Heart Sound Segmentation for Digital Stethoscope Integration

Thesis submitted to the University of Porto
for the degree of Masters in Medical Informatics
In the Faculty of Medicine
and
Faculty of Sciences

Author: Fabio de Lima Hedayioglu
Advisor: Miguel Tavares Coimbra
Co-advisor: Sandra da Silva Mattos

November 20, 2009

Contents

1 Introduction **4**
- 1.1 Auscultation Through History 4
- 1.2 Computers in Medicine 5
  - 1.2.1 Electronic Medical Record Systems 5
  - 1.2.2 Computer-Aided Medical Decision Systems 6
- 1.3 Contributions 7
- 1.4 Thesis Outline 7

2 Preliminary Knowledge and Medical Background **8**
- 2.1 Physics of Sound 8
- 2.2 Physiology of the heart 8
- 2.3 Heart Sounds 10
  - 2.3.1 Abnormal Heart Sounds 12
  - 2.3.2 Auscultation Areas 13

3 Mathematical Methods **15**
- 3.1 Introduction 15
- 3.2 Wavelets 15
  - 3.2.1 Wavelet transform 16
- 3.3 The Discrete Wavelet Transform 18
  - 3.3.1 The multiresolution respresentation 19

4 A Survey on Algorithms for Digital Stethoscopes **24**
- 4.1 Digital Stethoscopes 24
- 4.2 Heart Sound Analysis and Feature Extraction 26
  - 4.2.1 Heart Sound Segmentation 27
  - 4.2.2 Aortic Pulmonary Signal Decomposition 27
- 4.3 Automatic Pathology Classification 28
- 4.4 Conclusions 31

5 Data Collection System 33
5.1 Introduction . . . 33
5.2 Developed Prototype . . . 35
5.2.1 System's Overall Structure . . . 35
5.2.2 The Client Workbook . . . 35
5.2.2.1 The client tools . . . 36
5.2.3 The Server . . . 37
5.3 Results and Discussion . . . 39

6 Heart Sound Segmentation 41
6.1 Introduction . . . 41
6.2 Fundamental stages of a HSS algorithm . . . 44
6.2.1 Wavelet decomposition . . . 44
6.2.2 Envelogram . . . 44
6.2.3 Peak Picking . . . 46
6.2.3.1 Picking the Peaks . . . 46
6.2.3.2 Rejecting the Peaks . . . 47
6.2.3.3 Recovering the Peaks . . . 48
6.2.4 Cardiac Cycle Classification . . . 49
6.2.5 Materials . . . 51
6.2.6 Results and Discussion . . . 51
6.3 Improved Cardiac Cycle Classification . . . 52
6.3.1 Results . . . 54
6.3.2 Discussion . . . 56

7 Conclusions and Future Work 57
7.1 Conclusions . . . 57
7.2 Future Work . . . 58

Acknowledgments

I would like to acknowledge the Instituto de Telecomunicações for providing all the hardware used in this research, as well as the Real Hospital Português de Beneficiência em Pernambuco, specifically, the UCMF team for the effort in performing the data collection and the medical support necessary for this thesis.

Also, I would like to acknowledge the Mestrado em Informática Médica, not only for funding my publishing expenses, but also for their professional and effective way they guided me through the process of coming from Brazil to Portugal.

My special thanks to my advisor Miguel Tavares Coimbra, for all his support and guidance given along this thesis, in an academic and personal level.

My acknowledgments to my parents, who gave me all the necessary conditions to allow my intellectual, emotional and moral development.

I would like also, to express my gratitude to God for giving me the opportunity of having meeting people that helped me so much in my path of becoming a better human being (an endless task, I presume). Yet, I would like to thank Him for giving me wisdom to learn from the adversities and from the bad examples (and to give a chance to the ones who make those bad examples an opportunity to somehow help them).

Last, but not least, I would like express my gratitude to the Programme Alban, the European Union Programme of High Level Scholarships for Latin America, who supported my word under the scholarship number E07M402298BR.

Chapter 1

Introduction

1.1 Auscultation Through History

Since the early times, mankind knows that the working human body has its own sounds. Between 460 B.C. and 370 B.C., Hippocrates mentions cardiac sounds[1]. But it was only on the 17th Century that the clinician René Théophile Hyacinthe Laennec created a device that, later, would be one of the most remarkable symbols of the practitioner: the stethoscope.

At his time, to hear cardiac sounds, the clinician had to make use of a technique known as "direct auscultation", which consists in putting the ear directly on the patients chest. This method is as uncomfortable for the clinician as it is for the patient, and a great inconvenience on woman's examination. In hospitals, this method is impractical: due to the great corporal contact, the risk of infections increases significantly.

In 1816, Laennec (1781-1826) was requested to perform the examination of a woman with a high degree of obesity with general syntoms of cardiopathy. Direct auscultation was considered impractical by the woman's husband. Laennec, taking some sheets of paper, wrapping them up and putting one end of it on the patients chest, and the other end on his ear, could hear the heart sounds clearer than using direct auscultation [2].

Despite his fragile health, in 1819, Laennec published the result of his studies on the book: "Traité de L'Auscultation Médiate, et des maladies des poumons et du coeur" (Treatise on Mediate Auscultation and Diseases of the Chest and Heart). In this work, he introduced new terms to describe more accurately the heart sounds and murmurs as well as the triple rhythm. Inadequate physiologic knowledge at that time led to a faulty interpretation of heart sounds. This was corrected, however, by the middle of the 19th century.

The first recording of heart sounds was made by Hurthle (1895), who connected a microphone to an inductorium, the secondary coil of which stimulated a frog nerve-muscle preparation. At about the same time, Willem Einthoven [3] recorded phonocardiograms (the graphic representation of the sounds which originate in the heart and great vessels[4]), first by means of a capillary electrometer and then with a string galvanometer. The first monograph of phonocardiography was published by O. Weiss in 1909[5].

Since Laennec times until today, stethoscopes are used as the first cardiologic evaluation of patients. Its a non-invasive tool that detects a wide range of functional, hemodynamic and structural anomalies. This is an indicator of the effectiveness of the diagnose capabilities of the stethoscope, as a screening tool: the examination is fast, and the diagnostic capabilities are wide on cardiology.

1.2 Computers in Medicine

1.2.1 Electronic Medical Record Systems

Although the digital computers of today evolved from the ones developed during the World War II (Eniac and Colossus[6]) and mechanical computers existed from a long time before (the precursor of the modern digital computer was Charles Babbage's Analytical engine, designed in 1835 [7]), the first use of a computer on the medical field was after 1890, by Herman Hollenrith's development of electro-mechanical computer: a punched-card data-processing system, initially used for the US census of that year, and later adapted to epidemiologic and public health surveys[8].

Afterwards, computers began to be mostly used to create clinical decision support systems for physicians, as well as programs for hospital information systems (HIS). In the 1970s, two approaches for the development of those systems emerged: the first was the creation of integrated monolithic systems, composed by a large time-shared computer that would support the entire set of applications, running on their terminals. The second approach was the creation of distributed systems, in which every application is implemented in smaller, individual computers, allowing the independent evolution of the system's applications, since they were not part of an unique, wider and complex application. With the development of network technologies, the distributed model became practical, since now they could made use of fast and reliable communications among distributed and different type of machines[9].

By the 1980s the technology for electronic medical record system had evolved considerably: The minicomputers of the 1970s had evolved to cheaper, more user-friendly and widely spread microcomputers. Also, with the advent of Microsoft Windows in 1983 (although its widespread use did not occur until the release of version 3.1 in 1992) computers became more intuitive and accessible for laymen. Improvements on network systems were introduced on a large scale in the 1980s. This created a need for a data interchange protocol in healthcare. In 1987 The HL7 (Health Level 7) protocol was created, aiming to be a "comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. "[10, 11]

In the 1980s, a great emphasis was given to informatics research on expert system methodologies developed in the 1970s, to develop clinical decision support systems. Publications that showed that reminders incorporated into electronic medical records could decrease health care costs[12, 13, 14, 15] also appeared at that time. These clinical decision support activities focused on reducing medical errors related to overlooked patient information as well as improved access to medical knowledge.

1.2.2 Computer-Aided Medical Decision Systems

The early medical diagnostic decision support (MDDS) dates back to 1965, when research of general-purpose MDDS began[16]. Also, the first qualitative studies on analysis of medical images by computers began to be reported[17, 18, 19, 20, 21, 22]. Since computers and machines outperform humans in some tasks, at that time it was believed that computers could replace humans in detecting abnormalities[23]. Initially, due to limited processing power, lack of advanced image-processing techniques and difficulties in accessing digital images, this objective was not successful. The excessive high expectations from computers resulted in disbelief when the sheer difficulty of the task became apparent [23]. R. L. Engle's review article[24] on 30 years experience, states in its conclusion: "Thus, we do not see much promise in the development of computer programs to simulate the decision-making of a physician. However, after many years we have concluded that we should stop trying to make computers act like diagnosticians".

This approach changed in the 1980s: now it was realized that computer outputs could be used by clinicians, but not replace them. This concept is now known as computer-aided diagnosis[23].

From 1980 to 1990 the approach was to develop formal models that employ a higher mathematical rigor to the exploration of its associated heuristics. To overcome the limitations of the previous heuristic and simple Bayesian models, the new systems where based on a variety of mathematical theories, such as fuzzy sets, Bayesian belief networks, artificial networks[25].

Miller[25] states that those systems should not rely on the rigidity imposed by a single formalism, due to the multi-faceted analysis of a complex medical problem, since the physician often relies on his common sense in a clinical decision making process. He also states that: "Any model that in effect replaces a physician's reasoning, or does not allow the physician to modulate the performance of the system in a patient-specific manner, should be viewed with caution".

He also states that the interface, especially the graphical user interfaces (GUIs) is a key aspect for the acceptability of MDDS systems. The reasons for that statement ranges from simplifying data input to providing environments that give enhanced flexibility for user operations.

1.3 Contributions

Throughout this thesis, some contributions were made. The main ones are listed below:

1. A review of the state-of-the-art of heart sound segmentation and cardiac pathology classification algorithms.
2. A creation of a computer system for collecting remote patient clinical information and heart sounds.
3. The development of a segmentation algorithm suitable for implementation in embedded systems.

1.4 Thesis Outline

Chapter 1 Introduces a brief history of auscultation and computers in medicine

Chapter 2 Contains a brief summary on the physiology of the heart

Chapter 3 Describes the mathematical methods necessary for understanding the work done in this thesis

Chapter 4 Presents our survey on algorithms for digital stethoscopes

Chapter 5 Describes the data collection system developed for this thesis

Chapter 6 Covers the implementation and theory of a well known heart sound segmentation algorithm, proposing an improvement that is adequate for real clinical environments.

Chapter 7 Presents the main conclusions from this work, discuss some issues and presents possible future works on this field

Chapter 2

Preliminary Knowledge and Medical Background

2.1 Physics of Sound

A study on heart sounds is by no means complete without an introduction to the acoustic phenomena and an explanation about where everything starts. A sound is, roughly speaking, the motion of waves of alternative pressure generated by a vibrating object. This vibrating source sets particles in motion and in case of a sound with only one tone, the individual particles moves around their resting point, with the same frequency of that tone. In each movement, vibrating particles pushes nearby ones, putting them in motion, and therefore creating a chain effect, generating areas of high and low pressure. This alternation between low and high pressure areas moves away from the sound source and so does the sound wave. Usually those waves can be detected by their mechanical effect on a membrane (it could be a microphone's membrane or a stethoscope's diaphragm, etc.). A common way of describing a sound is by its intensity, frequency and duration[26].

2.2 Physiology of the heart

The heart is an organ mostly made of muscles, with two main functions: to collect oxygen-rich blood from the lungs and pump it to all tissues of the body and to collect blood with carbon dioxide from the tissues of the body and pump it to the lungs. Its located inside the thorax, in a space between the pleural cavities called the *middle mediastinum*, inside the membrane named pericardium, which involves the heart (Figure 2.1). This membrane has inner

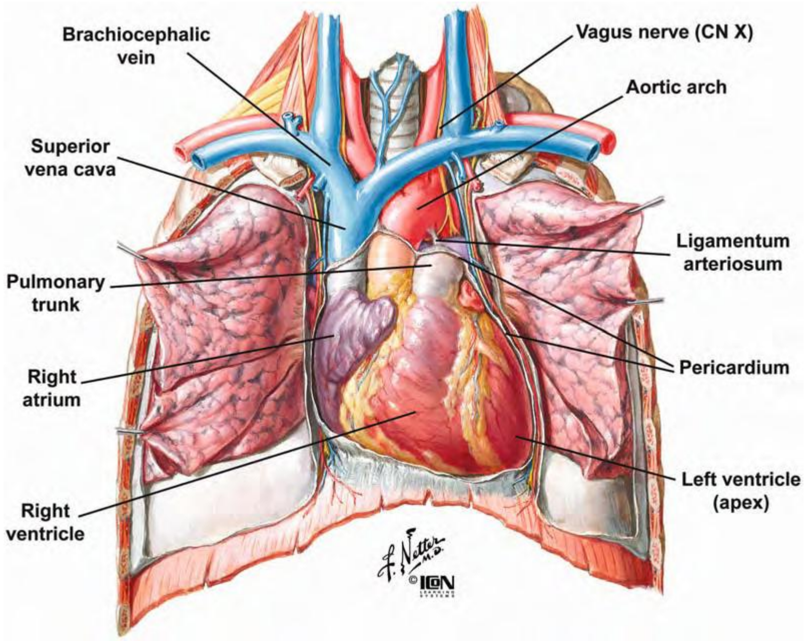


Figure 2.1: The position of the heart and the great vessels in the middle mediastinum (picture adapted from [27] p. 37).

and outer layers, with a lubricating fluid in between. The fluid allows the inner visceral pericardium to “glide” against the outer parietal pericardium[27].

The heart is composed by four chambers made of cardiac muscle or myocardium. The two upper chambers have the main function of collecting blood to, later on, inject it into the two lower chambers (ventricles), which are much stronger and works as a blood pump. The role of the right atrium and ventricle is to collect blood rich in carbon dioxide from the body and pump it to the lungs. The role of the left atrium and ventricle is to collect blood rich in oxygen from the lungs and pump it throughout the body. There is a one-way flow of blood through the heart; this flow is maintained by a set of four valves: The atrioventricular valves (tricuspid and bicuspid) allow blood to flow only from the atria to the ventricles. The semilunar valves (pulmonary and semilunar) allow blood to flow only from the ventricles out of the heart and through the great arteries, as depicted in the Figure 2.2.In general, the gross anatomy of the right side of the heart (also knows as right heart) is considerably different from that of the left heart; but the pumping principles of each are basically the same.

The ventricles are closed chambers surrounded by muscular walls, and the valves are designed to allow flow in only one direction. The cardiac valves passively open and close in response to the direction of the pressure gradient across them. The muscular cells (myocytes) of the ventricles are organized in a circum-

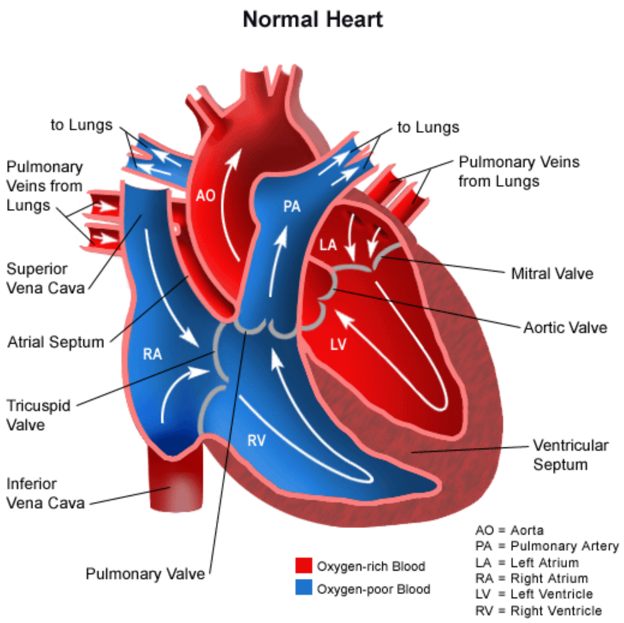


Figure 2.2: A normal heart (picture obtained from: http://www.childrenshospital.org/az/Site500/mainpageS500P0.html in 22/09/2009).

ferential orientation; hence, when they contract, the tension generated within the ventricular walls causes the pressure within the chamber to increase. As the ventricular pressure exceeds the pressure in the pulmonary artery (on the right heart) and/or aorta (on the left heart), blood is forced out of the given ventricular chamber. This active contractile phase of the cardiac cycle is known as *systole*. The pressures are higher in the ventricles than the atria during systole; hence, the tricuspid and mitral (atrioventricular) valves are closed. When the ventricular muscles relax, the pressure inside them falls below those in the atria, the atrioventricular valves open and the ventricles refill; this phase is known as *diastole*. The aortic and pulmonary valves are closed during diastole because the arterial pressures (in the aorta and pulmonary artery) are greater than the intraventricular pressures[28]. This is depicted briefly on the Figure 2.3

2.3 Heart Sounds

Heart sounds are caused by the dynamic events associated with the heartbeat and the blood flow. They are relatively brief and have different intensity (loudness), frequency (pitch) and quality (timbre). To better understand heart sounds, we need to know the physiology of the cardiac events. They happen as

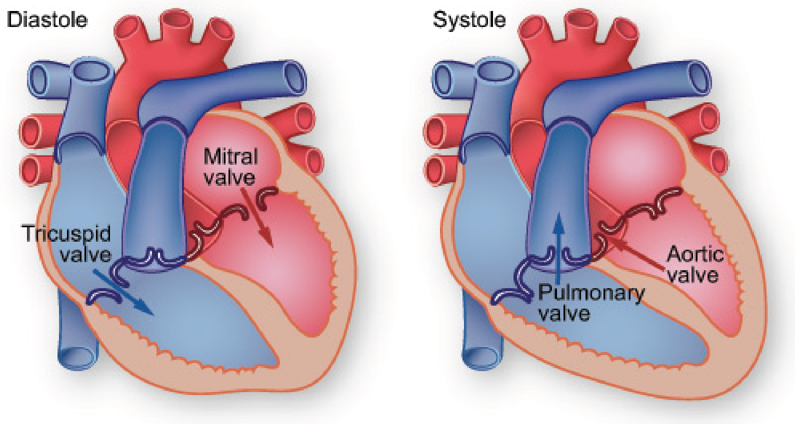


Figure 2.3: The Cardiac Cycle(picture obtained from: http://www.texasheartinstitute.org/HIC/Topics/Cond/ddisfunc.cfm in 22/09/2009).

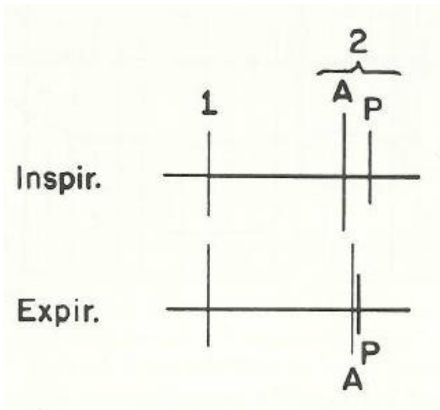


Figure 2.4: Physiological splitting of S2 (adapted from [29], p. 23).

follows: the electrical impulse for the contraction of the heart comes from the sinus node on the right atrium, making the right atrium to contract first. The Ventricular contraction results in mitral valve closure, and in this phase, the mitral valve closes before the tricuspid because the left ventricle is slightly faster than the right ventricle. Since the right ventricle has lower pressure, ejection begins first at that ventricle. Also, as the right ventricle finishes ejection first, it makes the aortic valve to close slightly before the pulmonary valve[27, 29].

The *first heart sound* (S1) is composed of the closure of the mitral valve (M1), then, slightly afterwards, by the closure of the tricuspid valve (T1).

The closure of the semilunar valves (aortic and pulmonary) produces the *second heart sound* (S2). These two components are closer in expiration and more widely separated in inspiration[27, 29, 30] (Figure 2.4).

The *third heart sound* (S3) happens slightly after A2. It is a low-pitched vibration caused by the fast filling of the ventricular in the diastole period. It is usually present in children, adolescent and young adults and considered as a sign of pathology if detectable after the age of 30 years[28].

The fourth heart sound (S4) is generated by the fast ventricular filling during atrial systole, which causes vibrations of the ventricular wall and the mitral apparatus. Its a soft and low-pitched noise, heard just before S1. S4 is normally heard in infants, small children and adults over 50 years of age.

2.3.1 Abnormal Heart Sounds

The normal blood flow in the heart is mainly laminar, therefore, does not produces murmurs. Whenever there is a pathology causing turbulent flow (valve regurgitation, stenosis, etc.), this turbulent flow creates vibrations that are heard as murmurs[27].The frequency range of the heart sounds and abnormal heart murmurs are within 10 to 1500Hz.

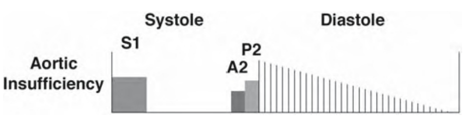
On a normal heart sound, the energy is distributed as depicted in Table 2.1.

Frequency in Hz	Energy
50-60	56%
60-70	27%
70-80	10%
80-90	4%
90-100	2%
100-110	1%

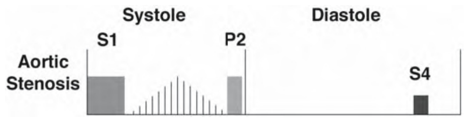
Table 2.1: Normal heart sound's energy distribution by frequency range

Cardiovascular murmurs are described based on their appearance on the cardiac cycle (early, mid, or late systolic, diastolic or continuous), its intensity, frequency, "shape", quantity, duration, or direction of radiation. Usually the loudness of a sound is classified in a scale from 1 to 6. Regarding the cardiac cycle, a common classification is as follows[27, 29]:

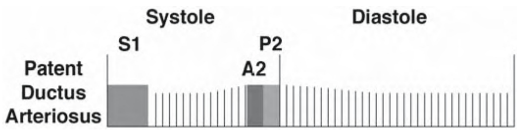
- *Early systolic murmur*: begins with S1 and ends before the midsystole.
- *Midsystolic murmur*: starts after S1 and ends before S2.
- *Late systolic murmur*: begins in the middle of systole and ends at S2.
- *Holosystolic murmur*: starts with S1 and continues for the duration of the whole systole.
- *Early diastolic murmur*: begins with S2.



(a) An early decrescendo murmur



(b) A crescendo-decrescendo murmur



(c) A plateau murmur

Figure 2.5: Examples of pathological heart murmurs (adapted from [27], pages 188 and 189).

- *Middiastolic murmur*: begins after S2.
- *Late diastolic murmur*: begins just before S1.
- *Continuous murmur*: continues during both systole and diastole.

Regarding the changes in intensity, the murmurs are classified as: crescendo (when its intensity increases), decrescendo (decreasing intensity, Figure 2.5a), crescendo-decrescendo (when its intensity first increases, then decreases. When it is symmetric, also called "diamond-shaped", Figure 2.5b) and plateau (when the intensity is constant, Figure 2.5c).

2.3.2 Auscultation Areas

The main classical auscultation areas[1] for hearing the different heart sounds and murmurs are depicted in Figure 2.6.

Auscultation sites may vary according to the patient anatomy (situs inversus, fat patients, etc) so Figure 2.6 should be used just as a model, and should be adapted for the sake of acquiring better auscultations.

Each area emphasizes certain components of the heart sounds. The aortic area is best to hear the components of the second heart sound (aortic and pulmonary valves). The pulmonic area is best suited for hearing the pulmonary component of the second heart sound and detecting murmurs related to pulmonary stenosis and pulmonary insufficiency. The mitral area is best for detecting the murmur of the mitral insufficiency or mitral stenosis and other murmurs like third heart sound and aortic component of the second heart sound. The tricuspid area is better suited for hearing the murmurs of tricuspid stenosis and

[1] Those areas may have different names throughout literature, such as "left ventricular area", "third left interspace", "right atrial area", "pulmonary area", etc.

inefficiency, as well as the third heart sound and the murmurs of pulmonary insufficiency and ventricular septal defect[27, 29].

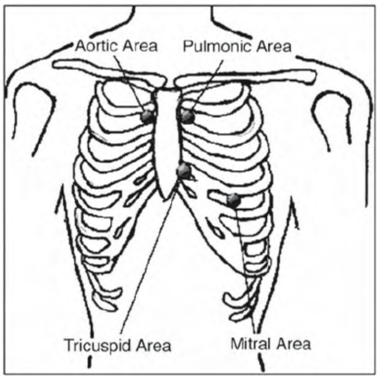


Figure 2.6: Auscultatory areas (adapted from [27], p. 186).

Chapter 3

Mathematical Methods

3.1 Introduction

As we could see in the previous chapter, specially in section 2.3.1, heart sounds have specific frequencies which are useful for preparing and analyzing heart sounds. In this chapter we will discuss a powerful technique capable of filtering and analyzing heart sounds: the Discrete Wavelet Transform.

Most of this chapter is based on chapters 2 and 3 of the *Illustrated Wavelet Transform Handbook* [31]. All information refers to this source, unless stated otherwise.

3.2 Wavelets

The wavelet transform is a remarkable mathematical method with the ability to examine the signal simultaneously in time and frequency, in a different way from previous mathematical methods. Wavelet analysis has been applied in a wide range of applications: from climate analysis, to signal compression and medical signal analysis. The application of wavelet transform analysis in science and engineering began to increase in the beginning of the 1990s, directly reflecting the interest of the scientific community[31].

Some of the more commonly used wavelets are depicted in Figure 3.1. We can notice that they have the shape of a small wave, localized on the time axis. Depending both on the signal we need to analyze and what characteristic we are analyzing, one wavelet can be better suited than others.

The wavelet function $\psi(t)$ has to satisfy certain mathematical criteria, such as:

1. Must have finite energy:

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \tag{3.1}$$

2. It must have no zero frequency component ($\hat{\psi}(0) = 0$), or in other words, if $\hat{\psi}$is the Fourier transform of $\psi(t)$:

$$\hat{\psi}(f) = \int_{-\infty}^{\infty} \psi(t)e^{-i(2\pi ft)}dt = 0 \tag{3.2}$$

3. It must hold the following condition:

$$C_g = \int_{0}^{\infty} \frac{|\hat{\psi}(f)|^2}{f} df < \infty \tag{3.3}$$

   The above equation is known as *admissibility condition* and C_g is known as *admissibility constant* and is dependent on the chosen wavelet.

4. An extra criterion, on complex wavelets, is that the Fourier transform must vanish for negative frequencies and also must be real.

3.2.1 Wavelet transform

Wavelet transform analysis uses 'local' wavelike functions to transform the signal under investigation into a representation which is more useful for the analysis of the desired feature (the features may range from corner detection to frequency analysis, depending on the wavelet and the transform itself). This transform is a convolution of the wavelet function with the signal.

The wavelet can be manipulated in two ways: it can change its location or its scale (Figure 3.2). If, at a point, the wavelet matches the shape of the signal, then the convolution has a high value. Similarly, if the wavelet and the signal do not correlate well, the transform results in a low value. The wavelet transform is computed at various locations of the signal and for various scales of the wavelet: this is done in a continuous way for the *continuous wavelet transfrom* (CWT) or in discrete steps for the *discrete wavelet transform* (DWT).

The operations over the wavelet are defined by the parameters a (for dilation) and b (for translation). The shifted and dilated versions of the wavelet are denoted $\psi[[t-b]/a]$. For sake of simplicity, let us take the mexican hat wavelet:

$$\psi(t) = (1-t^2)e^{-t^2/2} \tag{3.4}$$

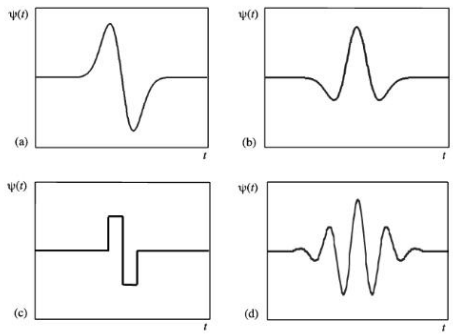


Figure 3.1: Example of wavelets: a)Gaussian wave (first derivative of a Gaussian). b) Mexican hat (second derivative of a Gaussian). c) Real part of Morlet (adapted from [31], p. 7).

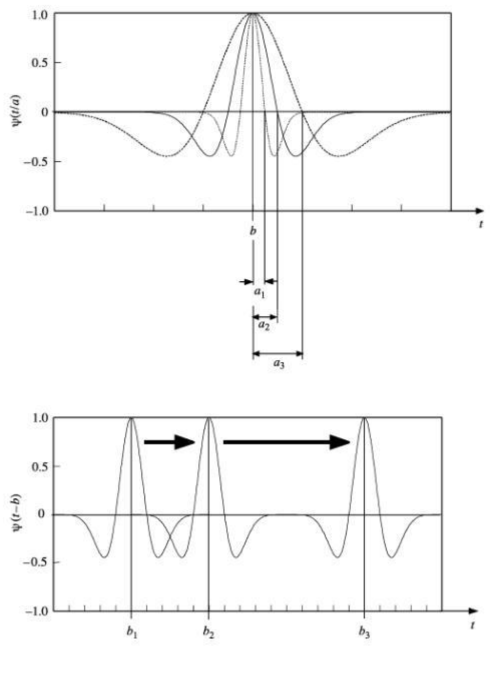


Figure 3.2: Two possible manipulation with wavelets: a) Translation (or location) and b) Scale (adapted from [31], p. 11).

The shifted and dilated equation for this version of the wavelet would be:

$$\psi(\frac{t-b}{a}) = [1-(\frac{t-b}{a})^2]e^{-\frac{1}{2}[(t-b)/a^2]} \tag{3.5}$$

The wavelet transform of a continuous signal with respect with the wavelet is a convolution given by:

$$T(a,b) = w(a)\int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-b}{a})dt \tag{3.6}$$

Where w(a) is a weighting function. Typically, for energy conservation purposes, $w(a)$ is set to $1/\sqrt{a}$, because it ensures that the wavelet would have the same energy on all scales.

3.3 The Discrete Wavelet Transform

To be of any practical use in a digital computer, we need to use the discrete version of the wavelet transform, namely, the discrete wavelet transform.

First we need to consider the discrete values of a and b over the wavelet, as the new discretized wavelet has the form:

$$\psi_{m,n} = \frac{1}{\sqrt{a_0^m}}\psi(\frac{t-nb_0a_0^m}{a_0^m}) \tag{3.7}$$

Where m and n are integers and the wavelet translation b_0 (must be greater than zero) and the dilatation step a_0 (must be fixed, greater than 1). Therefore, the discrete wavelet transform of a continuous signal using the discrete wavelet transform would be:

$$T_{m,n} = \int_{-\infty}^{\infty} x(t)\frac{1}{a_0^{m/2}}\psi(a_0^{-m}t - nb_0)dt \tag{3.8}$$

Here, the $T_{m,n}$is the discrete wavelet transform transform given on a m dilation and n scale. These values are known as *wavelet coefficients* or *detail coefficients*. The discrete sampling of the time and scale parameters of a continuous wavelet transform (as above), is known as wavelet frame. The Energy of the wavelet functions that composes a frame lies in the bounded range:

$$AE \leq \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}|T_{m,n}|^2 \leq BE \tag{3.9}$$

Where A and B are the frame bounds (where the wavelet is defined and nonzero), E is the energy of the original signal. The values of A and B depends on the values of a_0 and b_0 used on the selected wavelet. When $A = B$ the

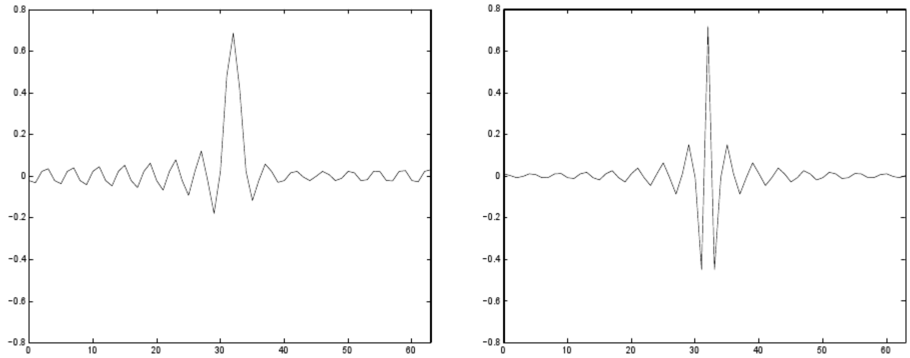


Figure 3.3: Shannon father wavelet (left) and Shannon mother wavelet (right). Notice that the father wavelet has components with lower frequency than the mother wavelet. (adapted from [32], p. 179).

wavelet family defined by the frame forms an *orthonormal basis.* The signal can be reconstructed by the following formula:

$$x'(t) = \frac{2}{A+B} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} T_{m,n} \psi_{m,n}(t) \tag{3.10}$$

When the wavelet family chosen is both an *orthonormal basis* and a *dyadic grid arrangement* (i.e.: the wavelet parameters $a_0 = 2$ and $b_0 = 1$), they are both orthogonal to each other and have unit energy, i.e., the product of each wavelet with all the others is zero. This means that this wavelet transform has no redundancy and allows the regeneration of the signal.

3.3.1 The multiresolution respresentation

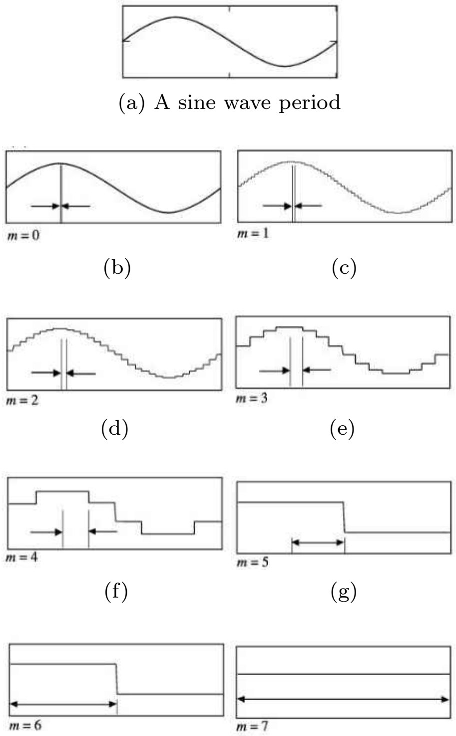
Let the piecewise smooth function$\varphi_{0,0}(t)$, also known as *scaling function* (or *father wavelet*), be an orthonormal base on our dyadic system, so that our wavelet can be defined [32] by:

$$\psi(t) := \frac{d^M}{dt^M}(t) \tag{3.11}$$

As our mother wavelet ψ is a differentiated version of the ϕ function, It also has higher frequency elements, if compared with the soothed ϕ wavelet, as we can see in the Figure 3.3.

The set of scaling functions is defined as same form as the wavelet:

$$\phi_{m,n} = 2^{-m/2}\phi(2^{-m}t - n) \tag{3.12}$$



(a) A sine wave period

(b) (c)

(d) (e)

(f) (g)

(h) (i)

Figure 3.4: a) original signal (sine wave), b) Sine wave on scale 0 (the horizontal arrow is the width of the approximation level, 2^m), from c) to i) approximation levels from 1 to 7 respectively. (adapted from [31], p. 71).

With the following property:

$$\int_{-\infty}^{\infty} \phi_{0,0}(t)dt = 1 \tag{3.13}$$

If the father wavelet is convolved with the signal (equation 3.14) then the approximation coefficients $S_{m,n}$ are produced.

$$S_{m,n} = \int_{-\infty}^{\infty} x(t)\phi_{m,n}(t)dt \tag{3.14}$$

A continuous approximation of the signal at a given scale m can be obtained by summing a sequence of father wavelets at that scale, factored by the approximation coefficients:

$$x_m(t) = \sum_{n=-\infty}^{\infty} S_{m,n}\phi_{m,n}(t) \tag{3.15}$$

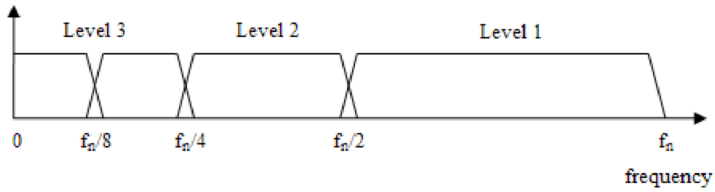


Figure 3.5: The frequency range on different levels (m values). The frequency cut-offs overlap due to the fact the filters that form the wavelet transform are not ideal filters (picture extracted from: http://commons.wikimedia.org/wiki/File:Wavelets_-_DWT_Freq.png, accessed on 22/09/2009).

Figure 3.4 shows a sine wave and several approximations using the decomposition (3.14) and approximation (3.15) equations. The scale used was set to a range of widths 2^0 to 2^7. The widths are indicated by the horizontal arrows in each figure. Note that these approximations are applied on the original signal (the sine wave) with different levels (m values)[32].

One should notice that, since the father wavelet has lower frequencies than the mother wavelet, the convolution of the signal with the father wavelet acts as a low pass filter, and the convolution of the signal with the mother wavelet acts as a high pass filter. Their frequency ranges are depicted in Figure 3.5.

A signal can be completely represented as a combined series expansion using the approximation and the detail coefficients:

$$x(t) = \sum_{n=-\infty}^{\infty} S_{m_0,n}\phi_{m_0,n}(t) + \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} T_{m,n}\psi_{m,n}(t) \tag{3.16}$$

We can see from the above formula that the signal is decomposed into an approximation and detail of itself at an arbitrary scale (m_0).

The contracted and shifted version of the father wavelet is as follows:

$$\phi(t) = \sum_k c_k\phi(2t-k) \tag{3.17}$$

Here, $c_k$is the scaling coefficient and k is the shift. Looking closely to this equation we can realize that one scaling function can be built up from previous scaling functions. Also, this function needs to be orthogonal (as it happens with the mother wavelet).

The coefficients for the wavelet function are as follows:

$$\psi(t) = \sum_{k=0}^{N_k-1} b_k\phi(2t-k) \tag{3.18}$$

From equations 3.12 and 3.17, at a given $m+1$ index, the next father wavelet becomes:

$$\phi_{m+1,n}(t) = \frac{1}{\sqrt{2}} \sum c_k \phi_{m,2n+k}(t) \tag{3.19}$$

Similarly, the next mother wavelet:

$$\psi_{m+1,n}(t) = \frac{1}{\sqrt{2}} \sum b_k \phi_{m,2n+k}(t) \tag{3.20}$$

One should notice that the scaling function at a scale is composed by a sequence of shifted scaling functions on a smaller scale, each one factored by their respective scaling coefficients.

Substituting equation 3.19 into equation 3.14 for the new indexes of the father wavelets yields the recursive form:

$$S_{m+1,n} = \frac{1}{\sqrt{2}} \sum c_k [\int_{-\infty}^{\infty} \phi_{m,2n+k}(t)dt] = \frac{1}{\sqrt{2}} \sum_k c_{k-2n} S_{m,k} \tag{3.21}$$

$$T_{m+1,n} = \frac{1}{\sqrt{2}} \sum b_k [\int_{-\infty}^{\infty} \phi_{m,2n+k}(t)dt] = \frac{1}{\sqrt{2}} \sum_k b_{k-2n} S_{m,k} \tag{3.22}$$

Here we can see that every coefficient on the detail (3.21) and on the approximation (3.22), are recursive until m_0, which is the signal itself. The above equations represent the multiresolution decomposition algorithm. By iterating those two equations, we are performing a lowpass filtering (3.22) and a highpass filtering (3.21).

To summarize: Consider the input signal $S_{0,n}$. Compute $S_{m,n}$ and $T_{m,n}$ using the decomposition equations (3.22) and (3.21). The first iteration would give us $S_{1,n}$ and $T_{1,n}$. Now we apply to the same approximation $S_{1,n}$ to the equations again to get the next coefficients $S_{2,n}$ and $T_{2,n}$ and so on until only one approximation is computed (On each level, the amount of samples on the signal decreases by half. This means we have a lower maximum frequency). Now we have an array with the detail coefficients on different resolutions and one approximation, on the last level. This process is depicted in Figure 3.6.

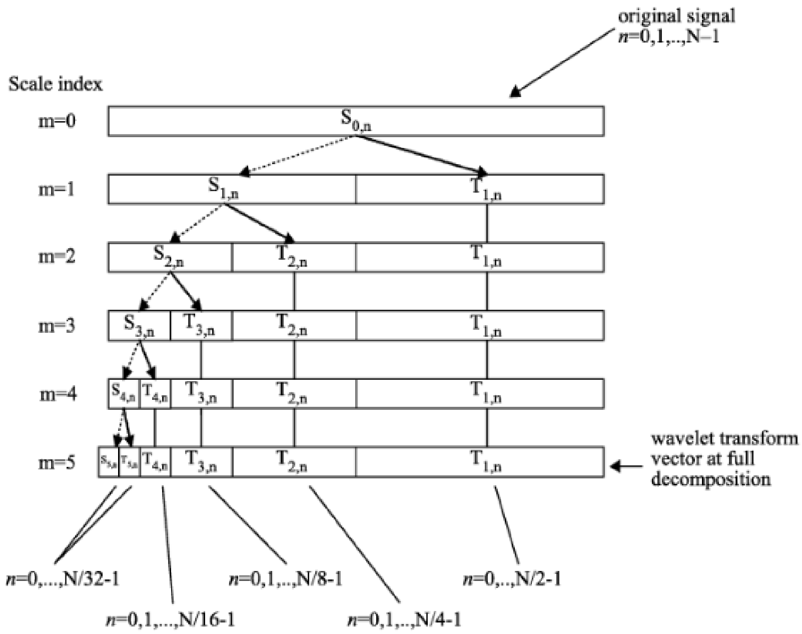


Figure 3.6: The wavelet decomposition using a filter bank: each filter receives the input from the previous levels approximation coefficients (adapted from [31], p. 82).

Chapter 4

A Survey on Algorithms for Digital Stethoscopes

4.1 Digital Stethoscopes

Throughout history, there have been several attempts to create electronically enhanced stethoscopes, with better sound amplification and frequency response. However, and according to Durand [33], their introduction into clinical practice has been hindered by factors such as their background noise, unfamiliar sounds to clinicians due to filtering or fragility and bad ergonomic design. Recent advances in electronics and digital circuits allow us to not only overcome these problems but also to exploit the benefits of digital signal processing for signal analysis and visualization.

Since there may be various interpretations for the name 'digital stethoscope', it is essential that we define this term. The traditional stethoscopes are based on the acoustic stethoscope developed by Laennec, and depend solely on acoustics to amplify and transmit the heart sounds to the clinician. The concept of electronic stethoscope arrives when electronic components were first used to amplify, filter and transmit the sound [33] (Figure 4.1).

There are various examples in literature regarding the development of digital and electronic stethoscopes. Bredesen and Schmerler [34] have patented an "intelligent stethoscope" designed for performing auscultation and for automatically diagnosing abnormalities by comparing digitized sounds to reference templates using a signature analysis technique. Several other electronically enhanced and digital stethoscopes have been developed and described in literature [33, 35, 36, 37].

Figure 4.2 shows a block diagram of an electronically enhanced stethoscope

Figure 4.1: Lab prototype of an electronically enhanced stethoscope developed by our group.

prototype developed by our group. The auscultation quality was considered satisfactory in clinical trials when compared to auscultation using acoustic stethoscopes, but clinicians still perceived differences in audio pitch, although this did not affect their ability to diagnose heart conditions. Our field experience confirms Durand's [33] opinion that audio enhancement alone is not enough for the clinical community to adopt this new technology. In order to make digital stethoscopes attractive to clinical cardiologists, we clearly need to address the numerous potential improvements provided by a fully functional, robust digital stethoscope: real-time acquisition, analysis, display and reproduction of heart sounds and murmurs. Digital stethoscopes must also open the doors for digital audio archives, simplifying the acquisition, storage and transmission process of cardiac exams and murmurs, potentiating functionalities such as the teaching of auscultation, telemedicine, and personalized healthcare.

For the analysis of the state-of-the-art on audio processing in cardiology, we have very loosely adopted some concepts of clinical systematic reviews. A rigorous systematic review of such a multi-disciplinary vast field is quite difficult to implement in practice due to the large number of papers retrieved by analysis of both engineering and medical scientific databases. Our review methodology was as follows:

- Considered that Durand's [33] excellent review paper fully covers this topic up to 1995.
- Consulted the IEEE Xplore (ieeexplore.ieee.org) database with the following query: "(((feature extraction)<in>metadata) <and> ((cardiology)<in> metadata))", obtaining 159 results after 1995.
- By title and abstract inspection, we kept only papers dealing with phonocardiogram data analysis, reducing this number to 19.
- We analyzed the references from all these papers, and selected all papers published after 1995 and with more than 10 citations, obtaining 20 results.

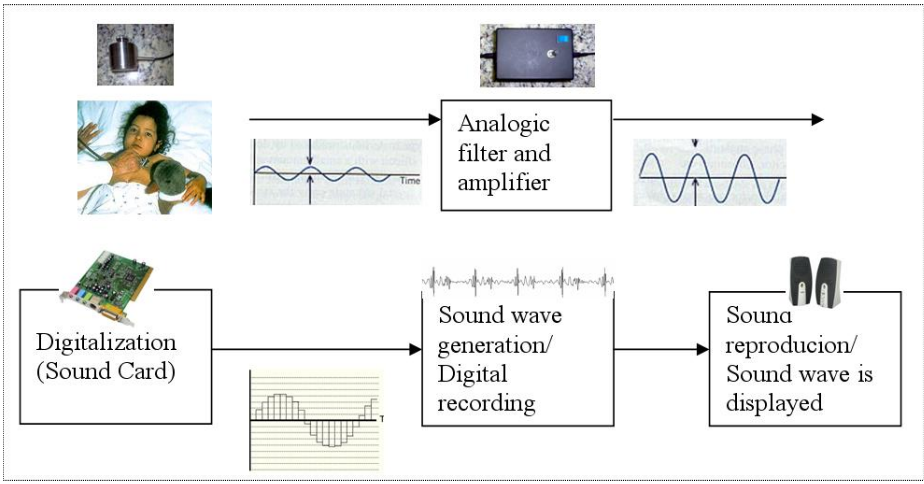


Figure 4.2: Block diagram of our electronically enhanced stethoscope prototype depicted in figure 4.1 and field-tested at Real Hospital Português de Beneficiência em Pernambuco in Recife, Brazil. Over 100 auscultations were performed during the clinical validation stage.

This enabled us to cover additional articles besides the ones published in IEEE journals and conferences, artificially expanding the scope of our review to other scientific databases.

- The total number of papers covered by this review is thus 39 (19+20).

Although we are certain that it is possible to miss some papers using this methodology, we feel that we have covered a sufficiently vast and interesting sample to draw some important conclusions, as described later on.

4.2 Heart Sound Analysis and Feature Extraction

As we had seen in the Chapter 2, the main constituents of a cardiac cycle are the first heart sound (typically referred to as S1), the systolic period, the second heart sound (S2) and the diastolic period. Whenever a clinician is performing an auscultation, he tries to identify these individual components, and is trained to analyze related features such as rhythm, timing instants, intensity of heart sound components, splitting of S2, etc [38]. This analysis allows him to search for murmurs and sound abnormalities that might correspond to specific cardiac pathologies. From a signal processing perspective, Heart Sound Analysis (HSA) is not only interesting by itself (allowing quantitative measures to be displayed automatically in a digital stethoscope), but is also an essential first step for the subsequent task of automatic pathology classification.

For sake of clarity, we will distinguish two sub-tasks of HSA: Heart Sound Segmentation (HSS) and Aortic Pulmonary Signal Decomposition (APSD).

4.2.1 Heart Sound Segmentation

In HSS we expect to identify and segment the four main constituents of a cardiac cycle. This is typically accomplished by identifying the position and duration of S1 and S2, using some sort of peak-picking methodology on a pre-processed signal. Liang [39] has used discrete wavelet decomposition and reconstructed the signal using only the most relevant frequency bands. Peak-picking was performed by thresholding the normalized average Shannon energy, and discarding extra peaks via analysis of the mean and variance of peak intervals. Finally, they distinguish between S1 and S2 peaks (assuming that the diastolic period is longer than the systolic one, and that the later is more constant), and estimate their durations. A classification accuracy of 93% was obtained on 515 periods of PCG signal recordings from 37 digital phonocardiographic recordings. The same authors further improved the statistical significance of their results by obtaining the same accuracy using 1165 cardiac periods from 77 recordings [38], and later attempted murmur classification based on these features and neural network classifiers, obtaining 74% accuracy [40]. Omran [41] has also studied this problem using normalized Shannon entropy after wavelet decomposition of the audio signal, but their experimental methodology is not so convincing.

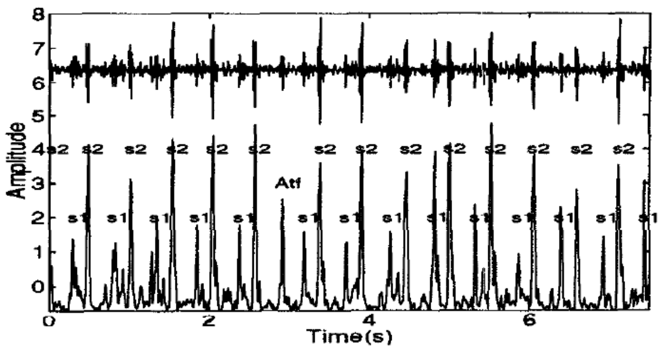
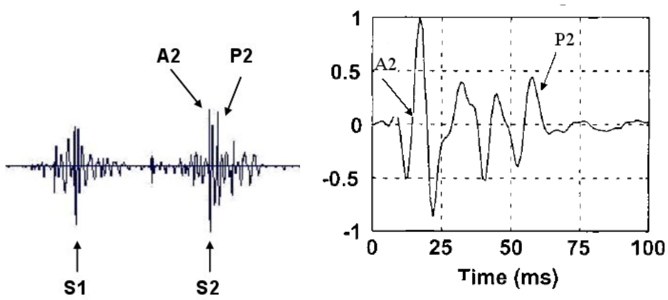


Figure 4.3: A heart sound segmentation example [39] (adated from [39]).

4.2.2 Aortic Pulmonary Signal Decomposition

Besides the four main components of the cardiac cycle, there is a clinical interest in the analysis of some of its associated sub-components [42]. It has been recognized that S1 may be composed of up to four components produced during



(a) (b)

Figure 4.4: The Aortic and Pulmonary components of S2. In (a) their location on the heart sound, in (b) a close-up on the two components (adapted from [42]).

ventricular contraction [33], although the complexity of this task has been a very difficult hurdle for the signal processing community. The S2 sound is more well known, being composed of an aortic component (A2), which is produced first during the closure and vibration of the aortic valve and surrounding tissues, followed by the pulmonary component (P2) produced by a similar process associated with the pulmonary valve [42]. Durand [42] demonstrated that it is possible to model each component of S2 by a narrow-band nonlinear chirp signal. Later [43] he adapted and validated this approach for the analysis and synthesis of overlapping A2 and P2 components of S2. To do so, the time-frequency representation of the signal is generated and then estimated and reconstructed using the instantaneous phase and amplitude of each component (A2 and P2). In this paper the accuracy evaluation was made by simulated A2 and P2 components having different overlapping factors. The reported error was between 1% and 6%, proportional to the duration of the overlapping interval. Nigam [44] also presented a method for extracting A2 and P2 components by assuming them as statistically independent. To do so, four simultaneous auscultations are analyzed using blind source separation. The main advantage of this method is the lower dependence on the A2-P2 time interval, although it needs a non-conventional 4-sensor stethoscope.

4.3 Automatic Pathology Classification

The vast majority of papers we have found regarding audio processing algorithms, adequate for the integration into a digital stethoscope, concern the detection of specific heart pathologies. This highlights the interest of the scientific community on this topic but, as our analysis shows, there are still some major flaws in most of them such as the absence of a clinical validation step and unconvincing experimental methodologies. Most papers use the well-established pattern recognition approach of feature extraction followed by a classifier. Bentley [45] uses Choi-Williams Distribution (CWD) as features, working with 45 normal/abnormal valve subjects. Some features were determined via visual inspection, others automatically from the CWD by simple rule-based classification. Later [46], the authors show that CWD is a better method to represent the frequencies in PCG and to get heart sound descriptors, than other time-frequency (T-F) representations. According to them, a simple description of the T-F distribution allows an analysis of the heart valve's condition. However, they highlight the need of a more comprehensive evaluation using a larger population of test patients. Wang [47] proposes a representation of heart sounds that is robust to noise levels of 20dB, using mel-scaled wavelet features. However, details regarding the used dataset are not clear enough for robust conclusions. Liang [48] developed an interesting feature vector extraction algorithm where the systolic signal is decomposed by wavelets into subbands. Then, the best basis set is selected, and the average feature vector of each heart sound recording is calculated. Neural Networks (NN) are used for classifying 20 samples after being trained with 65, obtaining an accuracy of 85%. NNs are also used by Abdel-Alim [49] for the automatic diagnostics of heart valves using wavelets feature vectors and stethoscope location information. They use two NNs: one for systolic diseases and the other for diastolic diseases. A total of 1200 cases were used: 970 cases for training and 300 for testing. The recognition rate was 95%. Turkoglu [50], Ozgur [51] and El-Hanjouri [52] also used wavelets as feature vectors for classification, although they provide too few details regarding the used data sets. Trimmed mean spectrograms are used by Leung [53] to extract features of phonocardiograms. Together with the acoustic intensities in systole and diastole, the authors quantified the distinctive characteristics of different types of murmurs using NNs. One of the few papers that is conscious about the important clinical validation step is from Kail [54]. The authors propose a novel sound representation (2D and 3D) and feature extraction algorithm using Morlet wavelet scalograms. After manual classification of the resulting graphs performed by two cardiologists on 773 subjects, they clinically validated the features as useful for sound and murmur extraction. Sharif [55] also proposes other features for classification systems based on central finite differences and zero crossing frequency estimation.

A summary of the techniques used on the above papers is listed in Table 4.1.

Author - Paper	Features Source	Classifier
Bentley - “Classification of native heart valve sounds using the Choi-Williams time-frequency distribution”	Choi-Williams Distribution	Rule-Based Classification
Bentley - ”Time-frequency and time-scale techniques for the classification of native and bioprosthetic heart valve sounds”	Choi-Williams Distribution	–
Wang - “Feature extraction based on mel-scaled wavelet transform for heart sound analysis”	Mel-Scaled Wavelet	–
Liang - “A feature extraction algorithm based on wavelet packet decomposition for heart sound signals”	Wavelet Decomposition	Neural Network
Abdel-Alim - “Heart diseases diagnosis using heart sounds”	Wavelet feature vectors and stethoscope location	Neural Network
Turkoglu - “An intelligent pattern recognition system based on neural network and wavelet decomposition for interpretation of heart sounds”	Wavelet Decomposition	Neural Network
Ozgur - “Classification of heart sounds by using wavelet transform”	Wavelets	Neural Network
El-Hanjouri - “Heart diseases diagnosis using HMM”	Wavelets	Hidden Markov Models
Leung - “Classification of heart sounds using time-frequency method and artificial neural networks”	Trimmed Mean Spectrograms	Probability Neural Network
Kail - “Internet digital phonocardiography in clinical settings and in population screening”	Morlet Wavelets Scalograms	–
Sharif - “Analysis and classification of heart sounds and murmurs based on the instantaneous energy and frequency estimations”	Central Finite Difference and Zero Crossing Frequency Estimation	–

Table 4.1: Summary of Automatic Pathology Classification most relevant papers used feature sources and classifiers

4.4 Conclusions

This chapter described the first contribution of this thesis. By covering the most interesting papers on audio-processing from a digital stethoscope perspective, we can make some observations regarding the state-of-the-art on this field, producing already important results regarding audio feature extraction. The S1 and S2 sounds can be robustly segmented and there is promising work regarding the extraction of secondary sounds such as A2 and P2.

The scenario is not so bright for automatic pathology classification. Reviewing some of the papers and simply observing the disparity in the number of publications when compared with the other challenges, we conclude that there is a strong interesting in this topic. However, in our opinion, there is still a long way to go before we can have robust automatic classification systems that can be introduced in the clinical routine of hospitals. We have identified three major problems that afflict most of the papers reviewed:

- Absence of a set of well-accepted features - We rarely found papers that selected the same features for pathology classification. Most acknowledge that the presence of S1 and S2 is important but there is no consensus of the scientific community on how these should be used. We have collected more than 25 different features with minimum overlap between papers. We clearly need more studies on the statistical significance and clinical importance of heart sound features, from an automatic pattern recognition perspective.

- Badly descript data-sets - It is not enough for authors to mention that they have worked with 300 cardiac cycles. Where were these obtained? From how many patients? In what conditions? Using which equipment? All these factors are vital in the analysis of a system's performance and robustness. Studies need to be much more rigorous on this topic so their results can be reasonably convincing.

- Absence of clinical validation - Almost no papers bothered to handle this vital task of assisted-diagnostic systems. No medical specialist will trust any kind of automatic system without it proving to be robust and accurate in real field testing. These conditions are very different from a typical biomedical engineering research lab, which can drastically affect results.

As a final conclusion, we can say that working towards next-generation 'intelligent' digital stethoscopes is highly desirable judging from the significant number of scientific publications on this topic but also examining the undeniable benefits that such systems can provide. There is already solid work regarding audio feature extraction and many unsolved challenges in this field such as the complex analysis of the sub-components of S1. Automatic pathology classification is still too undeveloped to be of any practical usage and we hope that the valuable lessons learned from this study can correct previous mistakes and provide a precious boost to the challenging field of audio processing for digital stethoscopes.

This Chapter resulted in the following publication: "A Survey of Audio Processing Algorithms for Digital Stethoscopes "[56].

Chapter 5

Data Collection System

5.1 Introduction

In order to work with heart sounds, we first need to have access to a large annotated heart sounds database. We identified a lack of massive public annotated databases available for cardiologic audio research. To solve this problem, we decide to create a data collection system, capable of gathering the auscultations themselves and the patient data potentially necessary for further computer assisted diagnosis.

Our data collection hardware should be able to perform auscultations in different sites (the clinical environment is quite dynamic, so the auscultation site could change from one room to another one), also, with mobility, we can later on do auscultations in parallel with other exams performed by heavier machines. Also, our experience with electronic stethoscopes shows that some noise that can defile the auscultation actually comes from noisy power supplies. We therefore chose to perform the auscultation with a notebook running on batteries to avoid this problem.

We used an external sound card, since it has an extra layer of electrical isolation from the computer, and additionally has a higher resolution (the model we used can go up to 24 bits resolution).

The stethoscope used (Elite WelchAllyn) was selected due to its shape being similar to the regular stethoscope, making its use more comfortable to the clinician.

The whole hardware system is depicted in the Figure 5.1

To accomplishing the objective of creating a system capable of gathering such clinically annotated database for cardiology signal processing research, the software should have the following features:

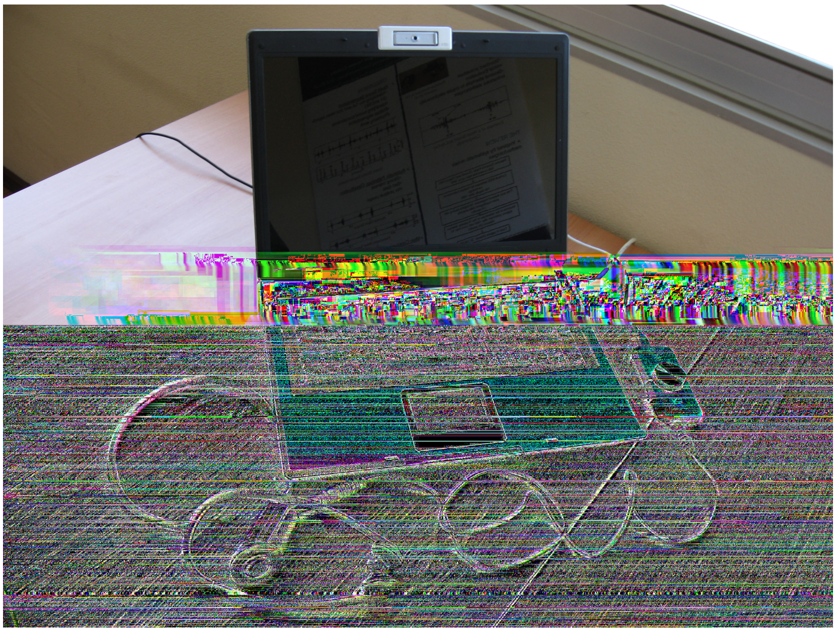


Figure 5.1: The hardware of the data collecting system

- *A database model for the collected data, including not only audio data but also patient record information*: The data collected by our system is heterogeneous and correlated (a patient might have auscultations done in different sites, also this information is correlated with the patent's clinical data, so this association should be recoverable). A database provides us with an efficient and natural way of keeping that relationship. It also gives us flexibility when dealing with a number of different hospital information systems in the future.

- *Design an effective data collection system that does not hinder typical routine Hospital work on cardiology*: Usually a clinician is a busy professional with a high demand for attending a large number of patients. Also, the familiarity level with technology varies widely between clinicians. Therefore, the choice of the technologies used in our system should create an environment as similar as possible with their working routine.

- *Secure protocols for data transmission*: As the hospital and the research unit are in different sites, the use of secure protocols for the data transmission is paramount to ensure the security of the system.

- *Patient secrecy*: Ethical research implies that the information that identifies the patient should not be transmitted to the University, but should also be identifiable by the clinician (and not to the research team). That 'traceability' of information is crucial for, later on, checking if the results obtained by the research are compatible with patient reality.

Our experience with such kind of systems shows that the above requirements should be meet to maximize the level of acceptance of our system by the clinicians.

5.2 Developed Prototype

5.2.1 System's Overall Structure

The chosen hospital for collecting data was the Royal Portuguese Hospital in Recife-Pernambuco Brazil, due to the fact that they spawned a high level of motivation in taking part on this research and some members of the team already worked together.

The distance between the Hospital and the University made us adopt a client/server structure. To ensure the *communication security*, the two modules of this system communicate via an SSH tunnel. Also, the system maintance is done using the *TightVNC* [57] software over the ssh tunnel.

The system overall structure can be seen in Figure 5.2.

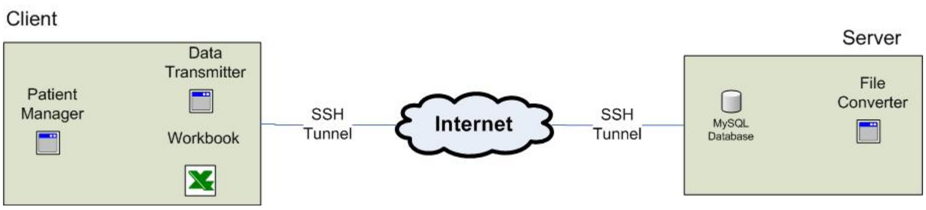


Figure 5.2: System Overall Structure

5.2.2 The Client Workbook

For the data collection part of the system, the first step was to have virtual interviews with the specialist on the hospital for gathering the important data necessary for a regular diagnosis. Also, they shared with us the interface the clinicians use to input the patient data. As they used Excel files to enter the patient data, we created our input interface using Excel and Visual Basic for Applications, as depicted in Figure 5.3.

The fields present on this workbook are described in Table 5.1. Fields 5, 9, 13, 14, 15 and 16 are pull-down inputs. Fields 17, 18 and 19 are button inputs (opening the default audio recording program for recording or playing auscultations, Figure 5.4). Field 20 is an internal code created to generate a file

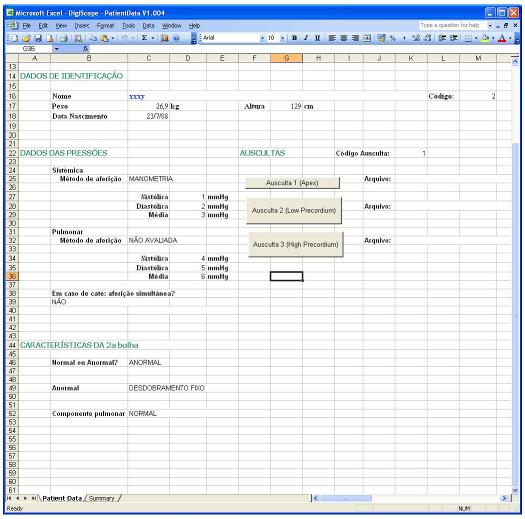


Figure 5.3: The workbook where the clinicians fill patient data and auscultations.

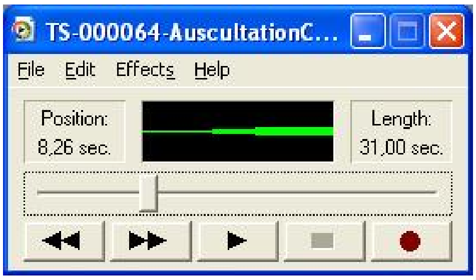


Figure 5.4: Audio recorder and player for collecting the auscultations

name for the auscultation. Field 21 is a code generated when the patient data is sent to our server. This code is generated on our servers and will be used later on to refer to that patient: this is a unique identifier for the patient on the research unit.

5.2.2.1 The client tools

Creating a workbook with the necessary fields and capable of recording and playing the auscultations for each patient is just one part of the data collecting process. We need a software component capable of anonymizing the patient information and send it throughout an SSH tunnel to the server, located at the University of Porto. The software created for this specific task can be run in two modes:

- As a scheduled system task: at a configured time, the program runs silently in the background, and then, transmits the data.

1-Name	12-Pulmonary average pressure
2-Weight	13-Simultaneous catheterism
3-Height	14-Second heart sound is normal
4-Birth Date	15-Abnormality in second heart sound
5-Systemic pressure measuring method	16-Pulmonary component status
6-Systemic systolic pressure	17-Apex auscultation
7-Systemic diastolic pressure	18-Low precordium auscultation
8-Systemic average pressure	19-High precordium auscultation
9-Pulmonary pressure measuring method	20-Auscultation code
10-Pulmonary systolic pressure	21-Patient code
11-Pulmonary diastolic pressure	

Table 5.1: Fields on the workbook

- As a GUI-based program: the user executes the program, and can see all the patients, and their status (sent or not sent). Also, at any given time, the user can start a transmission. The program interface is shown in Figure 5.5a.

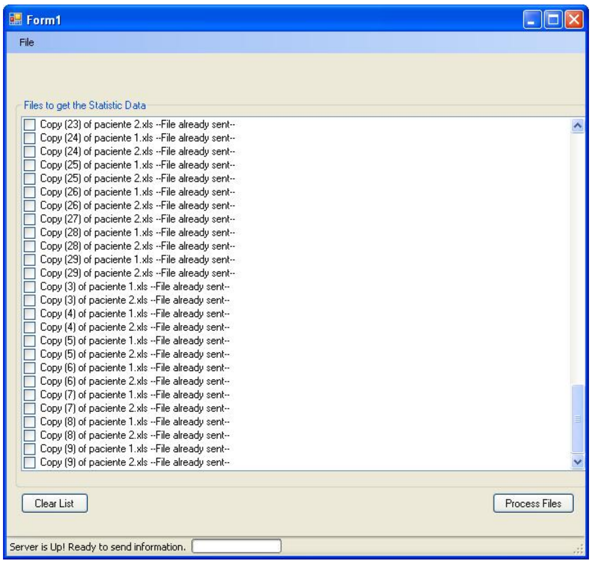
On both modes, the program checks for unsent patients and exams, connects to the server, sends the data, and, after checking the transmission, gets the patient code generated on the server, fills the code on the workbook and then, marks that patient as sent.

A module for managing the patients workbooks and allowing the user to easily create, edit, remove and check the patients codes was also created and is depicted in Figure 5.5b.

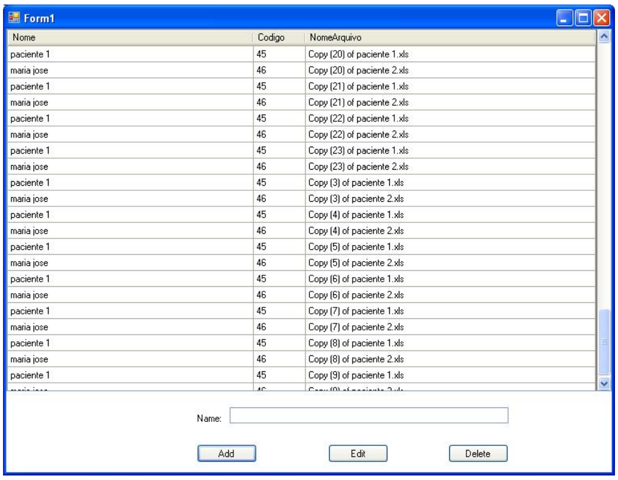
The creation of a manual audio annotation tool is a demanding task, due to the high complexity involved in synchronizing the audio, video and keeping track of all changes made by the user. So, given the scope of this master thesis, we opted to use the software Audacity [58] version 1.3.9 beta: a popular open source tool, capable of audio annotation. The process of sending the audio annotation files is currently manual, but this automation will be incorporated in our system.

5.2.3 The Server

The database used was the MySQL database. It is a reliable open source software widely used by the industry. Our server at the University of Porto runs the



(a) The GUI for the transmitter program module



(b) The patient manager program module

Figure 5.5: The Client Tools

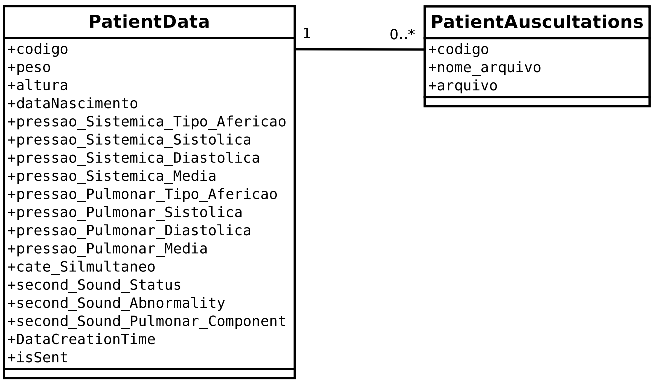


Figure 5.6: The database structure

database server 24h per day, so the clients might have maximum availability. To increase the security, the server only accepts connections from the SSH tunnel. The database structure is as depicted in Figure 5.6:

The 'PatientData' table is used to store patient textual information, as well as transmission data, useful to trace back eventual problems related to the data transmission.

The 'PatientAuscultations' table stores the auscultation files. The field 'nome_arquivo' has the original file name and the field 'arquivo' is a blob field with the file content itself. This structure is simple and flexible enough to allow the storage of different file types, in case we need to expand the program to collect different kinds of exams.

As a tool to be used on the server side, a program was developed to get the auscultations from the 'PatientAuscultation' table and re-create those files, so they can be used in our next step, the data processing phase of our research.

5.3 Results and Discussion

Until this moment, our data collection system has received 59 worksheets from different patients attended by the hospital's medical staff. Their ages range from 7 days to 15 years. A total of 96 auscultation files were received. During the development of this system, several difficulties were faced and overcome. This process taught us some important lessons about the development of software for clinical data gathering.

The first challenge was to choose which technologies we would use on our system. Instead of using new advanced technologies that our users are not familiar with, we decided that they are probably more productive if we use technologies that they are already familiar with as well as creating similar interfaces, so the learning phase can be as smooth and short as possible. This came from past experience with this same clinical team: They were used to write all the patient's ultrasound data and exams in separate Excel files. Our first attempt to create a unified repository for that data was to develop a stand alone application to perform such task. This was not successful since users were not familiar with the new interface and annotation concepts, hence they did not used the application properly and complaints were abundant. This approach was discarded and an Excel workbook with Visual Basic for Applications macros was created to collect data from all the patient's files and show them in a customizable way. That move itself made the users more comfortable in using the system and abated the error rate significantly, boosting productivity.

The second challenge we would like to address here is a technical difficulty not foreseen: some auscultations had a constant noise. After some study, we realize that this noise actually originated from the hospital's electric network: although the hospital had its own internal electrical sub-station, the power network was having too much interference. Once the auscultations where performed with the notebook unplugged from the power supply, the problem was solved.

On the field of hardware, we intend to test other type of stethoscope models that can transfer the auscultations to the computer using wireless technologies (such as Bluetooth or infrared). We expect that these stethoscopes could improve the signal quality and make the task of the recording auscultations easier.

Regarding the development of this system, we are planning to incorporate the audio annotation software with the patient record information. This feature is an important one that currently, our program does not have . With this feature incorporated, we could automate all relevant information about the patient and his auscultations.

Chapter 6

Heart Sound Segmentation

6.1 Introduction

As described in Chapter 2, when performing an auscultation, the clinician listens to the heart sounds, analyzes them and reaches his interpretation, which depends mainly on his experience and ability to recognize a wide variety of sound features (such as: hyperphonesis, fixed or variable splitting, ejection murmurs, etc.) and categorize these sounds and murmurs. Since this analysis is subjective, it is reasonable to assume that a computer program can help quantify these features and improve our ability to reach and reproduce a reliable diagnosis.

Looking more closely to the section regarding murmurs (section 2.3.1) we can clearly see that the clinicians categorize heart sounds based on their location on the cardiac cycle. The relative position of a heart sound component (S1, S2, etc.) or murmur, its shape, pinch, etc. may differentiate one heart disease or heart condition from another one. Therefore, we can conclude that, for a sensible heart sound diagnostic tool, its fundamental to have a good segmentation algorithm.

Our intention is to enable a digital stethoscope to have diagnosis capabilities, therefore, its algorithms should have low computational complexity. Also, according to our survey on algorithms for digital stethoscopes (Chapter 4), Liang's selected paper [38] is one of the most cited on heart sound segmentation, and has low computational complexity, therefore providing a good start point for our research on this field. Additional details on this work can be found here[39].

For these experiments, the authors used an electronic stethoscope developed on their laboratory[59], connected to a sound card using 16 bits resolution and sampling rate of 11025Hz. This is reasonable, since the heart sounds usually are lower than 1000Hz and have about 56% of their energy concentrated on the 50-60Hz range[29].

A total of 77 auscultations were recorded, with a duration of 6 to 13 seconds, from patients with pathological and physiological murmurs and ages from 0.4 to 15.4 years. The sounds where manually segmented by a pediatric cardiologist, who indicated the positions of S1 and S2.

An outline of the algorithm performed by this paper is given in the figure 6.1.

Throughout this chapter at the end of each subsection, I will present a small critical view of the decisions made by the paper's authors. At the end of this section I will give a critical precis about the paper as a whole. This will motivate the improvements proposed in this thesis for achieving robust HSS algorithms that can be embedded on current digital stethoscopes and are prepared for the uncontrolled clinical scenarios found in real hospitals.

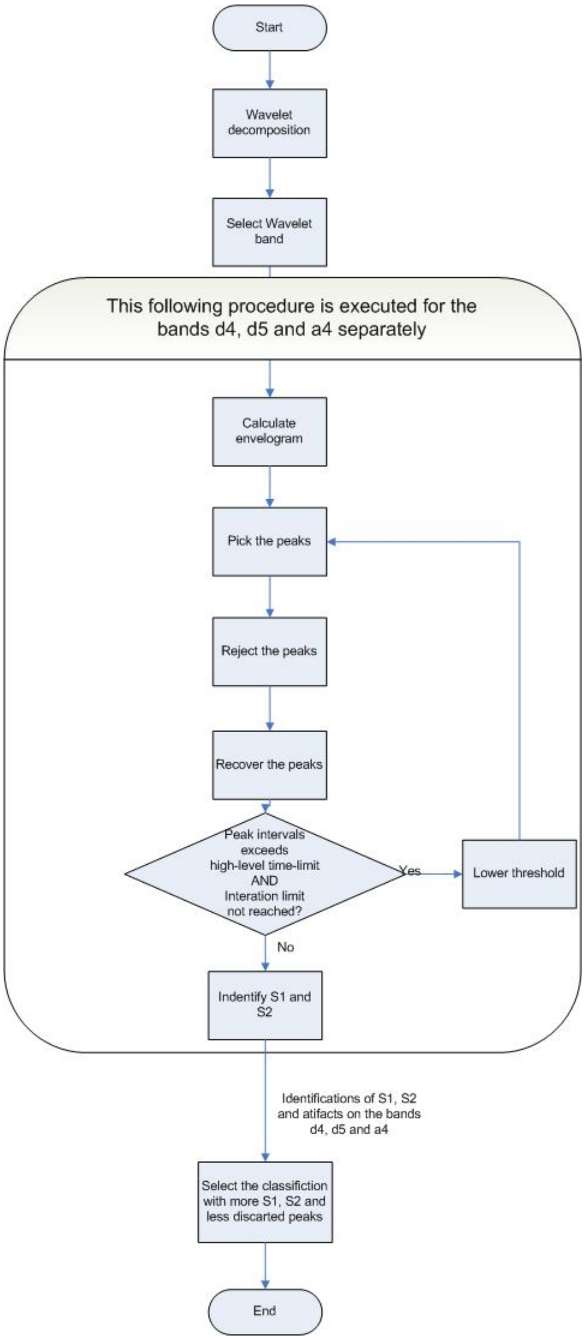


Figure 6.1: The outline of Liang's segmentation algorithm

Band	Frequency range in Hz
1st level detail	1102 to 2205
2nd level detail	551 to 1102
3rd level detail	275 to 551
4th level detail	138 to 275
5th level detail	69 to 138
4th level approximation	0 to 138
5th level approximation	0 to 64

Table 6.1: Details and approximations frequency range of the order six Daubecies wavelet

6.2 Fundamental stages of a HSS algorithm

6.2.1 Wavelet decomposition

First, the original signal is decimated down to 2205Hz. This operation does not lose any heart sound information, as we are still below the frequency of the majority of heart sounds (as seen in section 2.3.1 and A. Luisada[29]). A 6th order Daubechies discrete wavelet decomposition is used to obtain the bands used. The band's frequencies are shown in Table 6.2. One should notice that the frequency range on the detail bands is always the higher frequency of the previous band to half of that frequency. Detail 4, detail 5 and approximation 4 were selected for processing in the next steps.

The selection of these bands is reasonable given Table 2.1, the normal heart sounds frequency ranges [30, 29] and is compatible with the previous filters used for heart sound analysis by clinicians, mentioned on section 2.3.1, and by Lusiada [29] and Tavel [30].

6.2.2 Envelogram

Now its time to find an envelogram function that can lower the effect of low-amplitude noise and enforces the low intensity sounds.

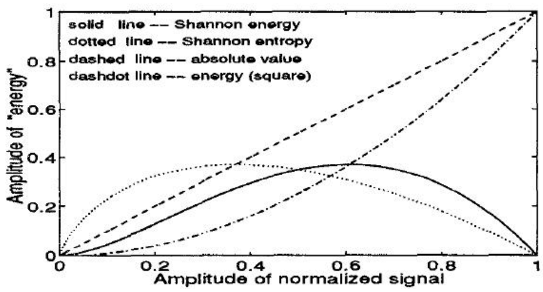


Figure 6.2: Different ways of calculating envelope (adapted from [39]).

The author compared different methods for calculating the envelogram (Figure 6.2), the definitions of each methods are described in Table 6.2.

Method name	Definition
Shannon energy	$E = -x^2 log(x^2)$
Shannon entropy	$E = -\|x\|log(\|x\|)$
Absolute value	$E = \|x\|$
energy (square)	$E = x^2$

Table 6.2: The equations for the different methods used to calculate the envelope

The energy (square) strengthens the high-intensity sounds, and weakens the medium and low-intensity ones. The Shannon energy strengthens the low-level sounds, therefore, increases the the low-level noise, generating a noisy envelope. The Shannon energy enforces the medium-intensity sounds and attenuates the low-level ones more than the high-level sounds, generating an envelope that is less noisy and capable of easier to find low-intensity sounds.

The average Shannon energy is then calculated using continuous segments of 0.02 seconds with an overlap of 0.01 seconds:

$$E_s = -\frac{1}{N}\sum_{i=1}^{N} x_{wBand}^2(i) log(x_{wBand}^2(i)) \tag{6.1}$$

Here, x_{wBand} is the selected wavelet band signal normalized and N is the number of samples representing 0.02 seconds, so, depending on the wavelet band we are using, the value of N varies. The normalized Shannon energy is calculated against the whole temporal axis:

$$P_a(t) = \frac{E_s(t) - \bar{M}(E_s(t))}{S(E_s(t))} \tag{6.2}$$

$M(E_s(t))$is the mean of $E_s(t)$ and $S(E_s(t))$ is the standard deviation of $E_s(t)$. From here on, the envelope $P_a(t)$ is used to find the peaks.

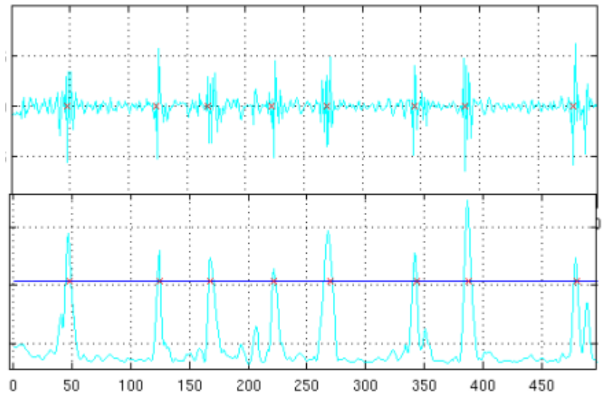


Figure 6.3: The wavelet band of the heart sound (upper) an the envelogram, the threshold and the peaks (lower)

In this step, the selection of a time-resolution of 0.01 seconds for the envelope is compatible with the report of Lusiada [29] regarding the human ear's perception: "The closest splitting that can be appreciated by the ear is that of two clicks made by the shutter of a camera at a 0.02 sec. interval (Johnston). (...) To be audible the interval between two components of the second sound is probably closer to 0.03 seconds". Therefore , the adopted time-resolution of 0.01 seconds seems reasonable for the purpose of heart sound segmentation.

6.2.3 Peak Picking

6.2.3.1 Picking the Peaks

After calculating the envelope, now a threshold is defined: an arbitrary value for selecting the peaks to pick. In our interpretation of Liang's paper[39], this initial value equals:

$$threshold = min(P_a) + \frac{max(P_a) - min(P_a)}{2} \tag{6.3}$$

Here, for each overshoot (a set of continuous values that are higher than the threshold), only one peak is picked per overshoot. A peak, by definition is a point (or set of continuous points with equal value) with the highest value inside a neighborhood. If an overshoot has more than a peak, then the first peak is chosen. In this step, peaks that represent noise are also picked up, but they should be discarded on the following steps.

An overshoot with more than one peak may represent a splitted heart sound, or noise. This situation should be detected on the following steps, with the analysis the statistical and energetic features of the peaks.

6.2.3.2 Rejecting the Peaks

On this step, peaks that may not belong to heart sounds are analyzed, detected and rejected.

First, the mean and standard deviation of peak intervals is calculated. Then, based on these values, the high-time limit and low-time limit values are computed. If a peak interval falls below the low time-limit, then one peak must be discarded.

If the interval is lower than 50ms, it is assumed that it is due to a splitted heart sound. In this case the first peak is kept and the second discarded if its energy is "not too small compared to that of the second one"[39], otherwise we get the second peak and discard the first one. In our implementation of this algorithm, we chose an energy difference greater than 90% to be "not too small".

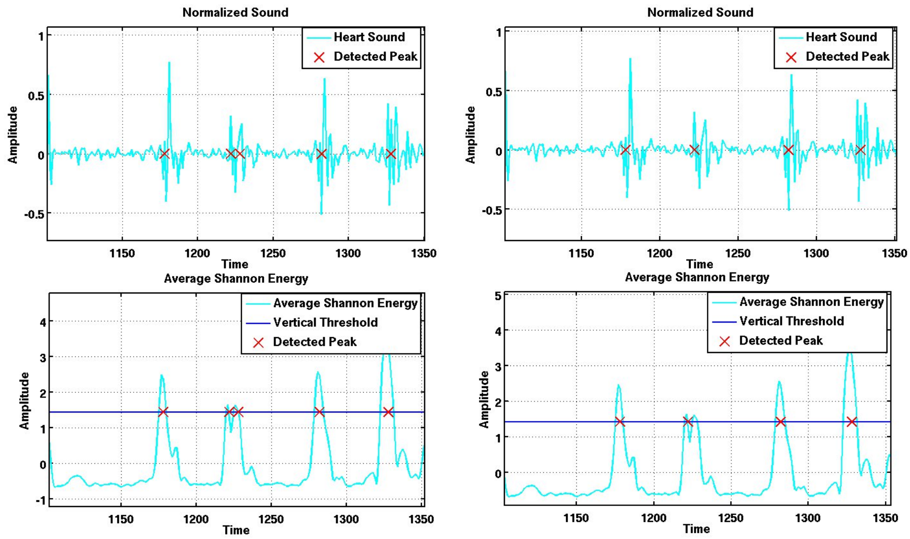
If the interval is between 50ms and the low-time limit, the peak's energies are compared: if the first peak has higher energy and is consistent (his variation with every second interval is within a certain variation), then this peak is kept and the second one rejected. Otherwise, the first peak is rejected and the second one kept. This procedure is depicted in the following pseudo code:

```
if (interval < lowerTimeLimit)
    if (interval<=50ms)
        if abs(peak(i)- peak(i+1)) <= 0.1*peak(i)
            reject peak(i+1)
        else
            reject peak(i)
        endif
    else
        if (peak(i) > peak(i+1))
           AND (consistency(peak(i-1),peak(i))
               reject peak(i+1)
            else
               reject peak(i)
            endif
    endif
endif
```

The function 'consistency' verifies if the interval peak(i-1), peak(i) is consistent with all second intervals. An example of this situation is observed in Figure 6.4. The picture shows a splitted heart sound and its rejection by the described procedure.

For the sake of clarity, it must be said that there are various details regarding this step that are not clear in the original paper[39, 40]. One is how the low and high-time limit are calculated. Also, the parameter that defines if an energy difference between two peaks is “too small” is loosely defined and is by no means specified by the author, as well as the tolerance of the variation of every second interval, performed by the method 'consistency' described on the above pseudo code.

The process of rejecting peaks makes the time-intervals wider, so we are eliminating the peaks that were detected excessively and, by rejecting them, we are widening the peak's intervals. Regarding the splitted heart sounds, we can see in [60], the wider splitted heart sound time is 50ms, so the decision of considering these two peaks has a physiologic justification. Also, the energy comparison between peaks below the low time-limit is a sensible way of verifying if we are dealing with a sound that belongs to another sound class (or even noise) or with a heart sound (S1 or S2).

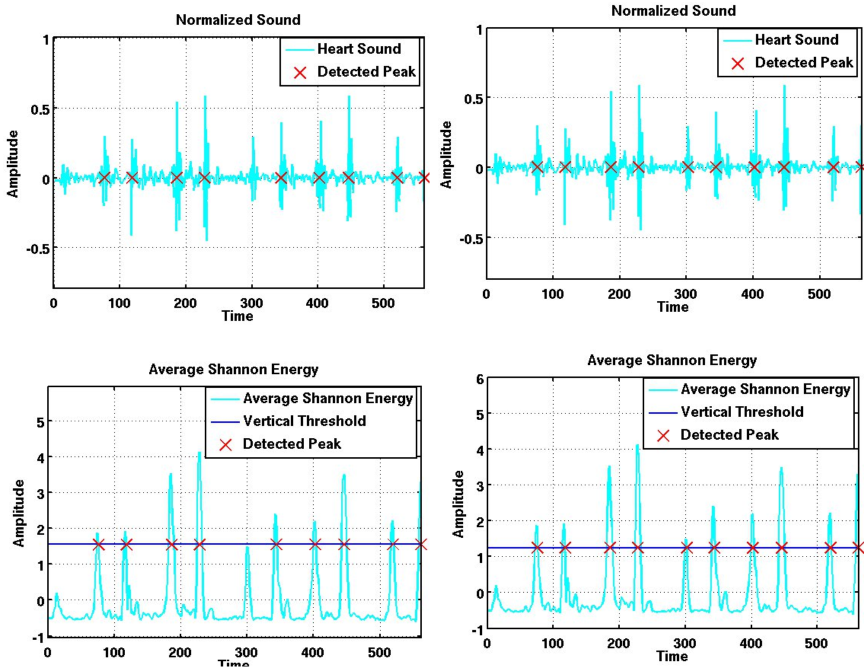


(a) (b)

Figure 6.4: A splitted heart sound with two peaks. In (a) we have two peaks detected between 1200 and 1250, in (b) one of the peaks is discarted.

6.2.3.3 Recovering the Peaks

If there are intervals that are higher than the high-time limit, it means that some peaks were missed (Figure 6.5). Usually the first heart sound (S1) has lower energy and is more commonly missed. To recover them, the threshold is lowered by a certain amount and the procedure is repeated from the Section 6.3, except that the new used threshold is equal to the current one decreased by a certain amount. This process stops either when all peaks are lower than the high time-limit or an iteration limit is reached. When this procedure is finished by putting all peaks below the high time-limit, it means that the intervals too short to be identified as heart sounds were eliminated and those S1's with low energy were found. Now the intervals should be either heart sounds or artifacts.



(a) (b)

Figure 6.5: A weak heart sound not reached by the threshold (on time 300). In (a) the threshold is too high to detect the heart sound, in (b) the threshold is lowered and the heart sound is correctly detected.

6.2.4 Cardiac Cycle Classification

This step will identify the intervals that are systoles, diastoles and the peaks that are neither S1 nor S2, therefore, considered artifacts.

The author makes his classification assuming the following:

- The systole is the transition interval from S1 to S2.
- The diastole is the transition interval from S2 to S1.
- The diastolic intervals are longer than the systolic ones.
- The systolic intervals are more constant compared to the diastolic ones.

The authors assume that, since the diastolic intervals are the longest on the heart cycle, the widest interval on the recording is assumed to be a diastole. Then, the authors create two parameters: $c1$ and $c2$. The parameter $c2$ represents the tolerance on the diastoles intervals (this variable is set empirically). In this implementation, we assumed that:

$$Diastole_{min} = \frac{Diastole_{max}(1 - c2)}{1 + c2} \tag{6.4}$$

Where $Diastole_{min}$ is the calculated minimum interval of a diastole, given our tolerance, $Diastole_{max}$ is the detected widest interval.

Although the paper does not state explicitly how the systole min and max times are calculated ($Systoles_{min}$ and $Systoles_{max}$ respectively), according to our understanding, we assumed that:

$$Systoles_{mean} + Diastoles_{mean} = 2 * mean(intervals) \tag{6.5}$$

Based on that, we can calculate the $Systole_{mean}$ and consequently, the minimum and maximum systole times, based on the tolerance parameter $c1$, also empirical.

Now we have both the minimum and maximum systoles and diastoles times, we can proceed with the classification of the intervals beginning from the widest one to the end of the auscultation, then from the widest one to the beginning of the auscultation. The intervals that destroys the consistency (two consecutive systoles or diastoles, intervals that are not inside the time-limits either of systole nor diastole), are marked as artifacts. The other ones, are marked as either S1 or S2 (Figure 6.6).

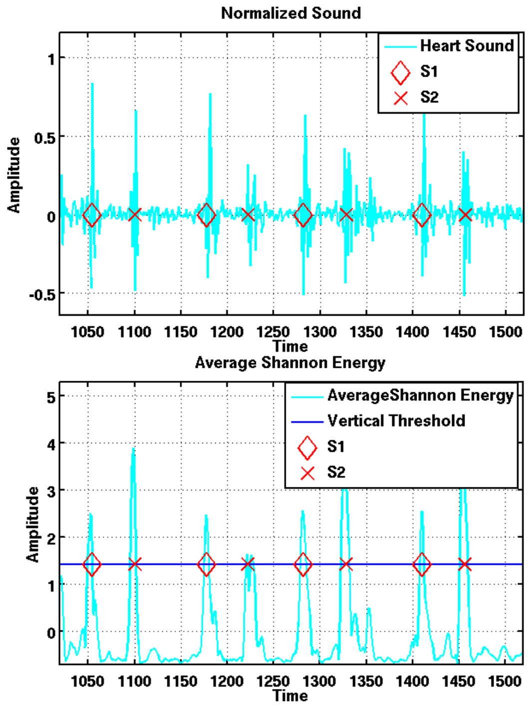


Figure 6.6: Identifying S1 and S2

6.2.5 Materials

A total of 96 auscultations collected by us, from 59 different patients. The algorithm was able to filter about 25 auscultations, from 19 different patients. Their ages ranges from 1 month to 288 months (24 years), with the average age of about 92 months (7,7 years). All the patients where described as having a normal auscultation (first and second heart sound normal, physiologic splitting of the second heart sound) and only one patient with second heart sound fixed splitting was reported. The auscultations were perfomed by the hospital's medical staff, in a regular office in the hospital, as part of regular office visit. A table with the patient's description can be found in Appendix A, Table 1. Most of our data has noise typical from this kind of environment, since we made the auscultations in a real non-controled hospital environment. Our intention is to make an analysis of how the state-of-the-art algorithm performs in such an adverse environment.

6.2.6 Results and Discussion

We compared the segmentation results from all the bands (d5, d5 and a4). Like in the original paper, we chose the band with more identified S1 and S2 and less discarted peaks. This method reached an accuracy of about 49% using our data set. An outline of these results can be seen in the Table 6.3

	Number of cycles	Percentage (%)
Correct	839	49.32
Missed	273	16.05
Incorrect	589	34.63
Total	1701	100

Table 6.3: Segmentation results

It was found that the algorithm could not cope with several auscultation recordings. A closer analysis showed that these were situations that were either too noisy even for a clinician (paediatric auscultation is particularly difficult: the patient may cry, move himself, etc.) or they where poorly recorded (either due to the operator inexperience, environment background noise, etc.). One must be aware that this is common in real environments so this type of algorithms must be prepared for them. As such, the task of filtering these auscultations needs to be investigated more closely in the future.

This paper has some good ideas on several points: the choice of Normalized Shannon Energy to calculate the Envelope seems to be a proper choice: It tends to emphasize heart sounds and has a good time resolution. Also, the 50ms rule for rejecting the peaks has a physiological justification, as well as the author's analysis of the variances of the systole and diastole time-intervals in a recording.

On the other hand, the paper is not quite clear on some methodological details. Some variables, like the method for calculating the average systole, are not mentioned, as well as the low-level and high-level time-limits. Also, the energy-difference by which the peaks are considered 'too small' is not defined. Some points could be improved, like the procedure for identification of S1 and S2 (the widest interval in the auscultation is assumed to be a diastole - This may not be true: the clinician may take out the stethoscope from the patients chest and then put it back again), and the calculation of the high and low-time limit (The author calculates the peaks intervals as if they belongs to an unique class, but on step 6.2.4, he uses two classes).

On the next section we present an improved method for classification of the cardiac cycle.

6.3 Improved Cardiac Cycle Classification

Accordingly to our experience, Liang's method for identifying S1 and S2 has the potential to be efficient in a controlled environment. However, our experience also shows that clinical environments are far from controlled (absence of silence, uncooperative patients, variety of potentially conflicting equipment in environments were auscultation is required). Furthermore, the algorithm presents a number of user-defined parameters that can interfere with the system's robustness. Motivated by that, we propose an improvement to boost the performance of Liang's computationally simple method for HSS.

More specifically we aim at an improved method for systole and diastole identification. By identifying intervals between peaks into systole and diastole, we are consequently, already identifying the peaks as S1 and S2.

For this classification, we will exploit some important physiological features and assumptions about the cardiac cycle:

- The systole is defined as the interval between S1 and S2, it is shorter than the diastole one, and has a lower standard deviation.
- The diastole is the interval between S2 and S1, it is larger than systole, and has a greater standard deviation.

Since the auscultation is composed mostly by systoles and diastoles, it is acceptable to assume that half of the cardiac intervals are systoles and half are diastoles.

Our first step is to calculate the median of the intervals. We assume that the intervals below or equal to the median are systoles and above the median are diastoles. An example of a histogram of an auscultation is given in Figure 6.7.

Then the average systole and diastole are calculated, as well as the minimum and maximum intervals:

$$Systoles_{min} = mean(SysIntervals) - std(SysIntervals) * SysTol \tag{6.6}$$

$$Systoles_{max} = mean(SysIntervals) + std(SysIntervals) * SysTol \tag{6.7}$$

Here the $std()$ is the function to calculate the standard deviation, $mean()$ calculates the average and $SysTol$ is the tolerance of our systolic interval, in standard deviations.

Similarly, we calculate the minimum and maximum diastole intervals.

In our implementation, we chose to include 1 standard deviation (68.8%) of the data. We are aware that this should be investigated more carefully in the future.

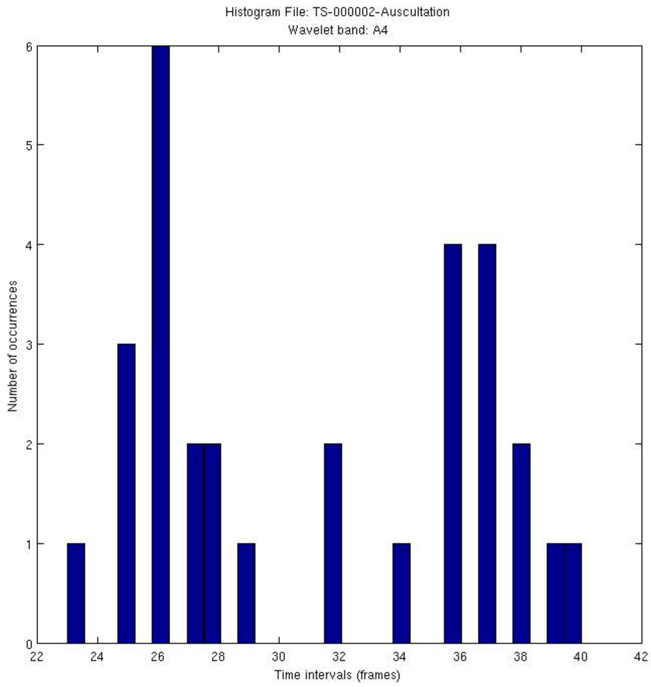


Figure 6.7: Example of an auscultation's histogram. The median in this case is 30.5 frames.

6.3.1 Results

To test this algorithm, we used the same dataset and decision criteria as before. Results comparing the performances are shown in Table 6.4.,

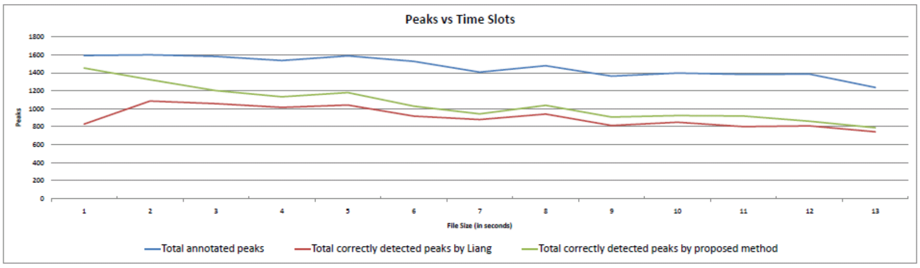
	Improved method Number of cycles (Percentage)	State of the art method Number of cycles (Percentage)
Correct	1052 (61.85%)	839 (49.32%)
Missed	279 (16.40%)	273 (16.05%)
Incorrect	370 (21.75%)	589 (34.63%)
Total	1701 (100%)	1701 (100%)

Table 6.4: Improved Segmentation results

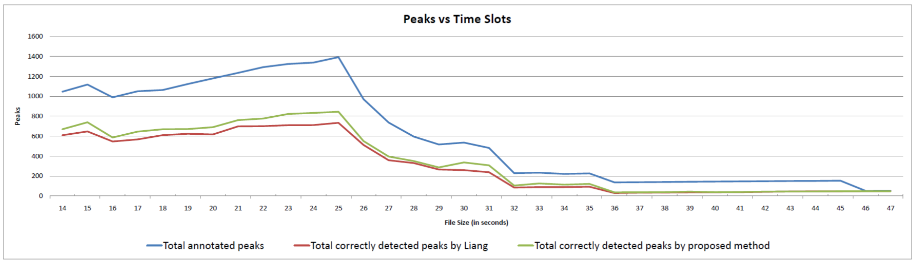
Results show that our algorithm performs better when faced with noisy auscultations obtained from real clinical environments, strengthening the validity of our improved analysis of the temporal distributions of the peaks and higher robustness to recordings with silence intervals. We also studied how both algorithms performs with different time-windows. In order to do so, we took all the auscultations, and divided them into smaller files with different sizes, from

1 second of duration to 47 seconds (the size of the biggest recording in our set), therefore, creating 47 sets, for each time-size (Figure 6.8).

It was expected that the improved algorithm would outperforms significantly in the longest sets. It was observed that, on average, the developed algorithm outperforms the state-of -the-art in almost all sets. However, in the sets that are longer than 25 seconds, the amount of files drastically decreases due to the fact that only a few recordings have such a long duration. As such, one must be careful with the conclusions drawn from this data. On the sets with the same time-range used on Liang's paper[40] (6 to 13 seconds), we can observe that the proposed algorithm outperforms in all sets (Figure 6.9).



(a)



(b)

Figure 6.8: Comparison between the two methods: accuracy vs file size (in seconds)

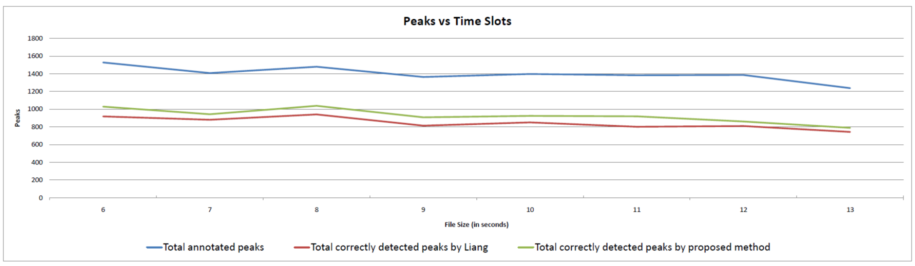
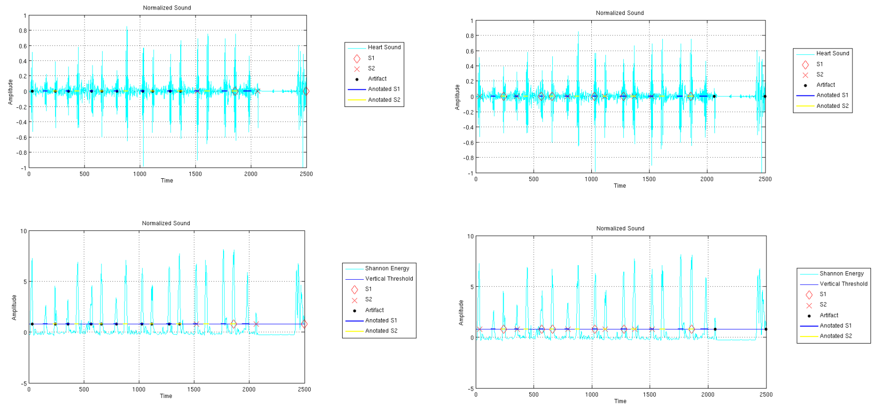


Figure 6.9: Performance of the two methods for Liang's temporal size range.

6.3.2 Discussion

The developed method has several advantages against the one discussed previously. The only parameters it does have are how many standard deviation values we want to choose for defining the duration of the systole and diastole. Also, it does not assume beforehand any classification for any particular interval. This means that even if the clinician lifts the stethoscope from the patent's chest, we can still perform HSS (Figure 6.10).



(a) (b)

Figure 6.10: Example of an auscultation with silence segmented by both algorithms: In (a) the state-of-the-art and in (b) the proposed method.

As we can see from Table 6.4, the developed method shows better accuracy than the current state-of-the-art for very noisy environments. Also by comparing the performance of both algorithms regarding the auscultation size in seconds (we used time-slots from 1 second to 47 seconds), on average, the improved method also outperforms the studied paper using an equally computationally light algorithm. We can conclude that being aware of the difficulties presented by real scenarios, and attempting to reduce the dependence of algorithms on fixed parameters are two key factors for having robust audio processing methods that can be embedded in digital stethoscopes in the future. Of course that for obtaining more solid conclusions we need not only to collect more data but to intensively research this type of algorithms. This was too ambitious for the work presented in this thesis and is planned as future work.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Based on our experience acquired throughout the development of this thesis, we can reach various conclusions.

On Chapter 4, we realized that, in spite of the great interest on this field by clinicians and engineers, most of the developed techniques are not yet suitable for real clinical environments. Most lack precise statistical descriptions and a vital clinical validation step. We strongly believe that a multidisciplinary team, where the components have a broader knowledge of the different fields (computer science, statistics, medicine) can greatly improve the quality of the research.

On Chapter 5, we developed an extensible data gathering software system capable of collecting clinical data, auscultations and transmit them anonymously from a remote hospital to our server. In this process, on the software perspective, we realized that our users felt more comfortable and productive when we designed the system to be as similar as possible to the systems they use daily. We could also improve user productivity and shorten the development time by using popular and widely accepted programs (Microsoft Excel, Windows Recorder) for the user interface, and free and open source software (Mysql, C#, openSSH) for data processing.

On the hardware perspective, we experienced that even a stethoscope connected to a notebook does not provide enough mobility and easy of use as we first thought. We believe that some difficulties still need to be addressed, such as: easy of use (i. e.: few buttons, less or even no adjustments, etc.), and improved mobility (absence of wires, battery autonomy, equipment's weight, memory, etc.). Once these features are implemented, the clinician would not be limited to a particular site to collect auscultations and the quality of data would be more robust, since now we have to deal mainly with acoustic noise itself.

Regarding Chapter 6, we improved the algorithms proposed by Liang et al.[39, 38]. Their research had some good ideas such as the envelogram for detecting the peaks but also had some room for improvement as the filtering performed by the algorithms is not suitable for real clinical noisy environments (only 34% of the collected auscultations were useful for the algorithm). We also realized that the pick-peaking process could be improved, since it does not analyze the frequency features of the peaks, and puts two statistically different components (S1 and S2) into the same statistical measure to calculate the low and high-time limit. Also, it has quite a large number of static parameters that need to be tuned. The difference between our results and the results obtained by Liang et al. can be due to the fact that their experiments seem to have been performed on a very controlled environment (the situation the authors report as reasons for incorrect classification, such as speech, crying, etc. are common on the uncontrolled environment of clinical scenarios).

7.2 Future Work

There are several areas to be explored and improved after this work.

On the field of hardware, other type of stethoscope models, such as the 3M Littmann electronic stethoscopes should be tested. These stethoscopes transfer the auscultations to the PC via infrared or Bluetooth, possibly increasing the signal quality and easing the recording process. Another possible future work would be to create and incorporate an audio annotation tool for the data gathering system, so all the software's components would be fully integrated in a single system's module, sharing all the transmission, storage and retrieval capabilities of the developed system.

Regarding heart sound segmentation, our work can be extended in four fronts:

1. Collect more auscultations, so we can strengthen the statistical analysis.

2. Development of a method for heart sound filtering: Our experience shows that regular filtering is not enough for efficiently filtering heart sound auscultations in a real clinical environment. Nearby conversations and patent's cries are among the commonest unfiltered noise found in our experiments.

3. Peak-Picking: Here there is room for some improvements, such as adding information from frequency signature of S1 and S2, and improving the statistical analysis of the selected peaks.

4. Heart cycle classification: Literature shows a variety of techniques that could improve the identification of S1 and S2, such as Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs)[61].

Also, as the sharper reader has noticed, from the medical description of heart sounds and murmurs, it is important to segment not only S1 and S2, but also A2 and P2. Our survey, on the other hand could not find any record of an automated method for this kind of segmentation. This would be a natural next step of the work here presented.

Appendix A

This section describes the data annotated by the clinicians regarding the patients whose auscultations where segmented by the algorithms presented in this thesis.

Age (months)	Weight (Kg)	Height (m)	1st and 2nd Heart Sound Remarks	Number of Auscultations
12	1	0.2	Normal, fixed splitting	1
108	10	0.9	Normal	2
-	-	-	Normal	1
-	-	-	Normal	1
-	14.8	0.87	Normal	1
60	19.8	1.11	Normal	2
60	21	1.15	Normal	2
96	30	1.26	Normal	1
132	43	1.49	Normal	2
8	3.6	0.485	Normal	1
288	48	1.63	Normal	1
84	-	-	Normal	1
-	62.5	1.86	Normal	1
-	23	1.19	Normal	2
48	-	-	Normal	2
264	30	1.38	Normal	1
156	52	1.63	Normal	1
1	3.7	0.52	Normal	1
168	54	1.69	Normal	1

Table 1: Description of the patients whose auscultations where segmented by the algorithm in chapter 6.

Bibliography

[1] V. A. McKusick, *Cardiovascular sound in Heart and Disease*. Baltimore: Wiliams & Wilkins, Jan. 1958, the History of Cardiovascular Sound.

[2] J. Forbes, *A Treatise on the Disease of the Chest*. Underwood, 1821.

[3] B. G. E. and D. Pasquale, *A history of electrocardiography*. Norman Publishing, 1990.

[4] M. E. Tavel, *Clinical Phonocardiography and External Pulse Recording*. Year Book Medical Publishers, 1972.

[5] O. Weiss and G. Joachin, "Registrierung von hertzönen und herzgeräuschen mittels des phonoskops und ihre beziehungen zum elektrokardiogramm," *Ztschr. klin. Med*, vol. 73, p. 240, 1911.

[6] B. Randell, *The COLOSSUS*. University of Newcastle upon Tyne, Computing Laboratory, 1976.

[7] L. Floridi, *Philosophy and computing: An introduction*. Routledge, 1999.

[8] E. H. Shortliffe, Ed., *Biomedical Informatics - Computer Applications in Health Care and Biomedicine*. Springer, 2006.

[9] D. W. Simborg and M. Chadwick, "Local area networks and the hospital." *Computers and Biomedical Research*, vol. 16(3), pp. 247–259, 1983.

[10] Hl7 home. [Online]. Available: http://www.hl7.org/about/index.cfm

[11] E. Berner, D. Detmer, and D. Simborg, "Will the wave finally break? a brief view of the adoption of electronic medical records in the united states:," *Journal of American Medical Informatics Association*, vol. 12, pp. 3–7, Feb 2005.

[12] R. Miller, M. McNeil, S. Challinor, F. Masarie, and J. Myers, "The internist-1/quick medical reference project-status report," *Western Journal Med.*, vol. 145, pp. 816–22, 1986.

[13] G. Barnett, J. Cimino, J. Hupp, and E. Hoffer, "Dxplain - an evolving diagnostic decision-support system," *JAMA*, vol. 258, pp. 67–74, 1987.

[14] W. H. Jr, "Iliad: moving medical decision-making into new frontiers." *Methods Inf Med*, vol. 28, pp. 370–2, 1989.

[15] C. McDonald, S. Hui, D. Smith, W. Tierney, S. Cohen, M. Weinberger, and G. McCabe, "Reminders to physicians from an introspective computer medical record. a two-year randomized trial." *Ann. Intern. Med.*, vol. 100, pp. 130–8, 1984.

[16] M. Korcok, "Computer diagnostics: Technology of the future," *Can. Med. Assoc. J.*, vol. 133, pp. 231–235, 1985.

[17] G. Lodwick, C. Haun, W. Smith, R. Keller, and E. Robertson, "Computer diagnosis of primary bone tumors: a preliminary report," *Radiology*, vol. 80, no. 2, pp. 273–275, 1963.

[18] P. Meyers, C. Nice Jr, H. Becker, W. Nettleton Jr, J. Sweeney, and G. Meckstroth, "Automated Computer Analysis of Radiographic Images." *Radiology*, vol. 83, p. 1029, 1964.

[19] F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," *Radiology*, vol. 89, no. 2, pp. 211–215, 1967.

[20] R. Kruger, J. Townes, D. Hall, S. Dwyer, and G. Lodwick, "Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors," *IEEE Transactions on Biomedical Engineering*, pp. 174–186, 1972.

[21] R. Kruger, W. Thompson, and A. Turner, "Computer diagnosis of pneumoconiosis," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 4, no. 1, pp. 40–49, 1974.

[22] J. Toriwaki, Y. Suenaga, T. Negoro, and T. Fukumura, "Pattern recognition of chest X-ray images," *Comput. Graph. Image Processing*, vol. 2, pp. 252–271, 1973.

[23] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.

[24] E. RL., "Attempt to use computers as diagnostic aids in medical decision making: a 30-year experience." *Perspectives in Biology and Medicine*, vol. 35, pp. 207–218, 1992.

[25] R. Miller, "Medical diagnostic decision support systems–past, present, and future: a threaded bibliography and brief commentary." *Journal of the American Medical Informatics Association*, vol. 1, no. 1, p. 8, 1994.

[26] G. Borden and K. Harris, *Speech science primer: physiology, acoustics, and perception of speech*, Baltimore, Ed. Williams & Wilkins, 1994.

[27] e. a. Paul A. Iazzo, *Handbook of Cardiac Anatomy, Physiology, and Devices*, P. A. Iazzo, Ed. Humana Press Totowa, New Jersey, 2005.

[28] W. G. C. S. Germann, *Principles of Human Physiology.* Pearson Education/Benjamin Cummings, 2002.

[29] A. A. Luisada, *From Auscultation to phonocardiograpy.* The C. V. Mosby Company, 1965.

[30] M. E. Tavel, *Clinical phonocardiography and external pulse recording.* Year Book Medical Publishers, 1972.

[31] P. S. Addison, *The Illustrated Wavelet Transform Handbook.* Institute of Physics Publishing Bristol and Philadelphia, 2002.

[32] M. W. Frazier, *An Introduction to Wavelets Through Linear Algebra.* Springer, 1999.

[33] L.-G. Durand and P. Pibarot, "Digital signal processing of the phonocardiogram: Review of the most recent advancements," *Critical Reviews in Biomedical Engineering*, 1995.

[34] M. Bredesen and E. Schmerler, "Intelligent stethoscope," US Patent 5,010,889, 1991.

[35] F. Hedayioglu, S. Mattos, L. Moser, and M. de Lima, "Development of a tele-stethoscope and it's application in pediatric cardiology," *Indian Journal of Experimental Biology*, vol. 45, 2007.

[36] D. B. M.E. Tavel and D. Shander, "Enhanced auscultation with a new graphic display system," vol. 154, p. 893, 1994.

[37] M. Brusco and H. Nazeran, "Development of an intelligent pda-based wearable digital phonocardiograph," in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Sepember 2005.

[38] S. L. H. Liang and I. Hartimo, "A heart sound segmentation algorithm using wavelet decomposition and reconstruction," in *19th International Conference - IEEE/EMBS*, Chicago, IL, USA, Oct., Nov. 1997.

[39] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," *Computers in Cardiology*, vol. 24, 1997.

[40] H. Liang and I. Hartimo, "A heart sound feature extraction algorithm based on wavelet decomposition and reconstruction," in *Proc. IEEE EMBS*, vol. 20, no. 3, 1998.

[41] M. T. Sherif Omran, "A heart sound segmentation and feature extraction algorithm using wavelet," in *Proc. of IEEE MWSCAS '03*, vol. 1, 2003, pp. 27–30.

[42] P. P. JingPing Xu, L.G. Durand, "Nonlinear transient chirp signal modeling of the aortic and pulmonary components of the secound heart sound," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, March 2000.

[43] J. Xu, L. Durand, and P. Pibarot, "Extraction of the aortic and pulmonary components of the secound heart sound using nonlinear transient chirp signal model," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 3, March 2001.

[44] V. Nigam and R. Priemer, "A procedure to extract the aortic and the pulmonary sounds from the phonocardiogram," in *Proceedings of the 28th IEEE EMBS Annual International Conference*, New York City, USA, Aug 30-Sep 3 2006.

[45] P. M. Bentley, J. T. E. McDonnell, and P. M. Grant, "Classification of native heart valve sounds using the choi-williams time-frequency distribution," *Biomedical Engineering, IEEE Transactions on*, 1995.

[46] P. M. Bentley, P. M. Grant, and J. T. E. McDonnell, "Time-frequency and time-scale techniques for the classification of native and bioprosthetic heart valve sounds," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, January 1998.

[47] P. Wang, Y. Kim, and C. B. Soh, "Feature extraction based on mel-scaled wavelet transform for heart sound analysis," in *Engineering in Medicine and Biology Society*, 2005.

[48] H. Liang and I. Hartimo, "A feature extraction algorithm based on wavelet packet decomposition for heart sound signals," in *Proceedings of the IEEE-SP International Symposium*, October 1998.

[49] N. H. Onsy Abdel-Alim and M. A. El-Hanjouri, "Heart diseases diagnosis using heart sounds," in *Radio Science Conference*, March 2002.

[50] I. Turkoglu and A. Arslan, "An intelligent pattern recognition system based on neural network and wavelet decomposition for interpretation of heart sounds," in *Proceedings of the 23rd Annual International Conference of the IEEE*, vol. 2, October 2001, pp. 25–28.

[51] Z. D. Ozgur Say and T. Olmez, "Classification of heart sounds by using wavelet transform," in *Proceedings of the Second Joint EMBS/BMES Conference*, vol. 1, 2002.

[52] M. El-Hanjouri, W. Alkhaldi, N. Hamdy, and A. Alim, "Heart diseases diagnosis using hmm," in *Electrotechnical Conference -MELECON*, 2002.

[53] T. Leung, P. White, W. B. Collins, E. Brown, and A. P. Salmon, "Classification of heart sounds using time-frequency method and artificial neural networks," in *Proceedings of the 22nd Annual International Conference of the IEEE*, vol. 2, July 2000.

[54] E. Kail, S. Khoor, B. Kail, K. Fugedi, and F. Balazs, "Internet digital phonocardiography in clinical settings and in population screening," *Computers in Cardiology*, 2004.

[55] Z. Sharif, M. S. Zainal, A. Z. Sha'ameri, and S. H. S. Salleh, "Analysis and classification of heart sounds and murmurs based on the instantaneous energy and frequency estimations," in *Proceedings of TENCON*, vol. 2, 2000.

[56] F. C. M. M. S. Hedayioglu, "A survey of audio processing algorithms for digital stethoscopes," in *HealthInf*, 2009.

[57] (2009, 9). [Online]. Available: TightVNC: VNC-Based Free Remote Control Solution

[58] (2009, 09) Audacity: Free audio editor and recorder. [Online]. Available: http://audacity.sourceforge.net/

[59] S. Lukkarinen, P. Korhonen, A. Angerla, A. Noponen, K. Sikio, and R. Sepponen, "Multimedia personal computer based phonocardiography," in *Proceedings of the 18th Annual International Conference of the IEEE*, 1996.

[60] J. Hurst, *The Heart, Arteries and Veins: arteries and veins*. Glencoe/McGraw-Hill School Pub Co, 1990.

[61] L. Rabiner., "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 1989.