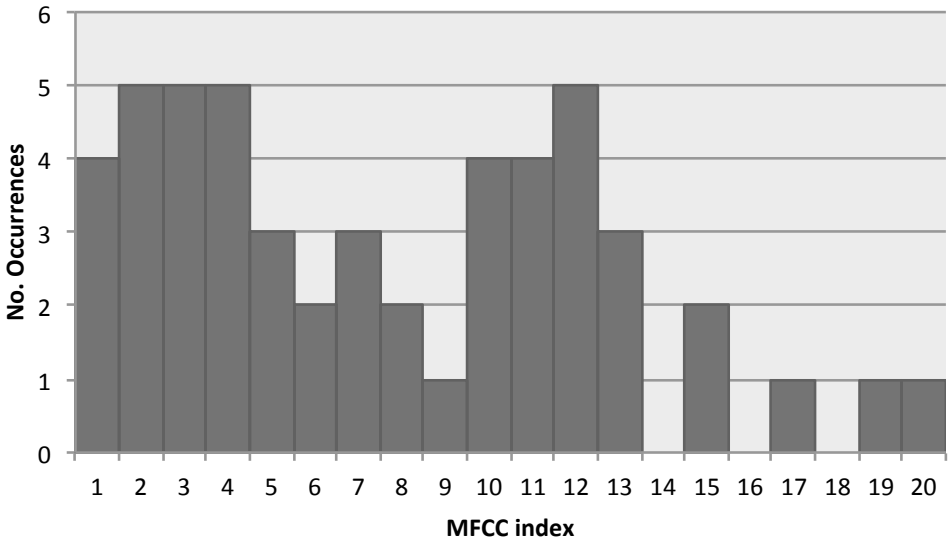

Figure 5.3 – Number of selections of each MFCC within the five segments possible.

Finally, the DWT feature selection yielded results similar to the feature selection performed to the HMM classification feature set. The min-max ratio was only selected once and only three out of ten features of the 5th detail level features were selected. However, in this case, there was no prevalence of the Shannon's Energy over the other methods of feature extraction, being the mean the most selected method. In terms of the segments with higher feature selection, there was no significant difference between the segments.

5.2.2 - Classification Error

The classification error was tested with six different feature sets: the whole feature set, the SFFS subset and other four subsets corresponding to the four analysis domains used. A dataset was constructed by extracting the features from each of the systoles identified. Each sample was labelled as presenting murmur or not according to whether the patient presented murmur or not.

Three different measurements were used to access the performance of the classification in each of the situations: the error, the sensitivity and the specificity. The first, the error of the classification, corresponds to the percentage of wrong classifications among the total. The sensitivity, already used above, corresponds to the percentage of segments/patients with murmur that were identified as such. Finally, the specificity, computed according to $\frac{TN}{TN+FP}$ that evaluates the percentage of segments/patients without murmur that were identified as such.

On a primary evaluation, the dataset was randomly divided into train and test sets in a 60-40 ratio. Because any classification method used would generate a classification per segment rather than a classification per patient, the percentage of segments classified as presenting murmur was used to infer the possibility of that patient presenting a murmur. A threshold would then have to be defined for this purpose. Under these conditions the results shown on **Figure 5.4** were generated. Only the results for the threshold that generates the least classification error are presented for each feature set.

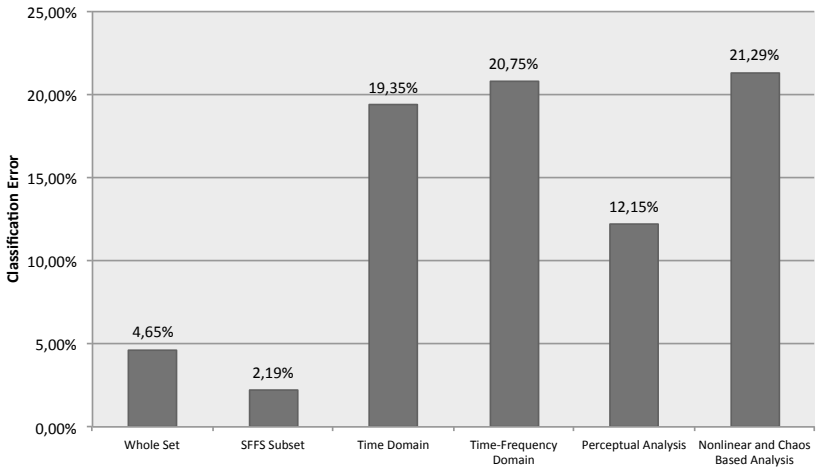

Figure 5.4 – Murmur classification error for the six different feature sets tested. Random test and train division.

These results clearly show the efficiency of the features with a 2,19% error for the SFFS subset for a sensitivity of 98,42% and a specificity of 97,21%. In comparison to the results shown on **Table 3.2**, obtained using this very same random division of data, this algorithm exceeds expectations in both the sensitivity and the specificity. However, it becomes clear that this approach to the problem is not correct even though it is commonly used and was used in each of the articles mentioned on **Table 3.2**. Because the division between train and test set was done to the whole dataset, samples from a certain patient that appear in the train set will most certainly also appear in the test set. This distribution eases the classification procedure due to the extreme similarity between two segments from the same auscultation spot from the same patient, which leads to the reduced error percentages observed. In a real situation however, the information about a new patient would not be present in the train set. Thus, a new division was made in which the patients were divided in a 60-40 ratio. The results for this classification are shown on **Figure 5.5** for the six different feature sets. Only the results for the threshold that generated the least classification error are shown for each feature set.

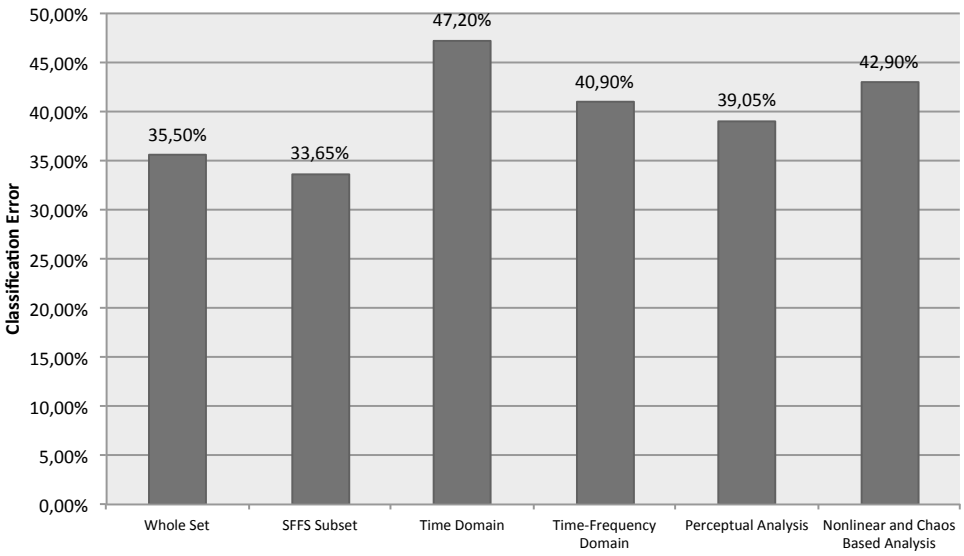

Figure 5.5 – Murmur classification error for the six different feature sets tested. Random test and train set division according to patients.

As expected, the error is much larger than in the previous division. However, this is the error that should be expected when classifying a new patient due to the patient variability that exists even in patients with the same pathology. Through analysis of the errors of the different domain subsets, one sees that the perceptual analysis is the one that gives the most reliable features and thus generates the smallest classification error. The remaining domain feature sets' results vary depending on the division used. For the division according to patients the worst feature set was the time domain feature set which had a classification error of 47,20% which is almost a random decision. This can be justified not only by the small number of features in this set but also by their high subjectivity to noise as these feature measure solely the energy of the PCG with total disregard for its structure or frequency. The SFFS subset presented little improvement in both divisions, circa 2%.

A ROC curve was plotted for the SFFS subset by varying the threshold defined earlier to classify a patient according to the percentage of segments identified as murmurs. This was performed by calculating the sensitivity and specificity for each case. The ROC curve generated is shown on **Figure 5.6**.

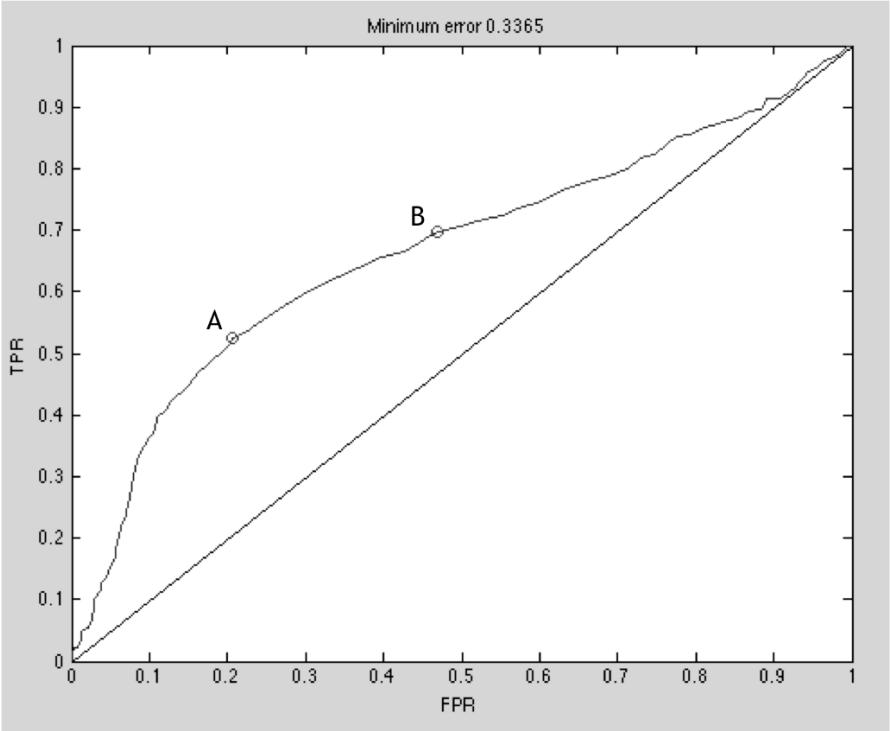

Figure 5.6 – ROC curve of the SFFS subset obtained by variation of the patient classification threshold. Point A marks the least classification error at 52,38% sensitivity and 79,40% specificity and Point B the ideal threshold for the problem at a sensitivity of 69,67% and 46,91% specificity.

Point A marks the threshold with a value of 0,49, which generates the least error. However, another threshold should be chosen given the context of the problem. It becomes clear that it is best if more false positives are attained rather than false negatives. The current value of sensitivity means that about 50% of the murmur cases would go undetected. If this value is diminished, more patients would be diagnosed as having a murmur but this is necessary evil. A false positive can be detected through sequential tests such as the echocardiogram; a false negative however may be sent home and thus remain with an undetected murmur. Point B was then chosen with a sensitivity of 69,67% and a specificity of 46,91% for a total error of 38,90%. This is achieved with a patient classification threshold of 0,37.

An identical division between train and test is implemented by E. Delgado–Trejos et al.. However, the loss of accuracy obtained wasn't nearly as large as the one obtained in this implementation. The accuracy reported by E. Delgado-Trejos et al. diminished for the time and time-frequency domains (94,35% to 87,92%) and for the perceptual domain (86,85% to 82,03%) and even rised for the fractal features (97,17% to 97,73%). Few reasons can be pointed out for the extreme differences between the two behaviours except for the fact that the databases are extremely different especially in the amount of patients. Situations 1 and 2 from the DigiScope database compose a total of 49 signals whereas 164 signals were used by E. Delgado-Trejos. The larger dimension of the database makes the similarity between the features of different patients much more common and may then justify, at least partly, the reduced loss of accuracy reported. Furthermore, the very PCG signals of the DigiScope database have a much larger variability between them and are most likely to be corrupted by noise. An example of this is the fact that no

age restrictions were imposed in the acquisition of the DigiScope database and many of the signals are from children. This impacts not only the variability of the database but also the amount of noise in these signals as children are more prone to move or speak during auscultation [7].

Chapter 6

Conclusions and Future Work

With the rising challenges of computer-aided auscultation, regarding the use of clinically acquired PCGs under unconstrained situations, the interest in the field is growing rapidly. In this dissertation, a new approach is presented both in terms of the segmentation of the heart cycle and of the murmur detection. Nearly every goal proposed was fulfilled, with room to improvement nevertheless. The main unachieved goal was the lack of the development of a diastolic murmur classifier. However, this was an impossible goal to the limitations of the database. If there were no limitations whatsoever the algorithm would be the same as the one applied to the systolic murmur classifier and was thus of easy application.

Promising results were obtained with the developed algorithms. The segmentation provided satisfactory results with the Pascal database. Nevertheless, it would be important to obtain labelled signals of the DigiScope database so that the performance of the segmentation algorithm could be evaluated for this database. This is especially true in what concerns the advantages of the larger length of the signal and noisy regions rejection. This would also be important to verify if the murmur detection is being implemented to the correct segments.

The HMM classification proved to be an effective method in spite of the error observed that is still too large. New and different features need to be developed and applied to ensure better results, as an error in this part of the algorithm will jeopardize the entire classification process. A duration-based feature could be applied to measure the difference between the sound intervals prior and after a heart sound. In spite of the fact that these variations are too small to be considered significant by themselves, hence the need to be classified with this HMM routine, their simultaneous use with time-frequency features may result profitable. Other methods of obtaining time-frequency features may be experimented with such as singular value decomposition. Other feature domains, such as nonlinear and chaos based can also be experimented with although there is no literature that supports the existence of a difference in the structure of S_1 versus S_2.

Regarding the murmur detection, the motivation of this project was to aid general practitioners in the detection of murmurs and thus, a comparison must be performed between the accuracy of

the developed algorithm and the accuracy of a general practitioner in the detection of murmurs. M. Lam et al. conducted a study that, among other things, evaluated the accuracy of physician trainees in the detection of murmurs. M. Lam et al. distinguished between types of murmurs obtaining accuracy values of 79,2% and 67% for pan-systolic and ejection systolic murmurs respectively and 28,3% and 23,6% for early diastolic and mid-diastolic murmurs respectively. These accuracy values are however, due to the way they were obtained, comparable to the algorithm's sensivity rather than to the algorithm's accuracy. By comparison to the obtained sensitivity value of 69,67%, one sees that it is within the range of the values obtained by M. Lam et al. for the physicians's normal performance. To make a difference in the clinical environment, a murmur detection algorithm should however have a superior accuracy to that of physical trainees. The potential to achieve correct murmur detection in different signals, which were not used in the train set, was proved and, thus, strategies must be conducted to improve the algorithm's performance. For one, there are many other ways to retrieve new and different information. The extraction of such features together with the application of a feature selection algorithm may, and most likely will, improve the performance as the ideal features are found. Examples of features that could be used are the simplicity of the signal, or the application of a Gaussian Mixture Model to the state space and retrieval of their characteristics. Both these elements have been used previously with satisfactory results. Another improvement to the performance would most likely occur if the database was enlarged. It was proved that if similar events exist in the train set, the error might be as small as 2,19% or even smaller. The larger the database, the more likely it is that similar events exist despite the inevitable differences from patient to patient.[47]

In regard to the diastolic murmurs, it becomes clear that it is much more complicated for physicians to detect them. This is probably caused by the same reason that led to the fact that only a classifier for the detection of systolic murmurs could be trained. Diastolic murmurs have much lower prevalence than systolic ones and thus physicians are probably less prepared to detect them. This makes it even more important to develop an algorithm to perform diastolic murmur detection and this can only be done by enlarging the database so that sufficient diastolic murmurs cases are present.

It would also be important to characterize murmurs after their detection. This would however be dependent of a more detailed labelling of the database in terms of, for example, the murmur's shape, intensity pitch and quality. These characteristics are commonly used by physicians and give additional information about the murmur's nature. The starting time of the murmur, which is an important characteristic to separate innocent from pathological murmurs, would also be important to retrieve. Features similar to the ones used in the murmur detection could be used for this very purpose in an additional step of the algorithm.

Additionally to these measures, a different approach should be experimented with. In this database the signals are composed of the four main auscultation spots, however, due to the fact that some signals aren't composed of the four spots, this could not be held into account. There are,

however, variations from spot to spot that may hinder the classification process. If the database was labelled in terms of auscultation spots, a method could be developed to separate the different auscultation spots. If this were possible, a classifier could be developed for each spot and thus the classification errors present from this fact would be nulled. Furthermore, when a murmur is present, it is more audible in some spots and less or even not audible in others. This fact surely confuses the training of the classifier as it is told there is a murmur in a certain segment when there is no proof of its existence. To solve this problem, a method would have to be used to label the segments according to whether the murmur is “audible” or not. However audible is a term too subjective. This has been done manually by cardiologists in other databases but the development of an automatic method for the distinction between cases would be important. This division into auscultation spots would have another advantage. As the different auscultation spots of a patient were identified as presenting a murmur or not, a specific pathology or possible pathologies could be pointed out as different murmurs are heard in different ways throughout the different auscultation spots.

References

1. Watrous, R.L. *Computer-Aided Auscultation of the Heart: From Anatomy and Physiology to Diagnostic Decision Support*. in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*. 2006.
2. Hanna, I.R. and M.E. Silverman, *A history of cardiac auscultation and some of its contributors*. Am J Cardiol, 2002. **90**(3): p. 259-67.
3. Constant, J., *Bedside Cardiology*. 5th Edition ed. 1999: Lippincott Williams & Wilkins.
4. Tavel, M.E., *Cardiac auscultation: a glorious past--and it does have a future!* Circulation, 2006. **113**(9): p. 1255-9.
5. Ghosh, A., *Mayo Clinic Internal Medicine Review*. 8th Edition ed. 2010: Mayo Clinic Scientific Press. 1130.
6. Bickley, L.S., *Bates' Guide to Physical Examination and History Taking*. 10th Edition ed. 2009, Philadelphia: Wolter Kluwer | Lippincott Williams & Wilkins.
7. Delgado-Trejos, E., et al., *Digital auscultation analysis for heart murmur detection*. Ann Biomed Eng, 2009. **37**(2): p. 337-53.
8. Debbal, S.M. and F.Bereksi-Reguig, *Frequency analysis of the heartbeat sounds*. Biomedical Soft Computing and Human Sciences, 2008. **Vol.13**(No.1): p. pp.85-90.
9. Naseri, H. and M.R. Homaeinezhad, *Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric*. Ann Biomed Eng, 2013. **41**(2): p. 279-92.
10. Arnott, P.J., G.W. Pfeiffer, and M.E. Tavel, *Spectral analysis of heart sounds: relationships between some physical characteristics and frequency spectra of first and second heart sounds in normals and hypertensives*. J Biomed Eng, 1984. **6**(2): p. 121-8.
11. Martínez-Alajarín, J. and R. Ruiz-Merino, *Efficient Method for Events Detection in Phonocardiographic Signals*. SPIE Proceedings, 2005. **5839**: p. 398-409.
12. Ahlstrom, C., et al., *Feature extraction for systolic heart murmur classification*. Ann Biomed Eng, 2006. **34**(11): p. 1666-77.
13. Liang, H., S. Lukkarinen, and I. Hartimo. *Heart sound segmentation algorithm based on heart sound envelogram*. in *Computers in Cardiology 1997*. 1997.
14. Ergen, B., Y. Tatar, and H.O. Gulcur, *Time-frequency analysis of phonocardiogram signals using wavelet transform: a comparative study*. Comput Methods Biomech Biomed Engin, 2012. **15**(4): p. 371-81.
15. Sharif, Z., et al. *Analysis and classification of heart sounds and murmurs based on the instantaneous energy and frequency estimations*. in *TENCON 2000. Proceedings*. 2000.
16. El-Segaier, M., *Digital Analysis of Cardiac Acoustic Signals in Children*, 2007, University of Lund: Lund.
17. Wang, P., Y. Kim, and C. Soh, *Feature extraction based on mel-scaled wavelet transform for heart sound analysis*. Conf Proc IEEE Eng Med Biol Soc, 2005. **7**: p. 7572-5.
18. Kumar, D., et al., *Heart murmur recognition and segmentation by complexity signatures*. Conf Proc IEEE Eng Med Biol Soc, 2008. **2008**: p. 2128-32.
19. Kantz, H. and T. Shreiber, *Nonlinear Time Series Analysis*, ed. C.U. Press. 2004, Cambridge: Cambridge University.

20. Osorio, I., et al., *Observations on the application of the correlation dimension and correlation integral to the prediction of seizures.* J Clin Neurophysiol, 2001. **18**(3): p. 269-74.
21. Raja, Y. and S. Gong. *Gaussian Mixture Models*. 1999 [18th January 2013]; Available from: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/RAJA/CV.html.
22. Kumar, D., et al. *Wavelet transform and simplicity based heart murmur segmentation*. in *Computers in Cardiology, 2006*. 2006.
23. Carvalho, P., et al., *A Framework for Acoustic Cardiac Signal Analysis*, in *BIOSTEC*2011.
24. Nigam, V. and R. Priemer. *Source adaptivity for cardiac sound separation*. in *Circuits and Systems, 2005. 48th Midwest Symposium on*. 2005.
25. Nigam, V. and R. Priemer, *Cardiac Sound Separation*. Computers in Cardiology, 2004. **Vol. 31**: p. p. 497-500.
26. Debbal, S.M. and F. Bereksi-Reguig, *Computerized heart sounds analysis*. Comput Biol Med, 2008. **38**(2): p. 263-80.
27. Gill, D., N. Gavrieli, and N. Intrator. *Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model*. in *Computers in Cardiology, 2005*. 2005.
28. Oskiper, T. and R. Watrous. *Detection of the first heart sound using a time-delay neural network*. in *Computers in Cardiology, 2002*. 2002.
29. Dokur, Z. and T. Ölmez, *Heart sound classification using wavelet transform and incremental self-organizing map*. Digital Signal Processing, 2008. **18**(6): p. 951-959.
30. Bentley, P., et al., *The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results*, 2011.
31. Yoder, N. *PeakFinder on MATLAB Central File Exchange*. 2009 [cited 2013 February 7].
32. Rabiner, L., *A tutorial on hidden Markov models and selected applications in speech recognition*. Proceedings of the IEEE, 1989. **77**(2): p. 257-286.
33. Kumar, D., et al. *A New Algorithm for Detection of S1 and S2 Heart Sounds*. in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. 2006.
34. Mehta, K. *mfcc on MATLAB Central File Exchange*. 2009 [cited 2013 February 7].
35. Sripathi, D., *Efficient Implementations of Discrete Wavelet Transforms Using FPGAs*, in *Electrical and Computer Engineering*2003, Florida State University.
36. Vikhe, P.S., S.T. Hamde, and N.S. Nehe. *Wavelet Transform Based Abnormality Analysis of Heart Sound*. in *Advances in Computing, Control, & Telecommunication Technologies, 2009. ACT '09. International Conference on*. 2009.
37. Castro, A., et al., *Heart Sound Segmentation of Pediatric Auscultations Using Wavelet Analysis*, in *Engineering in Medicine and Biology Conference, EMBC 2013*, I. Proceedings, Editor 2013: Osaka, Japan.
38. Pudil, P., et al. *Floating search methods for feature selection with nonmonotonic criterion functions*. in *Pattern Recognition, 1994. Vol. 2 - Conference B: Computer Vision & Image Processing., Proceedings of the 12th IAPR International. Conference on*. 1994.
39. Hassanpour, H., M. Mesbah, and B. Boashash, *Time-Frequency Feature Extraction of Newborn EEG Seizure Using SVD-Based Techniques*. EURASIP Journal on Advances in Signal Processing, 2004. **2004**(16): p. 2544-2554.
40. Wong, D., D.A. Clifton, and L. Tarassenko, *An Introduction for the Bispectrum for EEG Analysis*. IEEE PGBIOMED, 2009: p. 61-62.
41. Xiang, Y. and S.K. Tso, *Detection and classification of flaws in concrete structure using bispectra and neural networks*. NDT & E International, 2002. **35**(1): p. 19-27.
42. Gnitecki, J. and Z. Moussavi. *Variance fractal dimension trajectory as a tool for hear sound localization in lung sounds recordings*. in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*. 2003.
43. Kinsner, W. and W. Grieder. *Amplification of signal features using variance fractal dimension trajectory*. in *Cognitive Informatics, 2009. ICCI '09. 8th IEEE International Conference on*. 2009.
44. Suykens, J.A.K. and J.P.L. Vandewalle, *Nonlinear Modeling: Advanced Black-Box Techniques*. 1st ed. 1998: Springer. 273.
45. Fraser, A.M. and H.L. Swinney, *Independent coordinates for strange attractors from mutual information*. Physical Review A, 1986. **33**(2): p. 1134-1140.

46. Cao, L., *Determining Minimum Embedding Dimension from Scalar Time Series*, in *Modelling and Forecasting Financial Data*, A. Soofi and L. Cao, Editors. 2002, Springer US. p. 43-60.
47. Lam, M.Z., et al., *Factors influencing cardiac auscultation proficiency in physician trainees.* Singapore Med J, 2005. **46**(1): p. 11-14.