

# Computer Assisted Cardiac Auscultation: Probabilistic Modelling and Psychoacoustic Feature Extraction for Heart Sound Descriptions

Andries Meintjes



**AUT INSTITUTE OF  
BIOMEDICAL TECHNOLOGIES**

A thesis submitted to Auckland University of Technology in  
fulfilment of the requirements for the degree of  
Doctor of Philosophy (PhD)

2019

School of Engineering, Mathematics and Computational  
Sciences

## Abstract

Physicians have been using stethoscopes for over 200 years to listen to the sound produced by the heart, but the diagnostic accuracy of this practice has been called into question by studies that have found the clinical skills of doctors at all levels to be lacking. The development of electronic stethoscopes and advancements in the processing power of digital computers and the subsequent development of signal processing and machine learning methods has opened the door to the field of Computer Assisted Cardiac Auscultation (CACA). This field of study is concerned with increasing the diagnostic value of the heart sounds using computers and the myriad signal processing methods that these devices enable. This thesis represents an exploration of CACA from the viewpoint of probabilistic and psychoacoustic modelling.

Probability theory provides the framework with which we model the heart sounds, firstly, using an unsupervised machine learning method called Independent Component Analysis (ICA) and secondly, by expanding current work on the use of duration-dependent Hidden Markov Models (HMM). We also investigate heart sounds as perceptual phenomena using psychoacoustic models to arrive at descriptions of features of heart murmurs that correspond to those that an expert auscultator would listen for when auscultating. This enables the findings of the algorithm to be communicated in a form that is familiar and acceptable.

We present four case studies on the use of ICA in which the model can identify physiologically and diagnostically interesting features in heart cycles given an appropriately chosen number of sources. A probabilistic systolic murmur labelling model is developed as an expansion of previous work done in heart sound segmentation. The proposed algorithm achieves an F1-score of 93.6% compared to 90.6% achieved by the current state-of-the-art and can identify systolic murmurs with an area under the receiver-operator curve (AUC) of 0.90 as tested on a dataset of 56 heart sound recordings.

In the final part of the thesis, psychoacoustic models are developed for systolic murmurs. The perceptual qualities of 'loudness', 'pitch', and 'shape' are derived using psychoacoustic principles and compared to annotations made by expert auscultators. An online survey was developed and tested for the purpose of collecting expert annotations. The completion rate of the survey was 16%, perhaps in part due to the complex and time consuming nature of the task compounded by the online format of the survey. The collected responses show a percent agreement of 0.73 for 'loudness', 0.65 for 'pitch', and 0.35 for 'shape'. The proposed model showed strongest agreement with 'loudness' and some agreement with 'pitch', but there was little agreement on the 'shape' feature.

This thesis shows that the application of ICA, the explicit modelling of a systolic murmur state in heart sound segmentation, and models of psychoacoustic features increase the diagnostic value of heart sounds.

## Acknowledgements

Firstly, I would like to thank to my supervisors Andrew Lowe and Malcolm Legget for their support and guidance during this project. Andrew for his patience and encouragement when the project seemed unsurmountable, and the many inspiring conversations around concepts, often daunting, that helped me glean some understanding of what seemed completely opaque at first. Malcolm for the many meetings and discussions that provided such a unique and privileged view into the practice and theory of cardiology and medical sciences in general; meetings that have inspired and guided this work throughout. It would not be an overstatement to say that this work would not have been possible without them.

To my family, my parents, my brother and sister, and their families. Thank you for believing in me and encouraging me throughout the duration of this work. I believe I have been luckier than most being able to complete my studies in a foreign country while also being close to those that matter the most to me.

I would like to thank all the staff and students at both the Institute of Biomedical Technologies (IBTec) and the Sports Research Institute of New Zealand (SPRINZ). These two facilities have both acted as my home during these studies but would be lifeless buildings without the amazing people that inhabit them. Special thanks to Professor Ahmed Al-Jumaily from IBTec and Professor Patria Hume and Jono Neville from SPRINZ for the inspiration and sage advice that have helped me complete this project. Much appreciation also to Mr Tiff Day and his team at Marshall Day Acoustics for their sound advice and interest in this work.

And finally, to the friends and colleagues who have supported and stood by me during these studies. I thank you for crossing my path, especially during what has turned out to be a rather interesting (if not always pleasant) period of my life. To Ali and Saing for the conversations over coffee that have often carried me through difficult periods, Diana, Daniel, and Mia for their energy and presence, Melissa for introducing me to cheese curds and Daft Punk, and to Imi and Carine for all the inspiration and courage all the way through to the end.

## Glossary and Abbreviations

### Medical Terms

auscultation	The practice of listening to body sounds to assess the condition of internal organs.
cardiac	Related to the heart.
cardiovascular	Related to the heart and circulatory (blood) vessels.
fundamental heart sound	The heart sounds associated a normal functioning heart, sometimes referred to as 'lub-dub'.
systole	The period in which the ventricles of the heart are contracting, forcing blood out of the heart into the rest of the body and the lungs.
diastole	The period in which the ventricles of the heart are relaxing, allowing blood from lungs and the rest of the body to flow into the heart.

### Heart Sound Abbreviations

S1	The first fundamental heart sound, associated with the start of systole/end of diastole.
S2	The second fundamental heart sound, associated with the end of systole/start of diastole.
S3	The third heart sound/ventricular gallop.
S4	The fourth heart sound/atrial gallop.
M1	The part of the first heart sound associated with the closing of the <i>Mitral</i> valve.
T1	The part of the first heart sound associated with the closing of the <i>Tricuspid</i> valve.
A2	The part of the second heart sound associated with the closing of the <i>Aortic</i> valve.
P2	The part of the second heart sound associated with the close of the <i>Pulmonary</i> valve.
PCG	Phonocardiography, a specialist term for the recording of heart sounds and the resulting waveform

### Signal Processing Terms and Abbreviations

Time-frequency analysis	A decomposition of a signal into its time-varying frequency components
DSP	Digital signal processing. The manipulation of digital signals using computer algorithms.
SNR	Signal-to-noise ratio
STFT	Short Time Fourier Transform
WT	Wavelet transform
DWT	Discrete wavelet transform
CWT	Continuous wavelet transform

# Table of Contents

<b>Abstract .....</b>	<b>I</b>
<b>Acknowledgements .....</b>	<b>II</b>
<b>Glossary and Abbreviations .....</b>	<b>III</b>
<b>Table of Contents .....</b>	<b>IV</b>
<b>Attestation of Authorship .....</b>	<b>1</b>
<b>Preface .....</b>	<b>2</b>
<b>Chapter 1 Introduction.....</b>	<b>3</b>
<b>1.1 Rationale and significance .....</b>	<b>3</b>
<b>1.2 Outline of this thesis .....</b>	<b>7</b>
<b>Chapter 2 Auscultation of the Heart.....</b>	<b>9</b>
<b>2.1 Functional Cardiac Anatomy.....</b>	<b>9</b>
2.1.1 The Heart .....	9
2.1.2 The Heart Valves .....	10
<b>2.2 The Cardiac Cycle .....</b>	<b>12</b>
<b>2.3 Heart sounds.....</b>	<b>13</b>
2.3.1 Fundamental Heart sounds.....	13
2.3.2 Other Heart sounds.....	14
2.3.3 Murmurs .....	15
<b>2.4 Cardiac Auscultation .....</b>	<b>16</b>
2.4.1 Cardiac Auscultation in Context.....	17
2.4.2 Cardiac Auscultation Procedure.....	18
2.4.3 The Stethoscope .....	20
2.4.4 Recent Developments.....	22
<b>Chapter 3 Digital Signal Processing of Heart Sounds .....</b>	<b>24</b>
<b>3.1 Audio signals and Heart Sound Recordings .....</b>	<b>24</b>
<b>3.2 Digital Signal Processing Concepts.....</b>	<b>27</b>
3.2.1 Block-based Processing.....	28
3.2.2 The Fourier Transform .....	30
3.2.3 Short Time Fourier Transform.....	30

3.2.4	Wavelet Transform .....	32
3.2.5	Signal Envelope .....	35
3.2.6	Pre-processing Steps.....	36
<b>3.3</b>	<b>Digital Heart Sound Analysis.....</b>	<b>40</b>
3.3.1	Survey Articles .....	40
3.3.2	Noise Filtering .....	41
3.3.3	Heart Sound Segmentation.....	42
3.3.4	Feature Extraction.....	44
3.3.5	Classification .....	45
3.3.6	Summary of Heart Sound DSP review.....	45
<b>3.4</b>	<b>Research Questions.....</b>	<b>47</b>
<b>Chapter 4 Probabilistic Modelling for the Decomposition of Heart Sounds .....</b>		<b>48</b>
<b>4.1</b>	<b>Probabilistic modelling.....</b>	<b>48</b>
<b>4.2</b>	<b>A probabilistic view of the diagnostic process .....</b>	<b>53</b>
<b>4.3</b>	<b>Bayesian heart sound modelling.....</b>	<b>55</b>
<b>4.4</b>	<b>Independent Component Analysis.....</b>	<b>55</b>
<b>4.5</b>	<b>Probabilistic ICA .....</b>	<b>58</b>
4.5.1	Methodology.....	58
4.5.2	The Model.....	58
4.5.3	Source Model.....	60
<b>4.6</b>	<b>Methods .....</b>	<b>65</b>
4.6.1	Data.....	65
4.6.2	Pre-processing .....	66
4.6.3	Continuous Wavelet Transform.....	66
4.6.4	Independent component analysis of heart sounds .....	68
4.6.5	Heart sound source model.....	69
<b>4.7</b>	<b>Results .....</b>	<b>69</b>
4.7.1	Pre-processing .....	70
4.7.2	Heart Cycle CWT .....	72
4.7.3	Heart Cycle ICA.....	72
4.7.4	Number of Sources .....	84
<b>4.8</b>	<b>Discussion .....</b>	<b>86</b>
4.8.1	ICA Sources .....	86
4.8.2	Model assumptions.....	87

4.9	Future Work.....	88
4.10	Conclusion .....	89
<b>Chapter 5 Probabilistic Labelling and Segmentation of Heart Sounds.....</b>		<b>91</b>
5.1	Hidden Markov Model Parameters.....	93
5.1.1	Model Parameters .....	93
5.1.2	The Viterbi Algorithm.....	96
5.2	Methods .....	99
5.2.1	Dataset.....	99
5.2.2	Extensions to the Model .....	100
5.2.3	Qualities of systolic murmurs .....	101
5.2.4	Emission Modelling.....	102
5.2.5	Transition Probabilities .....	112
5.2.6	Duration Probabilities .....	112
5.2.7	Initial probabilities .....	118
5.2.8	Overall performance of the model .....	118
5.3	Results .....	122
5.3.1	Emission modelling .....	122
5.3.2	Heart sound Segmentation .....	122
5.3.3	Systolic Murmur Classification.....	123
5.3.4	Probabilistic Heart Sound Labelling .....	124
5.4	Discussion .....	128
5.5	Conclusion .....	131
<b>Chapter 6 Psychoacoustic Descriptions of Heart Sounds.....</b>		<b>132</b>
6.1	Fundamentals of psychoacoustics .....	132
6.2	The psychoacoustics of heart sounds.....	137
6.2.1	Methods.....	137
6.2.2	Subjective experience .....	138
6.2.3	Psychoacoustic training of auscultation.....	138
6.2.4	Psychoacoustic features of heart sounds .....	139
6.2.5	Intensity measures: loudness .....	140
6.2.6	Temporal Features: Shape, timing, and duration .....	141
6.2.7	Pitch .....	142
6.2.8	Systolic murmur description framework .....	143
6.3	Modelling approaches.....	143

6.3.1	Dataset.....	143
6.3.2	Loudness .....	143
6.3.3	Shape .....	150
6.3.4	Pitch .....	157
<b>6.4</b>	<b>Expert descriptions: Cardiac Auscultation Research Survey (CARS) .....</b>	<b>163</b>
6.4.1	Cardiac Auscultation Research Survey (CARS) design.....	164
6.4.2	CARS delivery .....	167
6.4.3	CARS results .....	168
6.4.4	CARS discussion .....	171
<b>6.5</b>	<b>Computer assisted cardiac auscultation: Automated descriptions of systolic murmurs .....</b>	<b>172</b>
6.5.1	Limitations .....	176
<b>6.6</b>	<b>Conclusion .....</b>	<b>176</b>
<b>Chapter 7</b>	<b>Conclusions.....</b>	<b>178</b>
<b>7.1</b>	<b>Summary.....</b>	<b>178</b>
<b>7.2</b>	<b>Final Conclusions.....</b>	<b>180</b>
<b>7.3</b>	<b>Future Work.....</b>	<b>182</b>
	<b>Bibliography.....</b>	<b>187</b>

## Attestation of Authorship

“I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.”

A handwritten signature in black ink, appearing to read 'Andries Meintjes', written in a cursive style. The signature is positioned above a horizontal line.

Andries Meintjes

## Preface

This work represents a continuation of my interest in human physiology and healthcare, specifically the application of science and engineering to address and solve challenges in this area. While mainly a biomedical engineering thesis, exploring mathematical modelling and digital signal processing techniques in medical applications, the surrounding anatomical and physiological knowledge and the potential clinical applications of the technology have formed the core of the exploration throughout. This is representative of my own undergraduate background, completing first a bachelors in physiology and then a bachelors with honours in electronic engineering with a focus on biomedical devices. It also represents the expertise and interests of my supervisory team, the amazing combination of an engineer with years of experience in the medical devices industry and in developmental research of medical technologies and an experienced cardiologist with years of clinical practice and experience in the development of state-of-the-art cardiac imaging technologies. This thesis is an attempt to create a multidisciplinary piece of work that is submitted as an engineering thesis but is firmly grounded in medicine and physiology. To that end I have attempted to use mathematical notation and jargon specific to certain disciplines sparingly, however it is a work in engineering and requires in many instances an exactness that written language alone cannot provide.



We start this thesis by first introducing the problem domain in the form of the rationale and significance of this work. This introduction aims to show that cardiac auscultation and the use of computers to increase the value of cardiac auscultation is a valid and interesting field of research.

### 1.1 Rationale and significance

The process of listening to the sounds produced by the heart, called cardiac auscultation, is a powerful yet inexpensive diagnostic technique. Heart sounds reveal a great deal about the functioning of the heart valves as they open and shut during each heartbeat, as well as providing valuable clues about the presence of structural abnormalities such as shunts or holes in the heart. Using a relatively inexpensive biomedical instrument (the stethoscope) a skilled physician can accurately predict the functioning of the heart valves, as well as many other structural and functional pathologies.

However, auscultation takes practice, experience, and dedication to master and current teaching strategies are proving inadequate. Recent research, performed on the cardiac auscultation skill level of medical professionals, has revealed these to be lacking. A study performed on the accuracy of clinical assessment of heart murmurs by general practitioners (GPs) in 1999 [1] concluded that their performance was “suboptimal” and that “educational strategies are needed to improve accuracy and reduce unnecessary referrals and misdiagnosis”. Another study, done in 2013 [2], on the same subject found that 22% of innocent murmurs were interpreted as abnormal across 106 primary healthcare physicians from a variety of experience levels. The lack of auscultation skills of general practitioners can lead to many healthy patients being referred to specialists and undergoing expensive imaging tests.

The shortfall in this important diagnostic skill has been attributed to a lack of emphasis on cardiac auscultation by medical schools [3] as well as a decrease in time spent on physical examination and an increase in time spent on electronic medical records and other paperwork [4], [5]. The traditional method of bed-side teaching is proving inadequate for the crowded classes of modern medical schools and hospital wards. There is also a trend in modern medicine to move away from the subtler, skill based

diagnostic techniques to the more technologically advanced imaging technologies, such as echocardiography, cardiac magnetic resonance imaging (CMRI) and cardiac computed tomography (CT) [6].

There is a debate over the future of cardiac auscultation, exemplified by the 1996 *Circulation* article by Tavel "Cardiac auscultation. A glorious past--but does it have a future?" [7]. This debate has recently entered the attention of the news media [8], [9] with two opposing sides emerging.

On the one side cardiac auscultation has been declared an obsolete practice, destined to be replaced by hand-held echocardiography (HHE). Kaul [10] argues that "physician inertia" is the main obstacle preventing the switch to the "accurate and reliable" HHE, citing a 2014 study by Mehta et al. [11] which concluded that HHE used by cardiologists provided more accurate diagnosis of cardiac conditions than physical examination alone. A similar argument is made by Frishman [12] who points out that the HHE is being introduced into medical school training programs with success [13]. Fuster [14] however argues that the two technologies do not overlap as much as the proponents of HHE suggest and advocates the unique value of skilful auscultation, using cases in which cardiac auscultation was able to identify factors that were not visible on echocardiogram as motivating examples. This view is echoed by Thompson [15] who mentions that cardiologists, the practitioners who most readily have access to both echocardiography, still make use of their stethoscopes daily and argues that they would be the first to "discard their stethoscopes" if there were "no correlation between heart sound and echo findings". Edelman and Weber [16] espouse the connection, in their words the "tenuous tether", that auscultation creates between the physician and the patient. They go on to encourage the use of technologies that expand the use of auscultation, for example providing playback of heart sounds through a speaker to allow all attending physicians and the patient to hear what previously only the auscultator himself could hear.

Instead of viewing auscultation and HHE as two opposing technologies, they can be regarded as complementary tools in the physical examination. As Tavel pointed out in his 2006 follow-up article "Cardiac auscultation. A glorious past--and it does have a

future!" [17], the emergence of electronic stethoscopes and the ability to apply signal processing and machine learning methods to these recordings has the potential to revolutionise the practice of cardiac auscultation. An important aspect to consider is how effectively these methods can be taught. Both echocardiography and auscultation require skill to perform and interpret and replacing one poorly taught skill with another is unlikely to lead to better outcomes. A pilot study by Legget et al. [18] investigated the combination of HHE and electronic stethoscopes as training aids for cardiac diagnosis. Motivated by the ability of HHE to effectively visualise heart murmurs at the bedside and electronic stethoscopes to allow recording and repeated listening of heart sounds, the results of the study suggest that these technologies have application not only in diagnostics but also in effective medical training. Barrett et al [19] and Finley et al.[20] take a similar viewpoint, shifting the focus of the debate to the techniques used to train medical students in the diagnostic practices.

Imaging technologies provide invaluable diagnostic information and the importance of further developing the science and engineering related to these cannot be stressed enough. However, the high cost and scarcity of these technologies limit their usefulness, especially in rural or underdeveloped locations. Recently, handheld echocardiography devices have been developed and some physicians see these as the future of cardiac diagnosis [21]. However, correctly performing and interpreting a cardiac ultrasound, as there is a significant likelihood of false positive and false negative findings in undertrained individuals. Operating and interpreting echocardiography requires extensive training, comparable to that of effectively auscultating [22]. Professional sonographers require at least 3 years of training and study as well as years of actual clinical experience to be able to effectively use these devices for diagnosis [23]. In New Zealand, if a physician suspects a detected murmur of being pathological, they will refer a patient to a sonographer for an echocardiogram as well as a consultation with a cardiologist. If the patient is an eligible permanent resident or citizen, these visits will be publicly funded, although the waiting lists can be months long, and many patients opt to receive the echocardiogram privately. Thus, a lack of accurate auscultation has a disproportionate effect on lower socio-economic groups who cannot afford to pay for more expensive tests and are at a greater risk of conditions like rheumatic fever which increases the likelihood of developing certain heart valve disorders [24].

Biomedical and software engineering can assist in the development of a solution to this problem. Digital signal processing (DSP) techniques can be implemented to assist both the training of auscultation skills as well as the heart sound analysis procedure, assisting the diagnostic process by clarifying information in the heart sound or through machine learning methods classifying the pathology (if any) present [25]. In the last two decades, a great deal of work has been done in the field of computer assisted auscultation, with a significant focus on the development of a *machine learner* that is able to accurately distinguish between different normal and abnormal heart sounds. These studies, discussed in the literature review, have largely followed the procedure of segmentation, feature extraction, and classification. So far, no studies have produced generalised, clinically validated results. The major difficulties appear to be badly described, small testing and training sets, no standardised set of features to extract, as well as disparity between studies on what pathologies to try to classify. Another issue with the current framework is acceptability; generally, the presented scheme attempts to identify pathological murmurs in the heart sound, although the information presented to the classification system (heart sound recording) does not contain much of the information that a physician would use to reach the same conclusion. This issue is fundamental in the design of any medical devices that aim to provide decision support. The limitations of the signal under analysis and the position of that signal within a greater context must be kept in mind in order to provide useful information. Due to this disparity between the diagnostic method of the attempted algorithms and the actual diagnostic method used in practice, physicians and the public might less readily accept the classification result especially with the current poor state of clinical validation.

The rationale for this work can be summarised as:

- Cardiac auscultation provides a relatively low-cost cardiovascular screening method.
- However, it is a clinical skill that takes effort and practice to master. Several studies have found the cardiac auscultation skills of physicians to be suboptimal.
- Computer assisted auscultation is a possible solution. Using electronic recording devices (electronic/digital stethoscopes) and computer algorithms to analyse heart sounds and serve as diagnosis support systems.

## 1.2 Outline of this thesis

In this research we have attempted to bridge this gap between computer assisted analysis and human expert analysis of heart sounds. A representation of the thesis chapters is shown in Figure 1.1.



Figure 1.1: An outline and navigation centre for the chapters that comprise this thesis. This figure also acts as an interactive navigation tool in the digital format of this document; try clicking the icons on the leaves and next to the chapter headings to navigate to the start of each chapter and back here.

In this chapter, Chapter 1, we have introduced auscultation and computer assisted cardiac auscultation as an active and beneficial area of research for biomedical applications. In Chapter 2 we explore the background of this field by first looking at the anatomy and physiology of the heart and then putting this knowledge to use in contextualising the clinical practice of cardiac auscultation. Chapter 3 introduces digital signal processing concepts and reviews developments in the application of signal processing techniques to heart sound recordings. We conclude Chapter 3 by introducing the perceived gap in the literature and the research questions this work answers to address this gap. Chapter 4 introduces the use of probabilistic models as a framework to perform heart sound analysis. To illustrate and test this methodology, a probabilistic independent component analysis (ICA) model is applied to examples of heart sounds. The probabilistic framework is further explored in Chapter 5 where we describe and expand upon previous work done in heart sound segmentation. Chapter 6 introduces the concept of “psychoacoustic” models where heart sounds are analysed considering the perceptual qualities of the human auditory system to form descriptions that match

the sensations they would elicit. Finally, in Chapter 7 the thesis is concluded, and future directions discussed.



We start our exploration at the origin of the signal we intend to study. The field of medical science is an ever-evolving domain of knowledge, with a previously unrecognised organ system being discovered as recently as 2018 [26]. The evolution of this field is in no small part due to the improvement in the instruments available to researchers and medical practitioners. In this chapter we first explore the organ system, that is the heart and great arteries or more succinctly, the central cardiovascular system, in which the signal we are studying, originates. After this we look at how the sounds produced in the cardiovascular system have been used by medical practitioners, a practice termed *cardiac auscultation*.

### 2.1 Functional Cardiac Anatomy

To design an algorithm that can analyse the sounds produced by the heart it is important to have a clear and accurate understanding of the functional anatomy of the heart structures as well as the physiological processes that produce those sounds. To this end the first part of this literature review will focus on current state of knowledge about the anatomy of the heart and associated cardiac structures.

#### 2.1.1 The Heart

The human *heart* is a large muscular structure that lies obliquely in the chest very slightly to the left of the sternum. The heart consists of two sets of chambers, the right and left *atria* and *ventricles* (Figure 2.1). Each set of chambers acts like a two-stage pump with the right atrium and ventricle pumping deoxygenated blood from the systemic circulation to the pulmonary circulation and the left atrium and ventricle pumping oxygenated blood from the pulmonary circulation to the systemic circulation. The right atrium receives blood via the superior and inferior vena cava as well as the coronary sinus and opens into the right ventricle through an atrioventricular valve which has three cusps - the *tricuspid* valve. The right ventricle opens into the pulmonary artery through the *semilunar pulmonary* valve. The left atrium receives the four pulmonary veins from the lungs and opens through to the left ventricle via the bicuspid, or more commonly: *mitral* valve. The left ventricle is the most muscular of the four chambers providing the necessary contractive force to maintain systemic blood pressure. The *aortic* valve is a

connective tissue structure formed out of three semilunar cusps that guards the opening between the left ventricle and the aorta, preventing blood from flowing back into the heart.

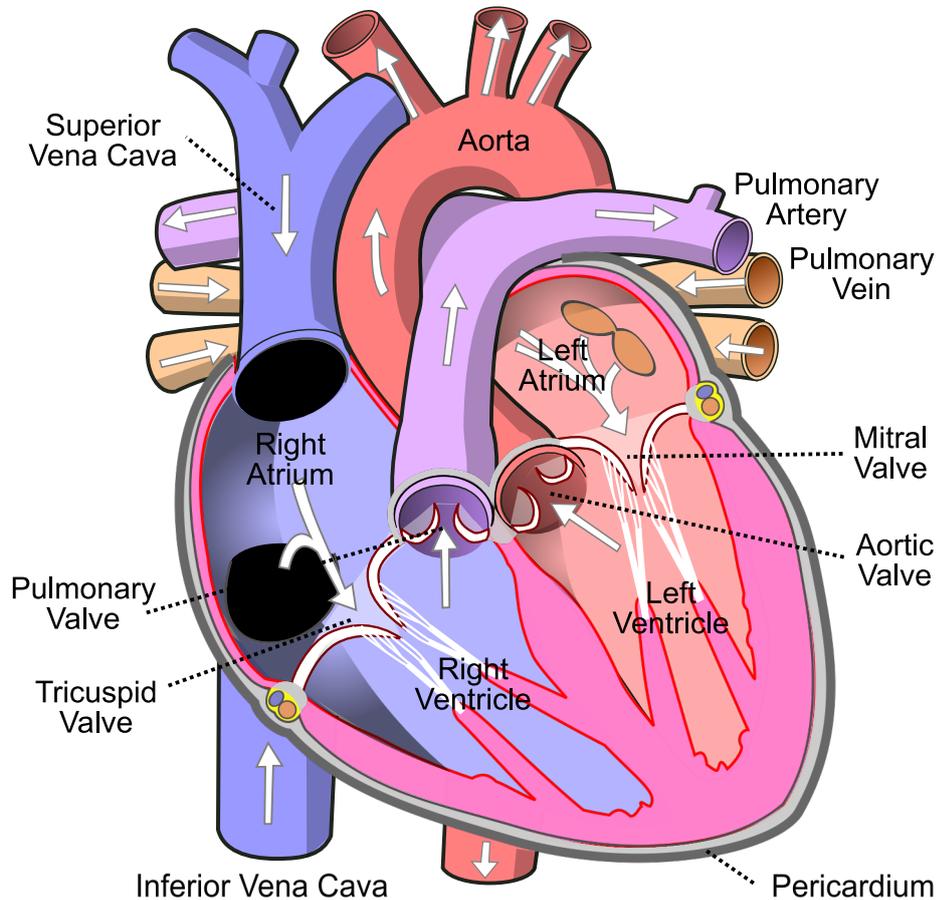


Figure 2.1: The Heart, with the heart valves and major arteries and veins labelled. The arrows indicate the direction of blood flow [27].

### 2.1.2 The Heart Valves

The heart valves, illustrated in a transverse cross-section of the heart in Figure 2.2, are connective tissue structures that regulate the movement of blood through the heart chambers. The atrioventricular valves prevent blood from flowing back into the atria during ventricular contraction (systole), while the semilunar valves prevent blood from flowing back into the ventricles during ventricular relaxation (diastole). The right atrioventricular (AV), or tricuspid, valve consists of 3 connective tissue flaps at the opening between the right atrium and ventricle. The pulmonary (right semilunar) valve guards the opening of the right ventricle to the pulmonary artery that connects to the lungs. The mitral (left AV) valve is formed by 2 cusps between the right atrium and right

ventricle. The aortic (left semilunar) valve connects the left ventricle and the aorta which allows blood to flow into the systemic circulation. Both AV valves are attached to papillary muscles of their respective ventricles by chordae tendineae, strings of connective tissue.

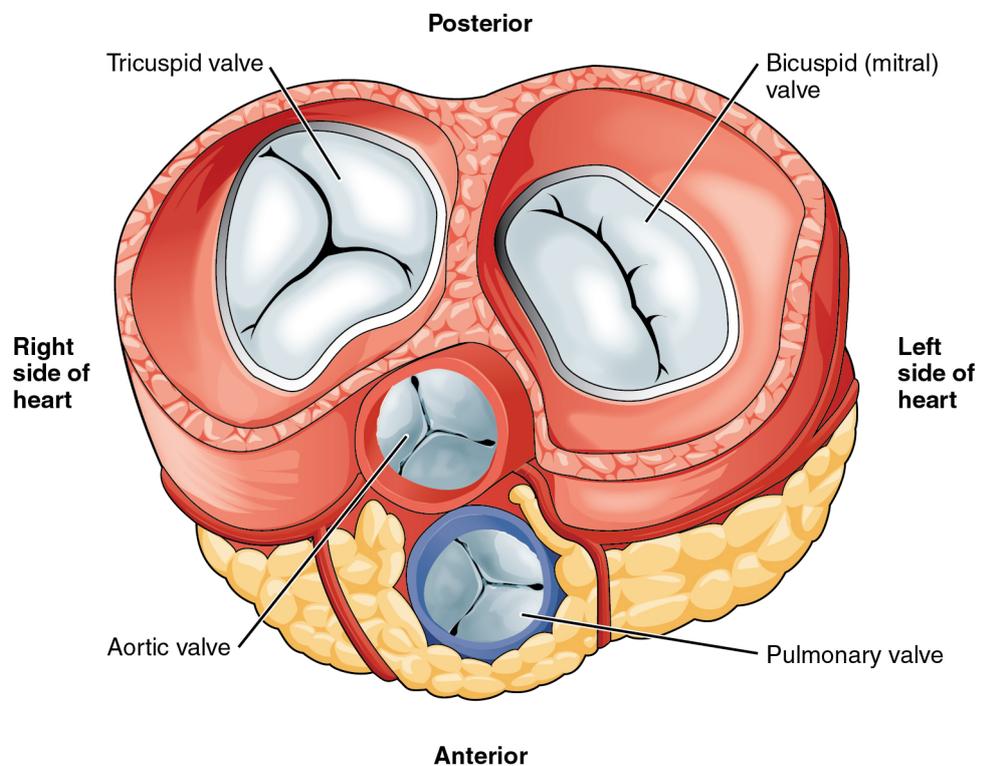


Figure 2.2: A transverse cross-section of the heart showing the heart valves [28]

If the heart valves do not function correctly the resulting condition is called heart valve disease [29]. Heart valve disease, also known as valvular heart disease/valvopathy, can present in two distinct ways, (1) valvular stenosis, and (2) valvular insufficiency. In valvular stenosis the heart valve tissue stiffens, causing the valve opening to narrow, thus impeding the flow of blood through the valve. In cases of severe valve stenosis, the amount of blood able to move through the narrowed valve may become insufficient for normal functioning. Valvular insufficiency is also called valve regurgitation or incompetence; such a valve may also be referred to simply as a “leaky valve”. In this condition the heart valve is not able to close completely and allows blood to leak back into either the atria or the ventricle after ejection. These two conditions are illustrated in Figure 2.3. Heart valve disease can be either congenital (present at birth) or acquired. Congenital heart valve disease affects about 0.1% of new-borns, with stenosis of either the pulmonary or aortic valve being most common [30]. Heart valve disease is classified

as acquired if the valve was structurally normal at birth. Several conditions can cause the heart valves to become diseased, including congenital and inherited conditions such as Marfans and Ehlers-Danlos syndrome, rheumatic fever, endocarditis, syphilis, and connective tissue disorders such as systemic lupus erythematosus.

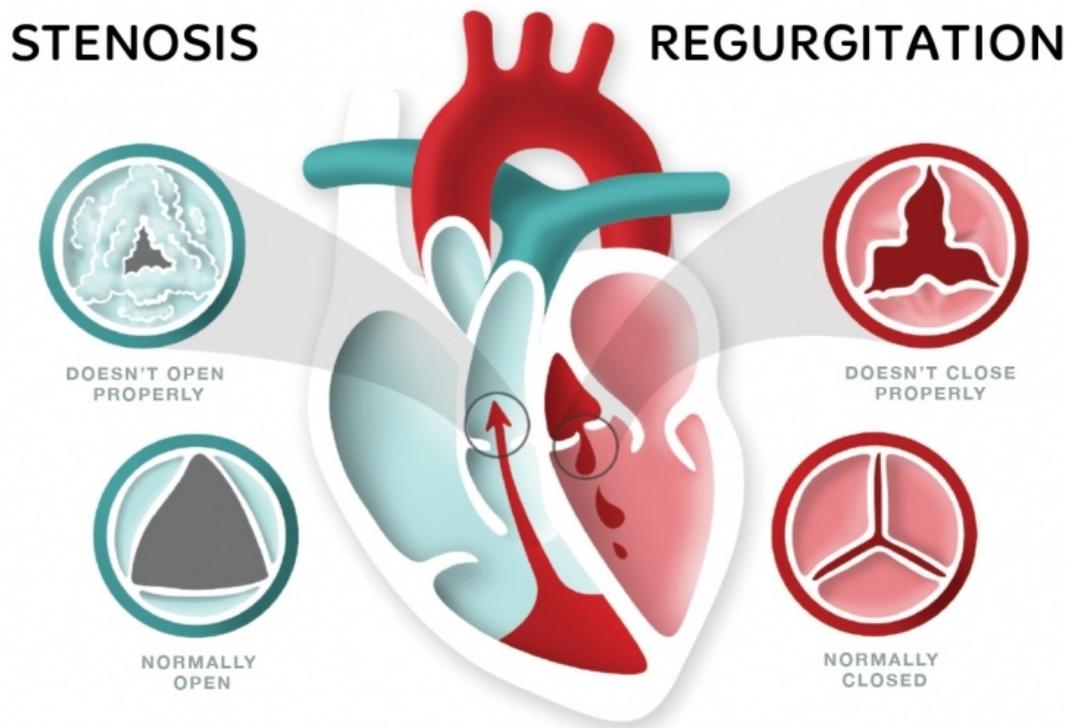


Figure 2.3: An illustration of heart valve stenosis and regurgitation, showing normal heart valves compared to those with impaired closure or opening [31].

## 2.2 The Cardiac Cycle

The heart rhythmically contracts and relaxes multiple times a minute in order to move oxygen and nutrients to the systemic circulation and toxins and carbon dioxide to the lungs for excretion; this process is referred to as the cardiac cycle. A complete cardiac cycle can be defined as the cardiac events initiated by atrial depolarization (observed as the P wave in an electrocardiogram, Figure 2.4) and continuing until the start of the next atrial depolarization (although, being cyclic, the selection of start/end is arbitrary). The cardiac cycle is divided into systole and diastole. Systole refers to events that are associated with ventricular contraction and ejection of blood, whereas diastole refers to the period of relaxation and filling. The two fundamental heart sounds, commonly referred to as S1 and S2 correspond to the start and end of systole. The first heart sound (S1) corresponds to the peak of the QRS complex of the electrocardiogram, while the second heart sound (S2) occurs after the peak of the T wave.

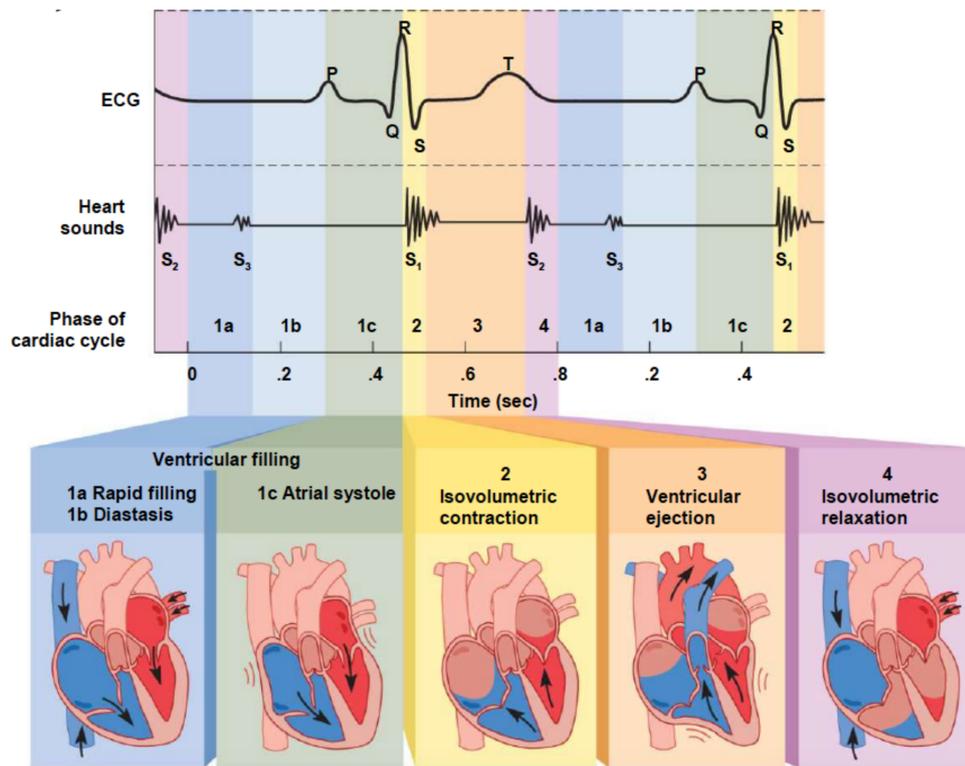


Figure 2.4: A nominal electrocardiogram and heart sound waveform and the associated phase of the cardiac cycle [32].

## 2.3 Heart sounds

### 2.3.1 Fundamental Heart sounds

When listening to a healthy heart, two distinct sounds (shown in Figure 2.5) are normally audible: S<sub>1</sub> also sometimes referred to as “Lub”, and S<sub>2</sub> (also called “Dub”).

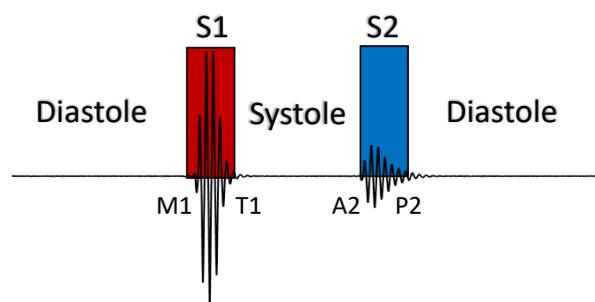


Figure 2.5: The fundamental heart sounds (S<sub>1</sub> and S<sub>2</sub>) at the apex area with the patient lying on their back. Recording from the University of Michigan's database of simulated heart sounds [33].

The first heart sound, S<sub>1</sub>, corresponds to the closing of the atrioventricular valves. While controversial, the most widely accepted theory for the origin of the sounds is vibrations of the atrioventricular valve structures, the connective tissue ring, valve cusps, and chordae tendineae caused by the abrupt deceleration of blood in the atria and ventricles

after the valves snap shut. The first heart sound is composed of the mitral closing sound (M1) and the tricuspid closing sound (T1).

The second heart sound, S2, occurs directly after the closure of the semilunar valves and has two audible components, namely the aortic closure (A2) and the pulmonic closure sound (P2). The second heart sound is the result of the sudden deceleration of blood causing vibrations in the semilunar valve cusps, as well as the walls and blood columns of the aorta and pulmonary artery as well as their respective ventricles.

### 2.3.2 Other Heart sounds

#### 2.3.2.1 The third heart sound (S3)

The third heart sound is a brief, low frequency beat that occurs in early diastole shortly after S2, at the time of maximal ventricular filling (Figure 2.6) [34]. Up until age around the age of 40 the third heart sound may be physiological, especially in young children, athletes, and pregnant woman, and should be judged by the presence or absence of significant heart disease. In older adults (>40 years) a third heart sound is usually pathological and an indication of ventricular dysfunction. The third heart sound is challenging to auscultate, since it is usually of very low intensity, it does not radiate widely over the chest wall, and usually the frequency content is near the lowest level the human ear can detect.

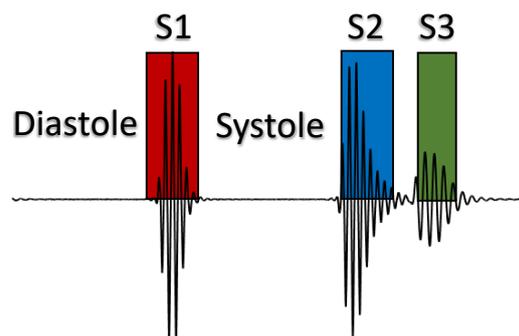


Figure 2.6: The third heart sound (S3) at the apex area with the patient lying on their left side. Recording from the University of Michigan's database of simulated heart sounds [1].

#### 2.3.2.2 The fourth heart sound (S4)

The fourth heart sound occurs late in ventricular diastole, shortly before the start of the first heart sound (Figure 2.7). It is a low frequency sound that coincides with atrial contraction at the end of diastole [35]. The fourth heart sound is the result of vibrations

generated within the ventricle after atrial contraction. Usually its presence indicates that there is a reduction in left or right ventricular wall compliance, and as such an increased resistance to filling. It is accompanied by a disproportionate rise in ventricular end-diastolic pressure, and often also by a short palpable outward movement of the chest wall right before systole.

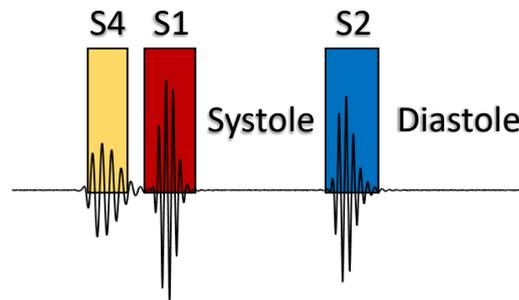


Figure 2.7: The fourth heart sound (S4) at the apex area with the patient lying on their left side. Recording from the University of Michigan's database of simulated heart sounds [1].

### 2.3.3 Murmurs

A murmur is a series of sounds of variable duration and frequency, audible with a stethoscope at the chest wall, that originates from the heart or great blood vessels (aorta, pulmonary artery) [36]. Murmurs are firstly differentiated based on their location in the cardiac cycle. Systolic murmurs occur in systole between the first (S1) and second (S2) heart sounds, while diastolic murmurs occur in diastole between the second (S2) and first (S1) heart sounds. Murmurs are indications of turbulent blood flow. According to the authors of "Clinical Methods" [37] a widely accepted theory of the generation of murmur sounds incorporates the concept of vortex shedding. This theory explains the sustained vibrations required to produce audible heart murmurs by the creation of "vortices" (or eddies) due to turbulence, the shedding of these vortices creating areas of relative stillness or "wakes", and the subsequent movement of the blood to fill in the wakes. The response of blood moving in to fill the wakes left by vortex shedding can explain the generation of sustained sound vibrations that are audible at the chest wall. Murmurs can occur for a variety of reasons and can be functional (non-pathological) or pathological, some common examples are shown in Figure 2.8.

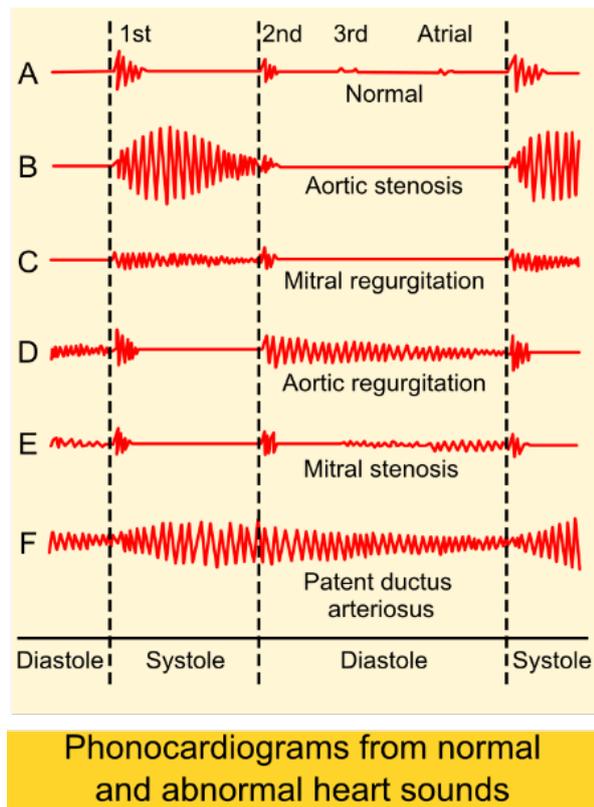


Figure 2.8: Representation of common murmurs and their associated causes [38]

## 2.4 Cardiac Auscultation

As early as 350 BC Hippocrates described the diagnostic value of listening to the sounds caused by the inner workings of the body [39]. The diagnostic value of auscultation remained relatively unchanged until 1628 when William Harvey published “On the Motion of the Heart and Blood” in which he provides the first description of the sounds produced by the heart, identifying the sound that can be heard within the chest as a pulse made when the heart transfers blood from the veins to the arteries, which he described as sounding like two clacks of a water bellows to raise water [40]. These discoveries established the fact that the condition and mechanical action of the inner workings of a person can be deduced by listening to the sounds produced.

Today we know that heart sounds are related to vibrations of the cardiac structures and the dynamics of blood flow through the heart. Medical professionals can use stethoscopes at different locations of the chest surface to listen to and identify normal and abnormal heart sounds. Normal heart sounds in the audible range include S1, associated with the closing of the atrioventricular valves, and S2, which results from the sudden closure of the semilunar valves at the end of systole [41]. Abnormal heart

sounds, called extra heart sounds or heart murmurs, are additional heart sounds that may be indicative of serious heart defects and valve lesions [42]. This process of listening to the sounds produced by the beating of the heart, known as cardiac auscultation, provides a low-cost method of screening for serious heart diseases.

#### 2.4.1 Cardiac Auscultation in Context

Cardiac auscultation commonly forms part of the *physical examination*, one aspect of the wider process of *medical diagnosis* as illustrated in Figure 2.9. Dr. Lester King described medical diagnosis as “a process by which a given individual is assigned to a given class, or the disorder from which he suffers is subsumed under a certain concept” [43]. According to Ledley and Lusted [44] the three logical concepts inherent in medical diagnosis are: (1) medical knowledge, (2) the signs and symptoms presented by the patient (which we could specify more broadly as any relevant data collected about the patient, e.g. patient history, test results), and (3) the final medical diagnosis itself. Thus, medical diagnosis is the process by which an observer (e.g. a physician) assigns a person (more specifically the health of that person) into a category based on the relationship between the knowledge and experience of the observer and the data (e.g. signs and symptoms, laboratory test) collected from the person. The aim of medical diagnosis is to find a diagnosis that allows a health care provider to determine a suitable and effective therapy.

*Physical examination* is one method commonly employed by physicians in collecting data about the health of a patient. As implied in the name, it consists of the physician physically examining the patient using their senses. *Auscultation* is then a part of the physical examination in which the physician uses their sense of *hearing* to listen to the sounds produced by the movements of the patient’s organs. The heart and cardiovascular system are in constant motion and so present one of the most prevalent sources of sounds in the body. *Cardiac auscultation* is the inspection of the cardiovascular system by way of the sounds produced by that system.

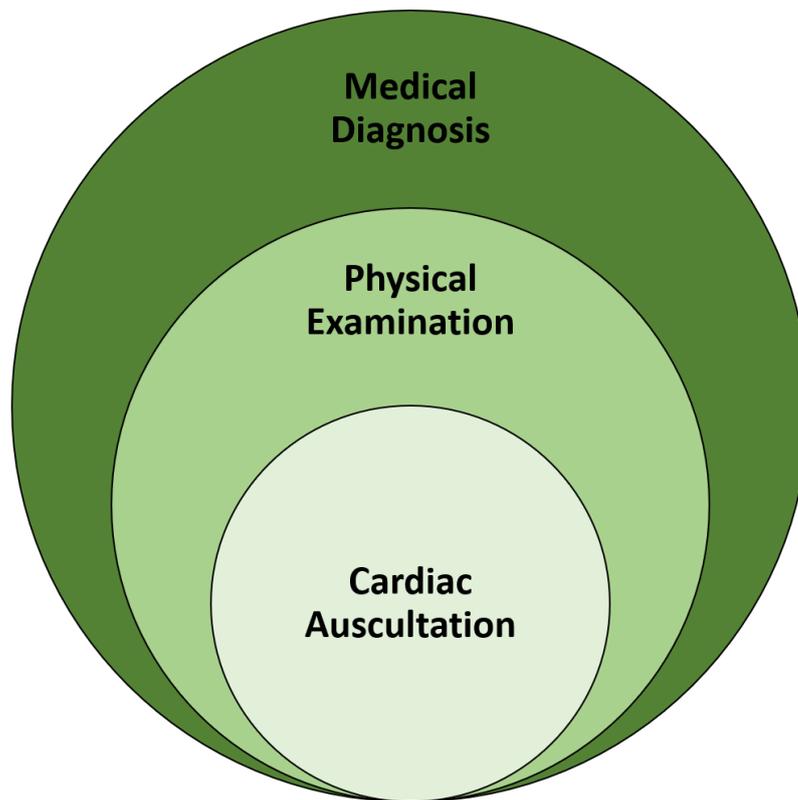


Figure 2.9: The position of cardiac auscultation within the greater goal of forming a medical diagnosis.

#### 2.4.2 Cardiac Auscultation Procedure

Cardiac auscultation is an essential and critical part of a physical examination. The investigating physician will perform cardiac auscultation to identify the normal heart sounds of S1 and S2 and to determine if any additional sounds are present [45]. To achieve this, a clinician will typically listen in a logical, sequential manner using both the high frequency and low frequency (bell and diaphragm) parts of the stethoscope at multiple “auscultation sites” on the chest wall. The locations of the most important auscultation sites are shown in Figure 2.10. The carotid pulse may be used to assist in identifying the timing of the different stages. Cardiac auscultation, as part of a physical examination and patient history, can be used to diagnose a variety of cardiac pathologies, especially those related to the heart valves as well as septal defects and shunts.

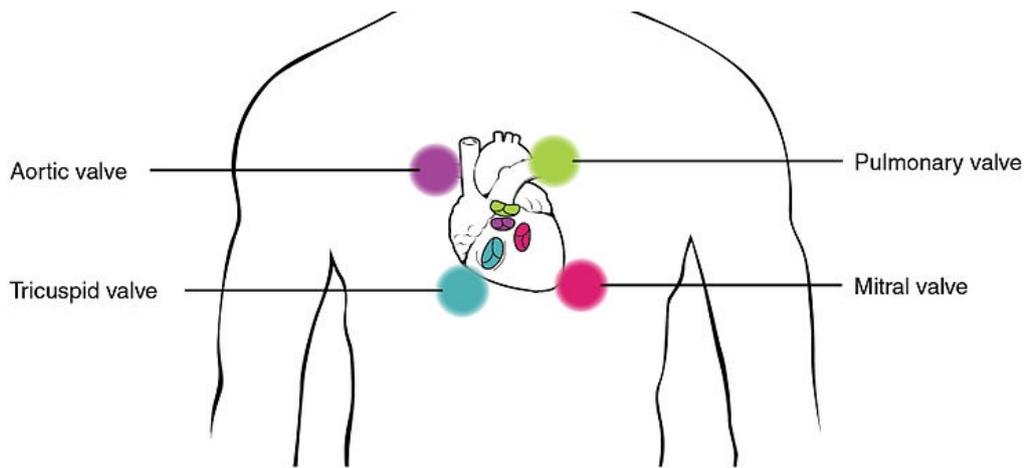


Figure 2.10: Cardiac auscultation sites for the various heart valves [46]

As a diagnostic tool, cardiac auscultation involves more than simply using a stethoscope to listen to the sounds produced by the heart. It is a step in a process of investigation that involves using previous knowledge about the patient's history and other symptoms, knowledge about the underlying functional anatomy and physiological processes, as well as a well-trained ear and well-developed sound perception abilities. All this knowledge and these skills are used to locate and identify sometimes very subtle sounds and alterations of sounds present in the heart sound signal transmitted from the heart through various levels of muscle, bone, skin, and stethoscope to the investigator's ears. Auscultating at different locations on the chest wall, with the patient in a variety of different positions (supine, standing, lying on their side, etc.), accentuates different sounds and provides more information. Heart murmurs can be heard differently at different locations of the chest and changing the patient's position may enhance the quality/intensity of the sound. Murmurs tend to radiate in the direction of the blood flow and the clinician may auscultate at alternative and remote locations in order to further identify heart murmurs [45], [47]. The relationship of murmurs to the respiratory cycle can provide further clues as to an affected valve. The intensity of murmurs produced by valves on the right side of the heart tend to increase during inspiration [48]. Systolic murmurs in particular may change in relation to the loading of the heart [6] and thus the information contained in the heart sound can be altered by actions of the patient, for example, changing their handgrip, or performing the Valsalva manoeuvre (forceful expiration against a closed glottis [49]). Cardiac auscultation, as a method of investigation, includes both listening with a stethoscope and using different techniques

to increase the audibility of the heart sound to a point where any underlying conditions becomes obvious.

Effective cardiac auscultation requires both knowledge and expertise in physical examination and the mastery of the technical skill of auscultation. Training the technical skill of auscultation involves acquiring the ability to recognise certain sounds using different features of those sounds, similar to how a student of music must learn to discern different notes and pitches. Research has shown that the mastery of auscultation requires hundreds of hours of repeated practice on recognising the different sounds [50]. The physician must also be trained in the necessary clinical skills to effectively auscultate. A successful physical examination requires the physician to recognise the available clues and direct the examination in the appropriate direction. They must know where to look and how to direct the patient, i.e. to perform a specific maneuver or assume a specific position that may accentuate or diminish certain aspects of the heart sounds under specific disease conditions. Thus, the physical examination does not represent a static data collection process but rather a dynamic procedure in which information is steadily revealed.

### 2.4.3 The Stethoscope

Up until the 19<sup>th</sup> century the only way for physicians to auscultate was through immediate auscultation, by directly applying their ears to the patient's body. This method had some severe technical limitations as well as being socially awkward and uncomfortable for both the doctor and the patient. In 1816 a French physician called René Laennec (1781 to 1826), inspired by the transmission of sound through a solid medium, decided to use a rolled-up paper applied to the region of the heart and the other end to his ear to listen to the heart sounds of a young woman presenting with symptoms of a diseased heart. Laennec discovered that the heart sounds could be heard much more clearly, and with much less discomfort for both him and the patient, by using "mediate auscultation" where an instrument is interposed between the ear of the examiner and the patient. Inspired by his success, he developed this idea further and named his invention the stethoscope.



Figure 2.11: A wood and brass stethoscope used by Laennec around 1820 [51]

The stethoscopes that Laennec used consisted of a long wooden tube and was monaural, which means that it had a single ear piece and thus the physician could only listen through one ear when using it (Figure 2.11). The binaural stethoscope was invented in 1851 by Arthur Leard. Modern stethoscopes, whether acoustic (mechanical) or electronic, are binaural, allowing users to auscultate using both ears at the same time. The traditional stethoscope consists of three main components: the chest piece, the tubing, and the earpieces as shown in Figure 2.12.



Figure 2.12: An acoustic stethoscope [52].

During auscultation, the chest piece is placed onto the patient's chest and acts as a medium for the different pressure waves produced by the body. The most common type of acoustic stethoscope in use today has a chest piece that consists of two distinct parts: the diaphragm and the bell. The diaphragm is a flat disk that is used to auscultate higher frequency sounds. The bell resembles a hollow cup and transmits lower frequency sounds than the diaphragm. Recently single piece chest pieces have been developed with "dual frequency" or "tuneable" diaphragms. These can switch between high and low frequency modes based on the amount of pressure that is being applied by the user.

Regardless of type, the aim of any chest piece is to transfer the vibrations produced by the body to the next part of the stethoscope, the tubing. The tubing is the air-filled hollow tubing that allows the sound waves to be transmitted from the chest piece to the earpieces. The earpieces play the important role of connecting the tubing, and chest piece, to the user's ear. In this way, the earpieces complete the closed system that allows sounds from the body of the patient to be transmitted directly to the physician's ears. The earpieces should fit comfortably, whilst also fitting tight enough to create the closed system.

Since its invention the stethoscope has become the most iconic and identifiable biomedical instrument. Routine physical examinations almost always include auscultation, due to the non-invasive nature of the procedure as well as the low cost and wide availability of the instrument. Recently the Glia project has developed an open source 3D printable stethoscope that can be printed for as little as \$5 USD, potentially further increasing the availability of the stethoscope [53]. Herein lies the stethoscope's greatest strength: it can provide knowledge of the inner workings of a person without expensive equipment or invasive and damaging procedures. The stethoscope allows doctors to use their sense of hearing to "look" inside the body of a patient and assess if there are problems. Although this method of investigation is limited to conditions that produce a sound wave or an alteration of a physiological sound, it provides a quick and relatively simple and inexpensive way to do so.

#### 2.4.4 Recent Developments

The advent of microelectronics and integrated circuits has led to the development of *electronic stethoscopes* [25]. As early as 1966 a patent was issued for a stethoscope that included an electronic circuit that could augment the frequency response of the device [54]. Since then electronic stethoscopes have become more common and the applications of electronics and signal processing in these devices have become more varied. These devices potentially have many advantages over traditional mechanical stethoscopes including the ability to amplify sounds, record and store/transmit heart sound recordings, and more precise control over the frequency response than acoustic stethoscopes.

A 1998 study conducted by Grenier et al. [55] compared three electronic stethoscopes available on the market at that time (Graham Field Labtron, Bosch EST40, and Starkey ST30) to acoustic stethoscopes based on user's (nurses and physicians) assessment. They found that most users preferred the acoustic stethoscopes and suggested that while the electronic stethoscopes allow amplification, they also introduced limitations especially prevalent at high amplification levels. They identified the following limitations (1) the introduction of electronic noise, (2) high sensitivity to impacts, movements, and ambient noises, (3) usage of electronic filtering terms that are not well understood by clinical users, and (4) discomfort resulting from poor ergonomic design. It should be noted that none of the electronic stethoscopes assessed in the 1998 study are still on the market in 2019.

More recent studies have focused on comparing the clinical and training effects of electronic and acoustic stethoscopes. Høyte et al. [56] found no significant difference between the performance of medical students training using an electronic stethoscope compared to those training with a traditional acoustic stethoscope after four months of training. However, a 2018 study by Legget et al. [18] suggested that the use of an electronic stethoscope during ward round teaching positively impacts the cardiac auscultation ability of students. A clinical study conducted in 2019 by Kalinauskienė et al. [57] found that the use of electronic stethoscopes increased the screening sensitivity of both cardiologists and resident physicians in assessing heart valve lesions in obese patients, although there was no difference in specificity between the two types of stethoscopes.

The increase in quality and availability, along with a decrease in the cost, of electronic stethoscopes has the potential to revolutionize cardiac auscultation and its usefulness as a diagnostic technique. Possible applications include tele-health monitoring and consultations, heart sound database creation (with cloud computing and machine learning applications), decision support systems, and computer assisted heart sound classification systems.

Potentially the most beneficial aspect of making digitised recordings of heart sounds is the ability to apply digital signal processing (DSP) techniques to these recordings. Since the development of digital computers, the processing power of these devices have been applied to signal processing tasks in which digitised observations of time-varying signals (e.g. the voltage across a capacitive sensor sampled at set intervals) have been manipulated computationally to reveal interesting and useful information in the signal. A ubiquitous example of DSP is the use of digital filters to filter out "noise" (signal components that are unwanted in a specific application) by attenuation (or weakening) and/or amplification (or strengthening) of different parts of a signal. The potential applications, and methodologies, of DSP are vast and it is necessary to discuss the specific methods investigated for heart sound analysis in this research.

We start by exploring how the motion of the heart produces oscillations that we can detect as sound at the body surface and how these oscillations are converted into representations that can be modified by computers (sequences of numbers). We then discuss some ways in which these "sequences of numbers" can be analysed and modified to provide insights about the nature of system responsible for generating them. The rest of this chapter gives brief overviews of different signal processing techniques that are relevant to the processing of digital heart sound recordings. First some fundamental definitions and concepts from digital signal processing are introduced, followed by a review of heart sound analysis in the literature.

### 3.1 Audio signals and Heart Sound Recordings

Sound waves are produced when matter is disturbed from a state of equilibrium. In an elastic medium these waves travel away from the disturbance as areas of *compression* and *rarefaction* in the surrounding matter. The sound waves can be observed as variations in the pressure of the medium they are travelling through. An audio signal, specifically a signal that would be perceived as sound by a human observer, can be described as a time-varying function of sound pressure level with frequencies in the range of around 20 to 20 000 Hz.

Microphones are transducers that are sensitive to the pressure waves that would be perceived as sound, converting the changes in sound pressure level into changes in electrical characteristics, i.e. current/voltage [58]. *Condenser* (or *capacitor*) microphones measure changes in pressure by detecting the changes in voltage that occur as pressure waves move the plates of a capacitor. *Piezoelectric* microphones use the property of certain materials called piezoelectricity that produce an electrical signal in response to mechanical strain (such as that caused by pressure waves).

Whatever the specific mechanism used to measure the sound waves the output is an analogue electrical signal. In order to represent and analyse the signals on computers we must convert the analogue signal into *digital signal* by means of an electronic system called an analogue-to-digital converter (ADC). The ADC samples and quantises the continuous voltage level into samples that are discrete in both level and time. The digital signal is then simply a sequence of numbers, each representing the level of the sound at a point in time.

The sampling rate can be expressed as the “number of times we sample the signal in one second” using the International System of Units (SI) unit *hertz* (Hz). One hertz is defined as one *cycle per second*. The discretization process necessarily introduces a degree of error into the signal, and it is important to sample the signal at a high enough rate (sampling rate) and quantize it at a high enough resolution (number of quantization levels) to not distort important aspects of the signal. Sampling theory states that frequencies above half the sampling rate are unrecoverable in the discretized signal [59]. The critical sampling frequency, or Nyquist rate, is the frequency at which a specific signal must be sampled so that all of its parts can be recovered and no error is introduced due to the aliasing (misinterpretation) of components higher than this rate [60]. The sampling process is illustrated in Figure 3.2 with 1 second of a heart sound recording sampled at 500 Hz.

As can be seen in the representation shown in Figure 3.1, adapted from “Textbook of medical physiology” by Guyton and Hall [41], the heart sounds have relatively low frequencies and low sound pressure levels. The heart sounds only have a significant overlap with the human auditory range between about 40 Hz and 500 Hz. Outside of this range the frequencies and/or the sound pressure level becomes too low for the

average person to perceive. Even inside this range the heart sounds are close to the threshold of audibility and would be perceived as soft sounds, not easily heard in the presence of significant other noises.

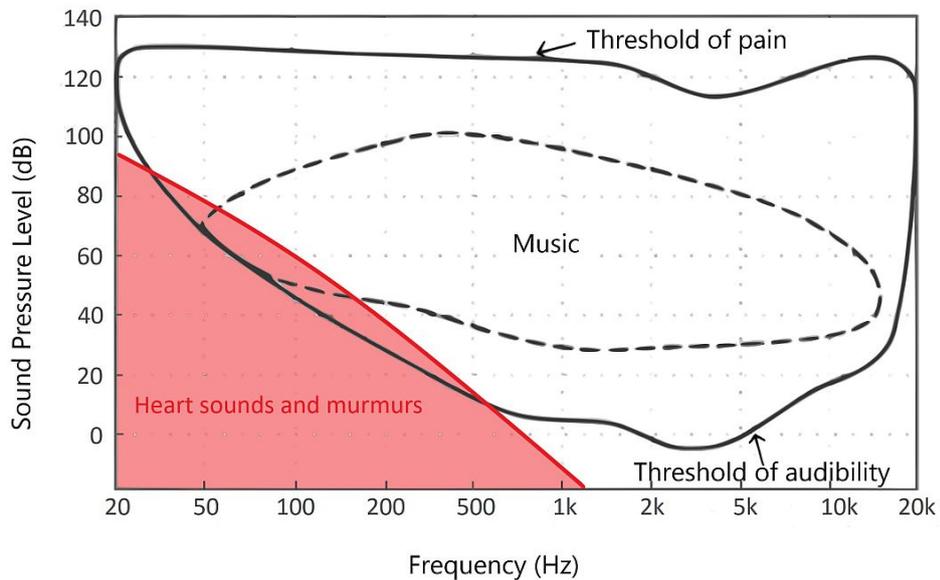


Figure 3.1: A representation of the human auditory range and approximate range of heart sounds. Adapted from Guyton and Hall [41]

The stethoscope allows a listener to more easily hear the heart by mechanically amplifying the sounds. Electronic stethoscopes address the problem using preamplification circuits that amplify slight changes produced by sound waves to make use of the complete input range of the ADC and thus minimise quantization error. Thus, the digitised signal that can be stored and transmitted is not in physical units, i.e. the absolute change in sound pressure level (e.g. Pascals), but rather in a quantised and amplified representation of the voltage changes as detected by a microphone. To relate the recorded signal to physical units would be a tedious task, requiring the exact specifications of all the stages of processing applied to the recording, in which each stage of processing would have to be reversed. The signal amplitude, intensity, and normalisation of these quantities will be discussed further in the following sections.

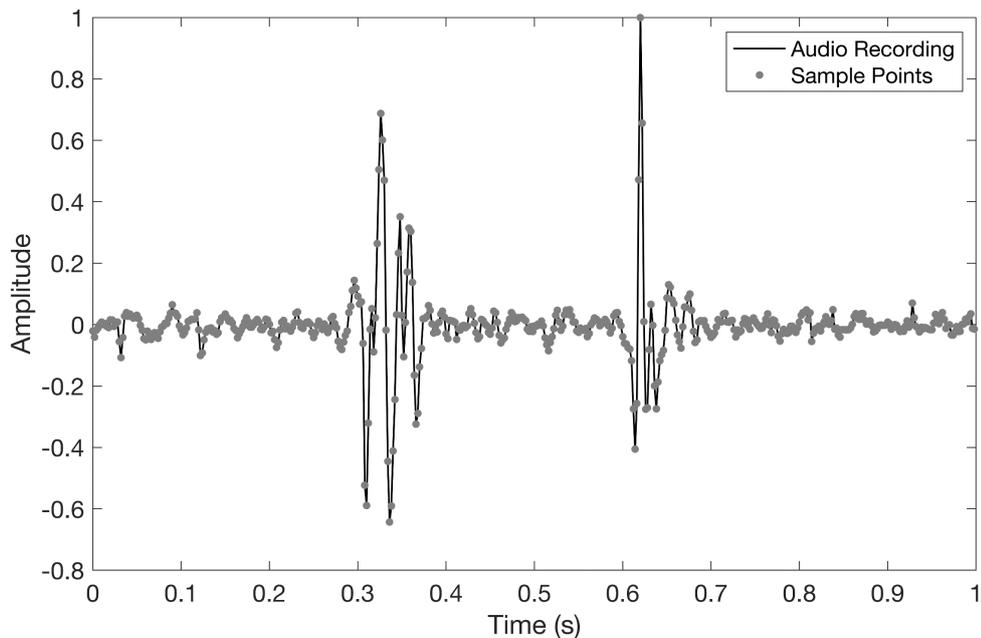


Figure 3.2: A digital audio recording of a heart sound sampled at 500 Hz and normalised to zero mean and unit maximum. The theoretical analogue signal is represented by the black line and the actual sample points are shown in grey. This heart cycle was taken from a recording from a healthy individual as part of the study described in Chapter 4.

Recordings of heart sounds are audio recordings of the sounds produced as blood is pumped through the heart. These sounds can be recorded in a multitude of ways, and the term phonocardiogram (PCG) is often used for recordings of cardiac sounds made with high-fidelity equipment in controlled conditions [61]. The recordings that we will be dealing with in this research have been made using physician-grade electronic stethoscopes by cardiologists, doctors, or medical students from real patients and made in realistic (clinical) situations. Therefore, in this research, the term PCG has been reserved to those recordings (if any) made using specialised, high-fidelity equipment for the explicit purpose of heart sound analysis.

### 3.2 Digital Signal Processing Concepts

At this point it will be useful to introduce some of the core concepts used in digital signal processing (DSP). This introduction is by no means comprehensive and does not cover the mathematical proofs and derivations; the interested reader is referred to the substantial amount of introductory literature on DSP, for example "Introduction to Digital Signal Processing and Filter Design" by B. A. Shenoi [62]. The discussion presented in this section provides the background theory for the DSP techniques applied in the rest of this work.

### 3.2.1 Block-based Processing

A digital heart sound recording, for instance the single heart cycle shown in Figure 3.2, can also be described as a *time-series*. This means that each of the data points is indexed by time and the order of the samples is a critical part of the information contained in the recording. When analysing time-series data we are often interested not only in what is happening sample-by-sample or in the entire signal all at once, but also in how the signal is behaving in *local regions*. For instance, in heart sound analysis we are very interested in how the frequency information of the signal is changing over time, as diagnostically important information can be derived from this. The concept of frequency does not make sense for individual samples and, as discussed in the following section, we lose all time information when we determine the frequency content of the entire recording. Block-based processing is one possible solution to this dilemma. Block-based processing is introduced here as a general DSP concept because it plays a fundamental roll in the methods described further on in this thesis.

In *block-based processing* signals are processed in consecutive blocks as opposed to sample by sample or the signal all at once. The signal is divided into blocks of *frame length*  $N$  and a specific number of samples between the start of blocks, referred to as the *hop size*  $H$ . To improve the time resolution the hop size  $H$  is generally chosen so that there is a certain amount of overlap between consecutive blocks. Hop sizes can also be expressed as an *overlap ratio* ( $\varphi$ ), an indication of how much overlap exists between consecutive blocks.

$$\varphi = \frac{N - H}{N} \quad (3.1)$$

The frame length  $N$  and hop length  $H$  need to be chosen appropriately depending on the application of the algorithm in question. Longer analysis blocks and smaller hop lengths will reduce processing time and give fewer output samples, but important details of the signal might be lost. Shorter analysis blocks and a greater amount of overlap between blocks requires longer processing times and give a higher number of output samples but provide a greater resolution of the signal being analysed. Some general heuristics are to choose the frame length to be at least the length of the smallest audio events of interest and to choose a hop size that is greater than or equal to half the

frame length [58]. That is, the overlap ratio should be at least 50%. Figure 3.3 demonstrates block-based processing on a short audio sample. In this figure only the 8th, 9th, and 10th, processing blocks are shown explicitly with their centre circles (time indices) shown with filled in circles. The root mean square of each of the blocks is shown at the bottom of the figure with the block indices and the newly calculated time indices, note that the root mean square is only used as an example and could be substituted for any number of metrics.

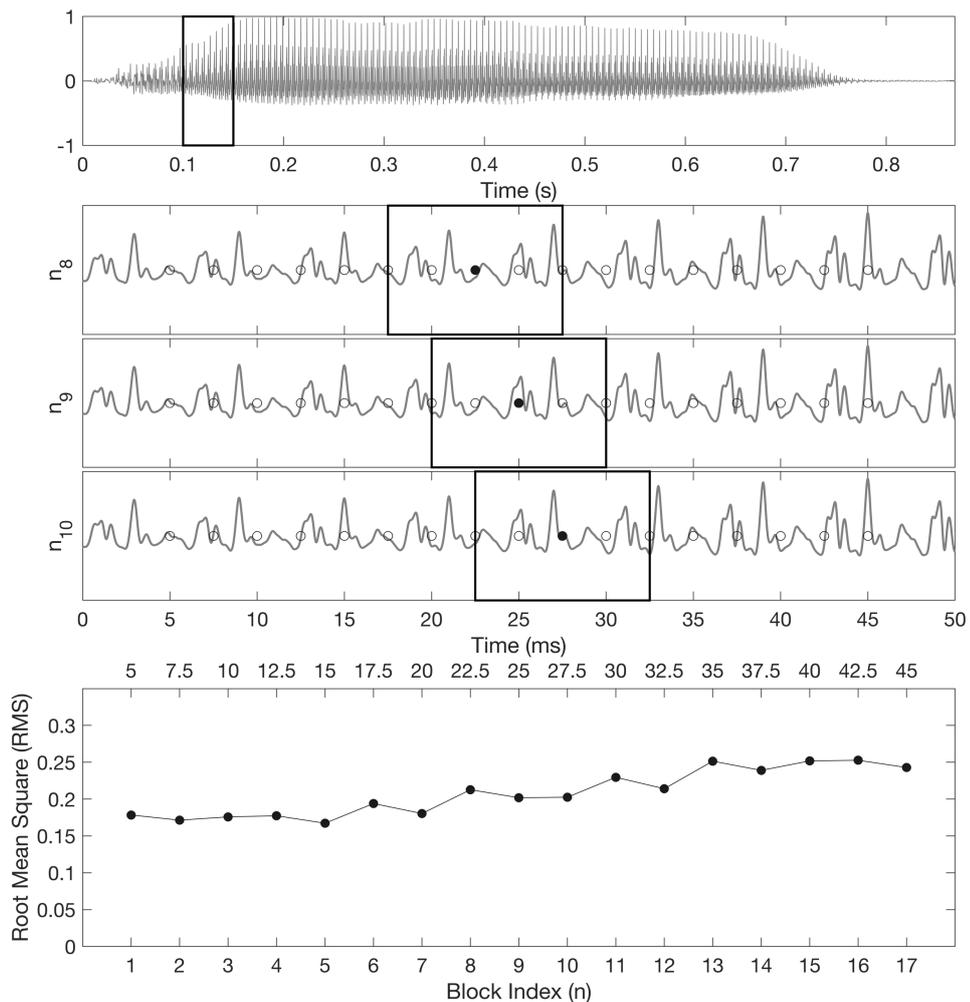


Figure 3.3: Block-based analysis of a trumpet note (shown at the top of the figure). A focused view of the marked 50 ms segment is used to illustrate block-based calculation of the RMS. The middle points of each of the analysis blocks are used as the time-indices for the RMS vector.

There are many technical and computational advantages of using block based processing [58], but from an audio processing standpoint this method is especially important in the computation of time varying features that do not have a natural time dependence [63]. Good examples of these are statistical features, for instance the

standard deviation, that will always equal zero if calculated sample by sample or a single number completely devoid of time varying information if calculated for the entire signal. Given appropriate frame and hop sizes, the way in which these features change over time can be accurately estimated and represented.

### 3.2.2 The Fourier Transform

When looking at a time varying signal, for example the trumpet recording shown in Figure 3.3, interesting (diagnostically important or significant in some other way) information is often not obvious in what can be called the *time-domain*, the representation the signal amplitude against time (or equivalently, samples). The heart sounds are one example of such a time varying signal and the time-frequency representations discussed in this section play a fundamental roll in the way that we will be visualising and analysing heart sounds in all the following chapters.

The *Discrete Fourier Transform* (DFT) is a fundamental technique used to determine the frequency contents of a digital signal; derivations, in-depth definitions, and discussion on the properties of this transform can be found in most textbooks on digital signal processing, for example the textbook by Shenoit [62]. The DFT moves a signal from the time-domain, in which the way it changes over time is represented, to the *frequency-domain*, in which the frequency components are represented. This allows us to view the frequency content (or *frequency spectrum*) of the transformed signal. An example of the Fourier transform is shown in Figure 3.4 (b), in which the *magnitude* of the positive frequency components of signal generated artificially for the sake of this example (Figure 3.4 (a) ). The frequency domain representation has no temporal information; there is no information as to *when* the frequencies occurred. Much of the useful and interesting information in a recording of heart sounds is to be found in the way that the frequency content varies over time; a so-called *time-frequency* representation is required.

### 3.2.3 Short Time Fourier Transform

The *Short Time Fourier Transform* (STFT) addresses this problem by using block-based processing, as described earlier, and computing the DFT for each of the (overlapping) blocks. The frame length of the analysing blocks has to be carefully chosen and the

uncertainty principle should be kept in mind; that is, time resolution is lost with increasing frequency resolution and vice versa, or more simply the closer we look at time intervals (shorter frame length) the less information we have to compute the frequencies. Higher overlap ratios lead to better time resolution at the cost of computational demand. The analysing blocks can also be shaped using different windowing functions, essentially weighting samples at the edges lower than those in the middle of the window, to decrease the effect of the sudden discontinuities at the edges of the analysing windows.

The STFT is illustrated in Figure 3.4 (c) and (d). Using the STFT we can obtain time varying information about the frequency of the signal under analysis. Figure 3.4 (c) and (d) show the effect of different window (or processing block) sizes and were chosen to have quite a large degree of difference (25 ms vs 200 ms). Note the large amount of spread in frequency at the lower frequencies of (c). This occurs because there are not enough samples in each window to accurately determine the frequency. There is however very fine time resolution and the discontinuities at the transitions between sine wave and chirps are clearly visible. In (d) the frequencies, especially 10 and 100 Hz, are more clearly revealed, although this comes at the cost of lower time resolution with the transitions and discontinuities becoming less clear.

Given appropriate windowing and overlap parameters the Short Time Fourier Transform provides an excellent representation of how frequency information within a signal changes over time. However, as shown in Figure 3.4, the block length needed to accurately discern low frequencies greatly reduces the time resolution of our analysis. This limitation, a result of fixed block sizes and the inherent time-frequency trade-off, makes the STFT less attractive for heart sound analysis, since the heart sounds are composed of low frequency components and rapidly changing high frequency ones.

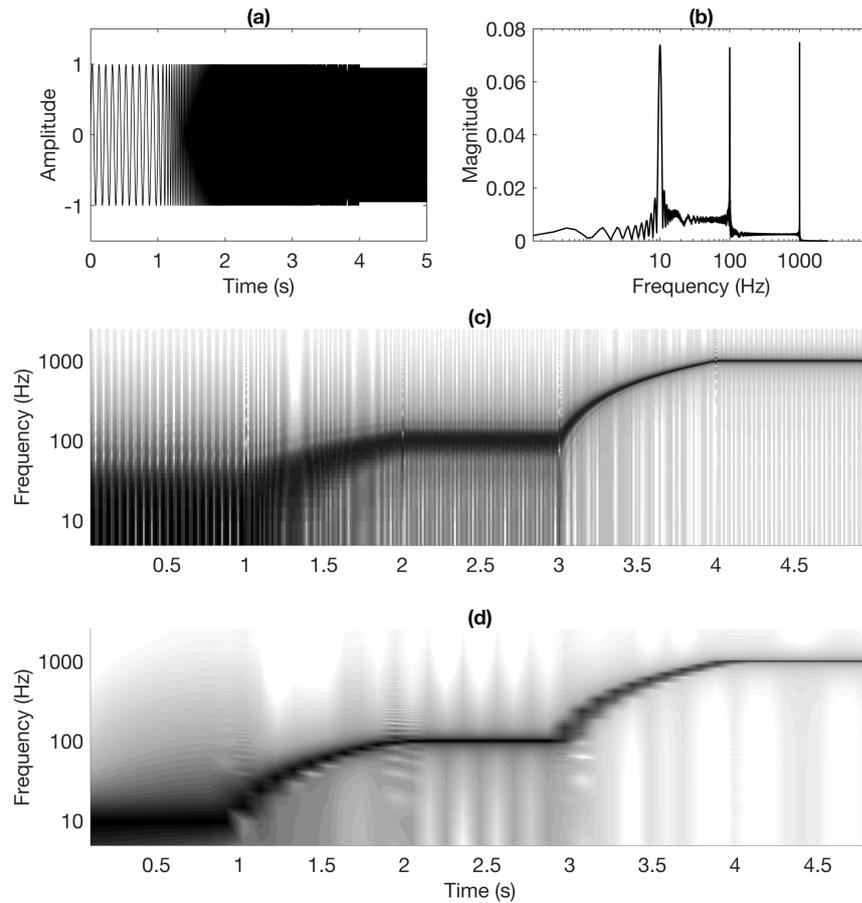


Figure 3.4: A waveform composed of one second intervals of 10, 100, and 1000 Hz sine waves connected by 1 second increasing chirp signals is shown in (a). The single sided frequency spectrum for the signal is shown in (b). STFTs of the signal are shown in (c) and (d) with window lengths of 25 ms and 200 ms respectively. In both cases an overlap ratio of 50%, 1024 frequency bins, and rectangular windows were used.

### 3.2.4 Wavelet Transform

Wavelet analysis addresses the limitations of the STFT by allowing the analysis window to “stretch” and “shrink” at different levels of frequency analysis as visualised in Figure 3.5. Wavelet analysis can be conceptualised as a generalisation of the STFT in two ways:

- The "window" can be any time-bounded function (wavelet)
- Analysis occurs at multiple *translations* and *scales*

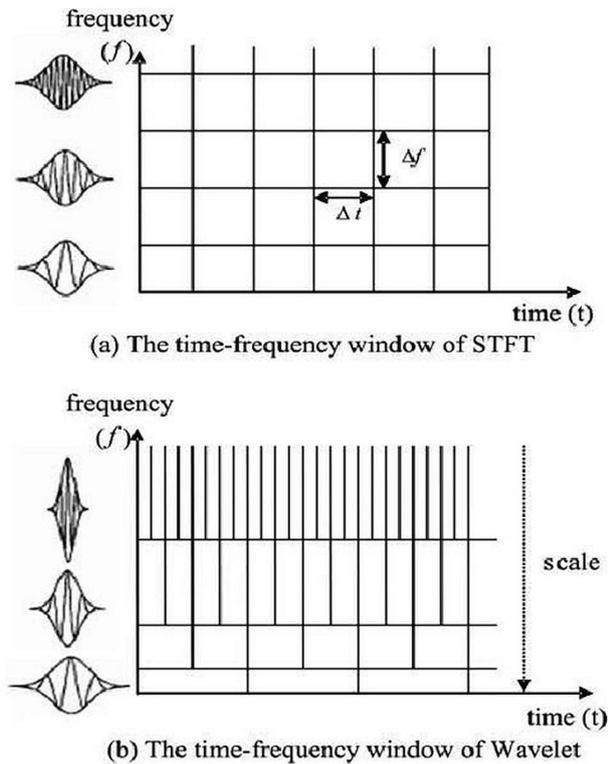


Figure 3.5: A comparison of how Wavelet analysis and STFT analysis windows are distributed in time and frequency [64].

Wavelets are short waveforms that must adhere to certain mathematical constraints in order to offer useful analysis. There are many different types of wavelets and they respond to different types and shapes of signals. A general guide is that wavelets that look like interesting parts of the signal being analysed (scale and translation invariant) will accentuate those regions of the signal. The wavelet is convolved with the signal being analysed at different *scales* and *translations*; at different time points and at different "extensions/dilations" of the wavelet. This corresponds to "sliding" (translating) the analysis window (the wavelet) across the signal and at each point calculating the convolution of the signal with the wavelet at different "stretches" (scales). In time-series analysis scale is analogous to frequency so we can think of the wavelet sliding across the signal and the convolution of the signal with the wavelet at different scales resulting in a time-frequency representation of the signal. In the continuous wavelet transform (CWT) the scale and translation parameters vary continuously, while in discrete wavelet analysis the convolutions are calculated at discrete values of the scale parameter. Due to the discrete nature of digital computing a true "continuous" wavelet transform is not possible, rather it is useful to think of the CWT as a discrete wavelet transform (DWT) with an extensive number of coefficients

calculated. Conversely, DWT can be considered a sparse version of the CWT where only the coefficients necessary for perfect (in a technical sense) signal reconstruction are calculated. For a comprehensive introduction to wavelet analysis theory and examples of applications the interested reader is referred to “The illustrated wavelet transform handbook” by Paul Addison [65].

Formally, the continuous wavelet transform is defined as

$$W_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left( \frac{t-b}{a} \right) dt \quad (3.2)$$

where  $a$  is the scaling parameter with  $a > 0$ ,  $b$  is the translation parameter and can assume any real number.  $x(t)$  is the signal under analysis, and  $\psi^* \left( \frac{t-b}{a} \right)$  is the complex conjugate of the “mother” wavelet  $\psi(t)$  scaled (stretched/shrunk) by  $a$  and translated (shifted) by  $b$ . For the CWT, when computed analytically, the scale ( $a$ ) and translation ( $b$ ) parameters are continuous (hence *continuous wavelet transform*). When computed numerically, though, we must specify a finite number of scales to analyse the signal at, this value depends on a parameter that is commonly referred to as the number of *voices per octave* and can be interpreted as the resolution of the resulting transform. Figure 3.6 illustrates the results of the CWT on the example signal first shown in Figure 3.4 using 2 different numbers of voices per octave.

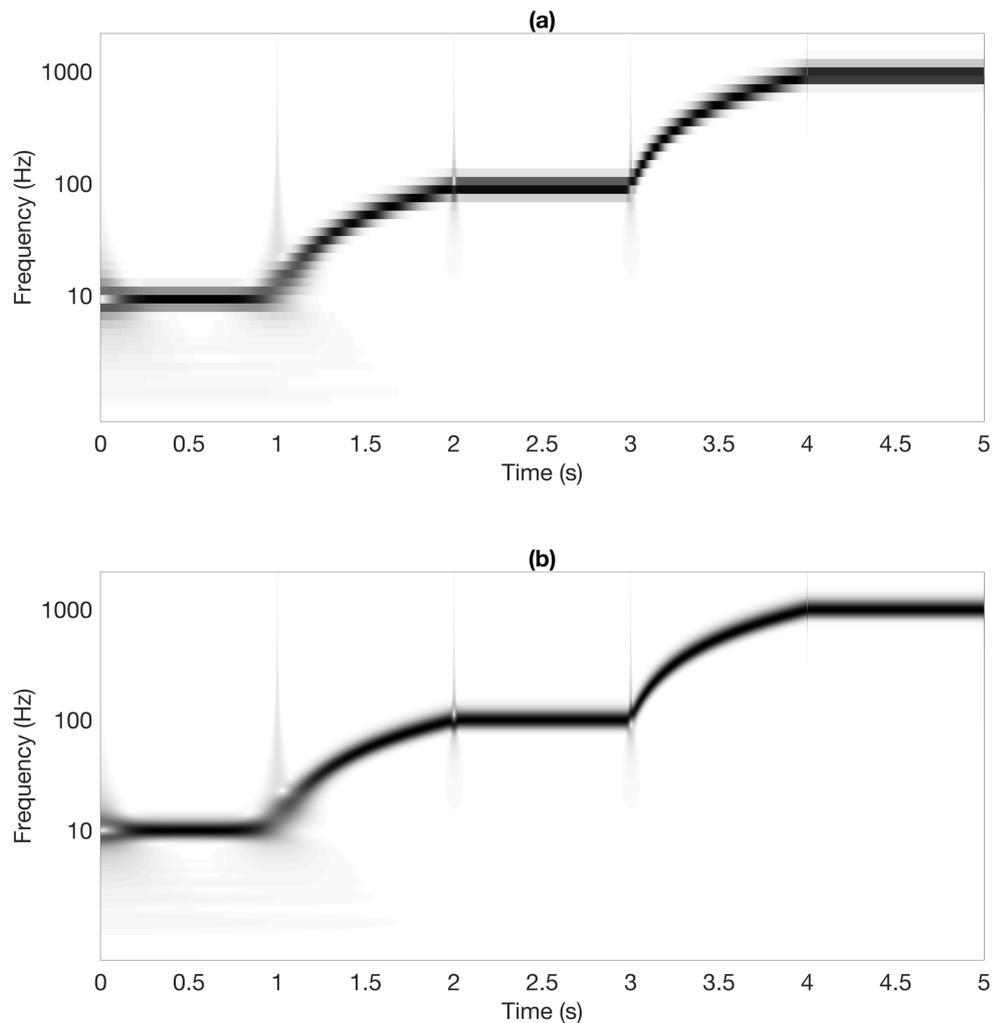


Figure 3.6: CWT of test signal shown Figure 3.4 (a). The top graph (a) was analysed using 4 voices per octave and the bottom graph (b) using 48 voices per octave. Both CWTs were performed using the Morse analytic wavelet between 0 and 2500 Hz. Frequency (y-axis) is plotted in logarithmic scale.

### 3.2.5 Signal Envelope

The envelope of a signal can be defined as the time domain boundary in which a signal is contained. The signal envelope could also be described as a simplified view of how the signal energy is changing over time. This type representation is especially useful in heart sound analysis to locate the fundamental heart sounds (S1, S2) and to determine how heart murmurs change over time. Block-based processing, discussed in 3.2.1, provide one method of estimating approximating the change of a signal over time, however these methods introduce a significant time distortion. Transformations that maintain an explicit dependency on time, e.g. the magnitude of the analytic signal, homomorphic envelope, or the absolute value of the signal are discussed in this section.

Real signals have frequency spectra that are evenly symmetric around zero, meaning that half of the signal energy is contained in negative frequencies (indicating ambiguity in the direction of rotation of the component sinusoids). For practical purposes it is preferable not to work with these negative frequencies and because of the even symmetry they do not contain unique information. The *analytic signal* is then a complex representation of the original signal which has a positive frequency spectrum. We can compute the analytic signal by first computing the Hilbert transform

$$H(t) = x(t) * (\pi t)^{-1} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(u)}{t-u} du \quad (3.3)$$

where  $H(t)$  is the Hilbert transform of the signal  $x(t)$  and  $*$  is the convolution operator. The analytic signal is then computed as  $x(t) + iH(t)$ ,  $i = \sqrt{-1}$ . The homomorphic envelope (HE) [66]–[68] is used to derive the signal envelope from the analytic signal by a process of demodulation. Homomorphic filtering is performed by low pass filtering the natural logarithm of the analytic signal and then computing the exponent. The analytic signal is used to avoid the discontinuity that occurs at the logarithm of zero. The entire process can be represented abstractly as

$$HE(t) = e^{F_{low}(\log H(t))} \quad (3.4)$$

where the low-pass filter has been represented as  $F_{low}$  and  $HE(t)$  is the homomorphic envelope of the input signal  $x(t)$ . The sequential process is represented graphically in Figure 3.7 using an example recording of a heart cycle taken from a participant presenting with a systolic murmur from the dataset described in Chapter 4.



Figure 3.7: Illustration of signal progression through the stages of HE calculation.

### 3.2.6 Pre-processing Steps

Pre-processing refers to signal processing steps that are performed at the onset of an algorithm. In general, the pre-processing steps are performed for several reasons, including:

- Improving Signal-to-Noise Ratio (SNR) by filtering out noise artefacts and isolating regions (usually in frequency, but possibly in other spaces) that are more likely to contain more of the signal and less of the noise.
- Normalisation of input data to account for different sources of data and to improve the generalisation of the algorithm.
- To improve computational performance by, for instance, down sampling the signal to decrease the number of samples the algorithm must process.

There exist many more applications of pre-processing that are more application specific, for instance in heart sound recording analysis the process of *heart sound segmentation* is usually performed as a first step and could easily be under the term pre-processing.

### 3.2.6.1 Filtering

Digital filtering is the act of attenuating (diminishing or reducing) signal components that are likely to be unrelated to the signal we are attempting to investigate. A wide range of different techniques can be classed under this label. Frequency filters attenuate certain frequency bands and can be classified as low pass, high pass, bandpass, or bandstop, based on the frequencies that they attenuate.

A common pre-processing step in heart sound processing is to bandpass filter the audio recording to limit the amount of out-of-band noise corrupting the signal. The frequency range of heart murmurs is less well defined and may range from around 100 Hz to over 1000 Hz. Thus, the selection of an appropriate low pass cut-off frequency for the band pass filter can be problematic. In this work the data is only high pass filtered using a 2<sup>nd</sup> order Butterworth filter with a cut-off at 20 Hz to remove very low frequency noise. An example of an audio recording heavily corrupted by low frequency (and almost completely inaudible) noise and the output from the high-pass filtering is shown in Figure 3.8.

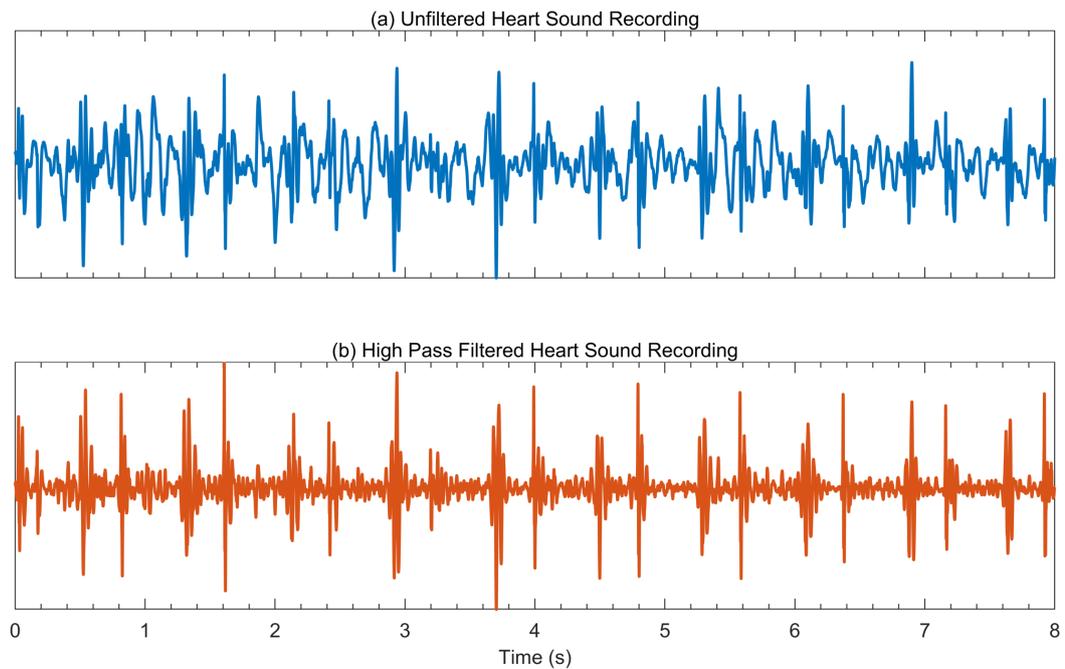


Figure 3.8: A motivating example for the application of high pass filtering as a pre-processing step for heart sound analysis. The audio recording, taken from the Physionet heartsound database [69], contains a large amount of low frequency noise that completely obscures the heart sounds in the recording. Filtering was performed using a 2nd order Butterworth filter with a frequency cut-off of 20 Hz. Phase distortions were corrected by applying the filter forwards and backwards, effectively resulting in a 4th order zero-phase filter.

### 3.2.6.2 Artefact removal

Another form of filtering, termed artefact removal, attempts to remove specific unwanted occurrences or events in the signal that are application specific. In the case of heart sounds, a common source of loud clicks and spikes in the audio recording is slight, abrupt movements of the stethoscope diaphragm, especially against hair on the patient's body. If the artefact has a higher intensity than the sound produced by the heart it can easily cause the feature extraction step to fail, especially if the signal is normalised to the peak intensity. A spike removal algorithm, proposed by Schmidt et al. [68] and by Springer et al. [70], uses the median of windowed portions and sets any spikes that have peaks that greatly exceed (greater than 3 times) the median of the largest peaks across the windows to zero with spike widths determined by zero crossings.

### 3.2.6.3 Normalisation

Many factors can influence the amplitude of the recorded sounds, the type of stethoscope used, the location on the chest wall where the recording was made, and physical characteristics of the patient's chest to name a few. It is advantageous to design

a processing system that can adapt to different input conditions. Normalisation allows us to scale an input signal to a pre-defined maximum amplitude. In this work the arithmetic mean (or DC offset) is removed in conjunction with the normalisation process; this has some mathematical consequences, such as the standard deviation and root mean square being equivalent.

A commonly used and simple method of normalisation is to divide all the samples of the signal by the absolute maximum of the signal, thus assigning a value of 1 to the largest sample and scaling all other samples to ratios of this value.

$$x_n(t) = \frac{x_i(t)}{\max(|x_i|)} \quad (3.5)$$

where  $x_n$  is the normalised signal, and  $x_i$  represents the input signal. This method is employed to ensure the signal occupies the full range of possible values used for audio playback. For instance, playback using the MATLAB function "sound" *clips* (reduces to 1) any samples that fall outside 1 and -1. Without normalisation any samples that fall outside this range would simply saturate the playback system and any relevant information in the recording (for example the melody of a song or spoken words) would be lost. However, scaling the signal in this way is sensitive to large amplitude noise artefacts such as clicks and scratches in which case the signal gets scaled to the amplitude of the noise instead of the signal. It should also be noted that all samples in the signal are scaled and this includes any background noise present in the signal.

#### 3.2.6.4 Standardisation

A related procedure to normalisation, from statistics, is called standardisation and involves subtracting the arithmetic mean ( $\mu$ ) and dividing by the standard deviation (the square root of the variance) of a dataset.

$$x_s = \frac{x_i - \mu}{\sqrt{\frac{1}{N-1} \sum_{i=1}^N |\mu - x_i|^2}} \quad (3.6)$$

The standardised values,  $x_s$ , are referred to as *z-scores* in statistical nomenclature and represent the amount of standard deviations a sample is from the mean of the distribution.

### 3.3 Digital Heart Sound Analysis

Using digital computers to analyse recordings of heart sounds was attempted as early as 1962 by Gerbarg *et al* [71], although the comparatively primitive digital computers of the day meant that practical usage was limited. Recently, advances in digital signal processing techniques and the enormous increase in the processing power and availability of computers have led to a substantial increase in the amount of work being done in the field of computer assisted auscultation. A typical flow diagram for computer assisted auscultation is detailed in Figure 3.9.

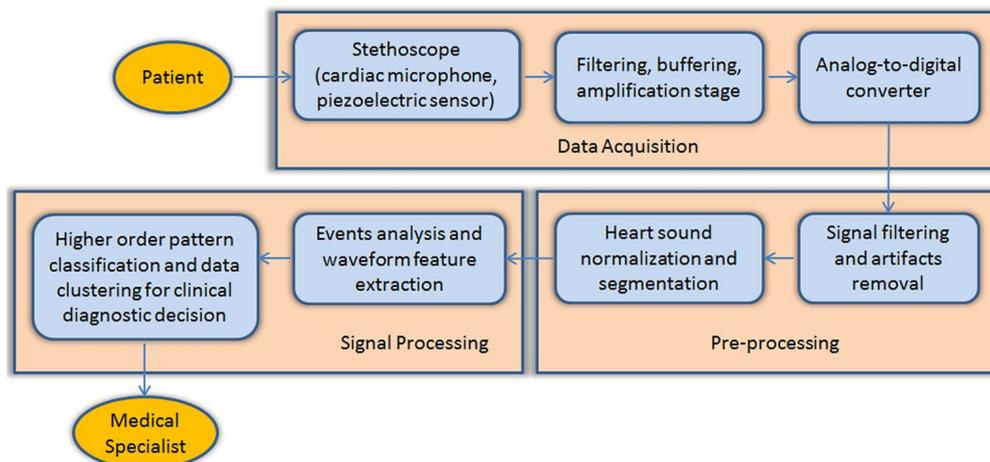


Figure 3.9: Flow chart detailing the typical approach to computer assisted auscultation [25]

As shown in Figure 3.10 this literature review has focussed on the typical signal processing steps of computer assisted cardiac auscultation literature. We start off by summarising some interesting findings of two literature survey articles published recently.

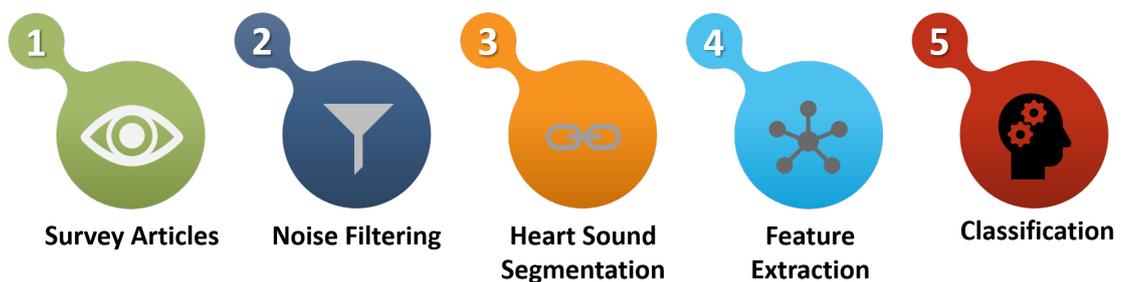


Figure 3.10: An outline of this digital heart sound analysis literature review.

#### 3.3.1 Survey Articles

A 2009 survey reviewed 39 papers published between 1995 and 2009 on PCG audio processing techniques [72]. This study, which focused on “heart sound analysis and feature extraction” and “automatic pathology classification”, identified three major

problems that affected most of the reviewed papers. These problems were the absence of a set of well-accepted features for extraction, badly described datasets, and the absence of clinical validation.

A 2015 review of the electronic stethoscope investigated recent advances in the field [25]. This survey focused on three main aspects, namely digital signal processing techniques, state-of-the-art electronic stethoscope products, and smartphone applications that interface with electronic stethoscopes. The survey of digital signal processing techniques included an in-depth summary and evaluation of articles pertaining to “heart sound denoising” and “heart sound classification”. The survey summarised the findings of nine studies in which Aortic Stenosis (AS), Aortic Regurgitation (AR), Mitral Regurgitation (MR), and Mitral Stenosis (MS) were classified using mainly support vector machine (SVM) and neural network (NN) classification techniques. The results of this meta-analysis are shown in Table 1 and although encouraging for the use of machine learning methods and digital signal processing techniques in heart sound analysis these cannot be considered as representative of the clinical performance of these classifiers as discussed below.

Table 1: Mean classification results for most commonly considered diagnoses [25]

<b>Disease</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>
<b>Aortic Stenosis</b>	89.8	98.0
<b>Aortic Regurgitation</b>	88.4	98.3
<b>Mitral Regurgitation</b>	91.0	97.5
<b>Mitral Stenosis</b>	92.2	99.3

### 3.3.2 Noise Filtering

The quality of heart sound recordings relies on many variables, the pressure applied to the stethoscope, ambient noises, movement of the stethoscope during recording, other physiological sounds such as digestive sounds and lungs sounds, etc. This means that heart sound denoising is a very important pre-processing step in the analysis of heart sounds. Falk and Wai-Yip Chan and Ramos et al. [73], [74] investigated modulation filtering as a method for separating noise and heart and lung sounds. Kumar et al. [75] suggested using the periodic nature of the heart sounds and physiological inspired criteria to construct a heart sound template signal which is then used in the second phase for non-cardiac sound detection. The first stage of this method is computationally

demanding. These methods have both been tested by injecting simulated noise that is similar to noise artefacts that would be encountered in a clinical setting into the recordings. The construction of a prototype heart sound signal is also investigated by Syed et al. [76] and Syed and Guttag [77] using clustering methods. This technique suppresses transient noises and provides increases robustness against time-warping in the signals being analysed.

### 3.3.3 Heart Sound Segmentation

Segmentation is the process of identifying the different phases of the cardiac cycle by identifying the fundamental heart sounds (S1 and S2) in a heart sound recording. Determining the locations of the fundamental heart sounds is very often the first step during cardiac auscultation because of the context provided by correct knowledge of the timing of sounds in the heart cycle. Segmentation methods can broadly be divided into methods that use an external signal to segment heart cycles (dependent segmentation) and methods that attempt to use only the heart sound signal and digital signal processing techniques (independent segmentation). Because of the complex and non-linear nature of heart sounds and the presence of many different sources of noise, successful segmentation of heart sound signals is not an easy task. The timing of any heart murmurs present in the heart sound provides very important information on any possible pathologies and the segmentation step is the most important step towards identifying the timing. Based on the review of the literature the gold standard for segmentation accuracy can be either the opinion of multiple experts presented with a heart sound recording, or simultaneously recorded high quality ECG recordings.

Dependant segmentation refers to methods of segmentation that involve an extra sensor to determine the timing of S1 and S2. In all the papers reviewed that utilised a dependant segmentation scheme the extra sensor used was an electrocardiogram (ECG). El-Segaier et al. [78] used the T-waves and R-waves from a simultaneously recorded ECG were used as references for the detection of systole and diastole. A 100% correct identification of S1 is reported, and 97% correct identification of S2. Ahlstrom et al. [79] aimed to increase the performance of an ECG-gated localization method using ensemble averaging (EA) to emphasize occurrences of S1 in PCG signals.

Methods that do not require an external signal to segment individual heart cycles have also been extensively investigated. A 1997 study by Liang et al. [80] reported 93% correctness in automatic identification of S1 and S2 when analysing 515 cardiac cycles from infants and children aged up to 14 years using a normalized Shannon energy based approach. Their results, although based on a small dataset from a paediatric population, are encouraging. A different approach used by Barabasa et al. [81] attempts to detect S1/S2 by adapting an algorithm originally developed for detecting the beat of music. The authors tested this method on 18 recordings consisting of 368 heart cycles and reported a 100% sensitivity and 97.3% specificity in identifying the first and second heart sounds. A summary of the two main categories of segmentation methods, including some possible advantages and disadvantages in terms of this project, is shown in Table 2.

Table 2: Comparison of dependant and independent segmentation methods

	Dependant Segmentation	Independent Segmentation
Overview	S1 and S2 locations determined using data from provided by an external sensor.	S1 and S2 locations are determined using only the data contained within the heart sound recording.
Specific Methods	ECG R-wave, Carotid pulse wave	Shannon Energy, Shannon Entropy, Simplicity/Complexity methods, Beat Tracking, Wavelet Envelope, Machine Learning Classifiers
Advantage	Provides very accurate and reliable segmentation.	Allows for simpler hardware design, heart sounds recordings can be segmented retroactively.
Disadvantage	Requires another signal to be simultaneously recorded and synchronised. More complex hardware design.	More complex software designs require longer processing times, while also not being as reliable as dependant systems

Presented as the current state-of-the-art segmentation algorithm, the 2016 Physionet heart sound classification challenge [82] released work done by Springer *et al.* [70], based on earlier work by Schmidt et al. [83]. This algorithm uses logistic regression and a hidden semi-Markov model (introduced extensively in Chapter 5) to determine the locations of the fundamental heart sounds and the phases of the cardiac cycle in a heart

sound recording. Simultaneously recorded ECG data are used to annotate the training set for the Markov model. Logistic regression, a statistical approach for modelling the relationship between a binary dependant variable and multiple independent, or explanatory, variables [84], is then used to train the model using features extracted from the homomorphic envelogram, the Hilbert envelope, and the Wavelet envelope. The Markov model is a probability model that assumes the system is always in one of several finite states, with all relevant events modelled as transitions from one state to another [85]. The Markov model used by Springer et al. [70] is referred to as a “semi” or “duration-dependant” Markov model, because it also takes into the account the amount of time that has been spent in the current state when determining the current state. As can be seen in Figure 3.11 S1, systole, S2, and diastole are the possible states that the Markov model can occupy. Transitions occur when the instantaneous probability of the system being in the next state exceeds that of it being in the current state.

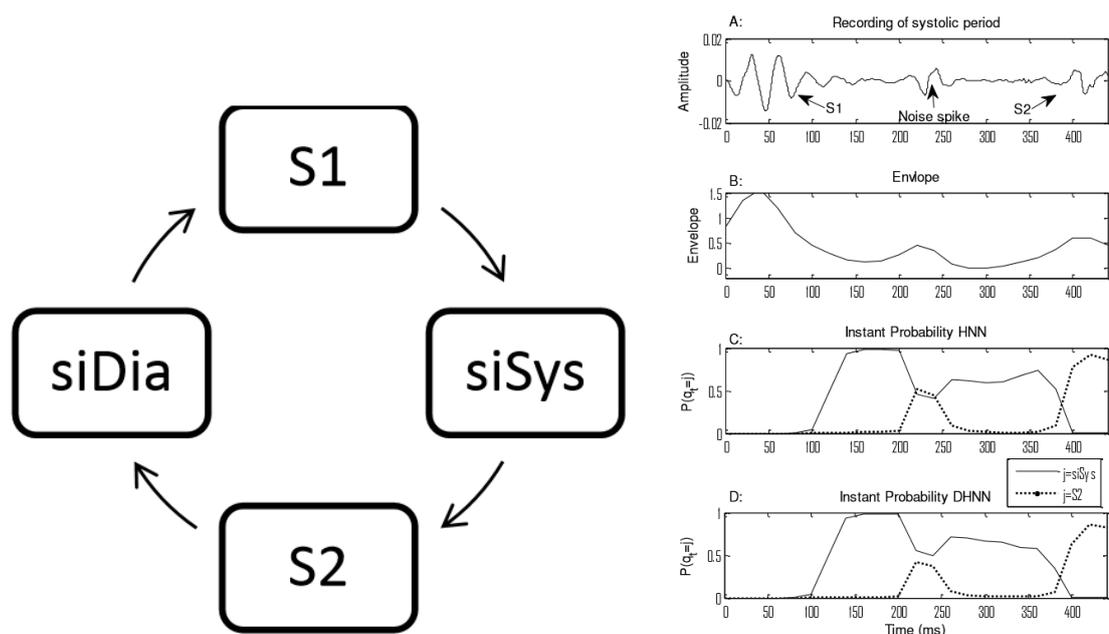


Figure 3.11: A Hidden semi-Markov model (HSMM), also called duration-dependent HMM [68], [70], [86]. The graphs illustrate that the semi-Markov model has a higher tolerance to noise artifacts than the standard Markov-model as evidenced by the fewer crossings between probability time-series, particularly at around time 220 ms.

### 3.3.4 Feature Extraction

Feature extraction is an important step in the correct classification of heart sounds. This step normally proceeds after signal denoising and individual heart cycle segmentation has been achieved, although some studies present methods that do not require

segmentation before feature extraction [87], [88]. In the articles that have been reviewed the most popular method of feature extraction is wavelet decomposition [88]–[92]. A comparative study of different feature extraction methods was conducted by Ahlstrom et al. [91]. A total of 207 different features were extracted. These features were reduced to a set of 14 features using an optimization algorithm and the classification performance, in differentiating between systolic murmurs, of the different feature sets were compared. Correct feature selection plays an extremely important role in classifier performance and as such this step is integral to the success of any heart sound classification. Current feature extraction methods do not generally approach the problem from a physiological perspective, but rather treat it as a computational problem, basing feature selection on classification results. This might be the best solution if a computer is meant to make the final diagnosis but may not be the best approach if the goal is to provide objective, relevant information to a physician.

### 3.3.5 Classification

Different types of classifiers have been used in the classification of heart sounds. Yuenyong et al. [88] made use of a neural network bagging predictor to classify heart sounds as normal or abnormal. They report an accuracy of 92% when assessing their algorithm using 10-fold cross-validation on a dataset of heart sounds from 57 individuals (12 normal and 45 abnormal). A support vector machine was used to classify heart sounds as normal, mitral stenosis, or pulmonary stenosis by Güraksin and Uguz [89]. The algorithm, which used discrete wavelet transform and Shannon entropy derived features, achieved a classification accuracy of 96.6% on a balanced dataset of 60 heart sounds. Grow-and-Learn (GAL) and Multilayer Perceptron-Backpropagation (MLP-BP) were used by Gupta et al. [90] to identify heart cycles as normal, systolic murmur, or diastolic murmur. The authors found the performance of the two networks to be similar, achieving 97% and 96.5% accuracy respectively when tested on a dataset of 201 recordings from 41 volunteer. The authors do not clarify whether recordings from the same participants appeared in both the training and testing sets.

### 3.3.6 Summary of Heart Sound DSP review

Overall the main goal so far in the analysis of heart sounds has been to produce a machine learning algorithm that is able to classify different pathological conditions

based on the heartsound recording. The literature reviewed regarding automatic heart sound auscultation algorithms report promising results for the automatic identification of pathological heart murmurs through phonocardiogram analysis. All the reviewed algorithms were tested in laboratory conditions, although on recordings made in clinical settings. Presently it is unclear if these methods would perform adequately in real world clinical situations.

The features most commonly used in the classification of pathological heart sounds in the reviewed papers are those derived from wavelet analysis, although other time-frequency methods (STFT, S-transform) have also shown success. Envelope methods, such as Shannon energy and Shannon entropy, have also proven to be successful features. There is no clear consensus on which features should be used and, even within a class of features e.g. wavelet, the exact features used vary wildly. Neural networks and support vector machine classifiers have been used with success in discriminating both between normal and abnormal recordings and between different classes of pathologies.

A major limitation in all the reviewed articles is the lack of clinical validation. None of the datasets used are representative of all possible heart sounds and the natural variability in the signal combined with the variability introduced by different recording techniques makes such a dataset infeasible. None of the articles encountered in this review have demonstrated wide generalisability and thus it is not precisely clear how they could be integrated into clinical support systems. This is not meant as a criticism of current research but rather an observation of the complex multi-layered nature of cardiac auscultation and medical diagnosis in general.

The diagnostic process of cardiac auscultation involves more than just the act of listening to the sounds produced by the heart. The investigating physician should, before even starting the process of auscultation, have made certain observations and have a hypothesis about what they are expecting to find in the heart sound. Any system that tries to emulate the procedure of auscultation needs to consider that the amount of information contained in the heart sounds, and certainly in a heart sound recording taken at only one site, may not be sufficient to identify the underlying cause of any abnormalities in that sound. The systems of automatic cardiac auscultation reviewed in the literature attempt to classify heart sound recordings into categories of pathologies

based simply on the information contained within the recordings. Based on the clinical approach to cardiac auscultation and physical examination this assumption is not correct; physicians tend to use a process of auscultation rather than listening just once at one auscultation location, while also taking other signs and symptoms into account. The research presented in this review does however point towards a significant amount of information being present in the heart sound recording and the value and utility of analysis of the heart sounds as part of the diagnostic process.

A novel approach to the overall analysis of heart sound recordings is proposed. Instead of attempting to directly ascertain a disease condition based solely on the heart sound recording, the proposed algorithm will attempt to show the acoustic features present in the heart sound. The approach that will be followed in this project is focussing the goal of computer assisted cardiac auscultation onto clarification of clinically relevant signs present in the heart sounds. Instead of acting as a decision-making system, the proposed system will attempt to clarify information contained within a recording of a heart sound using digital processing techniques and to communicate these findings in a way familiar to the user. Essentially the algorithm will attempt to perform, in an objective way, the subjective psychoacoustic part of the heart sound analysis and present the results to the clinician in a form that adds value to their overall diagnostic process. While clinical skills and intuition are not easily converted into a computer algorithm, the same is not true for pattern recognition.

### 3.4 Research Questions

Based on the review of the literature and the identified research gap, this research will address the following questions:

1. What are the features that expert cardiologists try to ascertain during cardiac auscultation and how can these be converted to mathematical models?
2. Can digital signal processing and machine learning techniques be used to extract these features from digital recordings of heart sounds?



In addition to the digital signal processing techniques discussed in the previous chapter we now introduce *Bayesian probabilistic modelling* as a framework for constructing computational models of the heart. *Probability theory* is a branch of mathematics that relates to the study of quantities that are not (or cannot) be represented as deterministic functions because of inherent randomness and/or uncertainty. Using probability theory these quantities (which we refer to as *random variables*) can be represented as probability distributions, a representation of the relative chance that the random variable has of assuming a given value. For example, the outcome of dice rolls or the weather in Auckland, New Zealand tomorrow theoretically involves too many variables and unknowns to construct useful deterministic models (in which the same input always produces the same output), rather we can represent the outcome of these two examples by establishing the chance that a given event will occur (e.g. that we roll a single six in five rolls of a dice, or that it rains during the afternoon).

#### 4.1 Probabilistic modelling

*Probabilities* are clearly defined mathematical quantities that must adhere to the axioms (fundamental rules) of probability. In formulating the probability axioms we use notation from *set theory* and probability theory. We start by defining a *set*  $S$  that contains all possible outcomes of the situation we are studying.  $S$  is comprised of subsets which we call *events*, denoted as  $E_1 \dots E_n$ ; events can be individual outcomes or combinations of outcomes (e.g. rolling a six in a single dice roll or rolling five sixes in ten dice rolls). Furthermore, we stipulate that each event  $E$  in set  $S$  is assigned a probability (i.e. a relative chance of occurring) denoted as  $P(E)$ . To clearly define the axioms we make use of the following notation commonly used in set theory:

- $\subset$  refers to a *subset*, as in  $A$  is a subset of ( $\subset$ )  $B$
- $\cup$  refers to the *union* of two sets, as in any elements in  $A$  or ( $\cup$ )  $B$
- $\cap$  refers to the *intersection* of two sets, as in the elements in both  $A$  and ( $\cap$ )  $B$

Now, using the set  $S$ , and the set theory notation we can state the axioms of probability as:

1. Probabilities are positive real numbers greater than zero

$$P(A) \geq 0 \text{ for all } A \subset S \quad (4.1)$$

2. The probability of the entire set of possible outcomes is 1

$$P(S) = 1 \quad (4.2)$$

3. If events are *mutually exclusive*, that is they have no intersection, then the probability of the union of the events is equal to the sum of the individual probabilities

$$\begin{aligned} \text{If } A \cap B = \emptyset, \\ \text{then } P(A \cup B) = P(A) + P(B) \end{aligned} \quad (4.3)$$

In simpler terms we could say that (1) probabilities must non-negative real numbers, (2) the probability of the entire set of possible outcomes must be 100%, i.e. considering all the possible outcomes something must occur, and (3) if events do not share any information their probabilities are additive. We represent the probability of an event  $P(E)$  using a *probability distribution* which is a function that maps all possible values the event could take to the corresponding probability of the event taking that value. Figure 4.1 shows example probability distributions for the rolling of one, two, three, and four six-sided dice. The x-axis represents the possible values that the total of the dice could take, and the y-axis represents the probability density of each of the possible values.

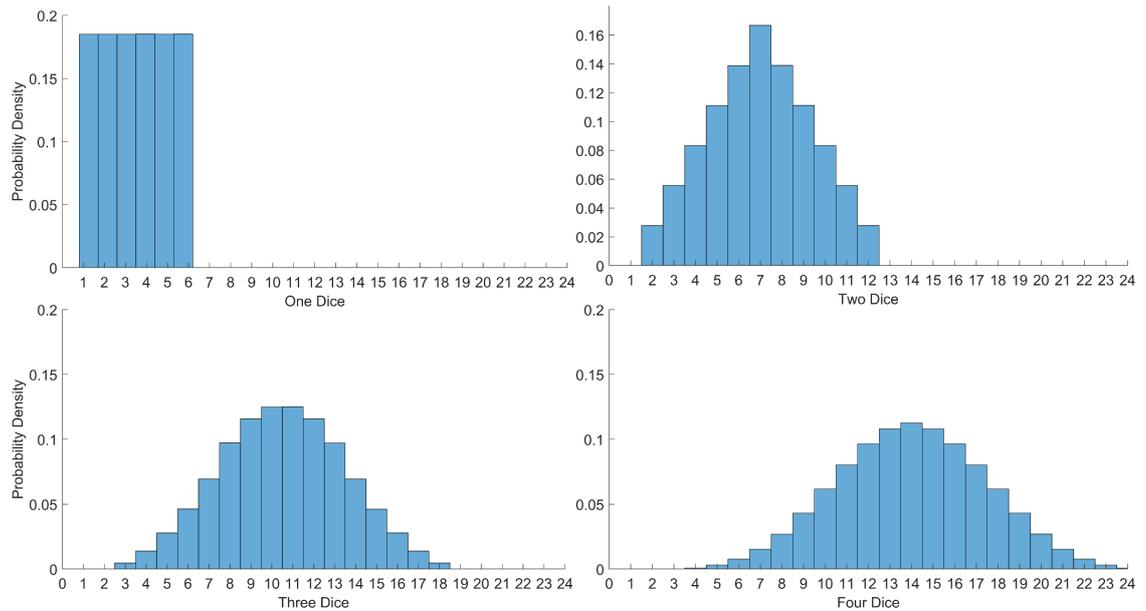


Figure 4.1: Examples of discrete probability distributions for the outcome of one and more dice rolls.

Bayesian probabilistic modelling extends this probabilistic thinking to include any uncertainty we have about a given phenomenon we are observing. For instance, we might want to know the airspeed (speed relative to the surrounding air) of an unladen swallow. The question itself is rather vague, for instance we have not specified whether we are referring to a European or an African swallow, and our lack of expert knowledge regarding swallows makes it impossible to give an “exact” answer (if such an answer even exists). We can, however, easily create a distribution of possible values using our understanding of the world. It does not make sense for swallows’ airspeed less than zero (they don’t fly backwards) and it seems unlikely that swallows travel near the speed of sound (based on the lack of sonic booms during migration). The more knowledge we have about swallows the better this estimate could be, knowing that swallows migrate about 10 000 km in around six weeks we could estimate the expected speed as a function of hours that the swallows are able to fly per day. Finally, if we had the right instrumentation we could perform experiments to directly measure the airspeed of a given sample of swallows and infer the population airspeed, but even in this case our instruments and research methodologies would have certain levels of error and there would be natural variation in swallow airspeed, thus we would still need to answer this question using a *distribution* of possible values.

Probabilistic modelling consists of quantifying our knowledge and uncertainty about a given process/system/phenomenon by constructing a probabilistic model (represented as a probability distribution) using both our *prior* knowledge about the underlying system and data collected from the system. This model can be referred to as a *generative* model as it represents (to a greater or lesser degree) the system that is responsible for generating the data. We can then use this model to draw inferences about the system we are investigating. This process is illustrated in Figure 4.2 for generalised statistical inference.

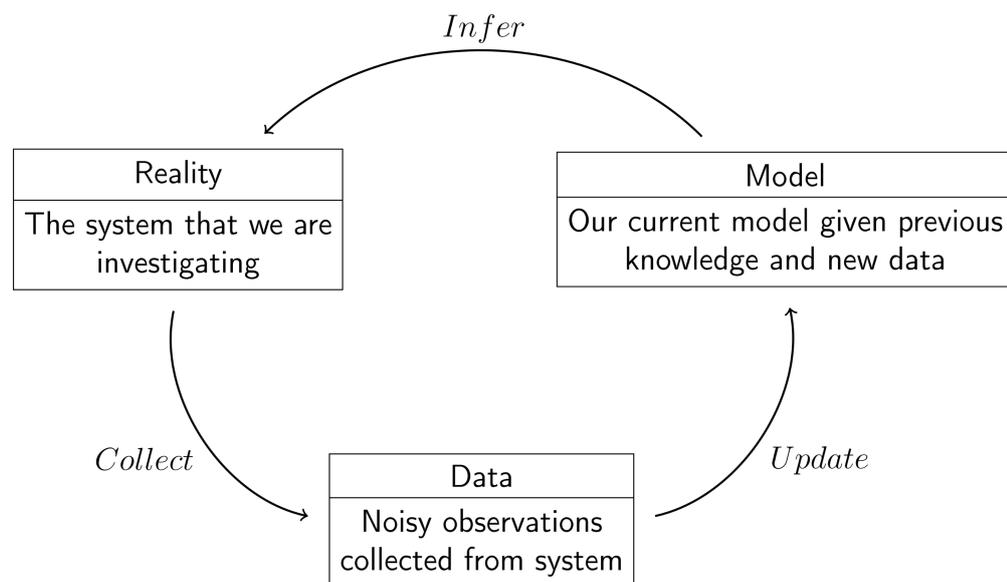


Figure 4.2: A model of statistical inference. In Bayesian modelling the model includes both observations and previous knowledge. Inclusion of this "previous knowledge", referred to as the *prior*, has been contentious due to the difficulties of arriving at philosophically acceptable and mathematically tractable "prior" models.

$$P(\text{model}|\text{data}) \propto P(\text{data}|\text{model})P(\text{model})$$

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}$$

Our knowledge after  
conducting the  
experiment.

The data from the  
experiment and our  
knowledge of the  
design of the  
experiment.

Our knowledge  
before conducting  
an experiment.

Figure 4.3: A representation of how Bayesian inference can be used to "update knowledge" using a simplified proportionality representation of Bayes' theorem.

Bayes' theorem, the namesake for this approach, is the rule by which we derive our updated model given the data. An illustration of a simplified representation of Bayes'

theorem as a proportionality is shown in Figure 4.3, although the formula is usually represented as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (4.4)$$

where  $A$  and  $B$  represent two dependent phenomena (that is to say  $B$  provides information regarding  $A$  and vice-versa). This formula, a logical consequence of the axioms of probability theory, has powerful epistemological implications if probabilities are allowed to represent "levels of certainty" about an event. It formulates mathematically how to change "current knowledge" in the light of new observations. Although quite acceptable in fields such as genetics and medicine where widely accepted "prior models" exists, the use of Bayes' theorem in this way is more contentious where the prior is influenced by subjective opinion to a greater degree. While classical statistics involve the development of techniques with as few assumptions as possible [93], the Bayesian paradigm is focussed on making assumptions as explicit as possible and analysing the data in light of these assumptions, possibly even comparing the outcome of multiple different sets of assumptions.

The Bayesian framework provides us with a methodology of creating and sequentially updating models of real-world phenomena in which there exists uncertainty and randomness. Probabilistic models are attractive in the design of complex diagnostic support systems as they allow us to estimate how much our (or the model's) level of certainty changes given new information. Probability theory provides the mathematical framework on which these models can be constructed and manipulated. In this work we explore the creation and application of Bayesian models through the development of an independent component analysis (ICA) model and the subsequent application of this model to different examples of heart cycles. For a more in depth and comprehensive introduction into probability theory, probabilistic modelling, and applications to artificial intelligence the reader is referred to "Pattern Matching and Machine Learning" by Christopher Bishop [94].

## 4.2 A probabilistic view of the diagnostic process

As an example of how an event or process can be imagined or modelled using a probabilistic framework, we can use the example of a medical professional examining a patient.

First, we define a "sample space"  $H$  that is representative of any particular person's possible health states. Conceptually, we could determine the "health" of a person by sampling (randomly choosing an outcome) from this space. More accurately,  $H$  represents our (or a doctor's) mental model about a particular person's health status and the relative area that a particular state (for instance "Healthy" and "Not Healthy") occupies in the space is representative of the probability that the model assigns to that state and thus we will refer to  $H$  as a *model*.

At the first meeting between a medical examiner and a patient the model  $H$  represents one of relative ignorance. In this framework we could say that the basic assumption of modern medical education is that the medical professional's training and experience has already "primed" their model (the sample space of possible "health" outcomes) beyond that of a layman. Thus, before first meeting (encountering data concerning the patient) the model represents the current knowledge of the examiner, which may include education and experience, as well as awareness of environmental conditions (e.g. "flu season", diseases specific to certain regions, etc).

As the examination progresses the examiner learns more information about the specific patient, this may include symptoms that the patient complains about, signs that the professional has been trained to look for, the outcomes of medical tests, the outcomes of treatment options, and nuances the examiner has acquired through experience of medical practice. If the information is relevant (or, more formally, if the patient's health status and the new information are dependent on each other) the model (the probability space describing the patient's possible health outcomes) changes. The new sample space represents the health status of the patient *given* that (conditioned on) some dependent condition being true, false, or some particular value. The mathematically sound way of updating the sample space given new information is what Rev. Thomas Bayes and Pierre-Simon Laplace (independently) discovered.

Using an updated model the medical examiner continues the diagnostic process. The model both guides the investigation and is updated by the findings of the investigation. This process continues until the model reaches a state where prescribing a particular treatment is justified, or stated from a probabilistic viewpoint, until the model is such that a condition with a known treatment has a probability high enough to justify the costs (financial, and risk of side effects) of that treatment. If the process progresses without a clear outcome, the use of more expensive and more invasive tests becomes justified.

As a more concrete example: a child goes to a school nurse complaining about a sore throat. The school nurse (being a trained professional) does not immediately begin to inspect the child's elbow, since their "model" of the child's health has already been updated with a symptom of "sore throat". Neither does the nurse immediately refer the child for a chest x-ray, even though a sore throat is a possible symptom of viral pneumonia and misdiagnosis could be deadly. Rather the least invasive and expensive test is performed first. The child's throat is examined visually ("say aaah") and their tonsils are observed to be inflamed. While the combination of the symptom of a sore throat and the sign of inflamed tonsils is not "conclusive evidence" of anything, the nurse's model has been modified sufficiently that the probability of "tonsillitis" is high enough to justify the prescription of rest, lots of warm fluids, throat lozenges, and perhaps further observation if symptoms do not improve within a week. There is, of course, the possibility that the child does have a malignant cyst in their elbow or severe lung infection, but the probability of these (or perhaps here we can risk saying the nurse's "level of belief") given the model (the new observations in light of previous knowledge and experience) is not sufficient to justify conducting the invasive and costly tests needed to confirm such diagnoses.

The models created by two individuals, and in particular the prior elements of the model, are very unlikely to be the same. Given two opposing individual models how do we know which model is "correct"? The short answer is that we don't, or perhaps it would be more accurate to say that we don't know which model is "more representative of the underlying reality", since both models, the result of finite information, are approximations and inherently incorrect. It is not a trivial task to formalise "mental models" into mathematical representations that are tractable while remaining

representative of the system being analysed. Although, in as far as we can make the models explicit, we can compare and criticise the conclusions drawn from opposing assumptions. In fact, the comparison and combinations of models provides a rich source of information and ideally a given situation should be assessed using a "many-models" approach [95] in which a phenomenon is examined in the light of different assumptions and different combinations of assumptions.

### 4.3 Bayesian heart sound modelling

From the perspective of probabilistic modelling we can view the task of signal processing as one of constructing a generative model representative of the 'source' of the signal we are analysing. We can then make inferences about this source based on the model. Applied to heart sounds the source becomes the heart and heart valves and the dynamics of the blood flow through the heart. The aim becomes to draw inferences about the state of a patient's heart given the generative model conditioned on examples of their heart sounds.

Similar to an actual heart, the "model of the heart" developed in this work consists of many different pieces and sub-models drawn from a variety of fields of study. This includes the signal processing techniques discussed in the previous chapter, as well as the machine learning and statistical tools discussed in this and the following chapters. In this chapter we introduce a probabilistic implementation of a statistical technique commonly called independent component analysis (ICA). This provides a foray into the practical application of probabilistic modelling to the signal processing problem at hand.

### 4.4 Independent Component Analysis

Independent component analysis (ICA) is an attempt to solve the blind sources separation problem [96], [97], illustrated in literature as the cocktail party problem: "Can a machine learner, given recordings of audio made of conversations at a cocktail party, identify the different speakers?". For this research we can contextualise the cocktail party problem as a "cardiac auscultation recording" problem: "Can a machine learner, given a recording of sounds made of a patient's heart sounds, identify sounds originating from different structures in the heart?".

ICA uses higher order statistical moments to find statistically independent components in a given dataset. Two events are statistically independent when observing one does not provide information about the probability of the other occurring. In a signal processing context, ICA attempts to find the  $L$  groups of samples in the signal, where samples from the same group have the least in common with samples from other groups, and equivalently the most in common with samples from the same group, where "in common" can be understood as "the value of one providing information about the value of the other". Comon [98] proposed *mutual information* as a natural measure of statistical independence, two events that have zero mutual information are completely statistically independent, and also showed that *negentropy* can be calculated as a substitute for the computationally expensive mutual information. The FastICA algorithm [99] approaches ICA by minimization of an approximation to the negentropy. Another approach, one implementation of which is the JADE algorithm [100], attempts to maximise the *kurtosis* of the groups of samples in order to maximise the statistical independence. *Information-maximization* or INFOMAX, an information transference optimization technique developed in the field of artificial neural networks, has also been used to perform ICA [101]. The INFOMAX approach can be also be thought of as maximum likelihood learning of a probabilistic latent variable model [102]. An alternative method is to infer the latent variables using the principles of Bayesian inference. The theory behind this approach is described in detail by Bishop [94], and specifically applied to the ICA problem by Lawrence and Bishop [103], Højen-Sørensen et al. [104], Winther and Petersen [105], and Choudrey [96]. The approach involves defining a generative probabilistic model and fitting the model to the data using Bayesian inference.

A fundamental issue in independent component analysis is the selection, or discovery, of the number of sources present in, or the latent dimensionality of, the dataset. The ICA algorithm will attempt to maximise the statistical independence between a specified number of sources, even if that leads to a clear departure from what we would recognise as a 'source'. Specifying too few sources can cause the algorithm to combine sources that are unrelated, while specifying too many could lead to signals from a single source being separated into many components.

One problem is perhaps the somewhat vague definition of what a 'source' is in the broader sense. In purely theoretical ICA the sources can be clearly defined as the statistically independent origins of the signals in the observed mixture. In a slightly less theoretical model, we acknowledge the limitations of our sensors and add a term that represents additive noise. From this purely mathematical view we can measure the suitability of our model by measuring how well it is able to represent our data. The variational negative free energy was used as a measure of model suitability by Choudrey [96], while the log likelihood and Bayesian information criterion (BIC) are used by Højensørensen et al. [104] and, Winther and Petersen [105]. These measures are both estimates of how well the model can represent the data and will generally increase monotonically with the number of sources until the model is able to generate the dataset optimally after which a plateau is reached. This approach was used by Choudrey [96] to accurately discover the number of sources that were used to generate an example dataset. While more complex models are able to fit the dataset more closely this comes at the cost of overfitting, leading to a loss of generalization [94]. The number of sources can here be thought of as the degrees of freedom of the model; more degrees of freedom will naturally lead to the model being able to generate the dataset more precisely. Modelling an increasing number of sources also leads to the model becoming more obscure and less interpretable.

In more practical applications the model assumptions are unlikely to be met exactly. The distinction between noise and source may be less clear and there are likely to exist some dependence (correlations and interactions) between sources. In short, the definition of a source, and the distinction between different sources, becomes less clear in more realistic situations. As an illustration consider the cocktail party problem. In a simplified view we are trying to distinguish the different individual speakers, but it is unlikely that these speakers are just standing around conversing with themselves. It is much more likely that the speakers are engaged in conversations with one or more other individuals. These conversations will have their own characteristics, for instance one conversation might be loud and argumentative, with speakers yelling over one another, while another might be slow and awkward with long pauses in between utterances. Now if we were to perform ICA on audio recordings recorded at  $M$  locations at the party, are we more likely to separate our dataset into individual speakers or into individual conversations?

Or put another way, which is more statistically independent: individual speakers or conversations?

## 4.5 Probabilistic ICA

In this section a probabilistic (or Bayesian) approach to modelling the independent components of heart cycles is investigated. The proposed model, based on work done by Choudrey [96], Højen-Sørensen et al. [104], and Winther and Petersen [105] is described in this section, followed by the methodology used to apply this model to time-frequency decompositions of heart cycles. This is followed by the results of the model on a small dataset of heart sounds containing various murmurs.

### 4.5.1 Methodology

The aim of ICA is to uncover the statistically independent sources that have combined to form the observed data. Thus, the underlying assumptions of ICA is that the observed data was generated by a finite set of statistically independent sources and that the observations are a linear, weighted mix of these sources. These assumptions are expressed and enforced by a probabilistic model. Additionally, we assume that the sounds produced by the mechanical action of valves and blood flow through the heart are the dominant sounds present in the recording. That is, we assume the 'heart sounds' are the most prominent sounds that have been recorded. This allows us to group the noises in the recording not associated with the heart into the additive noise term and thus assume that the discovered independent components are all heart sounds.

### 4.5.2 The Model

Let  $\mathcal{M}$  be a generative model. All distributions described in this section are implicitly conditioned on the model  $\mathcal{M}$  and acknowledge all the implicit assumptions in the construction of  $\mathcal{M}$ ; this conditioning is however omitted in further equations to simplify the notation.

The ICA model is represented as  $M$ -dimensional 'signals',  $\mathbf{x}$ , modelled as a linear combination of  $L$ -dimensional independent 'sources',  $\mathbf{s}$ , with added Gaussian noise

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \tag{4.5}$$

where  $\mathbf{A}$  is an  $M \times L$  matrix of mixing coefficients and  $\mathbf{n}$  is  $M$ -dimensional additive noise. The observations are a  $M \times T$  matrix  $\mathbf{x} = \{x_{mt}\}$ ,  $m = 1, \dots, M, t = 1, \dots, T$  with  $x_{mt}$  representing the sample at time  $t$  for signal  $m$ . The sources are an  $L \times T$  matrix  $\mathbf{s} = \{s_{lt}\}$ ,  $l = 1, \dots, L, t = 1, \dots, T$  with  $s_{lt}$  representing the sample at time  $t$  for the  $l$ th source. Data points are assumed to be independent and identically distributed (IID) i.e. that is there no inherent order to the data points. This is an obvious limitation of the model in terms of time-series analysis.

The sources  $\mathbf{s} = \{\mathbf{s}_1, \dots, \mathbf{s}_L\}$  are mutually independent i.e. the probability distribution of  $\mathbf{s}$  for data point  $t$  can be expressed as the product of the probability distributions of each of the sources

$$p(\mathbf{s}_t | \boldsymbol{\theta}) = \prod_{l=1}^L p(s_{lt} | \theta_l) \quad (4.6)$$

where  $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_L\}$  and  $\theta_l$  represents the parameters defining the  $l$ th source distribution, e.g. mean and variance.

The distribution of the noise is modelled as Gaussian, with zero mean and diagonal precision matrix  $\boldsymbol{\Lambda}$

$$p(\mathbf{n} | \boldsymbol{\Lambda}) = \mathcal{N}(\mathbf{n} | \mathbf{0}, \boldsymbol{\Lambda}^{-1}) \quad (4.7)$$

The likelihood of the observation vector  $\mathbf{x}_t$  at time  $t$  given the generative model is given by

$$p(\mathbf{x}_t | \mathbf{s}_t, \mathbf{A}, \boldsymbol{\Lambda}) = \left( \det\left(\frac{1}{2\pi} \boldsymbol{\Lambda}\right) \right)^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}_t - \mathbf{A}\mathbf{s}_t)^T \boldsymbol{\Lambda}(\mathbf{x}_t - \mathbf{A}\mathbf{s}_t)\right] \quad (4.8)$$

The aim of the ICA algorithm is then to uncover the latent sources  $\mathbf{s}$  that produced the observations  $\mathbf{x}$ . This can be expressed using Bayes' theorem as

$$p(\mathbf{s} | \mathbf{x}, \mathbf{A}, \boldsymbol{\Lambda}) = \frac{p(\mathbf{x} | \mathbf{s}, \mathbf{A}, \boldsymbol{\Lambda}) p(\mathbf{s})}{p(\mathbf{x} | \mathbf{A}, \boldsymbol{\Lambda})} \quad (4.9)$$

where  $p(\mathbf{x} | \mathbf{s}, \mathbf{A}, \boldsymbol{\Lambda})$  is the probability of observing  $\mathbf{x}$  given the model  $\mathcal{M}$ ,  $p(\mathbf{s})$  is the distribution of the sources  $\mathbf{s}$  described in the model  $\mathcal{M}$  prior to observing  $\mathbf{x}$ ,  $p(\mathbf{x} | \mathbf{A}, \boldsymbol{\Lambda})$

is the *evidence* for the model  $\mathcal{M}$ , and  $p(\mathbf{s}|\mathbf{x}, \mathbf{A}, \mathbf{\Lambda})$  is the distribution of the sources given the observations and the model  $\mathcal{M}$ .

### 4.5.3 Source Model

The latent sources that ICA attempts to discover are contained in the ICA model as the sufficient statistics (i.e. the statistics that can be used to completely describe the distribution) of the source probability distributions. Selection of appropriate source models is important to allow the generative model to effectively describe the components present in the data. It is also necessary to choose mathematically viable source models to ensure that the computations needed to fit the models remain tractable.

#### 4.5.3.1 Mixture of Gaussians

Mixtures of Gaussians (MoG), also called Gaussian mixture models (GMM), are probabilistic models in which each sample is assumed to have been generated by a superposition of  $K$  Gaussian densities [94]

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (4.10)$$

in which each Gaussian *component*  $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  of the mixture has its own mean  $\boldsymbol{\mu}_k$  and covariance  $\boldsymbol{\Sigma}_k$ . The parameters  $\pi_k$  are called the *mixing proportions* and represent the probability of sampling from the  $k$ th component.

For the purposes of ICA, we can consider a factorized MoG with  $L$  sources and  $K_l$  components per source. We introduce the variable  $z_{lk}$  for  $k = \{1, \dots, K_l\}$  and  $l = \{1, \dots, L\}$ , is a 1-of- $K_l$  indicator variable with  $z_{lk} \in \{0, 1\}$  and  $\sum_k z_{lk} = 1$ . The probability of  $z_{lk}$  assuming a value of 1 for a component of source  $l$  is equal to the mixing proportion of that component.

$$p(z_{lk} = 1) = \pi_{lk} \quad (4.11)$$

That is, the latent variable  $z_{lk}^t$  is equal to 1 for the component  $k$  of the  $l$ th source that is chosen to generate  $\mathbf{s}_t$  at time  $t$  with probability  $\pi_{lk}$ . Note that the mixing proportions

must be valid probabilities, that is  $0 \leq \pi_{lk} \leq 1$  and  $\sum_k \pi_{lk} = 1$ . The probability of  $\mathbf{s}_t$  given the model parameters  $\boldsymbol{\theta}$  is then

$$\begin{aligned} p(\mathbf{s}_t | \boldsymbol{\theta}) &= \prod_{l=1}^L \sum_{k=1}^{K_l} p(z_{lk}^t = 1 | \boldsymbol{\pi}_l) p(s_{lt} | z_{lk}, \mu_{lk}, \beta_{lk}) \\ &= \prod_{l=1}^L \sum_{k=1}^{K_l} \pi_{lk} \mathcal{N}(s_{lt} | \mu_{lk}, (\beta_{lk})^{-1}) \end{aligned} \quad (4.12)$$

where  $s_{lt}$  is the  $t$ th sample of source  $l$ , and  $\mu_{lk}$ ,  $\beta_{lk}$ , and  $\pi_{lk}$  are respectively the mean, precision, and mixing proportion of the  $k$ th component of the  $l$ th source.

The parameters of the model are denoted by  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_L\}$ , where  $\boldsymbol{\theta}_l = \{\boldsymbol{\pi}_l, \boldsymbol{\mu}_l, \boldsymbol{\beta}_l\}$  are the parameters of the  $l$ th source with  $\boldsymbol{\pi}_l = \{\pi_{l1}, \dots, \pi_{lK_l}\}$ ,  $\boldsymbol{\mu}_l = \{\mu_{l1}, \dots, \mu_{lK_l}\}$ , and  $\boldsymbol{\beta}_l = \{\beta_{l1}, \dots, \beta_{lK_l}\}$ . The joint probability of the source vector  $\mathbf{s}_t$  and the collection of all possible states of the indicator variable  $\mathbf{z} = \{z_1, \dots, z_K\}$ ,  $K = \prod_l^L K_l$  at time  $t$ , or stated otherwise, the probability of state  $\mathbf{z}_t$  generating source vector  $\mathbf{s}_t$ , can be expressed as a L-dimensional MoG with a total of  $K$  components [96].

$$\begin{aligned} p(\mathbf{s}_t, \mathbf{z}_t | \boldsymbol{\theta}) &= \prod_{l=1}^L p(z_{lt} = 1 | \pi_l) p(s_{lt} | z_{lt}, \mu_{lk_l}, \beta_{lk_l}) \\ &= p(\mathbf{z}_t | \boldsymbol{\pi}) p(\mathbf{s}_t | \mathbf{z}_t, \boldsymbol{\theta}) \end{aligned} \quad (4.13)$$

The likelihood of the data, assuming independent and identically distributed data,  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_t, \dots, \mathbf{x}_T\}$  given the MoG model parameters (including the latent indicator variable  $\mathbf{k}$ ) can now be expressed as

$$p(\mathbf{X} | \boldsymbol{\Theta}) = \prod_{t=1}^T \sum_{k=1}^K \int p(\mathbf{x}_t | \mathbf{s}_t, \mathbf{A}, \boldsymbol{\Lambda}) p(\mathbf{s}_t | \mathbf{z}_t, \boldsymbol{\theta}) p(\mathbf{z}_t | \boldsymbol{\pi}) d\mathbf{s} \quad (4.14)$$

where  $\boldsymbol{\Theta} = \{\mathbf{A}, \boldsymbol{\Lambda}, \boldsymbol{\theta}, \mathbf{z}\}$  is the collection of parameters for the model  $\mathcal{M}$ .

#### 4.5.3.2 Laplace

The Laplace, or double exponential, distribution has a heavier tail compared to that of the Gaussian distribution, this quality makes it suited to model the distribution of

interesting real world signals, such as speech [104]. Since the Laplace distribution is able to accurately model heavy tailed signals using a single Laplacian as opposed to at least 3 Gaussians for a MoG it has computational benefits above that of the MoG [106]. The Laplace distribution decays exponentially from the mean in both directions, assuming zero mean the probability density can be expressed as

$$p(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right) \quad (4.15)$$

In the context of ICA we model each source using a single Laplace distribution. Since the sources are independent, we can write

$$\begin{aligned} p(\mathbf{s}_t | \boldsymbol{\theta}) &= \prod_{l=1}^L p(s_t | b_l) \\ &= \prod_{l=1}^L \frac{1}{2b_l} \exp\left(-\frac{|s_t|}{b_l}\right) \end{aligned} \quad (4.16)$$

where the parameter  $b$ , a scale parameter referred to as the diversity, determines the spread of the distribution.

#### 4.5.3.3 Model Learning

Learning for the generative model described in the previous section is equivalent to calculating the optimal posterior probability density over the model parameters. How exactly to find the "optimal" posterior is an area of active research; an overview of different methodologies is provided by Choudrey [96]. One popular approach is mean field theory borrowed from statistical physics, specifically the variational Bayesian approach as described by Choudrey [96] as well as Højen-Sørensen et al. [104]. This approach is formulated for the MoG source model.

The variational methodology replaces the true, but unknown and intractable, posterior distribution with a tractable approximation. The aim of the learning process is then to minimize the difference between the true and the variational distributions. It can be shown, using Jensen's inequality, that the Kullback-Leiber (KL) divergence is minimized when the negative free energy is maximized [94]. Briefly,

$$\ln p(\mathbf{X}) = \mathcal{L}(\mathbf{X}) + KL[p^*(\mathbf{W})||p(\mathbf{X}, \mathbf{W})] \quad (4.17)$$

where  $p^*(\mathbf{W})$  is some approximation of the posterior  $p(\mathbf{W}|\mathbf{X})$ , and  $\mathbf{W} = \{\mathbf{A}, \mathbf{\Lambda}, \mathbf{s}, \mathbf{z}, \boldsymbol{\theta}\}$  is the collection of all parameters and hidden variables. The first term in (4.17) is known as the "negative variational free energy"

$$\begin{aligned} \mathcal{L}(\mathbf{X}) &= \int p^*(\mathbf{W}) \ln \frac{p(\mathbf{X}, \mathbf{W})}{p^*(\mathbf{W})} d\mathbf{W}. \\ &= E_{p^*(\mathbf{W})}[\ln p(\mathbf{X}, \mathbf{W})] + \mathcal{H}[\mathbf{W}] \end{aligned} \quad (4.18)$$

where  $E_{p^*(\mathbf{W})}[\cdot]$  is the expectation in terms of  $p^*(\mathbf{W})$  and  $\mathcal{H}[\mathbf{W}]$  is the entropy of  $p^*(\mathbf{W})$ . The second term is the KL divergence, a measure of pseudo-distance between two distributions. The KL divergence is strictly non-negative and so the negative free energy forms a *strict lower bound* on  $\ln p(\mathbf{X})$ . That is,

$$\ln p(\mathbf{X}) \geq \mathcal{L}(\mathbf{X}) \quad (4.19)$$

with  $\ln p(\mathbf{X}) = \mathcal{L}(\mathbf{X})$  if and only if the variational distribution  $p^*(\mathbf{W})$  is exactly equal to the true posterior  $p(\mathbf{W}|\mathbf{X})$ . This means that instead of having to deal with the unknown true posterior we can instead maximise the negative free energy, which only relies on variational distribution  $p^*(\mathbf{W})$  and this ends up being equivalent to minimising the KL divergence. In summary, to train our model we can use the variational free energy (4.18) as an objective function to be maximised.

The variational distribution  $p^*(\mathbf{W})$  must be selected such that it is computationally tractable, while also being flexible enough to form a good approximation of the true posterior distribution. The following factorisation is assumed

$$p^*(\mathbf{W}) = p^*(\mathbf{A})p^*(\mathbf{\Lambda})p^*(\mathbf{s}|\mathbf{z})p^*(\mathbf{z})p^*(\boldsymbol{\pi})p^*(\boldsymbol{\mu})p^*(\boldsymbol{\beta}) \quad (4.20)$$

Assuming that  $p^*(\mathbf{W})$  factorises over  $\mathbf{W}$  allows us to maximise each of the parameters individually. The source distribution is however conditioned on the indicator variable  $\mathbf{z}$  which implies a mixture model posterior source density for source  $l$ .

$$\begin{aligned}
p(s_{lt}) &= \sum_{k=1}^K p^*(z_{lk}^t = 1) p^*(s_l | \mathbf{z}_l) \\
&= \sum_{k=1}^K \gamma_{lt}^k \mathcal{N}(s_{lt} | \mathbf{M}_{lk}^t, (\mathbf{B}_{lk}^t)^{-1})
\end{aligned} \tag{4.21}$$

where  $\gamma_{lt}^k$  is the posterior probability of the mixing proportions and represents the *responsibility* that component  $k$  takes for explaining  $s_{lt}$ .  $\mathbf{M}_{lk}^t$  and  $\mathbf{B}_{lk}^t$  are respectively the mean and precision of component  $k$  of source  $l$ .

The second part we need is the joint probability of the observations and the model parameters, expressed graphically in Figure 4.4.

$$p(\mathbf{X}, \mathbf{W}) = \prod_t^T p(\mathbf{x}_t | \mathbf{s}_t, \mathbf{A}, \mathbf{\Lambda}) p(\mathbf{s}_t | \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\beta}) p(\mathbf{z} | \boldsymbol{\pi}) p(\boldsymbol{\pi}) p(\boldsymbol{\mu}) p(\boldsymbol{\beta}) p(\mathbf{A}) p(\mathbf{\Lambda}) \tag{4.22}$$

We can now substitute (4.20) and (4.22) into the equation for the variational negative free energy (4.18) to find the maximisation objective function for the model  $\mathcal{M}$ .

$$\begin{aligned}
\mathcal{L} &= \prod_t^T [E_{p^*(\mathbf{A})p^*(\mathbf{\Lambda})p^*(s_t|\mathbf{z})p^*(\mathbf{z})} [\ln p(\mathbf{x}_t | \mathbf{s}_t, \mathbf{A}, \mathbf{\Lambda})] \\
&+ E_{p^*(s_t|\mathbf{z})p^*(\mathbf{z})p^*(\boldsymbol{\pi})p^*(\boldsymbol{\mu})p^*(\boldsymbol{\beta})} [\ln p(\mathbf{s}_t, \mathbf{z} | \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\beta})] + \mathcal{H}[p^*(s_t, \mathbf{z})]] \\
&\quad + E_{p^*(\boldsymbol{\pi})} [\ln p(\boldsymbol{\pi})] + \mathcal{H}[p^*(\boldsymbol{\pi})] \\
&\quad + E_{p^*(\boldsymbol{\mu})} [\ln p(\boldsymbol{\mu})] + \mathcal{H}[p^*(\boldsymbol{\mu})] \\
&\quad + E_{p^*(\boldsymbol{\beta})} [\ln p(\boldsymbol{\beta})] + \mathcal{H}[p^*(\boldsymbol{\beta})] \\
&\quad + E_{p^*(\mathbf{A})} [\ln p(\mathbf{A})] + \mathcal{H}[p^*(\mathbf{A})] \\
&\quad + E_{p^*(\mathbf{\Lambda})} [\ln p(\mathbf{\Lambda})] + \mathcal{H}[p^*(\mathbf{\Lambda})]
\end{aligned} \tag{4.23}$$

The terms can now be further specified by finding functional forms for the expressions and substituting these into (4.23). This requires prior distributions to be specified over the model parameters. The chosen priors are shown as part of Figure 4.4 below. The resulting equations can be solved in an 'Expectation Maximisation (EM)'-like fashion [94]. In the variational equivalent of the expectation (E) step the model parameters are kept fixed and the *responsibilities* are estimated. In the subsequent maximisation (M) step these responsibilities are kept fixed and the variational distribution over the model parameters is re-computed. These steps are cycled until convergence. For a more in-depth formulation of the variational EM algorithms and its extensions, as well as the

complete derivation of the model update equations see Choudrey [96] and Winther and Petersen [105].

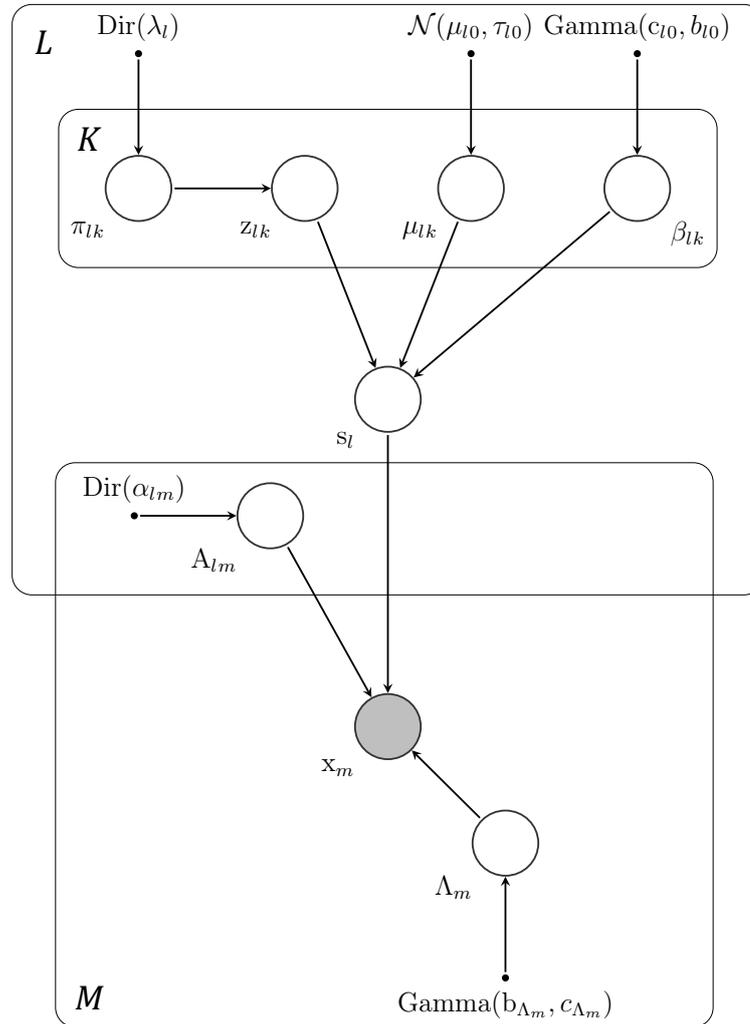


Figure 4.4: Graphical model for the generative model  $\mathcal{M}$ . Prior distributions of model parameters are shown as solid dots, model parameters as circles, and observed (unconditioned) variables as shaded circles. Plate notation is used to illustrate where the model parameters are defined for  $M$ -dimensions,  $L$ -sources, and  $K$ -components. An in-depth introduction into probabilistic graphical models is given by Bishop [94].

## 4.6 Methods

### 4.6.1 Data

To determine the utility of ICA in heart sound analysis, recordings of heart sounds were collected from volunteers. Examples of abnormal heart sound recordings were collected at the cardiac unit of Auckland City Hospital by an attending cardiologist, while a final year medical student collected heart sound recordings from non-cardiac patients as part of a heart sound data collection study. All heart sounds were recorded under ethics

approval obtained from the University of Auckland and Auckland District Health Board ethics committees, approval number UAHPEC 013321. Recordings were made with a 3M™ Littmann® Electronic Stethoscope Model 3200 [107] with a sampling rate of 4kHz. Diagnosis, where applicable, are based on the findings of the investigating cardiologist.

Heart sounds were collected from a total of 26 participants. Twenty of the recordings were collected from cardiac patients and the remaining six were collected from volunteers (who were not cardiac patients). To demonstrate ICA a subset of 4 heart sounds was chosen from the collected recordings. The subset was chosen by the researcher based on their judgement of the signal quality (i.e. signals that were too corrupted by noise were excluded) and the presence of interesting cardiac sounds (i.e. a representative set containing different examples of murmurs).

#### 4.6.2 Pre-processing

All recordings were first pre-processed using an ‘artefact removal’ or ‘de-spiking’ algorithm to remove obvious artefacts. The methodology followed closely resembles that proposed by Schmidt et al. [108] but was modified when the previous algorithm was not able to remove prominent movement artefacts in the data. In this work we assume a normal distribution for the peak amplitudes in the signal and identify peaks that exceed the mean peak intensity plus 3 times the standard deviation of the peaks, corresponding to a less than 0.3% chance of coming from the same normal distribution. The peaks considered are the local maxima of the absolute value of the normalised signal that fall above the 99th percentile of all sample intensities. Spike widths were determined by their prominence, a measure of the height of a peak relative to the surrounding samples, and the start and end points of the spikes determined using the width. Identified spikes were replaced using linear interpolation between the start and end points. After artefact removal the recordings were standardised to zero mean and unit variance.

#### 4.6.3 Continuous Wavelet Transform

A time-frequency decomposition of the audio recording was performed to form an input signal to the ICA algorithm from the one dimensional audio recordings. The audio data was analysed at  $M$  scales using the Continuous Wavelet Transform (CWT) with a

generalised Morse analytic wavelet [109], [110] with parameters  $\gamma = 3, \beta = 60$ . The Morse analytic wavelet with  $\gamma = 3$  provides a very nearly distinct frequency to scale mapping [110]. An example of the CWT of a heart cycle containing a late systolic murmur due to mitral valve prolapse is shown in Figure 4.5. This example recording was sampled at 4 kHz. The CWT was limited to between 50 Hz and 1.5kHz with 8 voices per octave. Voices per octave is a measure of how many scales (pseudo inverse frequency bands) are analysed within the frequency range being analysed (which can also be referred to as the 'number of octaves'). A higher number of voices per octave results in more frequency bands being analysed, effectively increasing the frequency resolution of the decomposition.

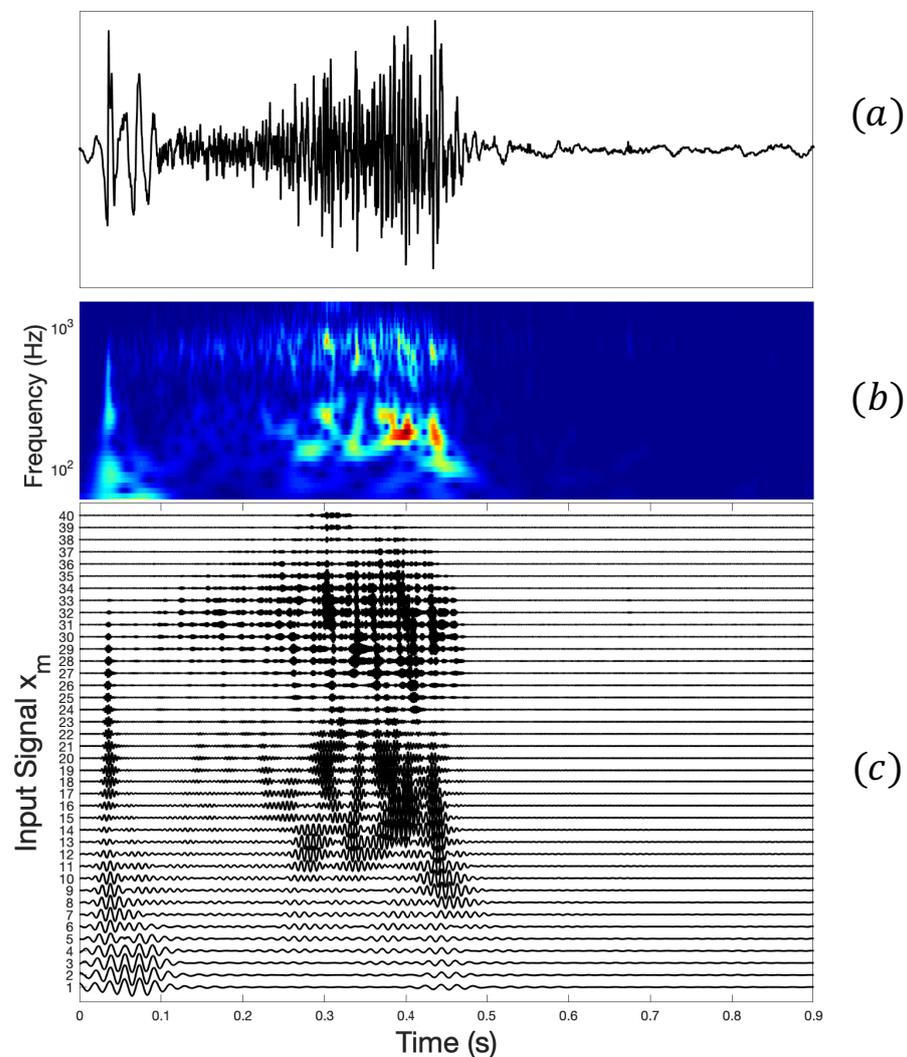


Figure 4.5: CWT decomposition of heart cycle shown in (a). The magnitude of the CWT coefficients is shown in the time-frequency domain in (b). In (c) the real parts of the decomposition, which forms the input signals for the ICA algorithm, are plotted as 40 ( $M = 40$ ) 1 dimensional signals.

#### 4.6.4 Independent component analysis of heart sounds

The following procedure was used to test the ability of the proposed model in separating components in heart sound recordings.

- Heart sounds are decomposed into  $M$  sets of 1 dimensional input signals using CWT.
- Heart sound CWTs are segmented into individual heart cycles in the order S1, systole, S2, diastole using both manual segmentation and the Hidden Semi-Markov Model based segmentation developed by Springer et al. [111]
- An individual heart cycle is extracted from the heart sound CWT to form an  $M \times T$  input matrix.
- The  $M \times T$  input matrix is used as a data set to train the model  $\mathcal{M}$  with the number of latent sources determined heuristically (discussed below).
- The  $L \times T$  source matrix of the trained model represents the learnt statistically independent sources.

In selecting the number of sources, it is important to keep in mind what system the data was generated by. It is also important to know what the ICA is being used for. In the case of using ICA as a tool to discover more about the system under investigation, the aim is to try and describe the system in the simplest but sufficiently comprehensive way possible. When trying to model a system as complex as the heart the model assumptions are necessarily not met, especially considering the variety of sources of noise. Trying to make the model complex enough to capture every detail of the real system it is trying to emulate is an unnecessary task. Such a task would require much more data than a simple recording of the sound produced as the heart beats. What should the model then try to capture? From a diagnostic standpoint a useful model would be able to separate diagnostically valuable sections or ‘events’ present in the heart sound. Adapting the blind source separation problem and more specifically ICA to the problem of heart sound analysis, we can restate the question as “are the diagnostically valuable aspects of a heart sound recording statistically independent and thus can ICA be applied to discover or clarify these aspects?”. The number of sources used to model each of the heart cycles was determined heuristically keeping this question in mind.

#### 4.6.5 Heart sound source model

In heart sound recordings, two distinct regions can be discriminated in terms of the distribution of their amplitudes. The heart cycle periods between the fundamental heart sounds consist of low amplitudes centred around the (zero) mean. Heart murmurs and other pathological sounds also occur in these regions. The second region is that of the fundamental heart sounds themselves. These regions contain higher amplitudes at lower frequencies and consequently have distributions with heavier tails compared to the Gaussian distribution.

#### 4.7 Results

Four heart sound recordings, one without any abnormal heart sounds and three with cardiac murmurs, were selected to demonstrate the proposed model's performance. Shown in Figure 4.6, recordings (a) and (b) are from two healthy volunteers and recordings (c) and (d) are from cardiac patients at the Auckland City Hospital. The amplitudes of all time-series plots have been normalised. The labels on the x-axis of the graphs have been excluded to increase visibility.

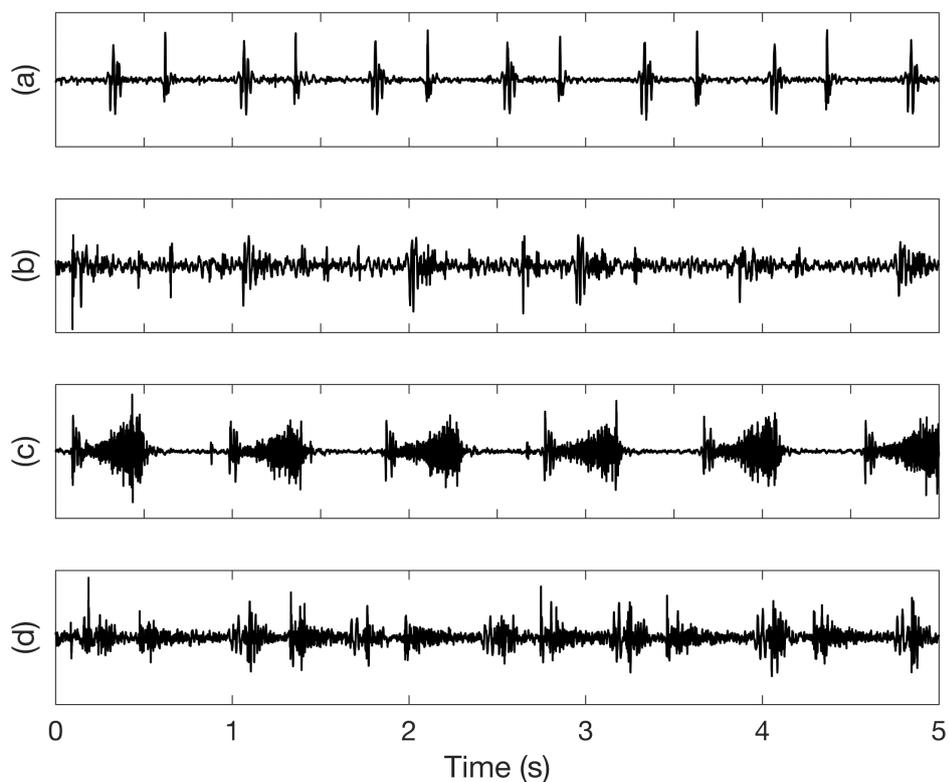


Figure 4.6: The heart sound recordings used to assess the proposed ICA model. (a) healthy heart sound, (b) early systolic murmur, (c) mid-to-late systolic murmur (d) early systolic murmur combined with early diastolic murmur.

Table 3: Summary of information available for heart sounds recordings used for ICA demonstration

Heart Cycle ID	Diagnosis	Recorded by	Auscultation site
(a)	N/A	Medical Student	Pulmonic Area
(b)	N/A	Medical Student	Aortic Area
(c)	Mitral Regurgitation	Cardiologist	Apex
(d)	Pulmonary Regurgitation	Cardiologist	Pulmonic Area

#### 4.7.1 Pre-processing

The proposed artefact removal algorithm is illustrated in Figure 4.7 on a normal heart sound recording with a significant noise spike just before the 6 second time mark. The algorithm proposed by Schmidt et al. [108] does not recognise and remove this spike which proves problematic when normalising the signal as all the heart sound signal amplitudes are reduced relative to the spike amplitude.

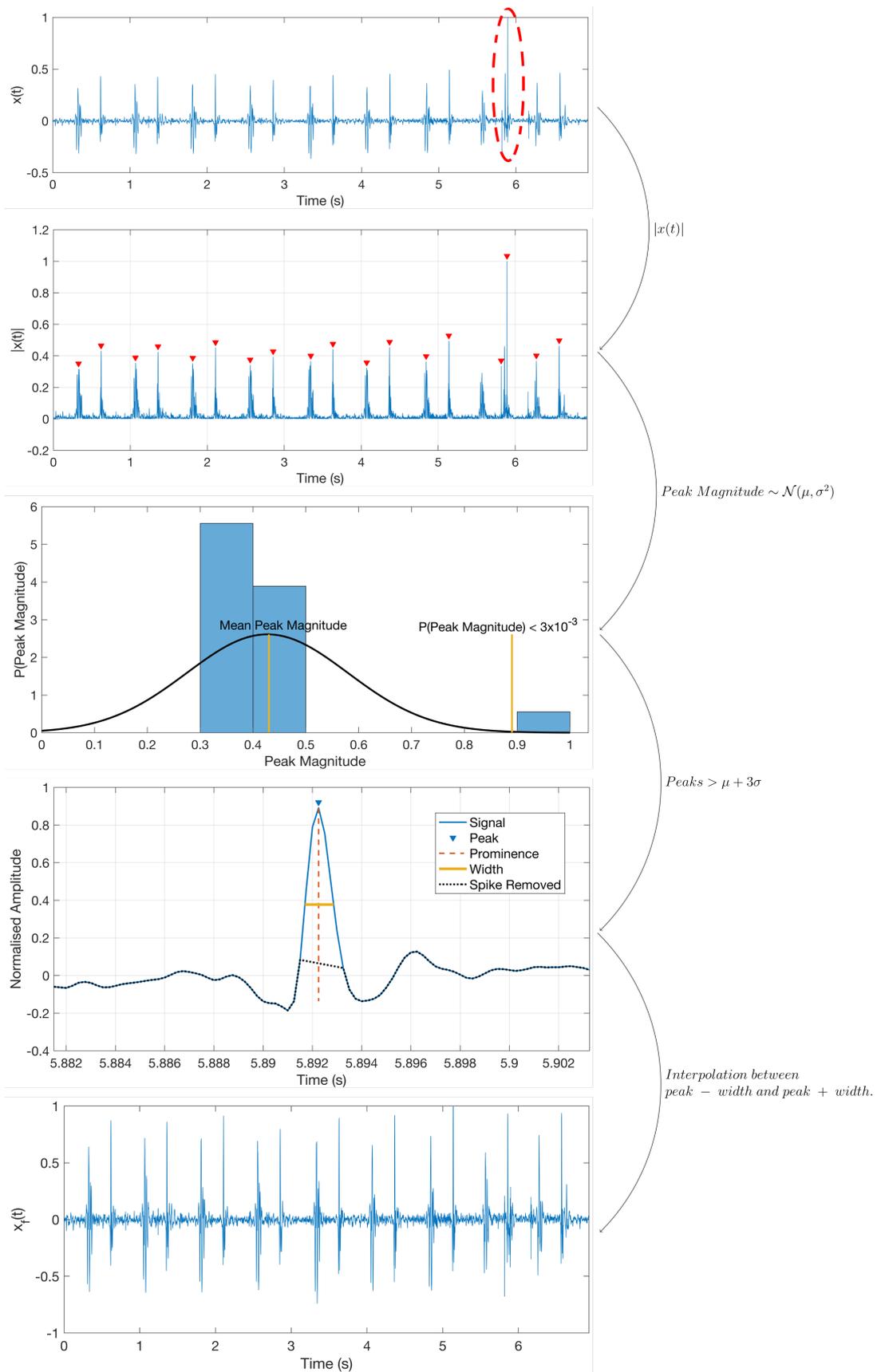


Figure 4.7: The methodology followed in this work to remove spikes from heart sound recordings.

### 4.7.2 Heart Cycle CWT

The CWT was performed using the Morse analytic wavelet with 8 voices per octave between 20 Hz and 1.5 kHz. This decomposition resulted in 50 input vectors for each of the heart cycles ( $M = 50$ ). Only the real parts of the resulting coefficients were used, and these were normalised to maximum one by dividing all values by the maximum absolute value of the coefficients. Selected heart cycles and their CWT are shown in Figure 4.8. To avoid edge effects the CWT was performed on the entire heart sound recording and the segmented heart cycle was extracted from the middle of this longer CWT.

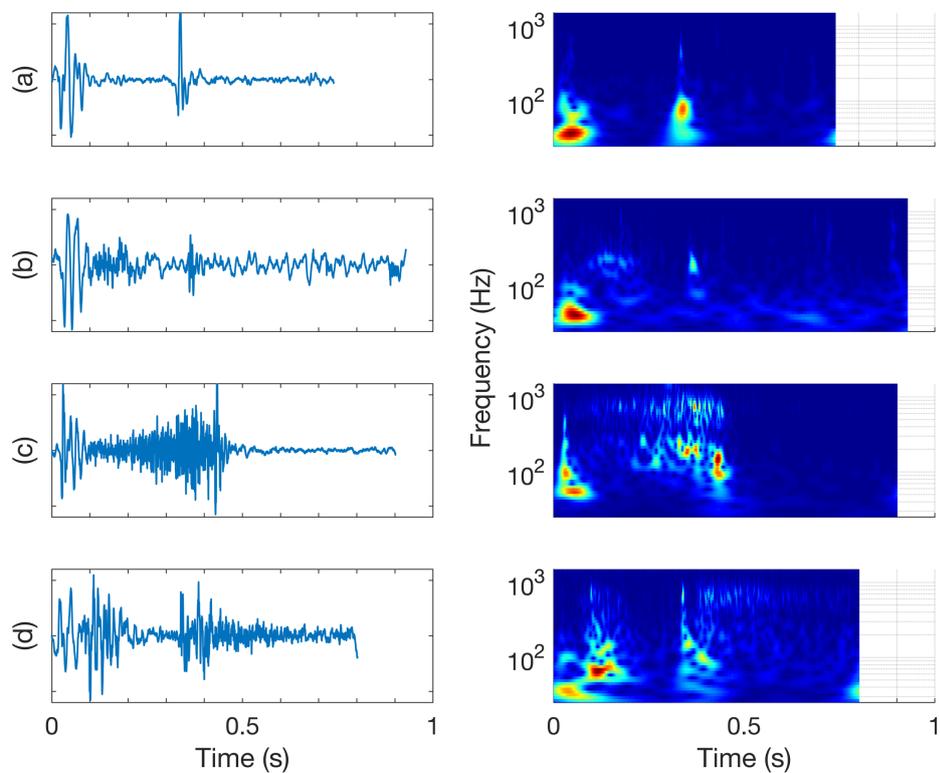


Figure 4.8: Individual segmented heart cycles of the corresponding heart sound recordings (a), (b), (c), and (d).

### 4.7.3 Heart Cycle ICA

The independent component analysis results are shown for heart cycle (a) in Figure 4.9, heart cycle (b) in Figure 4.11, heart cycle (c) in Figure 4.13, and heart cycle (d) in Figure 4.15. The first waveforms in each figure labelled (a), (b), (c), and (d) show the original heart cycle (left-hand plot) and its corresponding frequency domain plot (right-hand plot). The latent sources, or independent components, of the fitted model are shown

underneath. Note that these are numbered only for reference purposes and the independent components (IC) have no inherent order. All the ICs were standardised by subtracting their mean and dividing by their standard deviation. The frequency plots (right-hand side) show the magnitude of the frequency components between 20 Hz and 1 kHz plotted on a logarithmic scale. This required only 2 sources in the case of the healthy heart cycle (a) with a high signal-to-noise ratio, but at least 8 sources in the case of the dual murmurs heart cycle (d). In all cases, models with a larger number of sources, and thus increasing complexity, were able to represent the dataset more closely (as measured by log-likelihood), this is discussed in 4.7.4.

The frequency domain plots reveal the ICA of the CWT time-frequency decompositions has mainly separated out groups of components related by frequency. A possible conceptualisation of the action of the proposed ICA model is as generating (L) bandpass filters with bandwidths and centre frequencies determined by the model parameters learned from the dataset. In the rest of this section we assess the utility of the ICA by analysing the results of the ICA in light of the known underlying physiological condition. The application of ICA in this discussion is analogous to a researcher making use of a microscope. The ICA provides a tool with which to study the audio recordings, but it is up to the operator to control what level of 'magnification' to go to, and to interpret what he/she finds.

#### 4.7.3.1 Heart Cycle (a)

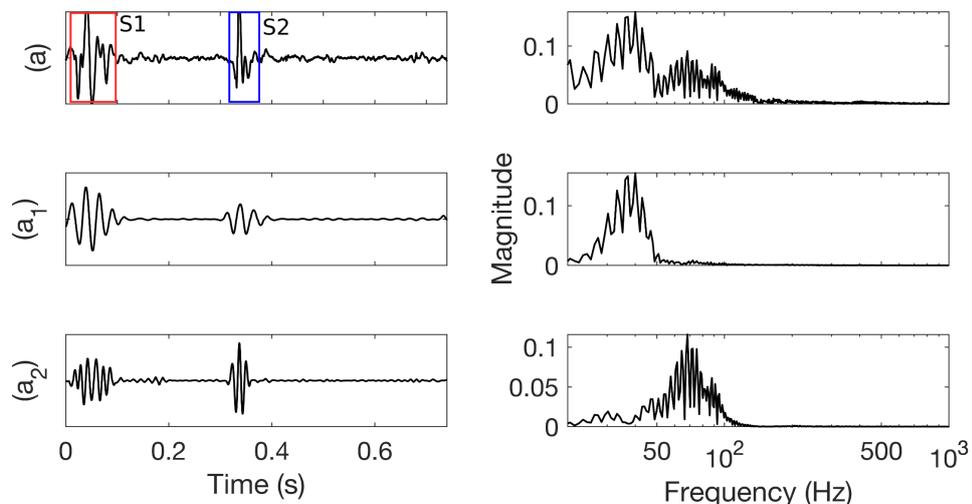


Figure 4.9: ICA of a heart cycle (a). An example of a heart cycle that does not contain any abnormal sounds.

In the figure above we can see that the proposed probabilistic ICA model has divided heart cycle (a) into two distinct components. The first component ( $a_1$ ) contains mainly frequencies below 50 Hz while the second component ( $a_2$ ) has prominent energies between 50 and 100 Hz. The acoustic events in the first component have slightly longer durations than those in the second component. The components that correspond to the first heart sound (0 to 0.1 s) in ( $a_1$ ) have a duration of around 110 ms and around 100 ms in ( $a_2$ ). For the second heart sound (around 0.3 to 0.4 seconds) the components last 100 ms in ( $a_1$ ) and 60 ms in ( $a_2$ ).

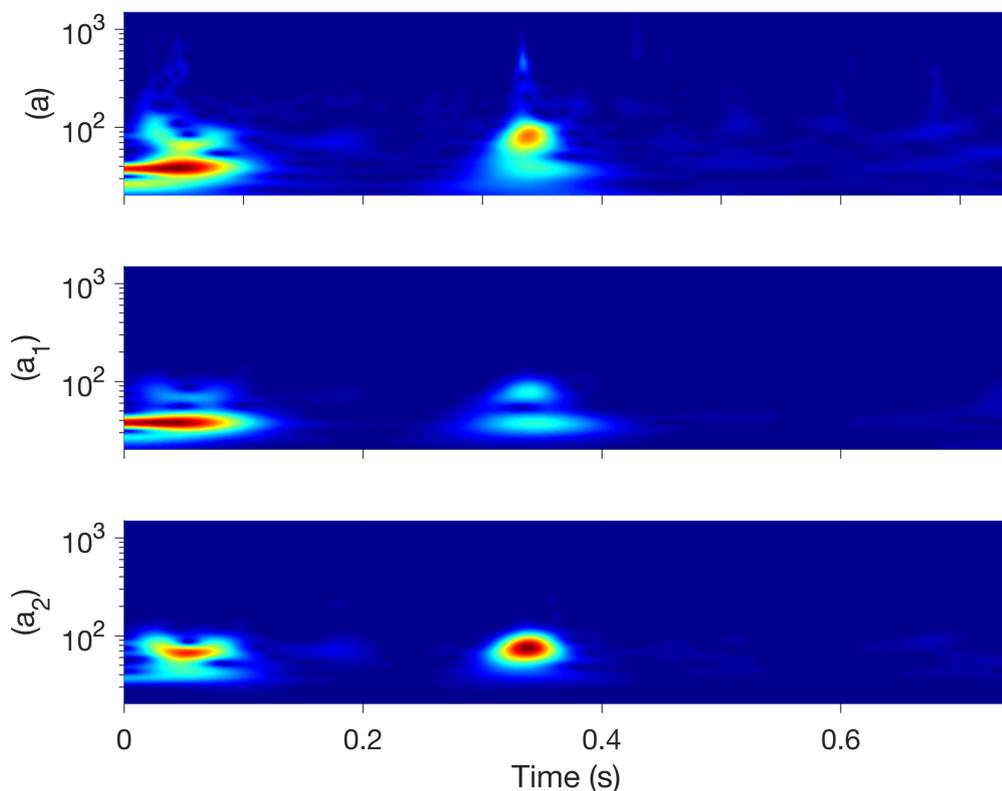


Figure 4.10: CWT of the independent components of heart cycle (a). Frequency is plotted on the y-axis on a logarithmic scale. Decomposition performed with 48 voices per octave between 25 and 1500 Hz.

The time domain and frequency domain representations of heart cycle (a) and the independent components ( $a_1$ ) and ( $a_2$ ) are shown in Figure 4.9. A time-frequency decomposition (using the CWT) is shown in Figure 4.10. This example does not contain any pathological sounds (i.e. it is 'normal'). The CWT clearly shows that the second heart sound has higher frequency components than the first heart sound; these higher frequency components have mainly been captured in ( $a_2$ ). This agrees with discussion in the literature regarding the pitch of the second heart sound compared to the first [112]. The sounds in the heart cycle are initiated by the closing of the four heart valves. We

can assess Figure 4.9 and Figure 4.10 with the assumption that the independent components correspond to the sounds produced by the different heart valves, or in other words that the sounds associated with different valves are maximally statistically independent. Sounds from the higher pressure left atrium (LA) and ventricle (LV) are higher pitched than those from the relatively low-pressure right atrium (RA) and ventricle (RV). Based on this we can infer that  $(a_2)$  corresponds to the closing of the mitral ( $M_1$ ) and aortic ( $A_2$ ) valves and that component  $(a_1)$  corresponds to the closing of the tricuspid ( $T_1$ ) and the pulmonary ( $P_2$ ) valves. Given the CWT decompositions as input, the ICA model was able to separate the lower frequency right-sided valve sounds from the higher frequency right sided sounds.

Clinically, the relative intensities and degree of splitting between the different fundamental heart sound components are very significant [112], [113] and thus an algorithm able to identify these would prove a valuable analysis tool.

#### 4.7.3.2 Heart Cycle (b)

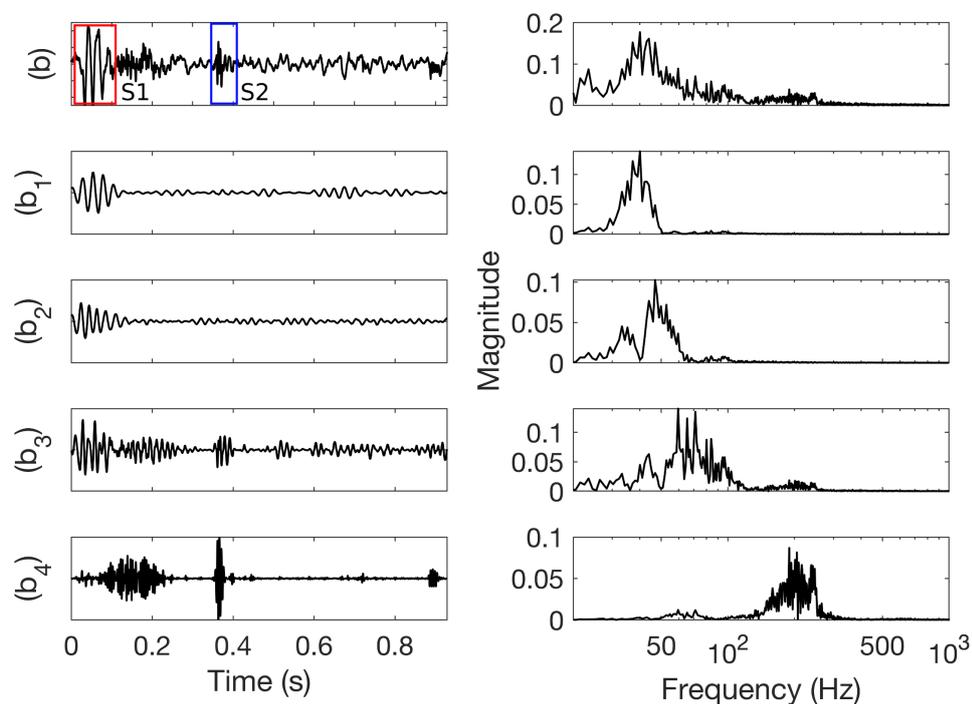


Figure 4.11: ICA of heart cycle (b). A heart cycle containing an early systolic murmur with no confirmed diagnosis.

Heart cycle (b), shown above in Figure 4.11, was modelled using 4 sources. This heart cycle contains a medium pitch murmur in early systole which can clearly be seen in  $(b_4)$

and  $(b_3)$ . The exact cause of the murmur in this recording is undiagnosed and it has not been associated with any valve pathologies. It can also be seen that the heart sound recording contains significant noise in the same frequency bands as the fundamental heart sounds. This noise can be observed in the original heart cycle  $(b)$  as well as the first three independent components. Component  $(b_1)$  contains mostly noise and the lower frequencies of the first and second heart sound; the second heart sound (which in this recording consists mostly of higher frequency components) is contained in  $(b_3)$  and  $(b_4)$ . The first heart sound is absent from the component  $(b_4)$  which has almost completely isolated the early systolic murmur from the rest of the heart sounds.

Figure 4.12 shows the CWT of heart cycle  $(b)$  along with the CWT of the independent components discovered by the probabilistic model. This figure presents the time-frequency information of the signals in Figure 4.11. Since this figure attempts to show more details of the separated heart cycle, the decomposition was performed using 48 voices per octave between 25 and 1500 Hz, resulting in a higher resolution (284 frequency bands) time-frequency representation. This method was followed for all the CWT figures of the independent components.

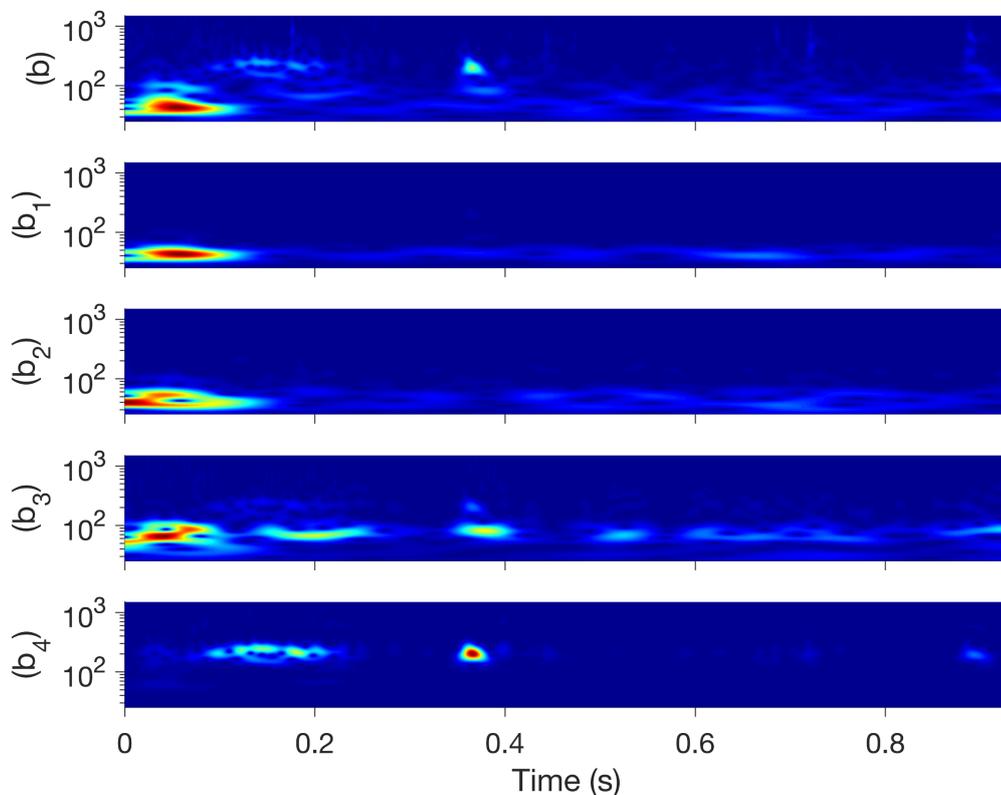


Figure 4.12: CWT of the independent components of heart cycle  $(b)$ .

The murmur, even this relatively simple one, greatly complicates the interpretation of the independent components and at this point it is necessary to stress that the following interpretation would in practice have to be integrated with more information about the patient in order to arrive at a more conclusive finding. The relatively simple approach used for heart cycle (a), comparing the frequency spectra of the two components and inferring that they originated from opposite sides of the heart, is much more ambiguous in the presence of noise and the murmur. We must then go to the time-frequency representations presented in Figure 4.12 to see if they can help us make sense of this heart cycle and the independent components discovered by our model.

The lower-frequency  $T_1$  sound has been captured in (b1) and (b2) meaning the ICA has separated components produced by the tricuspid valve into 2 separate components. The reason for this might be the presence of a large amount of low frequency noise in the same frequency band as the  $T_1$  sound, thus the model associates two parts of this sound more with parts of the noise signal. The remaining components, (b3) and (b4), are clearly associated with the murmur. In Figure 4.12 we can see that the mitral sound ( $M_1$ ) and the pulmonic sound ( $P_2$ ) have been grouped together with a low frequency component of the murmur in (b3), while the aortic sound ( $A_2$ ) has been grouped together with the high frequency components of the heart murmur in (b4). This could suggest that these sounds are associated with each other, i.e. the murmur is produced at the same valve with which it has been grouped, but to test such a hypothesis would require a greater level of information perhaps acquired by simultaneous echocardiogram recording.

The medium frequency early systolic murmur has been captured in component (b4). Based on the medium pitch, early systolic timing (component (b4)), and shape (crescendo-decrescendo) and the location of auscultation (aortic) of the murmur, it could be associated with mild aortic stenosis, mild aortic sclerosis, a murmur arising from the left ventricular outflow track, or an innocent flow murmur [36], [114], [115]. Since no pathological cause has been identified after expert investigation, it is likely that this is a *systolic flow murmur* (innocent murmur).

This heart cycle provides an example of the ability of ICA to isolate diagnostically interesting areas of a heart sound, as well as how ICA could serve as an aid to heart sound analysis in practice.

#### 4.7.3.3 Heart Cycle (c)

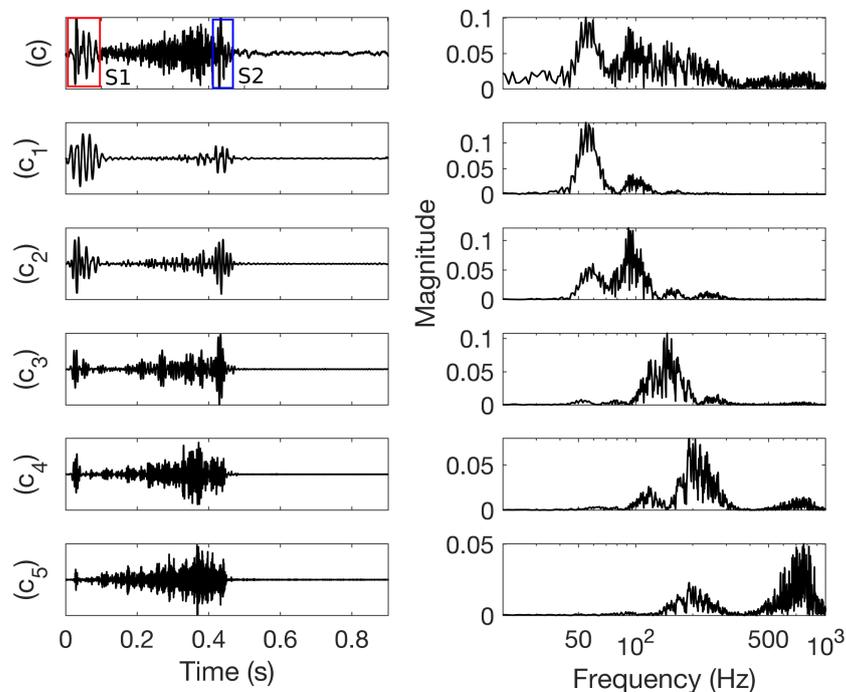


Figure 4.13: ICA of heart cycle (c). A heart cycle with a loud, mid to late systolic crescendo murmur that has been confirmed to be the result of mitral regurgitation due to mitral valve prolapse.

The murmur present in heart cycle (c), shown in Figure 4.13, is the result of mitral regurgitation due to mitral valve prolapse. It occurs during mid to late systole with increasing intensity (crescendo). The murmur contains frequencies from around 150 to 1200 Hz, overlapping with the higher frequencies of the fundamental heart sounds. The second heart sound is completely masked by the murmur when listening to the recording. The probabilistic ICA has largely managed to separate these different aspects present in the heart cycle. The heart sounds have been captured in (c<sub>1</sub>) and (c<sub>2</sub>), although it appears that (c<sub>2</sub>) contains some of the lowest frequencies of the murmur as well. (c<sub>3</sub>) and (c<sub>4</sub>) contain mainly the medium pitches of the murmur with some overlap from the upper frequencies of the fundamental heart sounds. The high pitch, and the most audible, part of the murmur is captured in (c<sub>5</sub>) and, to a lesser degree, in (c<sub>4</sub>).

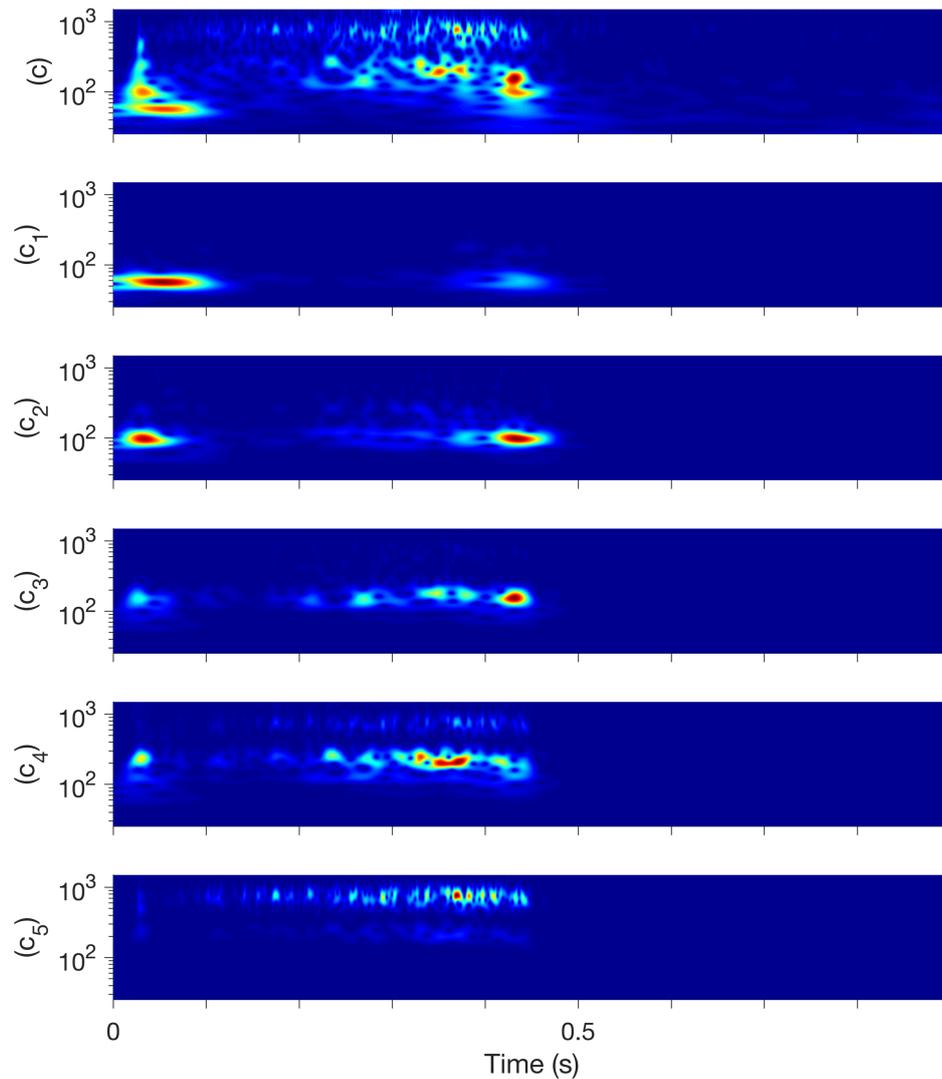


Figure 4.14: CWT of the heart cycle (c) and the independent components shown in Figure 4.13. Looking at the time-frequency CWTs of the independent components shown in Figure 4.14 it is clear that very little of the murmur has been grouped into component  $(c_1)$  and  $(c_2)$ . Rather all parts of the murmur are spread around the remaining components  $(c_3)$ ,  $(c_4)$ , and  $(c_5)$ . The timing and shape of the murmur (crescendo from early/mid-late systole), clearly seen in components  $(c_4)$  and  $(c_5)$ , agree with the description of *mitral regurgitation* [36].

It is of particular value to be able to listen to the isolated murmurs captured in components  $(c_4)$  and  $(c_5)$ . The proposed model provides an investigator the opportunity to listen, with adjustable volume and speed, to different aspects of the heart sound and to assess and analyse these in isolation.

#### 4.7.3.4 Heart Cycle (d)

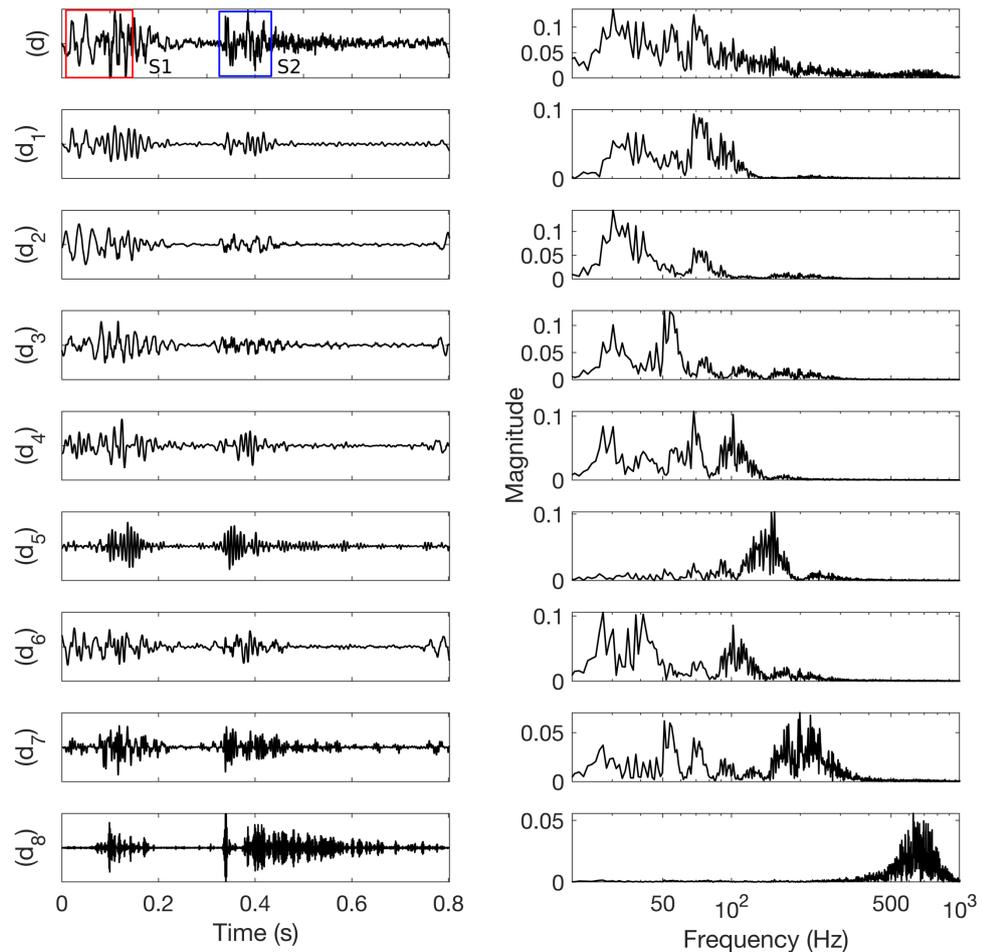


Figure 4.15: ICA of heart cycle (d). A heart sound with both an early systolic and an early diastolic murmur.

The final heart cycle examined presents the most challenge in terms of model complexity and the most diagnostically interesting aspects. The heart sound recording was segmented into what, at first, appeared to be S1-Systole-S2-Diastole. The S1 sound appeared to be significantly split in both the time plot of the heart cycle and when listening to the recording. After viewing the CWT and ICA results (Figure 4.8 and Figure 4.15) it becomes obvious that the very low frequency event right at the start of the heart cycle (around 0 – 100ms) does not fit into this pattern. This is especially visible in the frequency spectrum of components (d<sub>2</sub>) and (d<sub>6</sub>) in which most of this event has been captured. While it is possible that this event represents the lower frequency T<sub>1</sub> sound, this would then also mean that there is clinically implausible delay in the closing of the mitral valve (M<sub>1</sub>), since this sound normally occurs first. Inspecting the frequency spectra of components (d<sub>2</sub>) and (d<sub>6</sub>) it is clear that this event is lower in frequency even than the tricuspid component (as observed in heart cycle (a) for example). It seems more likely

that this event is a *fourth heart sound (S4)* [35] and this heart cycle has been segmented into S4-S1-Systole-S2-Diastole. Since the first heart sound indicates the start of ventricular systole the fourth heart sound technically occurs in diastole and thus a more appropriate segmentation of the heart sound would be S1-Systole-S2-Diastole-S4, as shown in Figure 4.16.

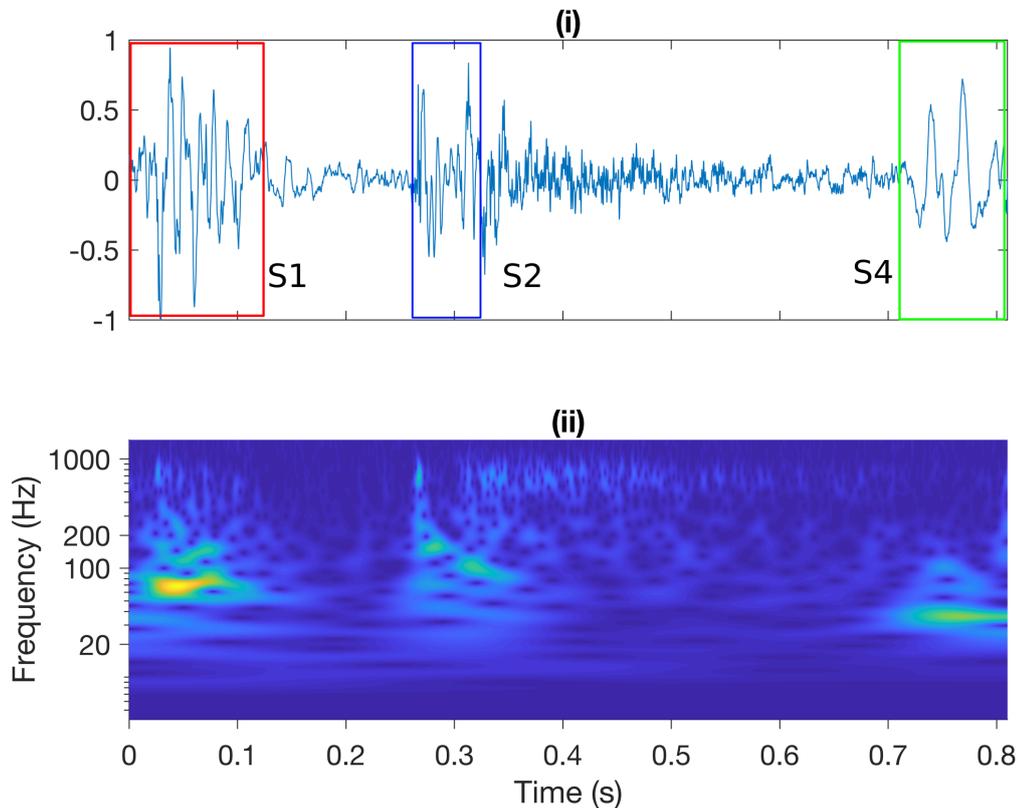


Figure 4.16: A re-segmentation of heart cycle (d) after analysis of CWT and ICA results showed the clear presence of an S4 sound. (ii) Time-frequency representation of the re-segmented heart cycle shown in (i). Decomposition performed using 48 voices per octave from 0 to 1500 Hz, specifically to show S4 sound more clearly.

After manually re-segmenting, the ICA model was fit to the new data. The CWT for these components, labelled ( $d_9$ ) to ( $d_{18}$ ) are shown in Figure 4.17. These components are very similar to the ones shown in Figure 4.15, but are easier to interpret since they are in the expected order of systole and then diastole. The S4 sound has been captured in component ( $d_9$ ) and to a lesser degree in ( $d_{10}$ ). There is a split in the second heart sound, especially visible when comparing ( $d_{13}$ ) and ( $d_{14}$ ), as well as in the difference between the high frequency spike of the second heart sound and the start of the heart murmur visible in ( $d_{15}$ ) and ( $d_{16}$ ). The medium and higher frequencies of the heart murmurs have been captured in ( $d_{15}$ ) and ( $d_{16}$ ).

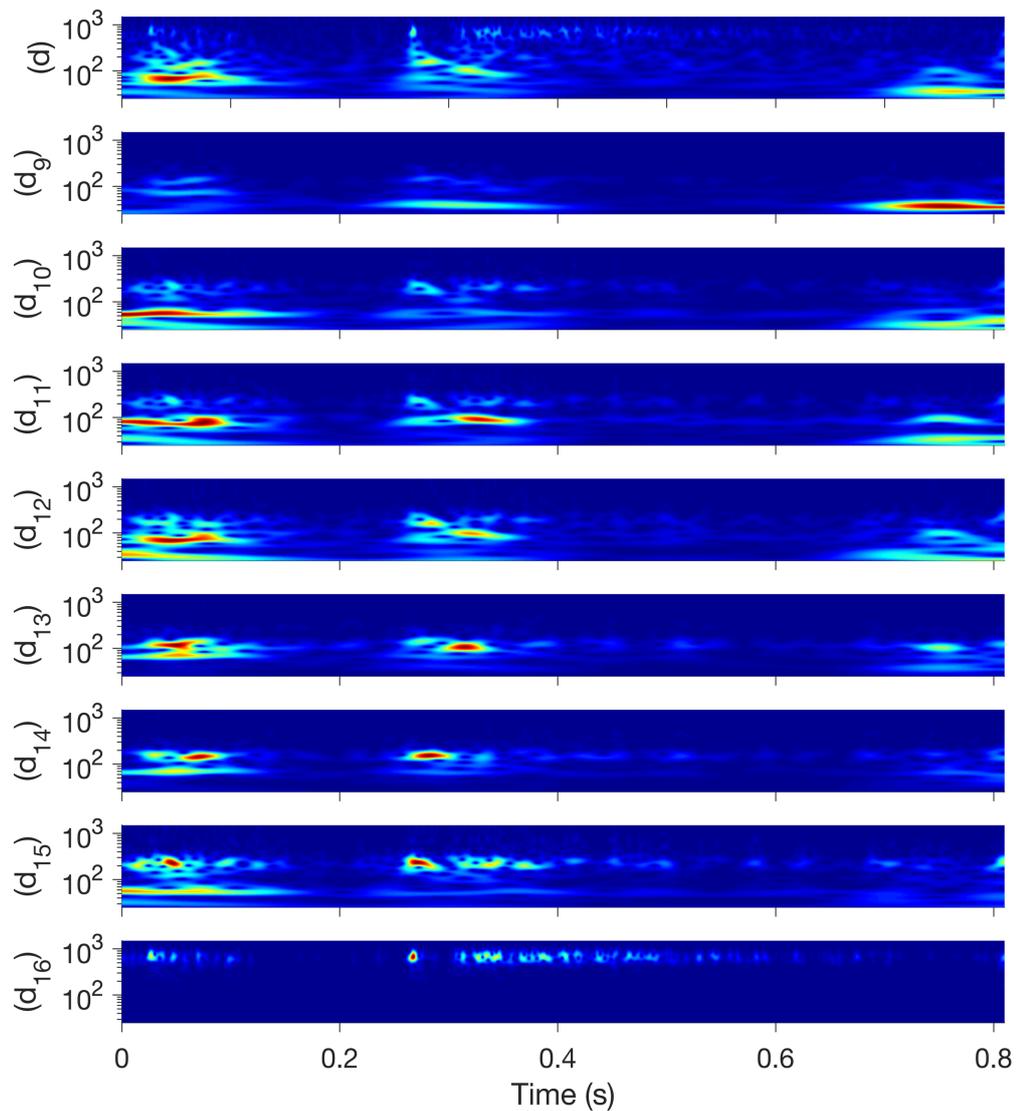


Figure 4.17: CWT of the re-segmented heart cycle (d) and the independent components shown in Figure 4.16. Although the components remain similar to those from Figure 4.15 the re-segmented heart cycle provides a visual representation that is easier to interpret, since the different heart sounds are in a more expected order (S1-Systole-S2-Diastole-S4)

In the case of heart cycle (d) the combination of CWT and ICA was able to assist in the heart sound pre-processing stage of segmentation. The S4 sound present in this recording was made clear in the time-frequency decomposition as well as in the resulting independent components. The physiological interpretation of the ICA results for this heart cycle will be developed using the CWT of the components as shown in Figure 4.17.

The component ( $d_{11}$ ) consists mostly of the S4 sound that was discovered during the initial CWT and ICA analysis. Components ( $d_{15}$ ) and ( $d_{16}$ ) have captured medium and high frequency content of the heart murmurs respectively. These two components

demonstrate a clear split in the second heart sound, especially visible between the initial high frequency spike in the second heart sound and the start of the diastolic murmur in ( $d_{16}$ ). Because of the timing and the high frequency content we can infer that this sound (between 0.25 and 0.3 seconds) is the  $A_2$  component of the second heart sound. The murmur clearly starts with the sound just following this, the lower frequency  $P_2$  sound. This, along with the decrescendo shape, high frequency, and timing of the murmur, suggests that this is a case of *pulmonary valve regurgitation* [116].

The brief, softer murmur at the start of systole occurs at the same time as the first heart sound and is indistinguishable from  $S_1$  when listening to the heart sound recording. Looking at components ( $d_{10}$ ) to ( $d_{16}$ ) the first heart sound in this heart cycle is quite complex, there are several sound events happening at the time of the first heart sound. The  $S_4$  sound occurs right before the onset of  $S_1$  and although very low frequency the components are quite high amplitude. The residual or continuing components of  $S_4$  are especially apparent in components ( $d_{10}$ ), ( $d_{12}$ ), and ( $d_{15}$ ). Looking at components ( $d_{11}$ ) and ( $d_{14}$ ) we can see two distinct peaks in the first heart sound, these likely correspond to the  $M_1$  and  $T_1$  components of  $S_1$ , suggesting that the ICA model has not split the sounds in this heart cycle based on their origin alone. It seems likely that due to the presence of other more complicated (or diverse) sounds these sounds exhibit more statistical dependence between each other than between the other sounds and thus get grouped together. This does complicate the interpretation of the high frequency murmur that coincides with the first heart sound. Taking into account the presence of the prominent diastolic murmur, what appears to be a systolic murmur is most likely a *pulmonary ejection sound* [117] caused by a structurally abnormal pulmonary valve and increased pressure across the valve.

There is also the possibility that the murmurs are generated as the result of turbulent flow across the aortic valve. The delay in the start of the regurgitant murmur could be the result of a delay of blood backflow into the ventricle as the valve remains closed at the start of diastole. After a moment the aortic valve loses integrity and allows blood to flow backwards from the aorta into the ventricle. This heart cycle would then be an example of *aortic valve regurgitation*. The location of recording (pulmonic area) suggests that this is more likely to be pulmonary regurgitation although this single heart

sound recording is not enough to base a diagnosis on, and further investigation and more information would be needed to confirm either of these conditions.

This heart cycle provided an excellent challenge for the proposed methodology. It was necessary to make use of all the information provided by CWT and the ICA of the CWT in order to arrive at a reasonable explanation for the sounds present in this heart cycle. The model was able to clarify the presence of a segmentation error by clearly distinguishing a fourth heart sound before the first heart sound. The obvious diastolic murmur and subtle and almost undetectable (when listening) ejection sound present in the recording were captured by the model and could be analysed in detail.

#### 4.7.4 Number of Sources

The most influential hyperparameter governing the results of the ICA are the number of hidden sources to be modelled. Figure 4.18 shows the marginal log-likelihood of the data of heart cycle (a) given the model parameters for an increasing number of hidden sources. To ensure the model remained tractable this experiment was performed on the simplest of the heart sounds and a reduced resolution CWT with 4 voices per octave was used; this resulted in 25 input vectors ( $M = 25$ ). Model learning was performed from 1 hidden source to a maximum of 25 hidden sources ( $M = L$ ). With an increasing number of sources, the log-likelihood (and thus the likelihood itself) increases, and levels out at around 17 sources. A plot of the independent components of heart cycle (a) with  $L = 17$  is shown in Figure 4.19.

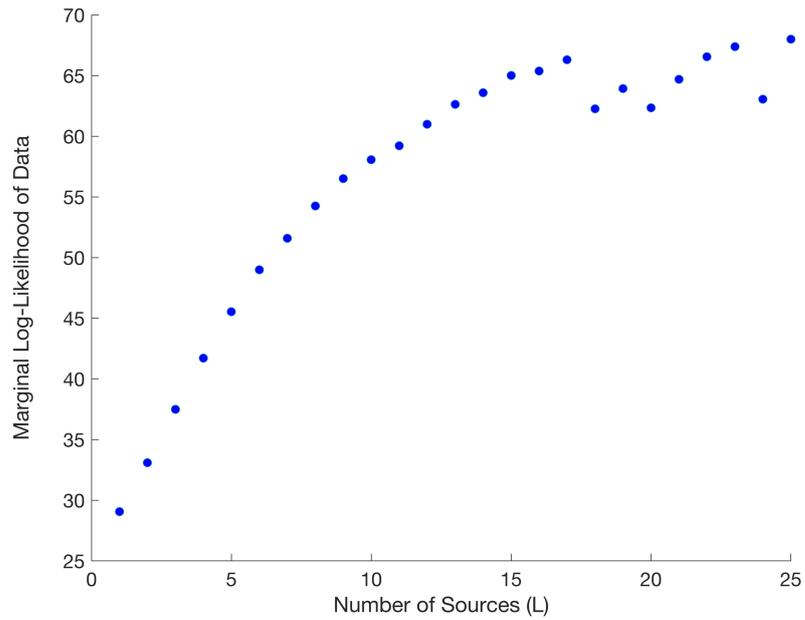


Figure 4.18: Final (converged) value of marginal log-likelihood of heart cycle (a) input data with different numbers of latent sources.

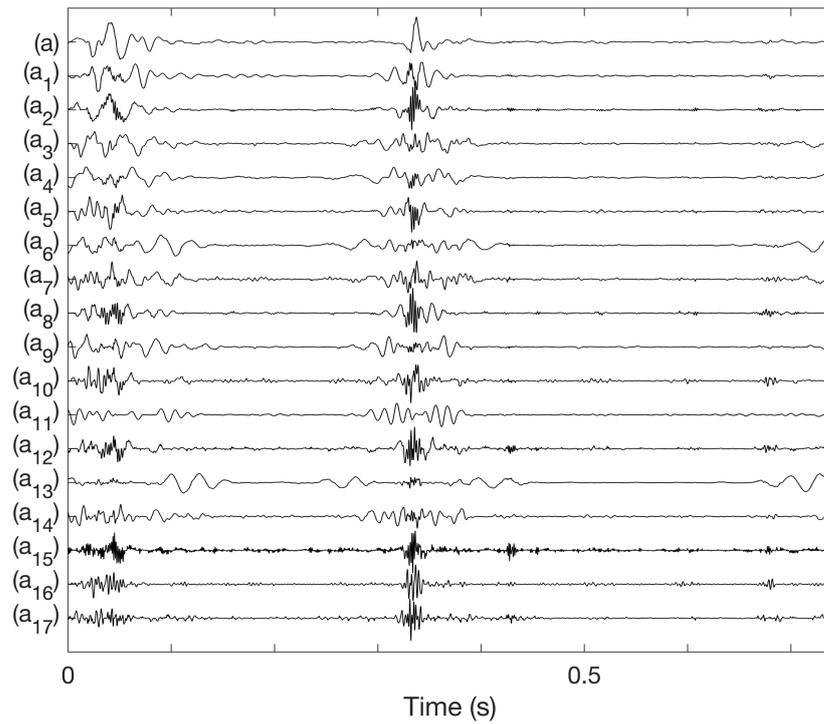


Figure 4.19: Independent components of heart cycle (a) with  $L = 17$  hidden sources.

Figure 4.19 represents a set of independent sources  $\mathbf{s}$  that, given the estimated mixing matrix  $\mathbf{A}$  and the sensor noise  $\mathbf{n}$ , can optimally reconstruct the heart cycle (a). ‘Optimally’ used here in the sense of how likely is (what is the probability of) the original heart cycle (a), given the set of independent sources,  $\mathbf{s}$ .

## 4.8 Discussion

The figures of the discovered sources and their associated frequency spectra show clearly that the proposed methodology can find and group interesting, and seemingly associated, aspects of heart cycles. Importantly, these separated components can be interpreted sensibly considering knowledge about the physiology and functional anatomy of the heart and the disease states that present with heart murmurs. In the discussion each of the heart cycles and their discovered independent components will be assessed in this manner.

### 4.8.1 ICA Sources

A major issue with applying ICA is the selection of an appropriate number of sources. The number of sources should provide a sufficient representation of all features present in the heart sound without overfitting the dataset, i.e. fitting so many components that small variations resulting from noise are captured, making the model and the results too obscure to interpret.

As shown in Figure 4.18, in terms of the likelihood of the data, heart cycle (a) produced by a model with  $L = 17$  is a better fit than with  $L = 2$ . Comparing Figure 4.19 to Figure 4.9 it is obvious that this higher likelihood does not translate into a more informative, or meaningful, representation of the heart cycle. Figure 4.18 and Figure 4.19 clearly demonstrates the overfitting discussed previously. The large degree of freedom provided by such a large number of sources allow the model to almost exactly fit all the aspects of the original signal, analogous to a 9<sup>th</sup> degree polynomial being able to exactly fit a dataset with 10 data points [94]. The model becomes representational only of the specific example. More specifically to the analysis of heart sounds, searching for 17 sources separates the heart cycle into components that are not recognisable as anything a person would hear when listening to the recording.

The issue remains as to how we can decide on a suitable number of sources. There is no simple number we can look at to say whether a certain set of sources are more suitable than another. In any case, 'more suitable' would vary according to the exact nature of the problem being addressed. In terms of heart sound analysis, the methodology followed in this work has been to increase the number of independent sources until all

components that are obvious in the original heart cycle, that is, all the components that are audible when listening to the recording and all components that are visible in the CWT of the heart cycle, are represented in at least one of the independent components. In this way the minimum number of components able to sufficiently represent the heart cycle are chosen. This methodology remains subjective i.e. it relies on the expertise of the investigator; it should be noted that the final test of any blind source separation (in which we don't know the actual generating sources) is whether or not an 'expert' would come to the same conclusion.

#### 4.8.2 Model assumptions

The model makes certain assumptions about the nature of the data being analysed. These assumptions need to be scrutinised to assess their reasonableness and possible side effects.

##### 4.8.2.1 Heart sound recordings

The ICA model presented in this chapter contains very little information specific to the system we are analysing. One key assumption that underlies the results presented in this chapter is that the heart sounds are the 'dominant' or 'most prevalent' signal in the recording being analysed. This assumption influences one of the key limitations of the model, the user input required to select an appropriate number of sources. Using the heart cycles presented in this chapter as examples, heart cycle (b) has a lower signal-to-noise ratio than heart cycle (c). In other words, heart cycle (c) conforms more to the assumption that the sound produced by the heart are the most prevalent signals in the recording. Looking at the independent components of the two cycles (Figure 4.11 and Figure 4.13) we see that there exists less overlap between the ICs of heart cycle (c) than those of heart cycle (b). Because there is less noise obscuring the heart sounds in heart cycle (c) the model does not identify associated sounds in the recording as being from different source distributions because of the presence of noise with similar characteristics. In high SNR recordings the model is more readily able to disregard noise and separate out interesting parts of the heart sounds.

##### 4.8.2.2 IID Data

The model assumes all samples are independent and identically distributed (IID) and thus all information about the relation of samples across time is disregarded. In other

words, the model assumes the data is generated instantaneously, or completely independently at each time point. This limitation is mainly due to the overwhelming complexity involved in incorporating correlations across time into the model. The probabilistic ICA model described in this section is already quite complex and extending the model to include information about correlations across time requires a significantly more complex model. Even so, modelling the temporal information contained in the heart sound could uncover information that the current ICA model is unable to.

#### 4.9 Future Work

A major issue that remains to be addressed is the selection of an appropriate number of sources to model. At this stage the recommendation is that this hyperparameter remain tuneable as it provides a measure of control over what could be viewed as the resolution of the resulting view of the heart cycle. It is thus beneficial to analyse heart cycles at multiple levels (number of sources) to reveal both coarse (explicit or obvious) and finer details present in the heart cycle. The minimum number of sources required to model all aspects of the heart sound could also serve as an interesting feature for measuring the relative complexity, and perhaps likelihood of pathology, of a heart cycle.

The methodology proposed in this chapter is an example of blind source separation. It is a form of machine learning/pattern recognition and it should be emphasized that there is no sense of “knowing” in the algorithm, only recognition and separation of the data statistics. The current methodology provides more information about the heart sound, but this information still must be interpreted by an expert and put into the greater diagnostic context to determine its clinical significance. Thus, the model would have to be incorporated into a greater framework to be able to automatically describe or classify a heart sound.

Future work could focus on developing more informative priors based on expert knowledge. The Bayesian (probabilistic) model allows us to incorporate our expectations and current knowledge about the system we are investigating (modelling). This is however not a trivial undertaking as it requires translating qualitative and subjective knowledge into probability distributions of often abstract aspects of the model.

The current ICA model also disregards any correlations across time. It has been demonstrated that heart sounds contain non-linear correlations [63] and determining the timing of the heart sounds and any extra sounds/murmurs is a very important step in cardiac auscultation [113]. The model can be extended to incorporate temporal information in several ways. Temporal information and non-linear correlations can be summarised by extracting the bispectrum from higher order moments calculated from lagged autocorrelations of the signal. Initial research by Ahlström et al. [63] has shown differences in bispectra of different disease states although features extracted from the bispectra failed to differentiate between these. Alternatively, the model could be extended to include latent (hidden) states which the system moves between. Hidden Markov models (HMM) have been applied to the problem of heart sound segmentation [68], [111] in which the expected duration that the system (heart) stays in each state was included as an explicit parameter. Choudrey [96] approached the problem of temporal information by incorporating HMMs into the ICA model both by replacing the Mixture of Gaussian source models of the ICA with Hidden Markov source models and by replacing the HMM observation models with ICA models, these methods were both developed under the title *Dynamic ICA*.

#### 4.10 Conclusion

In this section the use of independent component analysis has been explored as a tool to find interesting structures in time-frequency transforms of heart sounds. ICA attempts to maximise the statistical independence between subsets of the data set. The hypothesis that was tested in this section can then be summarised as "the interesting, or diagnostically useful, parts of a heart sound recording are statistically independent from each other and can be separated by independent component analysis." This was tested on a small but diagnostically interesting set of heart sounds containing a variety of examples of heart murmurs as well as other common cardiac sounds. The usefulness of the proposed ICA model fitted to CWT derived time-frequency representations of the heart cycles was explored by interpreting the results considering knowledge about the physiology and pathology of the heart. In each case the discovered model sources (independent components) could be interpreted sensibly under assumptions of physiological relevance. In conclusion it has been demonstrated that probabilistic

independent component analysis is able to separate heart cycles into parts that reveal diagnostically interesting features in the heart sound.

The independent components model provides a framework for introducing probabilistic modelling as a methodology for building a computational model of the heart sounds. We have shown that this model is able to separate heart cycles into interesting components. Because of relatively small amount of assumptions we had to impose, ICA provided an interesting starting point for exploring heart sounds, however this also limits the utility and the interpretability of the results. The interpretation of these results required *context* and *knowledge* that are not easily included. The most obvious, and critical, information about the signal that is not included in the model are the time dependencies. In Chapter 5 we explore modelling the temporal characteristics of heart sound recordings, moving another step closer towards a *cardiac auscultation algorithm*.

## Chapter 5 Probabilistic Labelling and Segmentation of Heart Sounds



Choosing any arbitrary starting time in the heart cycle, the heart sounds have an obvious order in which they occur. Modelling these temporal correlations adds a wealth of information to the analysis. Determination of the temporal information present in a heart sound signal is important enough that it is usually treated as an independent step of the processing. This step, commonly referred to as *heart sound segmentation*, involves accurately determining the locations of the first (S1) and second (S2) fundamental heart sounds. Once these sounds have been determined the heart cycle periods of ventricular systole and ventricular diastole can be distinguished from each other. Heart murmurs are fundamentally grouped into diastolic and systolic murmurs and much of the information about which valve is affected and what condition is the likely cause of the murmur is contained in the timing of the murmur in the heart cycle [36], [116].

The temporal information extracted by heart sound segmentation can be used on a finer level. There is important diagnostic information contained in the relative timing of the components of each of the fundamental heart sound, i.e. the mitral (M1) and tricuspid (T1) components of S1 and the aortic (A2) and pulmonary (P2) of S2. Within each of the FHS the component sounds can present with varying degrees of splits that contain clues about the functioning of the heart. For example, the second heart sound is normally split during inspiration but does not present with any splitting during expiration. Fixed splitting of S2, that is splitting during both inspiration and expiration, can be indicative of structural defects (i.e. atrial septal defect) or severe right heart failure [112].

Simultaneously recorded signals potentially have higher signal to noise ratio (SNR) have well understood and studied correlations with the mechanical events of the heart. Staying with the example of ECG, the start of ventricular systole necessarily follows the depolarisation of the ventricles which is indicated by the QRS-complex in an ECG. It is straightforward then to derive the locations of the first heart sound in a heart sound recording with simultaneously recorded ECG, especially considering the wealth of algorithms for R peak detection [118]. The main shortcoming of dependent segmentation is the increased complexity associated with simultaneously recording a

second physiological signal. Dependent segmentation is not applicable to the large open source databases available online and are much more complicated to attain than heart sound recordings that can simply be recorded and stored by cardiologists using electronic stethoscopes during their day-to-day practice. This is not to say that dependent segmentation is without use. When available, simultaneous ECG recordings, along with manual expert annotations, can serve as a valuable gold standard for training and testing independent segmentation algorithms.

Independent segmentation, that is, heart sound segmentation using only information from the heart sound signal itself, is an active area of research. The usual methodology is comprised of an envelope extraction step followed by a peak picking step in which the most likely locations of the fundamental heart sounds are identified, these are then labelled as S1 and S2 using the duration characteristics of systole and diastole. The focus of most researchers has been the use of different envelope extraction techniques in order to highlight the occurrences of S1 and S2 while diminishing noise that could obscure the peak picking process. Techniques of heart sound recording envelope extraction that have been investigated and reported in the literature include energy measures such as Shannon envelopes [80], [119], spectral features [78], [120], wavelet features [121], complexity measures [122], and recurrence statistics [91]. An extension to this framework, proposed by Gamero and Watrous [123], Gill et al. [66], and Ricke et al. [124], is to replace the peak picking step by a probabilistic model, specifically a Hidden Markov Model (HMM). This work was further developed, first by Schmidt et al. [68] to include explicitly modelled state durations (duration dependent HMM), and then by Springer et al. [70] who tested the algorithm on a larger dataset and further investigated the model parameters and envelope extraction methods. The work presented in this chapter is a natural extension of this previous work in which the model has been extended to label other interesting features of the heart sound. Specifically, the HMM model is extended in this work to identify systolic murmurs. The continuous wavelet transform is also used to derive spectral features that are investigated using random forest and logistic regression classifiers to discriminate between samples from each of the states.

## 5.1 Hidden Markov Model Parameters

Markov models are probabilistic models that explicitly model the dependency between successive samples. A first order Markov process assumes that the state a system is in at time point  $t$  depends only on the state the system was in at time point  $t - 1$ . Markov processes can be extended to include more previous state information (i.e.  $t - 2, t - 3, \dots, t - 4$ ) producing higher order Markov processes at the cost of exponential growth of the number of parameters. Hidden Markov models increase the usefulness of Markov models by modelling the state dependencies as latent variables that produce (emit) an observable signal and thus the state at time  $t$  depends both on the state at time  $t - 1$  and the observation at time  $t$ . Following Rabiner [125] we denote the state occupied at time  $t$  by  $q_t$ .

### 5.1.1 Model Parameters

Hidden Markov models can be described by the number of states in the model and three probability measures. In an HMM with  $N$  distinct states, the probability of transitioning between states is denoted by the  $N \times N$ ,  $\mathbf{A}$  matrix, with  $a_{ij}$  representing the probability of moving from state  $i$  to state  $j$ . The state emission probabilities, denoted by  $\mathbf{B}$ , represent the probability of an observation being produced by each of the  $N$  states. Finally, the probability of the system starting in each of the  $N$  states is denoted by the  $1 \times N$ ,  $\boldsymbol{\pi}$  vector. More formally these parameters can be defined as

$\mathbf{A} = \{a_{ij}\}$ , where

$$a_{ij} = p(q_{t+1} = j | q_t = i) \quad 1 \leq i, j \leq N \quad (5.1)$$

$\mathbf{B} = \{b_i(\mathbf{x}_t)\}$ , where  $\mathbf{x}_t$  is the observation vector at time  $t$  and

$$b_i(\mathbf{x}_t) = p(\mathbf{x}_t | q_t = i) \quad 1 \leq i \leq N \quad (5.2)$$

$\boldsymbol{\pi} = \{\pi_i\}$ , where

$$\pi_i = p(q_1 = i) \quad 1 \leq i \leq N \quad (5.3)$$

The complete set of model parameters can then be represented as

$$\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}) \quad (5.4)$$

A complete introduction on the use of hidden Markov models and examples of applications in speech processing is presented in a 1989 paper by Rabiner [125].

Following the example of Schmidt et al. [86], and Springer et al. [70], a first order hidden Markov model (HMM) is used to model the system that produces the cardiac sounds (i.e. the blood flow through, and mechanical action of, the heart). At all times it is assumed that the system is in one of four discrete states, the first heart sound (S1), the period of ventricular contraction (systole), the second heart sound (S2), or the period of ventricular relaxation (diastole) and moving through these states one after the other. In a hidden Markov model the probability of staying in the same state for  $T$  consecutive steps is an exponentially decaying function of  $T$  [94], [125]. This constraint is clearly not suitable for modelling the heart cycle states. Schmidt et al. [86] proposed a duration dependant HMM (see Rabiner [125] and Yu [126] for details) as a more suitable model for heart sound segmentation. In this model the state transitions are limited to a single following state, thus the state transition probabilities simplify to

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad (5.5)$$

This model can be represented using a unidirectional, nonergodic, finite state machine as shown in Figure 5.1.

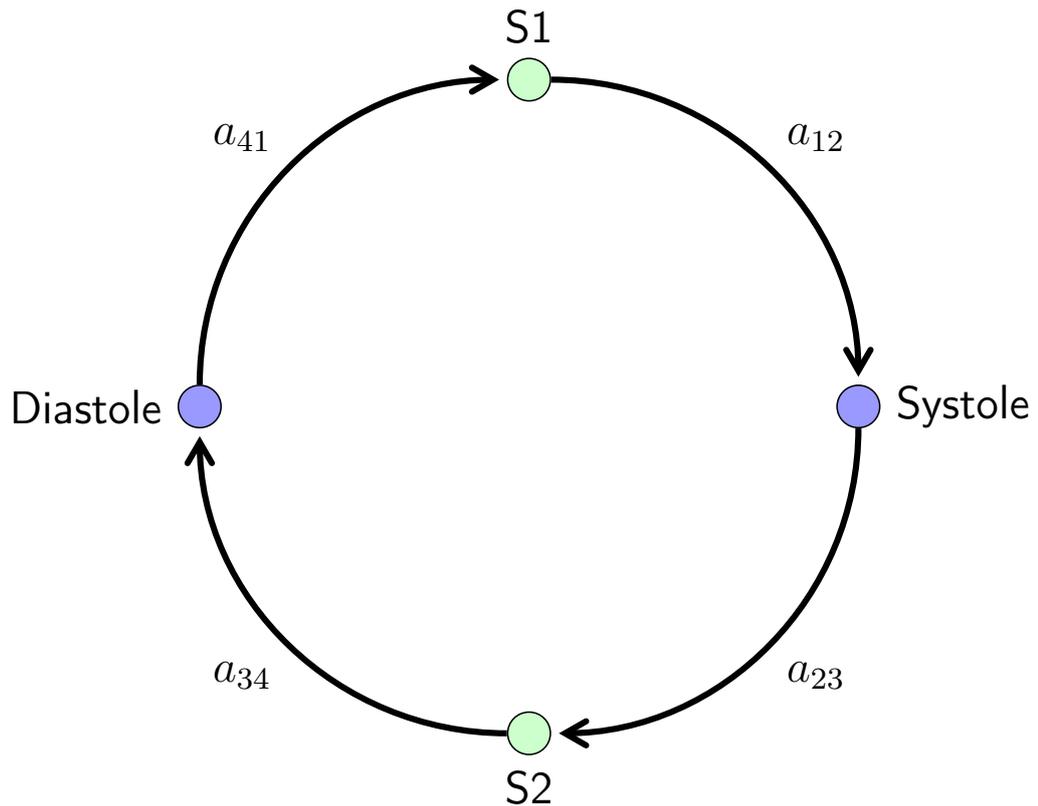


Figure 5.1: A state diagram illustrating the possible heart cycle states that the heart can occupy in the HMM as proposed by Schmidt et al.

This extension of the HMM explicitly models the duration distribution for each of the states and the model parameters can be expressed as

$$\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}, \mathbf{D}) \quad (5.6)$$

with the duration distribution parameters denoted by  $\mathbf{D}$ . This duration dependant model is represented in more detail in Figure 5.2. The figure illustrates the underlying assumptions of the model: As the system moves from state to state, with transitions governed by  $\mathbf{A}$ , and occupancy duration governed by  $\mathbf{D}$ , it produces an observable signal, the distribution of which is governed by  $\mathbf{B}$ . The emission is explicitly illustrated for the S1 state. For practical reasons the emitted signal is assumed to be envelope transforms of the actual audio recording, in this way noise can be mitigated and a higher degree of discrimination between states achieved.

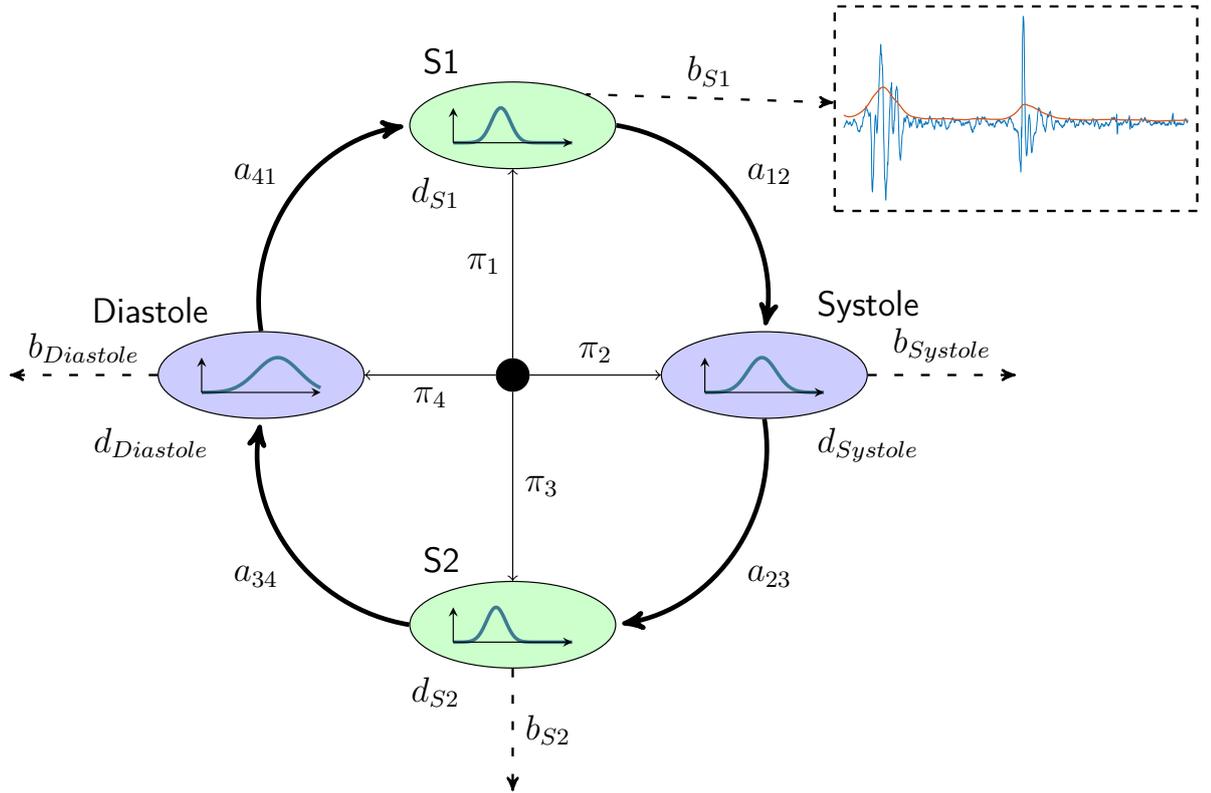


Figure 5.2: Model elements for a duration dependant heart sound segmentation hidden Markov model. The probability of the system starting in any of the given states is given by  $\pi$  and the duration of each state is explicitly modelled (e.g.  $d_{S1}$ ). The system can only transition to the next state in the sequence and thus all the  $A$  (e.g.  $a_{12}$ ) probabilities are equal to 1. While occupying each of the states, the system emits observations according to the probabilities in  $B$ , e.g.  $b_{S1}$ . In this work the emission probabilities are the probabilities of a state given the observations, derived using a classification system (e.g. random forest/SVM), for each sample.

### 5.1.2 The Viterbi Algorithm

For the model to be useful the "best" or "most appropriate" sequence of states needs to be identified. "Most appropriate" in the case of heart sound segmentation can be abstractly defined as "the sequence of states that most likely correspond to the actual states that the heart occupied when the observations were generated". To the model this would correspond to the state sequence that maximises the emission probabilities under the constraints of the transition and duration probabilities. To derive this sequence of states from the model parameters in a computationally tractable way we use a *dynamic programming* procedure termed the *Viterbi algorithm* [127]. The Viterbi algorithm finds the most probable state sequence by recursively calculating the product

of the emission, transition, and duration probabilities and keeping track of the state sequence that maximises this probability [94], [125].

The specifics of the Viterbi algorithm for a duration dependent HMM are shown in Figure 5.3. The most probable sequence that ends in state  $j$  at time  $t$  is defined as  $\delta_t(j)$  and is computed recursively for all time points and states. To compute  $\delta_t(j)$  it is necessary to keep track of the indices that maximise. These are represented by  $\Delta_t(j)$  and  $\psi_t(j)$  for time point  $t$  and state  $j$ . These quantities are calculated recursively for all time points and analysis window sizes. Finally, in the state path backtracking step, the state path is constructed from  $\Delta_t(j)$  and  $\psi_t(j)$ . The trellis diagram in Figure 5.3 illustrates the Viterbi decoding for a single heart cycle. All possible state paths (based on the transition matrix) are shown in dashed grey lines while the most probable state path is shown with a solid line and includes notations. The notation around the nodes show the products that are used to calculate the probabilities for each state and the notation inside the node represents the variable size analysis window for which the probabilities are calculated at each time step to find the most probable combination of duration and emission probability.

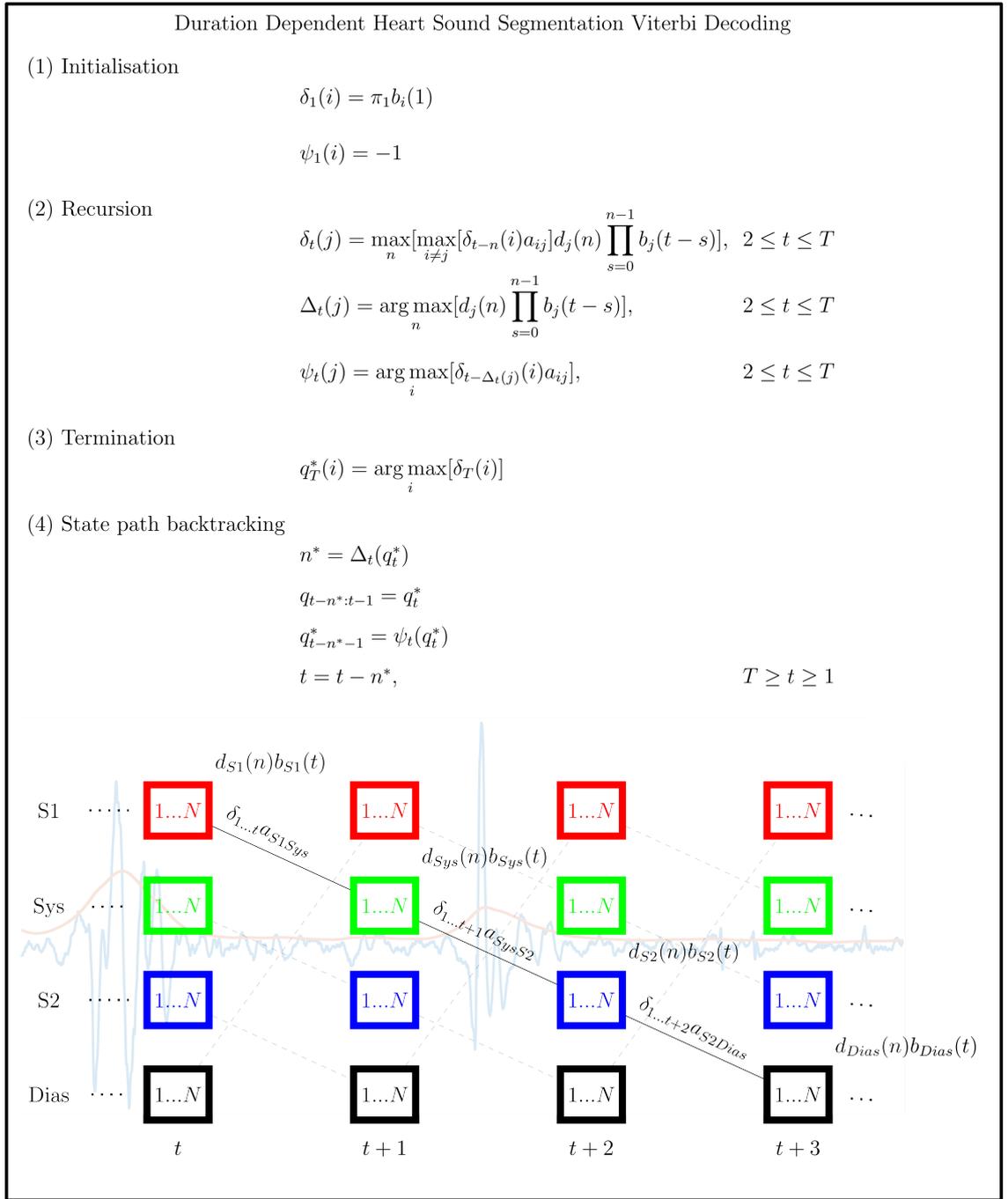


Figure 5.3: The Viterbi Algorithm for the duration dependent Hidden Markov Model used for heart sound segmentation along with a trellis diagram illustrating the algorithm for a complete heart cycle (shown in the background). The algorithm consists of 4 parts and allows the sequence that maximises the emission and duration probabilities to be computed in an optimal way. The states are indexed by  $i, j \in [S1, Systole, S2, Diastole]$ , the sequence is of length  $T$ , and the analysis window has maximum length of  $N$ , the maximum duration of any single state set equal to the duration of a complete heart cycle. The algorithm and its extensions are discussed more fully in the main text and in literature [68], [70], [126].

## 5.2 Methods

### 5.2.1 Dataset

The dataset for this work was assembled from the Physionet database [128] which in turn consists of several independent heart sound databases. Examples of systolic murmurs were selected based on the reported classification (normal/abnormal) as well as the presence of a murmur when listening to the recording and when viewing a time-frequency representation of the heart sound. The "true" locations of the fundamental heart sounds were annotated in 3 ways: manually based on audio and visually from envelopes and spectrogram representations, automatically using the heart sound segmentation proposed by Springer et al. and finally using ECG annotations when available. These three sources of annotations were synthesised manually where discrepancies occurred. Each of the selected recordings was from a different subject. Due to the rather intensive data preparation, in terms of manual segmentation and murmur identification, the number of training samples were chosen such that the addition of further samples did not significantly alter the distribution of samples. While this suggests that the training sample is, at least, representative of heart sounds (from a range of different recordings methods and locations) the additions of further, high quality data would, as always, improve the model.

#### 5.2.1.1 Training

The training set consisted of 118 recordings from 118 different subjects with 50 recordings containing systolic murmurs and 68 with no abnormal heart sounds present. The training set was chosen from the Physionet Heart Sound Database on the basis that each recording be from a different participant. The training set was used to train the emission probability model and as such the most relevant description of the dataset is that of the amount of time (or equivalently number of samples) that each heart sound state (S1, systole, S2, diastole, systolic murmur) represents in the dataset. The number of samples from each state was balanced before training of the emission probability model commenced by under-sampling all classes to the 12884 samples of the systolic murmur class. All recordings had a sampling rate of 4 kHz and feature vectors were downsampled to 50 Hz to ensure the model stayed computationally feasible. Thus, in

total the model was trained on 1288 seconds of heart sound data with an average of 11 seconds of data from each subject or 258 seconds of data per state.

#### 5.2.1.2 Testing

A separate dataset of recordings was used to test the results of the model. Recordings from 23 subjects presenting with systolic murmurs and 33 subjects without abnormalities in their heart sounds were used, again selected from the Physionet database. The 23 systolic murmurs represent a slightly more challenging (in terms of segmentation) subset of the Physionet database due to the presence of, in some cases very prominent, murmurs that obscure the locations of the fundamental heart sounds. Since these recordings were used to test the overall performance of the segmentation algorithm, the most relevant description is of the individual recordings. The total length of the dataset was 1076.9 s and the mean recording length was  $19.2 \pm 10.2$  s. Recordings with murmurs totalled 346.3 s and averaged  $15.1 \pm 0.5$  s with a total of 380 instances of S1 and 376 of S2. Recordings without abnormalities totalled 730.6 s and averaged  $22.1 \pm 12.5$  s with a total of 862 instances of S1 and 859 instances of S2. This imbalance in the testing dataset is more representative (although to a much lesser degree) of the imbalance in the population; most people do not have heart murmurs. Metrics reported for all but the emission modelling were tested exclusively on the testing dataset.

#### 5.2.2 Extensions to the Model

The HMM heart sound segmentation, as implemented by Springer et al. and Schmidt et al., can be conceptualised as a two-step classification process. In the first step the samples of the heart sound signal are first probabilistically scored into different classes by a classifier corresponding to the derivation of the emission probabilities. These probabilities are then combined with the state duration and state transition sequence information to form a more accurate picture of the heart sound.

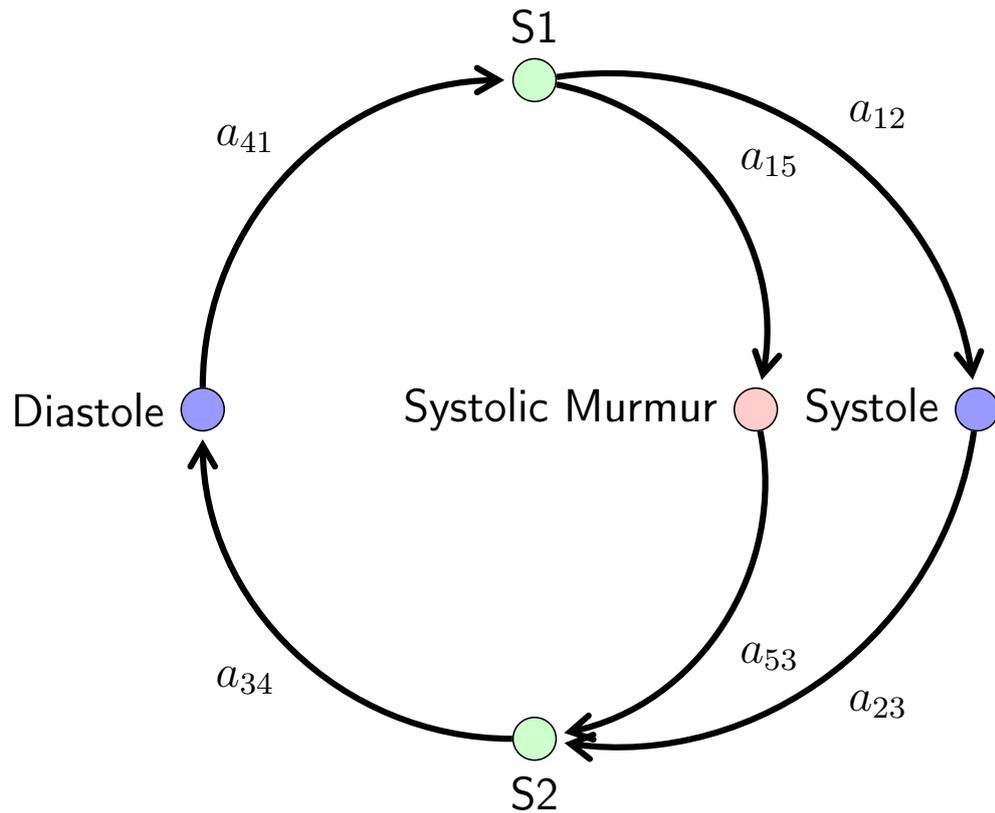


Figure 5.4: The proposed extension of the HSMM to include a systolic murmur state that exists in parallel to the normal systolic state.

Taking this conceptual viewpoint leads us to a natural extension of the model in which alternative, pathological states are also possible. For this research we have extended the model to recognise a systolic murmur state. This state exists in parallel with the state of systole as shown in Figure 5.4. The classifier in the first step is extended to score the probability of a sample being from a systolic murmur, or more correctly a systolic state that contains a systolic murmur, as defined in the training set, along with the other four states. This extension requires a more complicated classification scheme than logistic regression used in previous work [70], as well as features which accentuate the characteristics of systolic murmurs.

### 5.2.3 Qualities of systolic murmurs

The extension of the model to include a "systolic murmur" state calls for a discussion on the likely differences between a normal systole and one that contains a murmur. Murmurs are produced in the presence of disturbance of laminar blood flow. The classification "systolic murmur" refers to the murmurs that occur between the first heart sound (S1) and the second heart sound (S2). To improve the tractability of the model

the HMM is only allowed to transition to either the *systole* state or the *systolic murmur* state.

## 5.2.4 Emission Modelling

### 5.2.4.1 Feature Selection

The temporal information, the order and relative 'distance' between samples, is critical to this model. Feature selection for the emission probability modelling is approached as envelope extraction followed by downsampling. The most important criteria for selecting features is their ability to distinguish or discriminate between the different states. The first hurdle for the emission modelling is that the states of systole and diastole have very similar amplitude distributions, while S1 and S2 also exhibit similarities although to a lesser degree. The distributions of the homomorphic envelope samples for each of the states are shown in Figure 5.5. Systolic states that contain systolic murmurs exhibit a distribution more like the fundamental heart sounds than the periods of silence. This agrees with the intuition that there is more auditory energy in a systolic state with a murmur than without.

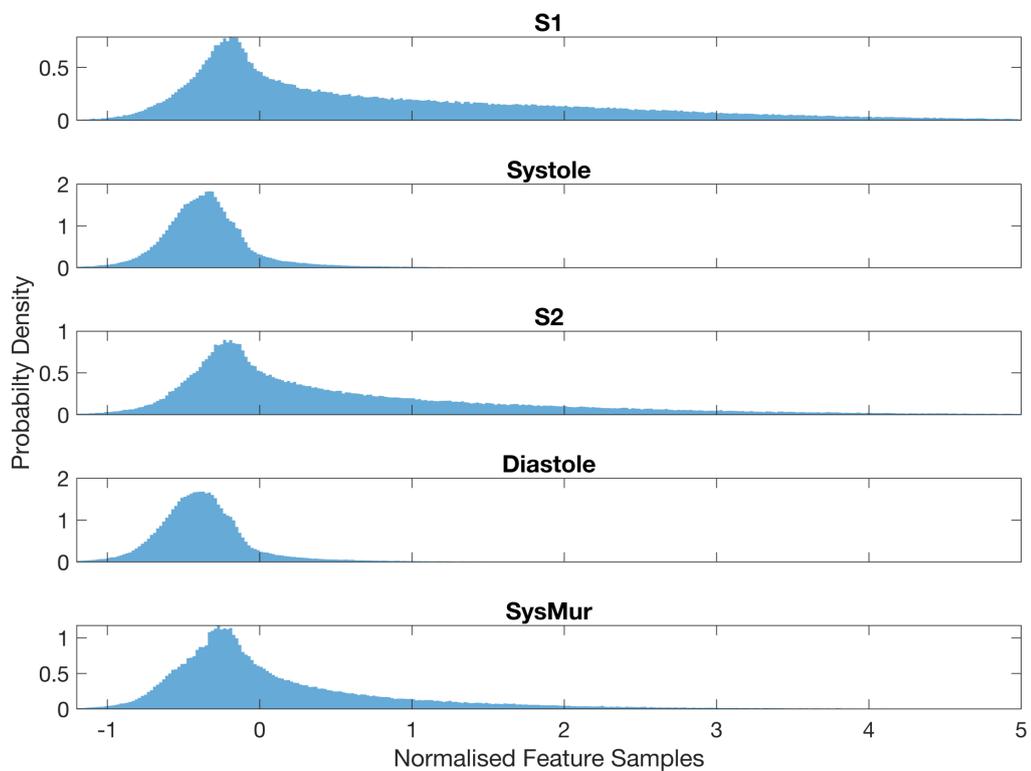


Figure 5.5: Distributions of training samples for each of the states of the model. Data taken from the training set described in section 5.2.1.1.

Many different features have been used for both heart sound classification (into disease categories or normal/abnormal) as well as heart sound segmentation (see introduction to this chapter). Since the model proposed in this work will perform both heart sound segmentation as well as detection of abnormal systolic states, features from both areas of research are of interest. To differentiate systolic murmurs in the presence of normal heart sounds and intermittent noise with overlapping frequency bandwidths presents several challenges. Firstly, as mentioned before, the amplitude distribution of murmurs is not wildly different from those of the fundamental heart sounds. Secondly, there is a wide variety of different murmurs that can occur during systole and it is difficult to find features that are suitable for all of these.

Time-frequency features, derived mainly by use of the wavelet transform [129]–[132] or the short-time Fourier transform [133]–[135], have been used in the literature to identify heart murmurs and more generally to classify heart diseases. In this work a selection of energy envelope, spectral, and statistical features were investigated. Several features are calculated using spectral information obtained by the continuous wavelet transform (CWT). Examples of the different feature envelopes calculated are shown in Figure 5.6 for the heart sound shown at the top of this figure. Shown in Table 4, a total of 21 features were extracted including 4 energy envelopes, 4 CWT derived features, 6 DWT multiresolution analysis features, 2 short time Fourier transform features, and 5 statistical features.

Table 4: The features used for emission probability modelling

Feature Class	Feature Names	Notes
<b>Energy Envelope</b>	Homomorphic Envelope Hilbert Envelope Shannon energy Shannon entropy	Normalised time dependent energy envelopes
<b>CWT</b>	S1 S2 Higher Frequency Spectral Flux	Continuous wavelet transform using the Morse wavelet
<b>DWT – MRA</b>	w1 w2 w3 w4 w5 s	Multiresolution analysis (MRA) constructed using the discrete wavelet transform with 'db6' as Mother wavelet.
<b>STFT</b>	LF HF	Power spectral density estimate calculated using Hamming window
<b>Block Based Statistics</b>	$\sigma$ $\sigma^2$ Min Max IQR	All statistics were calculated in blocks of 20 ms to maintain feature sampling rate (50 Hz)

#### 5.2.4.1.1 Feature Pre-processing

All audio recordings were high pass filtered with a 2nd order Butterworth IIR filter with a cutoff of 20 Hz. The filtering was performed forward and backwards to correct for any phase distortion. Artefacts were removed from the recordings using the artefact removal algorithm described in Chapter 4. The feature vectors were low pass filtered with a zero-phase anti-aliasing filter before being downsampled to 50 Hz (resulting in 50 sample values per feature per second) to increase the computational performance of the algorithm. Finally, feature envelopes were normalised to zero mean and unit variance by subtracting their mean and dividing by their standard deviation, formally

$$x_{normalised} = \frac{x_{feature} - \mu_{feature}}{\sigma_{feature}} \quad (5.7)$$

#### 5.2.4.1.2 Signal Envelopes

The homomorphic and Hilbert envelopes were calculated using the methodology discussed by Schmidt et al. [68] and in Chapter 3 of this thesis. The Shannon energy and Shannon entropy were calculated as

$$S_{Energy}(t) = -x_n(t)^2 \log(x_n(t))^2 \quad (5.8)$$

$$S_{Entropy}(t) = -|x_n(t)| \log|x_n(t)| \quad (5.9)$$

where  $x_n(t)$  is the original audio recording signal normalised to [-1,1] and  $\log(a)$  is the natural logarithm of  $a$ .

#### 5.2.4.1.3 CWT features

The CWT features were calculated with 4 voices per octave using the analytic Morse wavelet. The different CWT features correspond to 3 different frequency bands: CWT - S1 from 40 to 80 Hz, CWT - S2 from 90 to 130 Hz, and CWT - Higher Frequency from 150 to 500 Hz. Each feature envelope represents the standard deviation of the frequency bands at each time point, that is

$$x_{cwt}(t) = \sqrt{\sum_{k=k_{low}}^{k_{high}} (X(k, t) - \mu_t)^2} \quad (5.10)$$

where  $X(k, t)$  is the  $k_{th}$  frequency bin of the  $t_{th}$  sample of the CWT coefficients,  $k_{low}$  and  $k_{high}$  represent the frequency band edges, and  $\mu_t$  is the mean of the frequency bins at time  $t$ . The CWT - Spectral Flux ( $x_{SF}$ ) is a measure of the change in the frequency content over time [58] calculated from the CWT magnitude spectrum. Formally,

$$x_{SF}(t) = \sqrt{\sum_{k=0}^K (|X(k, t)| - |X(k, t-1)|)} \quad (5.11)$$

where  $K$  is the total number of frequency bins that the signal is decomposed into. The normalisation constants have been excluded for brevity in both equation (20) and (21) since all feature vectors are normalised to zero mean and unit variance and thus do not contribute any information.

#### 5.2.4.1.4 DWT Features

The DWT features were calculated using the 'db6' Daubechies mother wavelet and decomposition was performed to 5 levels. A maximum overlap discrete wavelet transform (MODWT) was performed and the results used to compute a multiresolution analysis (MRA) representation of the audio signal. The 'db6' wavelet was chosen because an orthogonal wavelet is required for MODWT and it has previously been used with success for heart sound analysis [129], in part perhaps due to the similarities in appearance between the wavelet and heart sounds. Importantly, especially for heart sound segmentation, the MRA acts as a zero-phase filter and the results are time-aligned with the original waveform. The interested reader is referred to Percival and Walden [136] for an in-depth introduction to the DWT and MRA with MODWT. The absolute values of all five of the resulting projections of the signal onto wavelet subspaces (MRA w1 - w5) and the scaling space (MRA - s) were included as features.

#### 5.2.4.1.5 STFT Features

The power spectral density (PSD) features were computed using an STFT with a Hamming window function of length of 20 ms and 50% overlap between windows. This windowing function, used to decrease edge effects, can be expressed as,

$$w(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N}\right), \quad 0 \leq n \leq N \quad (5.12)$$

where  $N = L - 1$  and  $L$  is the window length in samples. The PSD of input signal  $x$  is then be estimated as the square magnitude of the  $k_{th}$  frequency bin of the  $t_{th}$  block, formally,

$$PSD(k, t) = \left| \sum_{i=i_s(t)}^{i_e(t)} w(t)x(t)e^{-jk\omega(i-i_s(t))} \right|^2 \quad (5.13)$$

where the normalisation constant has been excluded for the sake of brevity and frequencies are assumed to be in radians per sample for the same reason. Each PSD feature represents the mean of the PSD between 40 - 80 Hz, for PSD - LF, and between 150 - 500 Hz, for PSD - HF. The PSD feature for time  $t$  and frequency range  $k_{low}$  to  $k_{high}$  calculated can be expressed as,

$$x_{PSD}(t) = \frac{1}{k_t} \sum_{k=k_{low}}^{k_{high}} PSD(k, t) \quad (5.14)$$

with  $k_t$  the total number of frequency bins in the range  $k_{low}$  to  $k_{high}$ .

#### 5.2.4.1.6 Block Based Statistics

Statistical features were calculated using a window length of 20 ms with no overlap between windows resulting in 50 sample values for each second, effectively downsampling to 50 Hz. For each of the frames the standard deviation ( $\sigma$ ), variance ( $\sigma^2$ ), minimum, maximum, and inter-quartile range (IQR) was calculated. The standard deviation for the  $t_{th}$  frame ranging from  $i_s(t)$  to  $i_e(t)$  can be expressed formally as,

$$x_{\sigma}(t) = \frac{1}{N-1} \sum_{n=i_s(t)}^{i_e(t)} (x(n) - \mu_x(t))^2 \quad (5.15)$$

where  $\mu_x(t)$  is the mean of the input signal  $x$  for the  $t_{th}$  frame. Each of the other statistical features were calculated in a similar fashion with the standard deviation replaced by the relevant measure.

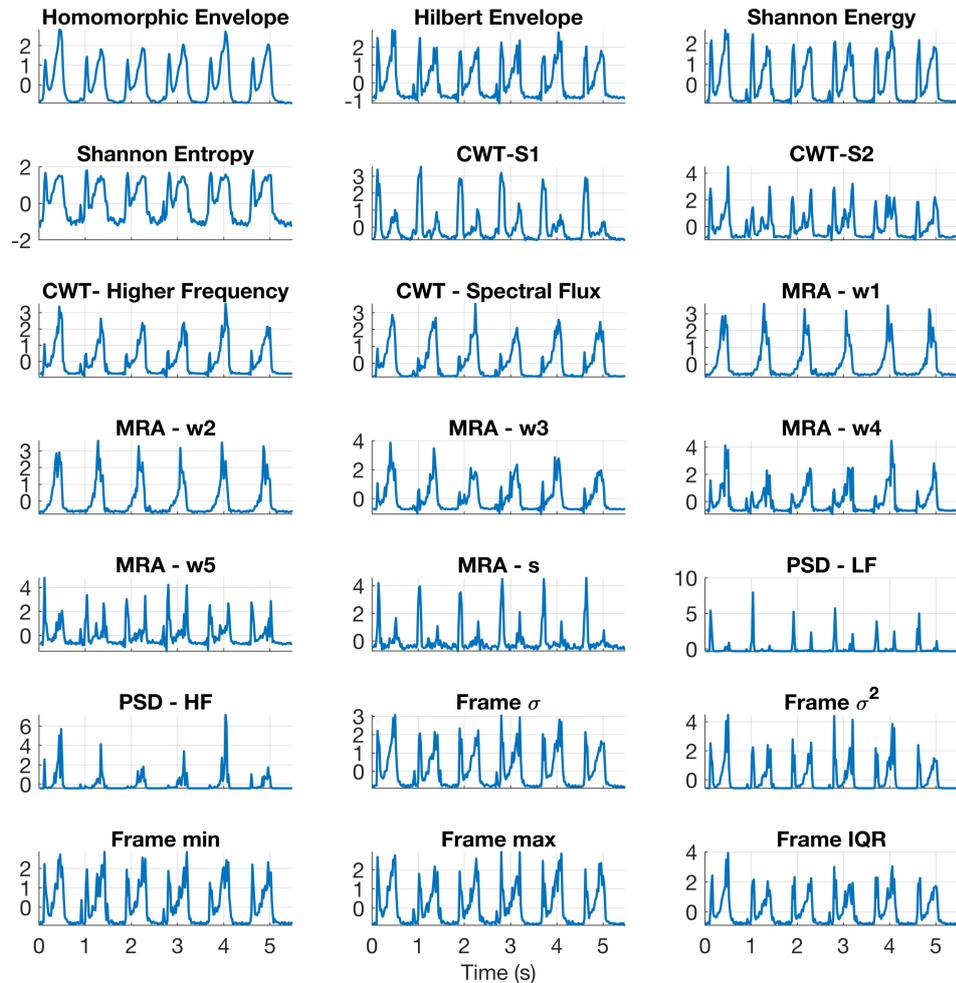
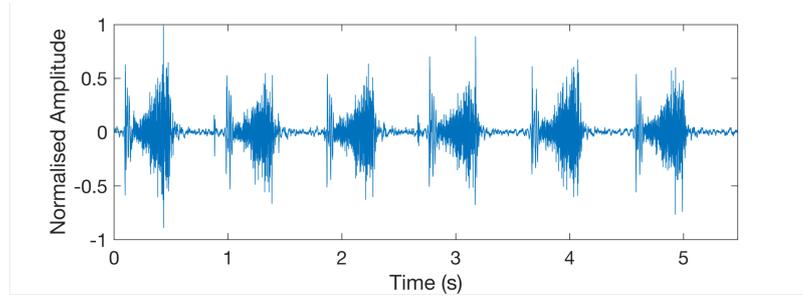


Figure 5.6: Feature vectors for the 21 features that were extracted to perform emission modelling. The original heart sound recording is shown at the top and all y-axis are in arbitrary normalised units. The actual input feature vector will have dimensions  $T \times 21$ , where  $T$  is the downsampled length of the audio recording.

#### 5.2.4.2 Emission Probability Models

In theory, any classification scheme that can provide a probabilistic score for each of the possible classes could be used in the model. The bulk of research in heart sound classification has focused on discriminating between pathological and normal heart sounds, the largest single sources of which are perhaps the 2016 Physionet Computing in Cardiology Challenge [137], [138] as well as the 2012 "Classifying Heart sounds

Challenge" [139]. The research generated by these challenges is of interest in this work since the methodologies developed by these researchers were scored on their ability to discriminate between healthy and unhealthy heart sound recordings. Thus, the features and classifiers that were best able to discriminate unhealthy heart sounds, provide an excellent starting point for the selection of an appropriate classifier for this work.

The classification method that achieved the most success in the 2016 Physionet challenge combined Convolutional Neural Networks (CNNs) and an ensemble of AdaBoost classifiers [140]. This method achieved the highest accuracy (86.02%) in the competition, with a high percentage of false positives (specificity = 77.8%, sensitivity = 94.24%). The methodology in second place used an ensemble of support vector machines (SVMs) [141], which achieved an accuracy of 85.9% with a much smaller amount of spread between specificity (84.9%) and sensitivity (86.91%). This indicates that although the SVM method was less successful in correctly identifying all the actual cases it had less false positive cases. Which method provides the "best" result depends on the specific use case. If the algorithm were to be used as a screening tool it is perhaps preferable to have a high amount of false positives compared to false negatives. Other classification methods that proved successful in this challenge included neural networks [142], k-Nearest Neighbour [143], and random forests [144].

In this work a random forest (RF) with 100 decision trees and random feature subsampling (feature bagging) is used to estimate the emission probabilities for the HMM. The RF methodology is considered since it has achieved success in heart sound analysis in previous work. For example, Nabhan Homsy et al. [144] report a sensitivity of 88.5% and specificity of 80.5% in classifying between unseen examples of abnormal and normal heart sounds. Decision trees, the building blocks of random forests have also shown promise in heart sound analysis; Pavlopoulos et al. [145] was able to correctly distinguish 45 out of 50 cases of mitral regurgitation and aortic stenosis using decision tree classifiers. Finally, the intuitively motivated random forest/decision tree classifier provides an interesting parallel to the more mathematically rigorous logistic regression classifier. The use of a random forest is compared with a logistic regression model similar to that used by Springer et al. [70] examined in the next section.

### 5.2.4.3 Logistic Regression

The output of a regression analysis using a logistic function (logistic regression) provides a natural probability score and as such is well suited for the emission probability modelling of the HMM. A one-versus-all approach was applied in which a binary classifier was trained with samples from one class as the positive class and samples from all the other 4 classes as the negative class. This resulted in a total of 5 binary logistic regression models. For this binary logistic regression problem, the posterior probability of the positive class,  $C_1$ , given the feature vector  $\phi$  can be written as

$$p(C_1|\phi) = \sigma(\mathbf{w} \cdot \phi) \quad (5.16)$$

where  $\mathbf{w}$  is the vector of model parameters and,

$$\sigma(\mathbf{w} \cdot \phi) = \frac{1}{1 + e^{(w_0 + \sum_{i=1}^M w_i \phi_i)}} \quad (5.17)$$

the *logistic sigmoid* function for an  $M$  dimensional feature space. From probability theory it follows that the probability for the negative class (or in this case that the sample was produced by a different state) is then simply  $p(C_2|\phi) = 1 - p(C_1|\phi)$ . The binary learners are organised using an error-correcting output code (ECOC) model [146] and the results are determined by aggregating the results from each of the binary classifiers. This approach, similar to the approach used by Springer et al. [70], was implemented to provide a baseline comparison while allowing for more predictor variables (a higher dimensional feature space) than the methodology employed by Springer et al.

### 5.2.4.4 Random Forest Classifier

To introduce the concept of a *random forest* (RF) classifier it is first necessary to provide a definition for the machine learning concept of a *decision tree* (DT), the primary building block of random forests. A decision tree (more specifically a *binary* decision tree) can be thought of as a sequence of binary decision points (e.g. yes/no questions, greater than/less than thresholds) that a piece of information (a sample) is presented to and that result in a final classification (or category). In the terminology of decision trees, a sample is passed from the *root* (at the start of the tree) through the decision nodes (the binary outcome questions) until it reaches a *leaf* node (the final category) that

corresponds to the finding of the DT. Decision trees provide a natural and intuitive way to codify a decision-making process. As they correspond to a series of binary decisions on the input variables, decision trees are easily interpretable by human observers. Unfortunately, decision trees also tend to overfit the data they are trained on and may provide suboptimal decision boundaries since the boundaries they create are always aligned to the feature axis [94].

Random forests attempt to improve the performance (and decrease the overfitting) of the decision tree algorithm by using an ensemble of decision trees fitted to different random subsets of the input variables (feature bagging) [147]. The final classification result is determined by the mode of the results from the decision trees and the probability of a class as the ratio of decision trees that "voted" for the class in question. This rather natural interpretation of the results as probabilities makes the random forest algorithm well suited for the modelling of emission probabilities.

#### 5.2.4.5 Emission probability performance metrics

The performance of the emission probability model was assessed using 10-fold cross validation. Samples from the training dataset were randomly partitioned into 10 training and validation sets and trained and tested in turn on each of these sets. The reported metrics for these models is then the performance of the model on the unseen samples. It is very likely that the training and validation set may contain samples from the same subject; this limitation was considered acceptable as the emission models are meant to determine the probability of a given sample coming from each of the heart sound states (*S1*, *Sys*, *S2*, *Dias*, *SysMur*) and not from a particular subject or disease state.

Since the classifiers are used for modelling the emission probabilities, the overall classification accuracy is not the primary measure of concern; it is more important for the classifier to give a high probability score to the correct class. For example, if the classifier correctly classifies a majority of 'S1' samples but assigns a very low probability for the 'S1' samples that it misclassifies, the overall performance of the algorithm will be lower than a classifier with similar accuracy but a greater amount of "certainty" in its scoring. This is referred to as the *classification margin*, the amount of difference between the classification score of the true class and the maximum classification score in the false classes. In this work the classification margin is calculated as

$$m = yf(x) \tag{5.18}$$

where  $x$  is an observation,  $y = \begin{cases} 1 & \text{if } x \text{ is correctly classified} \\ -1 & \text{if } x \text{ is incorrectly classified} \end{cases}$ , and  $f(x)$  is the score (probability in this work) assigned to the *correct* class. The classification margin ranges from -1 (observations that are assigned to an incorrect class with probability 1) to 1 (observations that are assigned to a correct class with probability 1), with a classification margin of 0 indicating that the classifier assigned equal probability to the correct class and an incorrect class.

### 5.2.5 Transition Probabilities

The extension of the HSMM to include a systolic murmur state requires a modification to the transition probabilities to allow entry into the new state. In this duration dependent version of the HMM the transition probabilities play the role of providing a structure of which transitions are possible and which are not. Since the systolic murmur state is essentially an alternative systole or stated otherwise it occurs in parallel to the normal systolic state, the same transition rules apply as to the "Systole" state. Formally, for state order  $S_q = \{S1, Systole, S2, Diastole, SysMur\}$  the transition matrix  $A$  is set to

$$A = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{5.19}$$

### 5.2.6 Duration Probabilities

In the duration dependent version of the HMM the state durations are made explicit parameters of the model. As shown in Figure 5.7 the durations are modelled using normal distributions. These probability distributions govern how much time is spent in each state. More specifically the duration probabilities limit the possible duration of states and make certain state durations more likely than others.

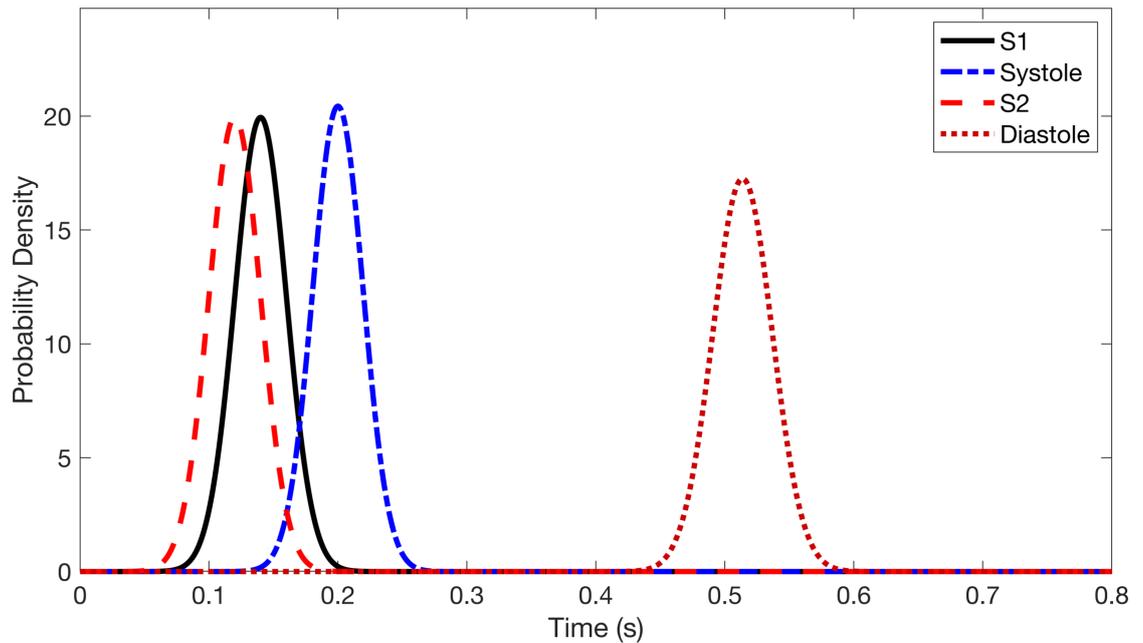


Figure 5.7: Example duration distributions for a heart rate of 65 bpm and an estimated systolic time interval of 235 ms.

#### 5.2.6.1 Fundamental Heart sound Duration Models

The possible duration of the fundamental heart sounds is not always obvious. Both the heart sounds are composed of multiple sounds that may be noticeably split under both normal and pathological situations. A split represents a noticeable pause between components of the heart sounds and in split cases the heart sounds would have longer duration than ones without splits. The splitting of the second heart sound occurs with inspiration in healthy subjects due to changes in pulmonary pressure, while fixed splitting (occurring during both inspiration and expiration) is possible in certain disease states, for instance atrial septal defect [112]. The first heart sound is split under normal conditions and may present with even wider splits in certain disease conditions, or may present without any noticeable split under other conditions [113].

Exact modelling of the fundamental heart sound durations is problematic as well as not being essential to the operation of the model. The durations of the fundamental heart sounds are approximated using a normal distribution based on mean and standard deviations of heart sound durations reported in literature [68], [70]. In this work the durations from the dataset are used to update the means reported in literature. The means and standard deviations used in previous work (shown in Table 5) were used as the prior distribution and updated using state duration data from the dataset described in this chapter using Bayes theorem. Formally,

$$p(\text{state durations}|\text{data}) = \frac{p(\text{data}|\text{state durations})p(\text{state durations})}{p(\text{data})} \quad (5.20)$$

The posterior distribution  $p(\text{state durations}|\text{data})$ , which is the state durations model after the new observations, was estimated using variational inference [94] to infer the new means and standard deviations of a normal distribution. The resulting posteriors, and thus the fundamental heart sound duration distributions used in the model, are shown in Figure 5.8, along with the prior distributions and histograms of the durations measured from the dataset.

Table 5: Fundamental heart sound duration distribution update

	Prior (s)	Likelihood (s)	Posterior (s)
S1 duration	$\mu = 0.122$ $\sigma = 0.022$	$\mu = 0.134$ $\sigma = 0.011$	$\mu = 0.134$ $\sigma = 0.012$
S2 duration	$\mu = 0.094$ $\sigma = 0.022$	$\mu = 0.113$ $\sigma = 0.010$	$\mu = 0.113$ $\sigma = 0.011$

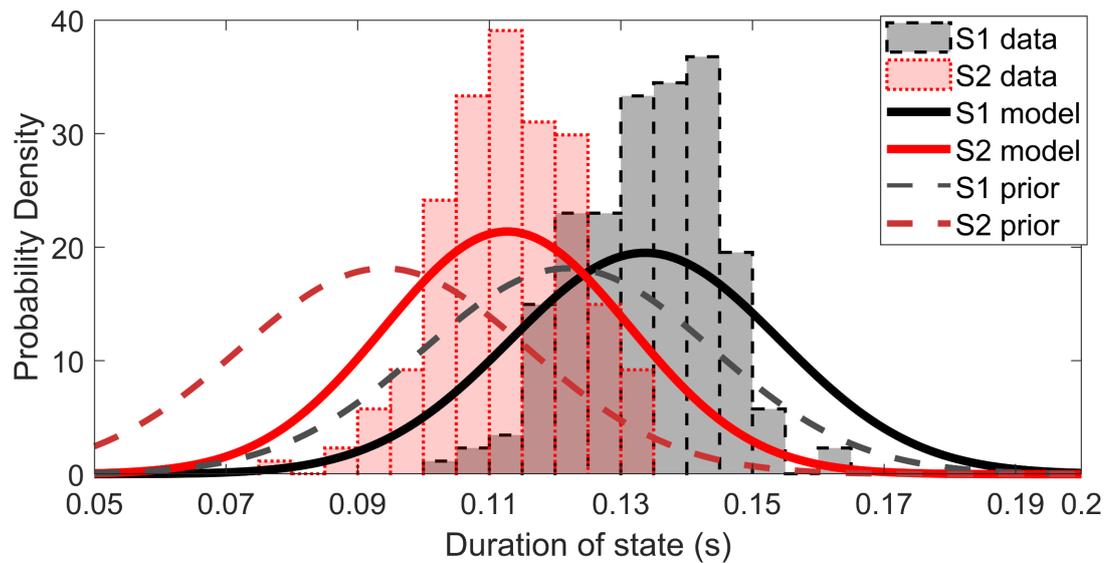


Figure 5.8: The posterior, prior, and likelihood distributions of the first and second heart sound durations as modelled in this work.

### 5.2.6.2 Systolic and Diastolic Duration Models

Systole and diastole represent the periods in the heart cycle when the ventricles of the heart are, respectively, contracting and relaxing. Thus, these are also the periods in which blood flows through the heart valves, either out of or into the ventricles. As can clearly be seen in Figure 5.9 the distribution of systole duration is much more peaked and lower than that of diastole. The mean durations are also plotted against heartrate showing a clear negative correlation between diastolic duration and heartrate

(correlation coefficient = -0.79) and a weaker negative correlation between systolic duration and heartrate (correlation coefficient = -0.55).

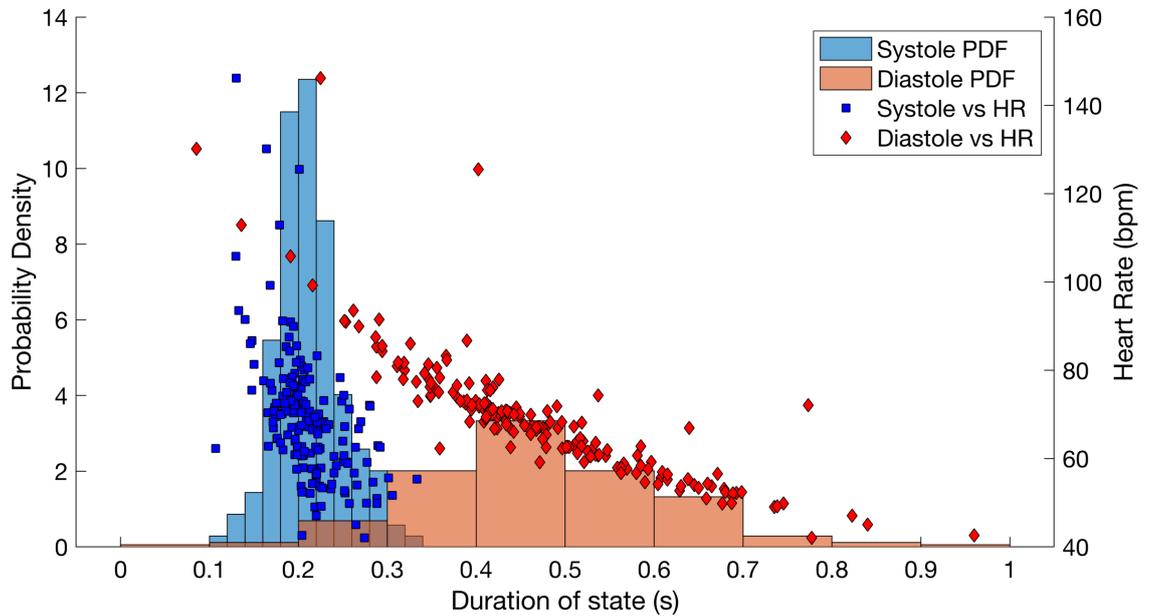


Figure 5.9: Histograms of the mean systolic and diastolic state durations from 118 heart sound recordings (see text for further details on data), along with scatter plots of the state durations and the estimated heart rate of the recordings.

In this work the diastolic state duration is modelled using a normal distribution with a mean estimated as a power function of heart rate and a standard deviation modelled using a log-logistic distribution. The two components of the diastolic duration model are shown in Figure 5.10. The parameters of the power function and log-logistic distribution were estimated from the durations measured in the dataset using maximum likelihood and are likely to overfit the data available in the dataset. However, the impact of the error introduced by this overfitting is expected to be minimal if the dataset is in fact a representative sample of the "true" distribution of diastolic durations, which seems likely (see section 5.2.1 for further discussion).

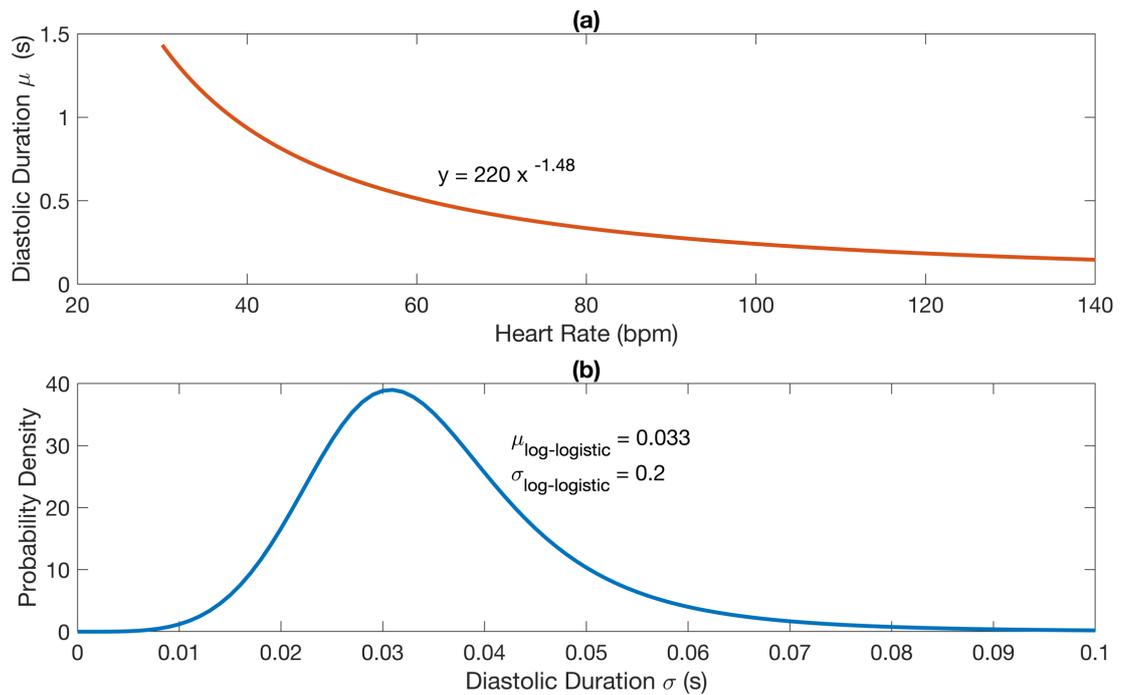


Figure 5.10: Illustrations of the models used to determine (a) the mean, and (b) the standard deviation of the diastolic duration.

For modelling of the systolic duration the method employed by Schmidt et al. [68], which estimates heartrate and systolic time interval by finding peaks in specific time windows of the autocorrelation of the audio recording, was adapted. While effective, this method relies on the heart sounds being the most prominent features of the recording (S1 for heartrate, and S2 for systolic time interval), which is not always the case in recordings with loud murmurs. When S2 is very soft or not present in the heart sound the algorithm is not able to accurately estimate the duration of the systolic state. The algorithm also struggles when a recording contains ectopic beats (beats that occur out of rhythm) which may occur in otherwise normal hearts.

In this work the heart rate estimation is performed with narrowband signal filtered to between 25 and 125 Hz to limit the amount of non-FHS sounds in the autocorrelation of the signal, a peak finding algorithm ("findpeaks" from the MATLAB Signal Processing Toolbox) is used to find the most likely peak that represents S1 in the autocorrelation of the signal. The heart rate estimation is also limited to between 40 and 140 bpm to increase the robustness of the algorithm. The systolic time interval To limit the impact of the previously mentioned challenges to systolic time estimation, the window of time in which the peak that indicates the systolic interval duration is expected to occur was

windowed using a normal distribution with a mean of 209 ms and a standard deviation of 70 ms.

$$x_{SI}(t) = \mathcal{N}(x(t)|\mu = 0.209, \sigma^2 = 0.0049) \quad (5.21)$$

where  $x(t)$  is the autocorrelation of the audio recording and the parameters of the normal distribution  $\mathcal{N}$  are given in seconds. This modification raises the probability of the peak occurring near the expected mean while still allowing prominent peaks further away from the expected mean to impact the estimation of the systolic time interval. The mean and standard deviation of  $\mathcal{N}$  were estimated from the training data. Similar to the diastole duration model the standard deviation of the systolic duration was modelled using a log-logistic distribution with parameters estimated from the dataset. The window is shown in Figure 5.11 (a) along with the log-log distribution used to model the standard deviation of the systolic time interval.

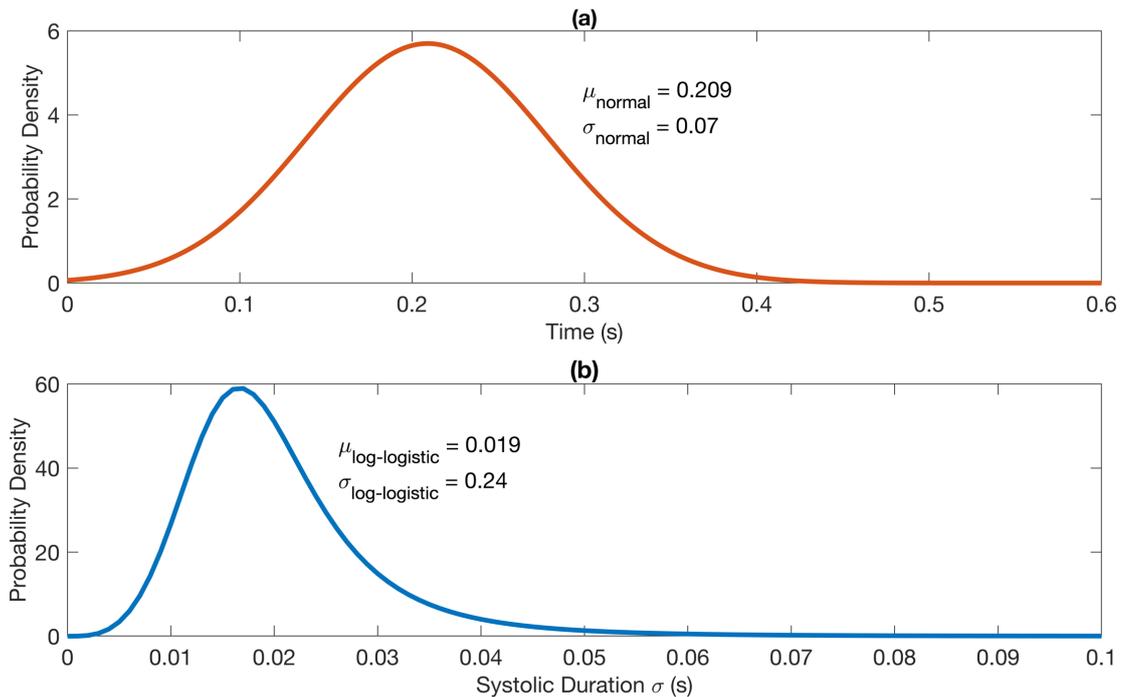


Figure 5.11: Probability density functions for the systolic duration model. The PDF shown in (a) is used as a windowing function to modify the signal autocorrelation when looking for a peak that represents the systolic time interval. The standard deviation of the systolic duration is modelled by the log-logistic PDF shown in (b) with maximum likelihood parameters estimated from the dataset.

### 5.2.7 Initial probabilities

Based on the estimates of the duration probabilities the initial probability vector  $\boldsymbol{\pi}$ , is set to

$$\boldsymbol{\pi} = \left[ \frac{\mu_{S1}}{HC_{dur}} \quad \frac{\mu_{sys}}{2HC_{dur}} \quad \frac{\mu_{S2}}{HC_{dur}} \quad \frac{\mu_{dias}}{HC_{dur}} \quad \frac{\mu_{sys}}{2HC_{dur}} \right] \quad (5.22)$$

Where  $\mu$  represents the mean of the duration distribution for a given state and  $HC_{dur}$  is the expected duration of the average heart cycle in the recording at hand. The probability of the model starting in either the systolic or systolic murmur state is considered together and set to be equally likely given the expected duration of systole.

### 5.2.8 Overall performance of the model

To determine the overall performance of the model as a heart sound labeller in a way comparable with previous work the model was tested in different ways. Firstly, the ability of the model to accurately identify the fundamental heart sounds to within a tolerance was tested; this can be called the *segmentation* ability of the model. Secondly, the model was tested for its ability to correctly identify systolic murmur states; this is referred to as the *classification* ability of the model. Lastly, the model is assessed in its ability to provide informative and accurate "labels" to different parts of the heart sound recording. Based on the results of the emission modelling cross validation cross validation a random forest classifier was used for the rest of the testing.

#### 5.2.8.1 Heart sound Segmentation

To determine the heart sound segmentation ability of the model we first determine its ability to accurately find S1 and S2 individually; this is done by comparing the model's results to the "true" label sequence as derived from ECG and manual annotations. The following procedure is explained for the first heart sound (S1), but the same steps were followed for the second heart sound (S2).

The segmentation ability of the proposed model was tested by specifying a  $\delta$  ms window around the midpoint of each of the "true" labelled S1. We denote this window as  $\Delta_i$  where  $i$  indicates the  $i$ th sequence labelled as S1 in the "true" label sequence.

$$\Delta_i = \left[ \phi_i \pm \frac{\delta}{2} \right] \quad (5.23)$$

where  $\phi_i$  is the midpoint of the  $i$ th sequence labelled as S1 rounded up to the nearest sample. For each of the given S1 sequences in a recording (each of the series of samples marked as S1) we postulate three possible outcomes for the model: either the model has correctly identified the S1 region (True Positive), the model has identified a S1 but not within the bounds of the true label (False Positive), or the model has not identified a S1 region at all (False Negative). Note that we do not define the notion of "True Negative" for this task. We use the notation  $\eta_i$  to indicate the outcome for the  $i$ th S1 in the *model* label sequence. Next, we denote each of the individual sequences of samples that the model has labelled as S1 with  $\lambda_i$  where  $i$  indicates the  $i$ th S1 in the *model* label sequence and we use  $\lambda_i$  to indicate the midpoint of the  $i$ th S1 sequence. The heart sound states identified by the model are then marked as true positive (TP) if  $\lambda_i$  falls within the window  $\Delta_i$ . If  $\lambda_i$  is not within in the bounds of  $\Delta_i$  then it is marked as a false positive (FP). Using the defined notation, we can write,

$$\eta_i = \begin{cases} TP & \text{if } \lambda_i \in \Delta_i \\ FP & \text{otherwise} \end{cases} \quad (5.24)$$

Since false negatives are S1 sequences not detected by the model we make the assumption that if the true label sequence contains more S1 sequences than the model's sequence the difference is the number of false negatives.

$$FN_{total} = \begin{cases} N - M & \text{if } N > M \\ 0 & \text{otherwise} \end{cases} \quad (5.25)$$

where we use  $FN_{total}$  to indicate the total number of false negatives in a recording. Multiple values of the tolerance parameter  $\delta$  are reported to show the segmentation ability of the model at difference tolerances. An example of how the difference values of  $\delta$  look relative to an actual heart sound is shown in Figure 5.12.

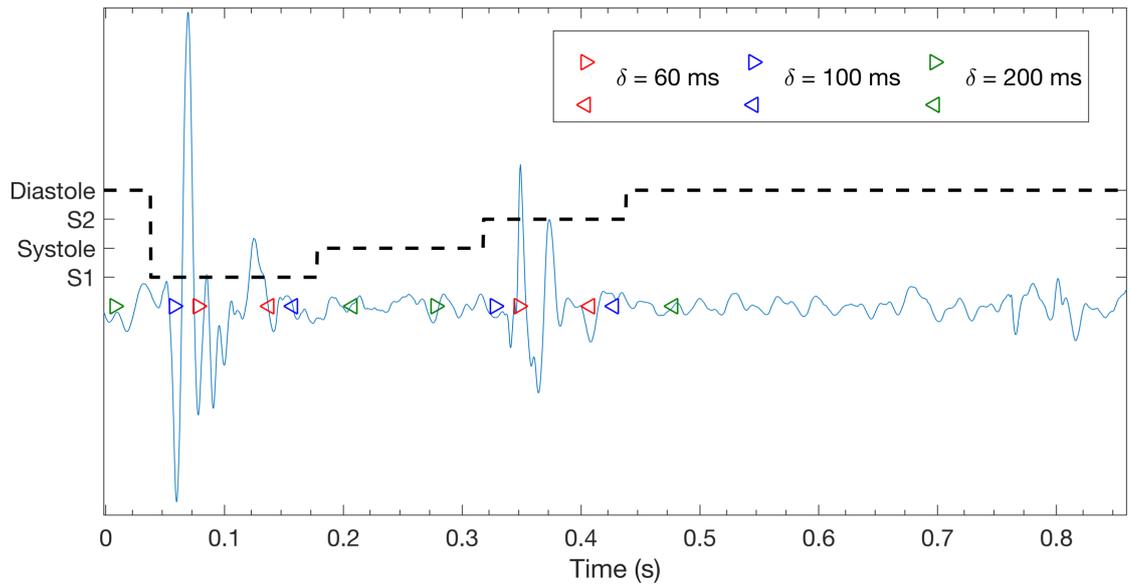


Figure 5.12: Illustration of the different segmentation tolerances. A fundamental heart sound was determined to be correctly identified if the middle of the region marked as S1/S2 fell inside the region indicated by the markers for each of the values of  $\delta$ .

The segmentation problem is similar to an information retrieval problem (e.g. document retrieval) in which it is important to retrieve all relevant events (*recall*) but also important that the retrieved events are actually relevant (*precision*). Using confusion matrix metrics, the recall is calculated as

$$Recall \text{ (Sensitivity)} = \frac{TP}{TP + FN} \quad (5.26)$$

and the precision is calculated as

$$Precision \text{ (Positive Predictive Value)} = \frac{TP}{TP + FP} \quad (5.27)$$

In simpler terms, the question "does the model identify all the heart sounds in the recording?" is addressed by the reported recall and the question "are all the sounds the model has identified in the recording actually correctly identified fundamental heart sounds?" is answered by the precision metric. These two metrics are combined in the  $F_1$  score, calculated by taking the harmonic mean of the recall and precision, which provides a sense of the overall performance of the model and provides a metric that is comparable to previous work. Since the precision and recall both represent ratios, the harmonic mean provides an intuitive average of these two values. *Accuracy* does not

provide a suitable measure for unbalanced datasets [148], and is not reported in this work.

### 5.2.8.2 Systolic Murmur Classification

The classification ability of the model was tested as a binary classification problem. Each of the "Systolic Murmur" and "Systole" states identified by the model were compared to the annotated state at that time and scored as a true positive (TP), false positive (FP), true negative (TN), or false negative (FN). To determine the ability of the model to distinguish between normal systolic states and those with murmurs, the model was tested at different settings of *prior class probability*, or equivalently, the *misclassification cost* for the "Systole" and "Systolic Murmur" states. Thus, the levels of *sensitivity* and *specificity* of the murmur can be tuned to an acceptable level of false positives/false negatives. The *informedness* (Youden's J for the binary case [149]) of a classifier can be calculated as

$$I = TPR + TNR - 1 \quad (5.28)$$

where TPR and TNR respectively are the true positive rate and true negative rates of the classifier. The informedness of a classifier quantifies how much knowledge a classifier demonstrates about the underlying true state when specifying a sample as either positive or negative. An informedness of +1 indicates that the classifier correctly identifies each positive state as a positive and negative state as a negative while an informedness of 0 indicates that the classifier does not perform better than randomly assigning samples to classes. An informedness of -1 indicates a completely "perverse" classifier that classifies all positives as negatives and vice versa, of course in this case by inverting the output we are left with perfect classification. The MCC is a measure of classification quality that is suitable for classification tasks on unbalanced datasets [150]. As a confusion matrix metric, it is defined as,

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5.29)$$

## 5.3 Results

### 5.3.1 Emission modelling

The distributions of the classification margins for the random forest and ECOC logistic regression classifiers are shown in Figure 5.13 and the means and 95% confidence intervals over 10 folds are reported in Table 6. The mean classification error (also referred to as loss) over the 10-fold cross validation is a measure of the proportion of samples that are misclassified. The *classification edge* represents the mean of the classification margins; higher values indicate that the model assigns higher probabilities to the correct class.

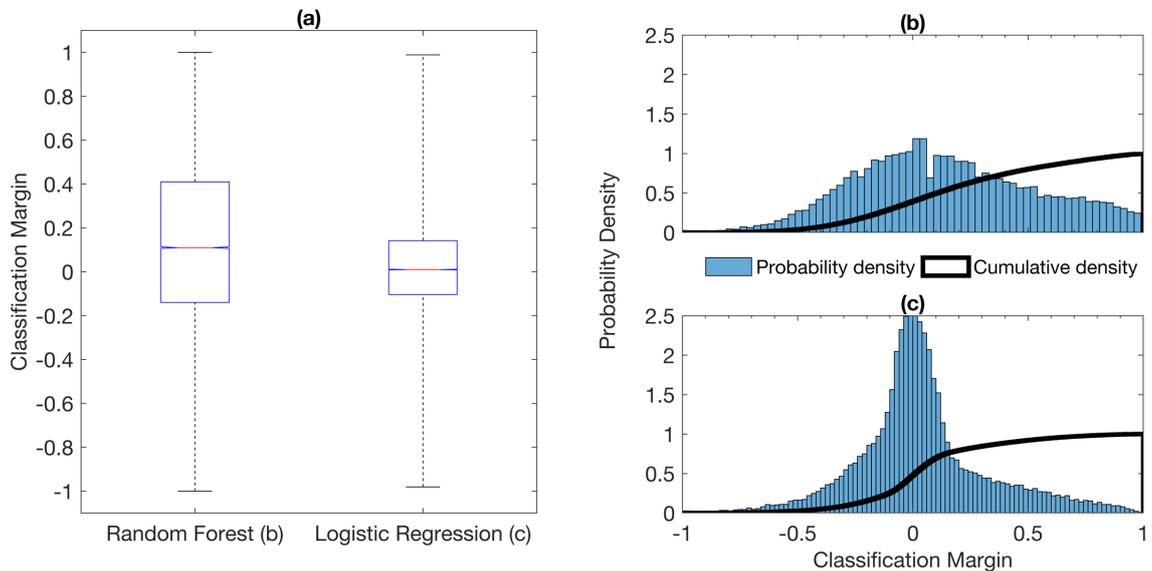


Figure 5.13: Classification margin distributions from the classification results of the 10-fold cross validation for the random forest and ECOC logistic regression classifiers. The boxplots in (a) show the median (red line in centre) and 25th/75th percentiles (boxes). Figures (b) and (c) show the probability and cumulative distribution of the random forest and logistic regression respectively.

Table 6: Average performance of the classifiers used for emission probability modelling

Method	Mean Classification Error (% Loss)	Mean Classification Margin (Edge)
Random Forest	$39.64 \pm 0.350$	$0.1357 \pm 0.002$
ECOC Logistic Regression	$46.87 \pm 0.297$	$0.031 \pm 0.001$

### 5.3.2 Heart sound Segmentation

To account for the random elements in the proposed model Table 7, 8, and 9 show the mean and 95% confidence intervals for 10 iterations of the model. Table 7 shows the

segmentation results on the entire testing dataset described in section 5.2.1.2 given different tolerance values.

Table 7: Performance metrics for heart sound segmentation for different tolerances

Tolerance( $\delta$ )	Recall (%)	Precision (%)	F1-score (%)
<b>60 ms</b>	99.2 $\pm$ 0.1	80.9 $\pm$ 0.5	89.1 $\pm$ 0.3
<b>100 ms</b>	99.3 $\pm$ 0.1	90.2 $\pm$ 0.3	93.9 $\pm$ 0.3
<b>200 ms</b>	99.2 $\pm$ 0.1	97.1 $\pm$ 0.3	98.2 $\pm$ 0.2

We compare the results of the proposed labelling algorithm on the testing dataset with the results of the publicly available Springer heart sound segmentation algorithm [70] for a tolerance level ( $\delta$ ) of 100 ms. This comparison is shown in Table 8. The Springer algorithm does not contain any random elements and so no variance between iterations is expected.

Table 8: Comparison of heart sound segmentation results on the complete testing set

Algorithm	Recall (%)	Precision (%)	F1-score (%)
<b>Springer HSS (LR-HSMM)</b>	97.3	87.2	90.6
<b>Probabilistic Heart Sound Labelling</b>	99.3 $\pm$ 0.4	90.2 $\pm$ 0.3	93.9 $\pm$ 0.3

The addition of the systolic murmur state is expected to improve segmentation performance in recordings that contain systolic murmurs. In Table 9 we report the results of the two algorithms for only the recordings in the testing dataset that contained systolic murmurs (a subset of the testing dataset containing 23 recordings).

Table 9: Segmentation results for heart sounds containing *systolic murmurs*

Algorithm	Recall (%)	Precision (%)	F1-score (%)
<b>Springer HSS (LR-HSMM)</b>	93.4	68.7	77.1
<b>Probabilistic Heart Sound Labelling</b>	98.8 $\pm$ 0.9	86.2 $\pm$ 0.7	91.4 $\pm$ 0.8

### 5.3.3 Systolic Murmur Classification

The resulting ROC curve is shown in Figure 5.14, along with two operating points and their corresponding confusion matrices. The first operating point, indicated by the red circle (a), is the point at which the *informedness* of the classifier is maximised. The second operating point, indicated by the green asterisk (b), is the point at which the Matthew's correlation coefficient (MCC) is maximal (in this case point (b) also corresponds to the maximum value of the accuracy and the F1-score).

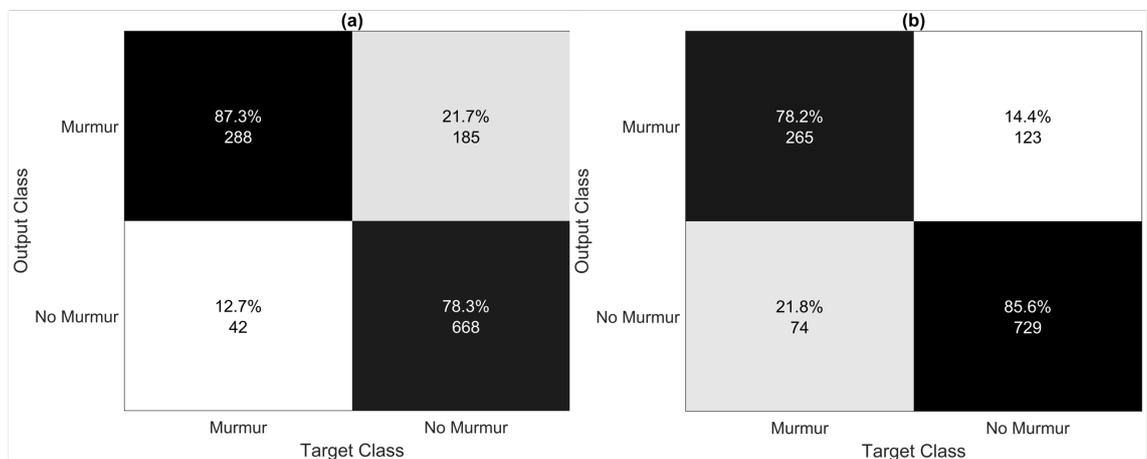
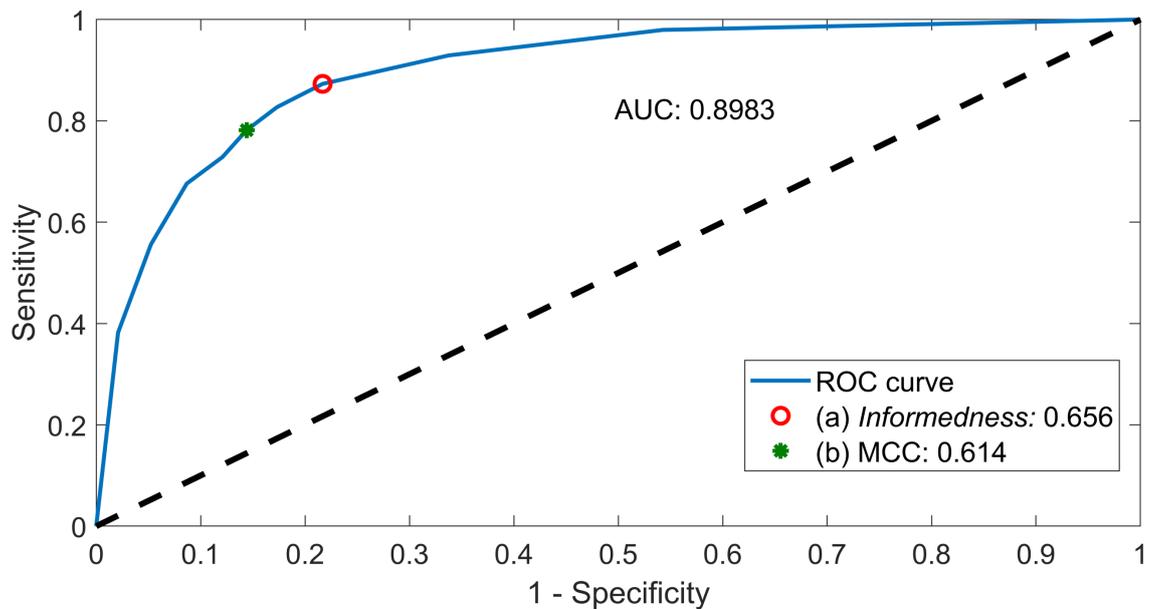


Figure 5.14: Receiver-Operator Characteristic (ROC) curve for systolic murmur classification. The marked operating points correspond to (a) maximum *Informedness*, the point at which the tradeoff between sensitivity and specificity is lowest, and (b) maximum MCC, a confusion matrix metric used to give a balanced indication of overall classification quality (see text for more details).

### 5.3.4 Probabilistic Heart Sound Labelling

In the previous sections the performance of the different aspects of the model were assessed individually and while this provides interesting and important information about these functionalities of the model, in reality, the model does not perform these functions in isolation and in fact the proposed model is not a classifier attempting to find the "correct category" for a sample. The systolic murmur state acts as another possible state the duration dependent HMM has access to when discerning the most probable sequence of heart states that produced the observable sequence.

The model not only provides us with systolic murmur label but also (more fundamentally in fact) provides a score for how probable the murmur is given the model parameters. Since this score is computed in a sample by sample manner through the emission model, we have opted to report the expected value (mean) of the individual samples of each murmur state across the duration of that state to represent the probability that the model has assigned to that murmur. The performance metrics in the previous sections provide interesting quantitative measures of the proposed mode. To give an idea of the labelling function of the model this section presents specific examples (the heart sounds from which the heart cycles used in Chapter 4 were taken) and allows the reader to get a qualitative sense of the performance of the proposed model.

Figure 5.15 shows the output of the proposed heart sound labelling algorithm on a normal, healthy heart sound recorded at the pulmonic area. The red line represents the state labels (as indicated on the right-hand axis). The numbers in black above the systolic state labels shows the average probability assigned to the 'SysMur' state by the random forest emission model. In this example the model has successfully segmented the heart sounds and correctly identifies each of the systolic states as not containing any murmurs. The second systolic state has been assigned a probability of 0.37 of containing a murmur, the highest of any of the states. This is likely due to a higher amount of noise in this systolic period; this also seems to be the case for the systolic period that occurs at around 5 seconds with an average murmur probability of 0.34.

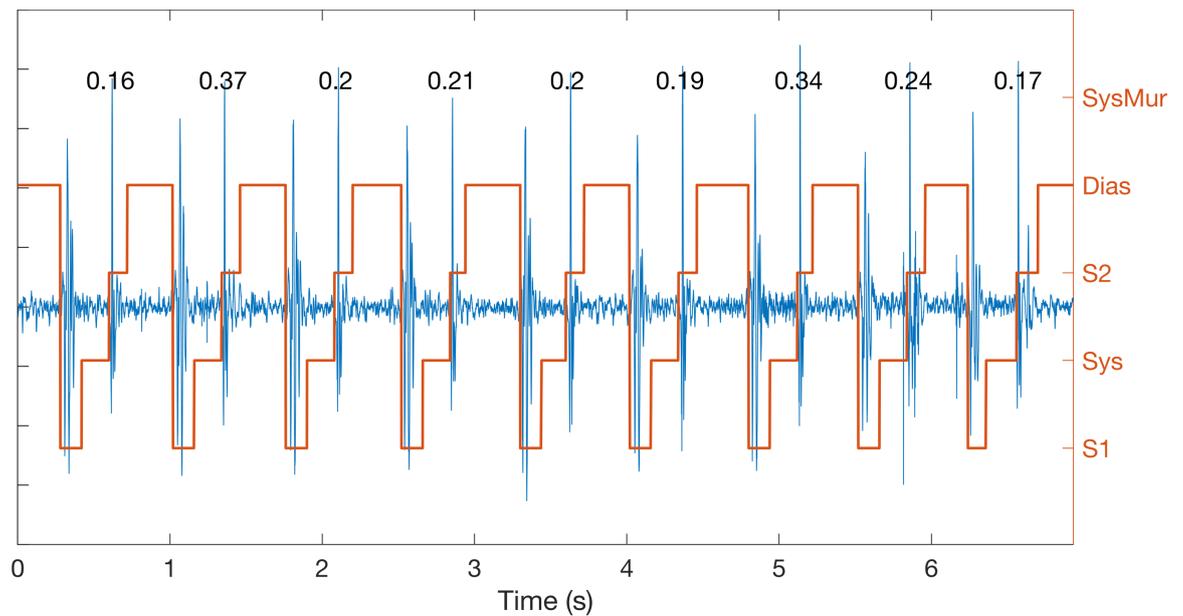


Figure 5.15: Heart sound (a). A normal, healthy heart sound recorded at the pulmonic area. This heart sound has been successfully labelled and relatively low representative probabilities assigned to the presence of systolic murmurs.

The next example, shown in Figure 5.16, presents more of a challenge to the algorithm and, consequently, interesting results. This example contains significant noise artefacts and an undiagnosed, asymptomatic, systolic murmur recorded at the aortic area. The algorithm incorrectly labelled the first heart cycle and the first region labelled *SysMur* is in fact a first heart sound. Besides this error the rest of the recording is segmented correctly. Two anomalies occurred with the labelling of the systolic murmur state: the first, at 4 seconds, had an average murmur probability of 0.64 but was not assigned to the *SysMur* state by the Viterbi algorithm, the second, just before 6 seconds had a lower average of 0.43 but was assigned to the *SysMur* state. These anomalies most likely arise because the product, and not the mean, of the emission probabilities are used by the model to determine the state probabilities. This suggests that perhaps the geometric mean of the murmur state probabilities would give a more intuitive summary of the probability assigned to a given murmur by the model.

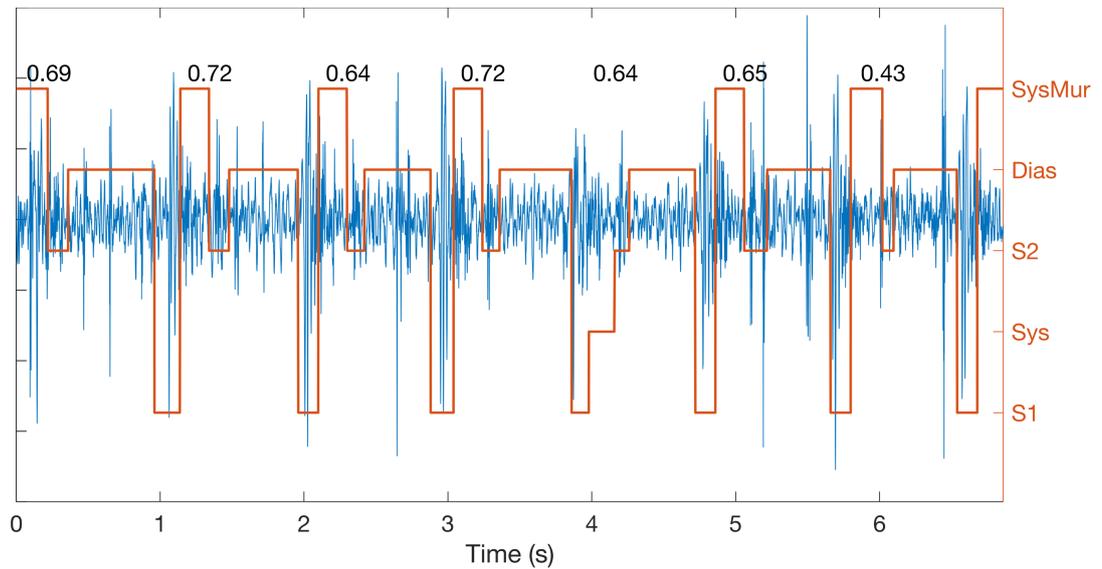


Figure 5.16: The state labels of heart sound (b) shows some interesting anomalies. Although the probability score of the murmur at around 4 seconds (0.64) is relatively high the decoding of the state sequence has labelled it as a normal systolic state, while murmur at around 6 s with a lower score of 0.43 has correctly been labelled as a systolic murmur state.

The third example heart sound recording is shown in Figure 5.17. This recording was made at the apex and contains a prominent mitral regurgitation murmur due to mitral valve prolapse. This recording contains very low level of noise relative to the heart sounds and was labelled without any significant issues. All the systolic states were labelled as *SysMur* with high mean emission probabilities ( $>0.8$ ) for all six of the systolic murmurs.

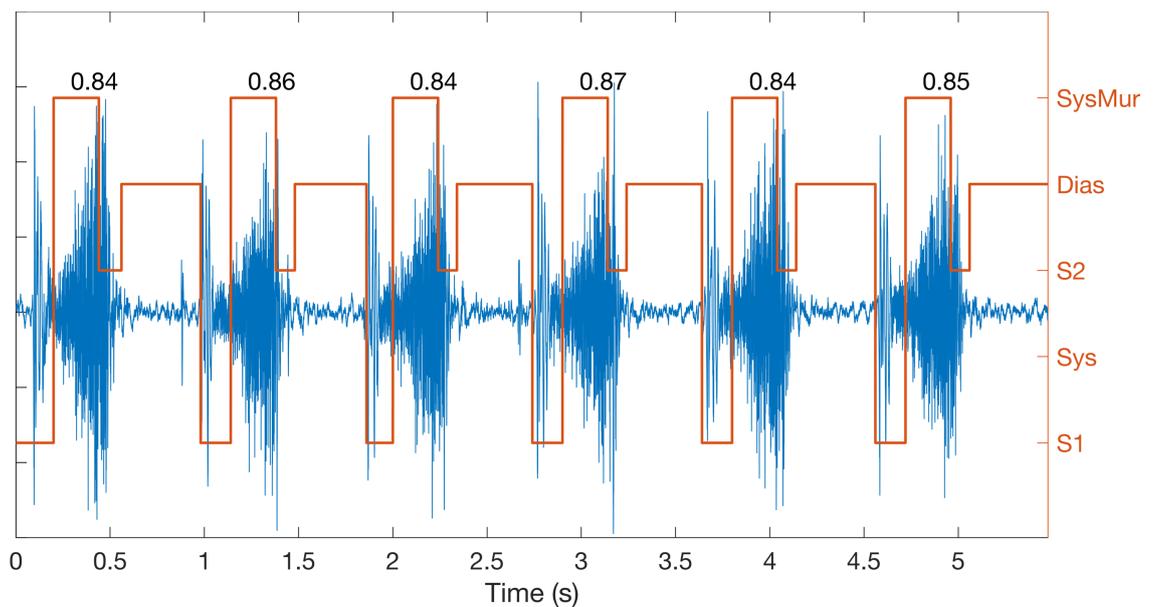


Figure 5.17: Heart sound (c) with a prominent systolic murmur and a very high SNR that has been correctly labelled by the proposed model. The algorithm is quite certain of the systolic murmur in this recording, as shown by the high probability score for the murmur states.

The final example, heart sound (d) from the dataset used in Chapter 4, is shown in Figure 5.18. This recording presents another challenging and interesting case for the proposed algorithm. The recording, made at the pulmonic area, presents with a fourth heart sound (S4), a prominent early diastolic murmur, a slight murmur at the start of systole, and slightly ectopic (out of rhythm) heart cycles. The heart sound segmentation of this recording has largely been successful despite the erratic heart rhythm in the first 3 seconds of the recording, although the third and the final S2 sound labels are slightly delayed. The means of the emission probabilities are relatively low, an interesting result given the soft and short nature of the murmur. The segmentation error of the third S2 may have included some of the diastolic murmur into the analysis of the systolic state, explaining the high probability given to the third systolic murmur. The relatively low score given to the second systolic murmur is more puzzling; one possibility is that slightly more of systole was included at the end of S1 and thus less of the short, early systolic murmur was included in the analysis of the murmur probability.

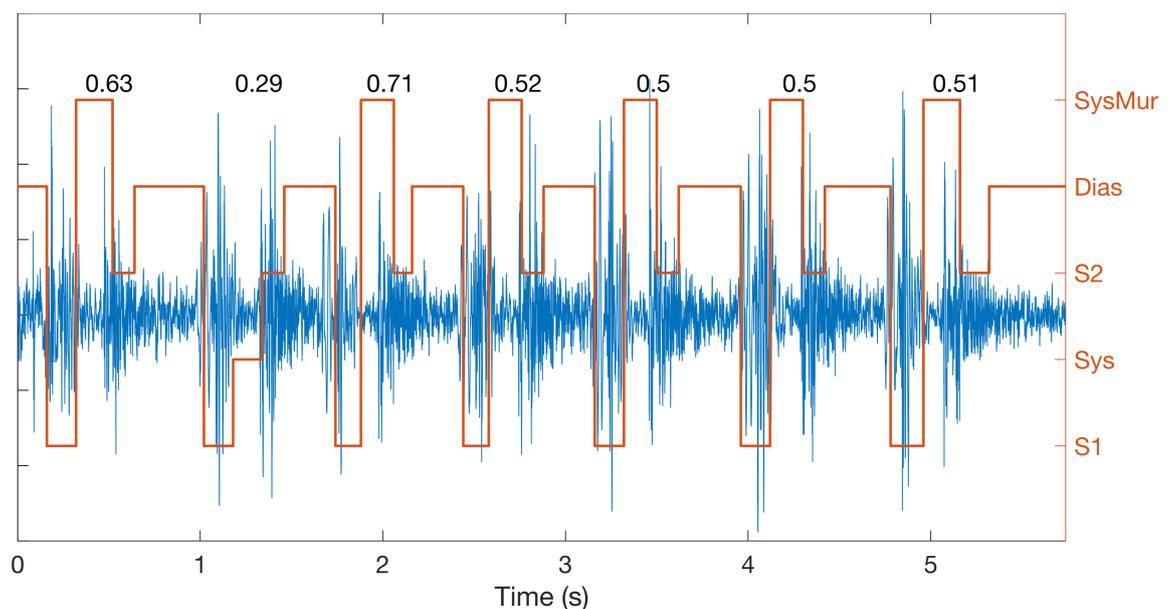


Figure 5.18: Heart sound (d) presents another interesting case. As discussed during the ICA of this example, while this heart sound has a prominent diastolic murmur, the systolic murmur is subtle and occurs directly after the first heart sound. The model has correctly labelled the majority of the systolic states as containing a murmur, albeit with relatively low probability scores.

## 5.4 Discussion

In its current guise the model described and constructed in this chapter can be conceptualised as a two-stage classifier. In the first stage the samples are treated as

independently and identically distributed (I.I.D.) data and scored by the emission probability model. In the second stage the temporal characteristics of the signal, modelled by the duration, transition, and initial probabilities of the HSMM are used in conjunction with the emission probabilities to find the most likely sequence of states (that the heart occupied) that led to the observations (the audio recording).

The derivation of the emission probabilities was tested using two contrasting methodologies: the non-parametric random forest with more intuitive than mathematical underpinnings and a parametric logistic regression model. In this case, the random forest (RF) classifier outperformed the ECOC logistic regression (LR) classifier with a lower cross-validated classification error and, more importantly, a higher classification edge (mean classification margin). The classification margin (illustrated in Figure 5.13) indicates that the random forest, on average, gives a 0.1357 higher probability score for the correct class, while LR averages 0.031. The mean classification loss over ten folds for RF was 39.64% and for LR was 46.87% both lower than what would be expected from a random guess (80% for a 5 class classifier), but, as can be seen when viewing the cumulative distribution of the classification margin, most of the correct classifications that LR achieved are located close to zero margin, meaning that the probability assigned to the correct class was only slightly higher than that assigned to the second highest scored class. Looking at the cumulative distribution for RF we can see that the probability scores are much more spread out between 0 and 1, further illustrating the result suggested by the classification edge. Since, as stated before, the emission probability modelling is not actually a classification problem but an uncertainty modelling one, a fully Bayesian treatment of the emission probabilities, that is modelling the emission probabilities using a completely probabilistic generative model, could potentially provide more realistic estimates of the emission probabilities than the statistical classification tools used in this work. The construction and training of such a model is non-trivial and outside the scope of this work but could provide interesting future research.

The application of other classification techniques also holds promise for emission probability modelling. Convolutional neural networks (CNN), a neural network architecture that has been proven to be effective in image recognition tasks [151], have recently also been applied to the task of emission probability modelling [152]. A similar

methodology employing CNNs has been shown to be effective in distinguishing between isolated time-frequency representations of the fundamental heart sounds [153]. It is feasible that a classification scheme that is powerful enough, or in other words, able to correctly identify the "true" origin of sound events could perform both the tasks of the emission probability model and the HSMM and research has been done in this direction [88], [141]. The proposed model, and its conceptualisation as a two-stage classifier, has the advantage of transparency; an important advantage in the case of a diagnostic aid.

The performance of the second stage of the model was measured by its segmentation and classification ability. The model was able to accurately segment heart sound recordings, even in the presence of prominent murmurs obscuring the fundamental heart sounds. These findings agree with previous finding on the effectiveness of duration dependent HMM for heart sound segmentation [154]. As shown in Table 9, the model outperforms the Springer segmentation algorithm [70] on a challenging subset of testing data, the 23 systolic murmur recordings, at a tolerance level of 100 ms while maintaining a high level of performance overall.

The model was able to identify most systolic murmurs in the testing dataset. Regardless of operating point, most of the false negatives (as shown in the confusion matrices of Figure 5.14) were from two heart sound recordings that presented with murmurs that were short in duration and low in amplitude. One example is shown in Figure 5.18 in which the model has segmented the heart cycles without error but has not detected the heart murmur. The duration of the murmur is the most likely suspect for these false negatives as it seems probable that even if the murmur samples were identified as such, the number of "normal systole" samples may have overwhelmed the probabilities during the Viterbi state sequence decoding. Murmurs such as these might be addressed with an extension to the state transition model which makes a serial connection between the Systolic Murmur and Systole possible, although such an addition would increase the complexity of the algorithm.

In adjusting the classification cost/prior probability of the systolic and systolic murmur state it is also important to consider the specificity. In assessing the classification ability of the model, it is important to note that the proposed model does not attempt to make any diagnostic claims about the underlying pathology or cause of the systolic murmur.

It simply tries to determine if a state sequence containing what it "understands" to be systolic murmurs is more likely than one that does not, based on the content of the heart sound recording. Diagnosis is left to the operator who can use the results from the algorithm as a supplement to their overall investigation.

## 5.5 Conclusion

In this chapter a model was proposed that extends the process of heart sound segmentation to include the identification of systolic murmurs. The temporal correlations modelled by a duration dependent hidden Markov model were combined with a feature extraction and classification methodology to automatically label interesting regions in heart sound recordings. The results of the model further demonstrate that there is significant potential in the use of these techniques in automated cardiac auscultation.

The addition of further cardiac states, a "diastolic murmur" and a "noise" state would greatly increase its value. Whether or not the added complexity of adding more and more states to the HMM would lead to the effectiveness of the model decreasing remains to be seen. Evaluation of the model also becomes more involved as the complexity increases and as the workings of the model becomes more and more opaque the main advantage of transparency is lost. The expansion of the model is left up to future work and for now systolic murmurs remain its only pathological target.

The next step in this work is to make use of the labels derived from the model developed in this chapter to further develop a system able to evaluate and describe systolic heart murmurs in a way that is intelligible and acceptable to the user. This description relies on psychoacoustic analysis of systolic heart murmurs, and this is explored and modelled in the next chapter.

## Chapter 6 Psychoacoustic Descriptions of Heart Sounds

This chapter starts with an introduction to some fundamentals of human sound perception, after which specific signal processing techniques for extracting perceptual features are discussed.

### 6.1 Fundamentals of psychoacoustics

Psychoacoustics is the study of sound and human perception of sound; of the physical properties of sound and the perceptual response to these properties [155]. It is a branch of psychology that attempts to answer questions regarding how our auditory system senses, analyses, and interprets acoustic waves. The field of psychoacoustics goes well beyond the scope of this thesis and only the small part applicable to the analysis of heart sounds is discussed here. For a more in depth and expansive introduction and explanation of psychoacoustics the reader is referred to “Psychoacoustics: Facts and Models” by Zwicker and Fastl [156].

#### 6.1.1.1 Why psychoacoustics?

Psychoacoustic modelling was first introduced as a possible approach to the analysis of heart sounds by a 2011 conference paper by Patil et al. [157]. The authors suggested the use of psychoacoustic models as a framework for deriving the "quality" of a given heart sound using descriptors such as 'loudness', 'sharpness', 'tonality', 'strength', and 'roughness'. This work aims to extend this idea to the specific analysis and description of heart murmurs. Firstly, we aim to firmly motivate the use of psychoacoustics in heart sound analysis.

For a diagnostic instrument to be useful it needs to be accurate in its data collection (i.e. its ability to detect, sample, and discriminate the relevant signal correctly) and it also needs to be able to report these findings in a way that is interpretable by physicians or technicians. This second aspect can be called the *usability* of the instrument. One way of understanding usability,, specifically in terms of diagnostic aids, is as a measure of how well the instrument translates diagnostically important information into a form that the user can understand and integrate. A good example of the importance of usability is functional Magnetic Resonance Imaging (fMRI), a powerful imaging technology that has been used, among others, in the field of neuroimaging. Problematically though, fMRI

results contain too much information for a human being to interpret and have to be analysed using specialised statistical software packages [158]. Recently though, the results reached by the software packages have also been called into question [159], [160]. This highlights the importance of the analysis instrument being able to explain the way in which it draws conclusions and not to be seen as a "black box" that delivers the correct answer without question. The importance of psychoacoustics in this work is then to translate the findings of *computer assisted cardiac auscultation* into concepts and terms known to anyone familiar with, but perhaps not an expert in, *traditional cardiac auscultation*. In this way, the proposed models agree more with the perception of the heart sounds producing results that are understandable and reasonable to a user, therefore increasing the acceptability of the algorithms as a decision support tool.

#### 6.1.1.2 Perception of sound: The human auditory system

The external human auditory system, the outer, middle, and inner ear, is shown in Figure 6.1. What is commonly referred to as the ear, the auricle, forms the *outer ear* together with the ear canal. The *middle ear* is composed of the tympanic membrane (eardrum), and the three ossicles, the malleus, incus, and stapes. Finally, the *inner ear* consists of the spirally coiled cochlea and the auditory nerve that innervate sensitive "hair-like" cells in the cochlea.

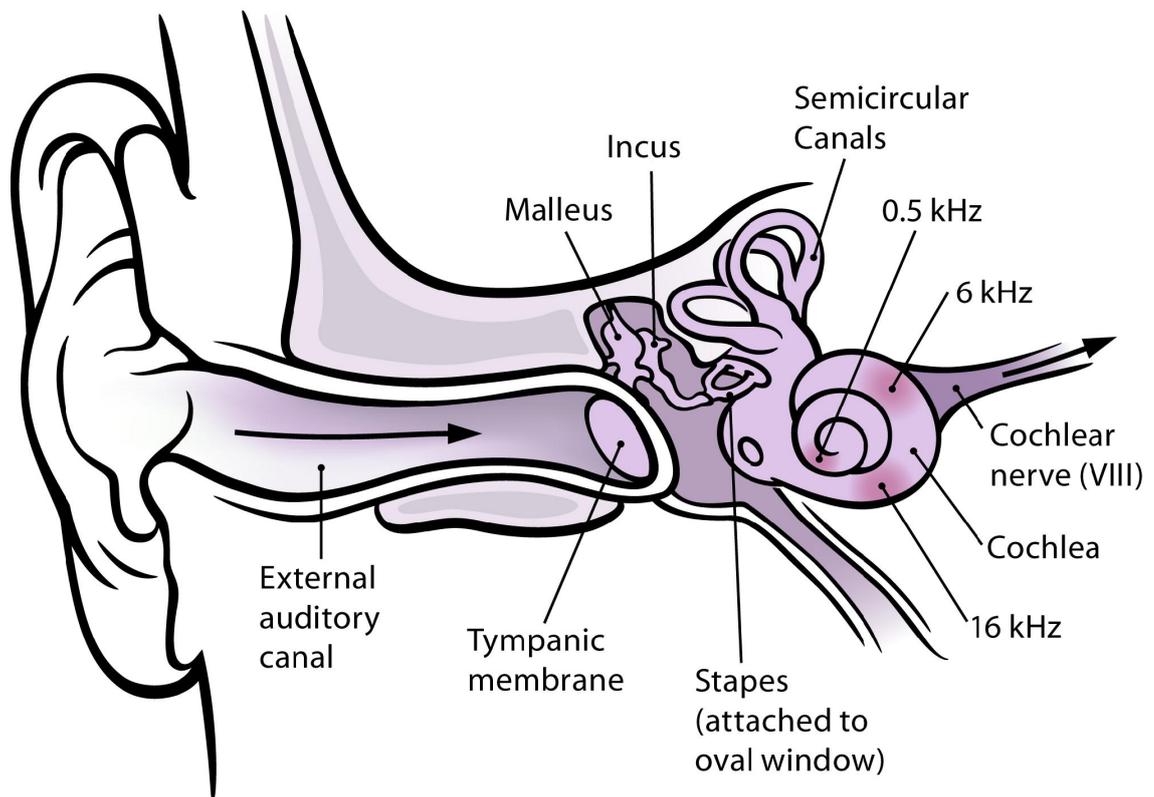


Figure 6.1: Graphical depiction of the outer, middle, and inner ear. The frequency annotations correspond to the approximate regions of the cochlea that are sensitive to these frequencies [161].

The auricle of the outer ear collects sound energy and transmits it to the ear canal and ear drum. The ear canal acts as a transmission line for the collected/received sound energy allowing the sensitive middle and inner ear to be protected in the hard temporal bone while also being placed nearer to the auditory processing centres of the brain. The ear canal also acts as a frequency filter, exhibiting a pronounced sensitivity at frequencies around 4 kHz.

The tympanic membrane, ossicles, and oval window of the middle ear act as a mechanical impedance matching system. The motion of air particles colliding with the tympanic membrane, with large displacement and small force is transferred into large force, small displacement motion of the fluid in the inner ear. The malleus (hammer) is securely attached to the tympanic membrane and transfers any motion to the footplate of the stapes (stirrup) through the incus (anvil). The ratios of the lengths of the ossicles along with the ratio of the area of the tympanic membrane to the area of the oval window results in an almost perfect match of the impedances of the air outside the ear and the body fluid inside the ear at around 1kHz in human beings.

Conversion of acoustic waves into electrical action potentials occurs in the cochlea of the inner ear (Figure 6.2 and Figure 6.3). When the stapes is displaced by acoustic waves interacting with the tympanic membrane it in turn causes the attached oval window to move. The oval and round windows form the two ends of the perilymph-filled scala vestibuli and scala tympani which are connected at the apex of the cochlea (helicotrema). Movement of the oval window causes displacement in the perilymph which in turn moves the basilar membrane and endolymph in the cochlear duct (scala media). The base of the basilar membrane is stiff and narrow compared to the apex (shown as relative length of membrane fibres in Figure 6.3). This lengthwise variation in structure causes sound waves of different frequencies to have peak displacement at different points on the basilar membrane. The amount that the basilar membrane is displaced is related to the intensity of the stimulus sound. As shown in Figure 6.3 using a visualisation of an expanded cochlea, high frequencies cause maximum displacement near the oval window while progressively lower frequencies peak nearer to the apex. In effect then the basilar membrane can be thought of as a series of bandpass filters, with progressively lower pass bands as we move closer to the apex. Transduction of the mechanical displacement of the basilar membrane into neural signals is performed by the *hair cells* of the Organ of Corti. These cells reside in the endolymph filled cochlear duct that connect into the auditory (Cochlear) nerve that travels to the central nervous system where higher order processing occurs.

This very brief introduction to the auditory system, although providing some important insights into the psychoacoustic findings and models presented in the rest of this chapter, has barely scratched the surface and the interested reader is referred to open source text books by OpenStax [162] and the Open University [163] (sources and images provided under a Creative Commons license [164]).

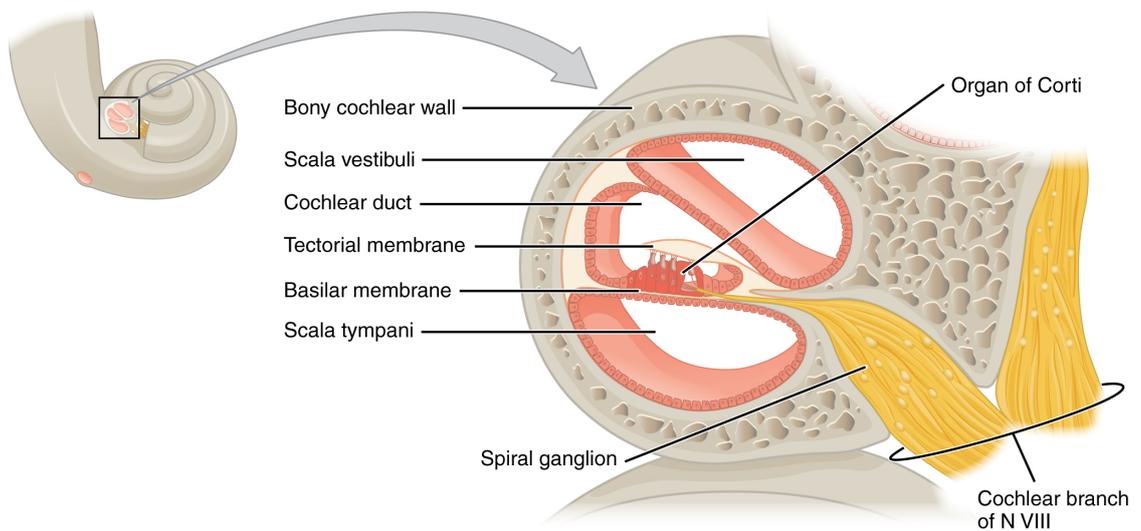


Figure 6.2: A cross section of the fluid filled cavities of the cochlea [165].

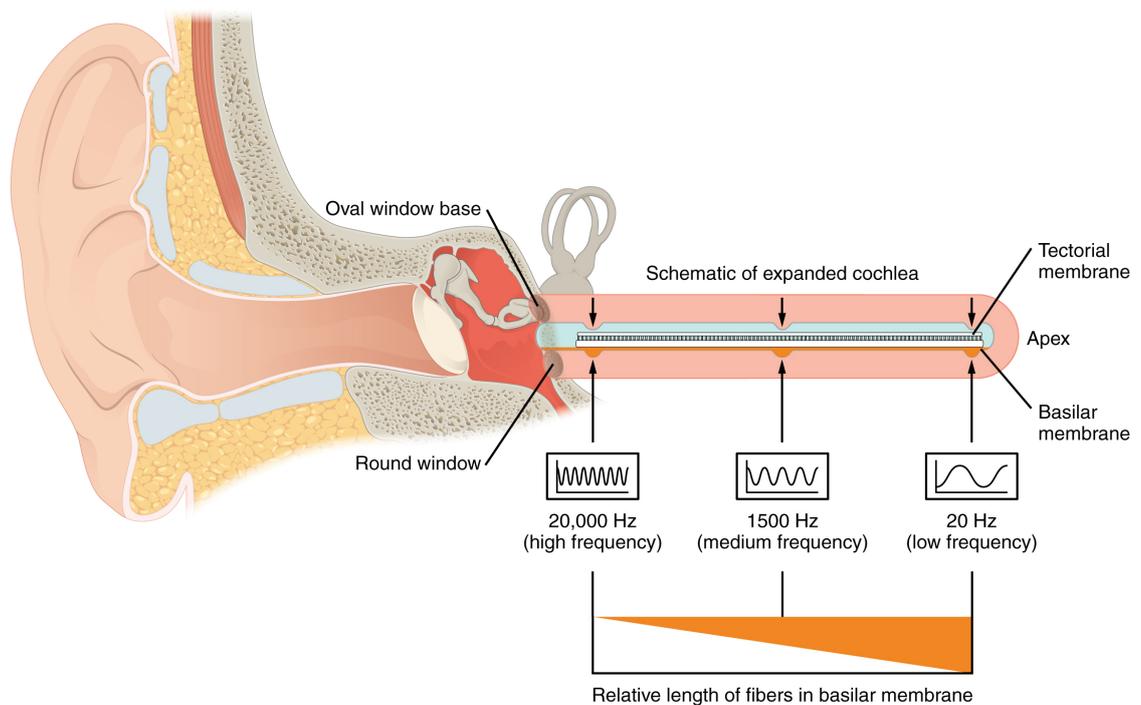


Figure 6.3: A illustration of the variation in the frequency response of the basilar membrane (unrolled in this diagram) along its length. The relative length of the fibres, shown at the bottom of the diagram, is representative of the "stiffness" of the fibres. [166]

### 6.1.1.3 Masking and critical bands

An important concept in psychoacoustics is that of masking. One sound is said to "mask" another if the other sound is not perceptible as a result of the masking sound. The *masked threshold* is the sound pressure level of a test tone necessary to be just audible when a masking effect is present. Along with simultaneous masking there are two

interesting time-related effects of masking, namely, pre-masking and post-masking. Pre-masking occurs before a masking sound, i.e. a sound that is masking another sound, is switched on (up to 20 ms) and can be understood to be the result of the comparatively rapid build-up time of a loud masking sensation compared to that of a softer test tone. In other words, the human auditory system reacts more rapidly to a loud sound than a soft one. The effect of post-masking decays for up to 200 ms after the masker has been removed and depends on the duration of the masker [156].

A related and central concept is that of *critical bands* which is related to the frequency mapping of the basilar membrane of the cochlea discussed in the previous section. The human auditory system is not sensitive to frequencies on a linear or logarithmic scale but rather on a scale related to the physical structure of the basilar membrane; this scale is what the critical bands attempts to approximate. Critical bands are a quantization of the frequencies and associated bandwidths that the human auditory system is sensitive to and help to explain masking phenomena. The Bark [156] and ERB (also referred to as Cam) [167], [168] scales are examples of approximations of the "auditory filters" that the critical bands represent. These "filters" demonstrate saturation effects, thus sounds that occupy the same critical band (i.e. are "close enough" to each other in frequency) "compete" for the same hair cells and will not produce the same sensation of loudness compared to sounds (tones) that are far apart in frequency.

## 6.2 The psychoacoustics of heart sounds

To arrive at an understanding of how heart sounds and heart murmurs are described and understood we turn to the wealth of literature available on cardiac auscultation.

### 6.2.1 Methods

Common psychoacoustic features used when describing heart sounds and murmurs in literature were identified. The findings from this review were used in the creation of a *heart murmur description framework*; representing a set of features that can be used to present what we will call a *psychoacoustic description* of a heart sound which could be conceptualised as the output of performing cardiac auscultation.

Sources that focus on cardiac auscultation and on the fundamentals of practicing cardiac auscultation were reviewed and features that are commonly used to describe heart

murmurs identified. An important limitation to note is that we only have access to the audio as recorded by an electronic stethoscope. This means that important features such as "point of maximal intensity" and "location of radiation" that cannot be inferred from the waveform alone are not included in this framework. Another important set of features that unfortunately must be excluded are those related to the performance of various physiological maneuvers that change the characteristics of murmurs. An example is the Valsalva maneuver, in which the patient is asked to forcefully exhale against a closed airway which modifies the effect of inspiration on systolic murmurs [36]. One of the possible benefits of taking a descriptive rather than a classification approach to heart murmurs is that a physician could use the algorithms to find descriptions of a murmur under a variety of conditions, thus providing information from which to draw diagnostic conclusions under different conditions.

### 6.2.2 Subjective experience

It is important to acknowledge that we are faced with the problem of subjective experience as well as the presence of unconscious processing performed by the human brain. Cardiac auscultation has been described as a "technical" skill as opposed to an "intellectual" one [169]. It seems likely that even an experienced auscultator would be unable to put into exact terms the features that he or she has identified when auscultating and even the act of dividing the sounds experienced during auscultation into some set of discrete features would potentially have an effect analogous to *quantization noise* that could diminish the diagnostic value of these data. An auscultator might not actually be using exactly the features they are consciously aware of to arrive at a diagnostic conclusion. Such questions, perhaps more within the bounds of psychological studies, are left to medical researchers and practitioners and in this work we continue under the assumption that there is significant diagnostic value in a description of psychoacoustic features present in a heart sound; thus that these features, when "correctly" attained, are representative of the underlying reality of the heart and diagnostically important.

### 6.2.3 Psychoacoustic training of auscultation

A remarkable feature of the auditory system is that we can augment our perception of sounds through training. In fact the effect of training exercises on the physiological

representation of sound, as measured by electroencephalogram (EEG), has been observed in both human and animal studies [170]. Research also suggests that different aspects of auditory perception can be trained. A listeners' perception of subtle differences in pitch [171] and of subtle differences in the temporal characteristics [172] of a sound can be altered through training. Likewise for cardiac auscultation a study by Barrett et al. [173] showed that student's ability to recognise cardiac murmurs improved after training. Based on this, it can be conjectured that repeated training of the auditory system allows a listener to recognise more subtle differences within and between sounds. This *plasticity* of the perception of sound is important to consider as we delve into a discussion of psychoacoustics. Studies in psychoacoustics normally focus on what could be called a "statistically normal" individual, that is, they attempt to find the average ranges of auditory attributes by the collection of many samples. The auditory systems that we are trying to model in this work is, however, not statistically normal. An *expert auscultator* is someone who has adapted their auditory system to specialise in the analysis of heart sounds. Thus, the heuristics and thresholds described in psychoacoustic literature, while useful as guidelines and limits, will be less representative of the auditory system of expert auscultators.

#### 6.2.4 Psychoacoustic features of heart sounds

In reviewing literature on the practice of cardiac auscultation three groups of features relevant to psychoacoustic descriptions of cardiac murmurs appeared most prominent. The first, and most prominent, group of features are those related to the *intensity* of the murmur. The second group is related to the *temporal characteristics* of the murmur, or in other words how the murmur changes (or does not change) over time as well as the temporal region of the heart cycle it occupies. The last group of features used to describe heart murmurs are those related to the frequency or *pitch* of the murmur. We have classed descriptions of the "quality" and "timbre" or musical nature of the murmur in this third category although they are perhaps the results of a more complicated relationship between the three groups. Figure 6.4 shows a word cloud constructed from the sources reviewed for this section. Key words associated with the identified groups have been colourised. This word cloud was created using NVivo 12 [174]. A cut-off of 4 letters was chosen and the 100 most common words occurring in all sources cited in



graded in reference to the fundamental heart sounds in the recording. Instead of six distinct levels a murmur is rated as either softer, the same, or louder than the fundamental heart sounds. The scale proposed by Keren et al. [178] is especially relevant to this work because it provides an internal reference with which to compare the murmur loudness. The original Levine scale does not elaborate on concepts such as loud and faint and trusts an auscultator to develop a sense of these levels by experience. This is infeasible for a computer assisted auscultation system unless the method of recording is strictly controlled so that the intensity of sound recordings is in an absolute, or physical, scale. In other words, we cannot compare two different recordings and say one is louder than the other unless we are sure that they have been recorded by the same device in the same manner. The methodology proposed by Keren et al. provides a more feasible approach. Since the fundamental heart sounds in each heart sound recording are necessarily recorded in the exact same manner as any heart murmurs in the recording the levels of these are directly comparable.

Table 10: Systolic murmur loudness rating systems

Rating	Levine Description [177]	In reference to FHS [178]
I	Faint. Heard only after careful auscultation	Clearly softer than the heart sounds
II	Faint murmur heard immediately	Approximately the same level as the heart sounds
III	Moderately loud murmur	Clearly louder than the heart sounds
IV	Loud murmur	-
V	Very loud murmur. Heard even when only the edge of stethoscope is in contact with skin	-
VI	Loudest possible murmur. Can be heard with stethoscope next to by not touching the skin.	-

### 6.2.6 Temporal Features: Shape, timing, and duration

The shape, timing, and duration are all descriptions of the temporal characteristics of a given heart murmur. *Shape* (also referred to as *configuration* [179]) describes the way in which the loudness of a murmur changes over time and is readily divided into 4 categories. The perceived loudness of a murmur can either be increasing, decreasing, increasing then decreasing, or not changing over time. These 4 categories are commonly referred to by their musical nomenclature as crescendo, decrescendo, crescendo-

decrecendo, and plateau respectively. The *timing* of a murmur is used here to refer to any information regarding the heart cycle state in which it occurs as well as the more exact positioning of the murmur within that heart cycle state. For systolic murmurs the possible timings can be specified as early, mid, or late systolic. The terms "ejection" and "regurgitant" have also been used [36], [180], relating whether or not a murmur is the result of blood flowing out of the heart (ejection) or back in (regurgitation), although these descriptions encompass multiple aspects of the timing all at once and their specification is left to future work.

The timing information can also refer to the peak of the murmur, the point in systole at which it is loudest, and a murmur could, for instance, be described as "early peaking". The *duration* describes the amount of a given heart cycle that the heart murmur occupies. A murmur might be "pansystolic" meaning the murmur lasts the entire duration of the systolic state. Murmurs may also be described as "brief" or the description may even overlap directly with the *timing* characteristic as described above; for instance a "mid-systolic" murmur implies both that the murmur occurs in the middle of systole and that the duration of the murmur is limited to a short period in the middle of systole. In this work we have limited the determination of these characteristics to the description of the shape of the murmur.

### 6.2.7 Pitch

The pitch of a murmur is related to the velocity of blood flow. In general the higher the velocity of the blood flow the higher the pitch of the murmur [36], [179]. Murmurs are commonly designated as low, medium, or high pitch. The association of pitch with blood flow velocity can provide some clues on the origin of a murmur. High pitch murmurs are generally associated with areas at which blood flows from areas of high pressure to areas of lower pressure. For instance the high pitched diastolic murmur of aortic regurgitation (aortic insufficiency) is the result of blood leaking back into the relaxing ventricle out of the relatively high pressure aorta [175]. The perception of pitch in heart murmurs has not been extensively studied, and usually pitch related descriptions are associated with certain disease conditions. For instance, the murmur of rheumatic mitral valve regurgitation is described as "high pitched" by Alpert [36]. A primary assumption we will make in this work regarding pitch perception and description is that auscultators

tend to describe the pitch in relation to the fundamental heart sounds, and thus that the descriptions of "low", "medium", and "high" are relative to the frequencies of the FHS. The description of "low" could then be further explained as "approximately the same frequency as the heart sounds", "medium" can be expanded to "slightly higher frequency than the heart sounds", and "high" as "clearly higher frequency than the FHS". It is of course also possible that expert auscultators, similarly to expert musicians, have acquired the ability to recognise pitches in some absolute sense or relative to an "internal library" of murmur pitches, but even so the common descriptions used to describe pitch still seem to be in relation to some constant reference, the most obvious being the fundamental heart sounds which have a relatively constant frequency spread.

### 6.2.8 Systolic murmur description framework

These three groups of features, which we will label *loudness*, *shape*, and *pitch*, form the basis of the systolic murmur description framework which details the features that the algorithms developed in this chapter will attempt to discover. In the following section models for these groups of features are developed based on psychoacoustic principles in combination with digital audio signal processing techniques.

## 6.3 Modelling approaches

### 6.3.1 Dataset

Data from subsets of the Physionet/Computing in Cardiology Challenge 2016 [138] dataset, presented in detail by Liu et al. [128], were used in this work. The probabilistic labelling murmur subsets, described in Chapter 5, were again used in this chapter.

### 6.3.2 Loudness

Although sometimes used interchangeably in common speech there exists a clear difference between the *intensity* and *loudness* of an audio signal. The *intensity* of a sound is a physical quantity, measured in watts per meter squared, and defined as the product of the sound pressure and the velocity of the medium the sound wave is travelling in. As discussed previously, microphones generally measure the time varying sound pressure and not the actual sound intensity. The sound pressure level (SPL) is the sound pressure converted to a decibel (dB) scale which is more commonly used in order to account for the wide range of audible air pressures. Similar to microphones, the

human auditory system also senses changes in sound pressure rather than directly measuring sound intensity; thus in this work we will not consider intensity (formally defined) any further, instead it will be used in its common usage to denote the *amplitude/volume* of a sound.

*Loudness* is a perceptual entity that can be defined as the perception of the intensity of a signal and can only be measured by responses of observers [58]. The intensity of a sound is the primary correlate of the loudness of the sound [155], but loudness constitutes a complex relationship between the sound waves and the auditory system of the listener and along with sound intensity is also a function of the frequency, duration, and bandwidth of the sound. Many attempts to quantify loudness have been made in the last century, see for instance the work of Zwicker and Fastl [156], a review and summary of the state of the art of loudness measurement and modelling can be found in the work by Florentine et al. [155].

Loudness and sound pressure level are not linearly related. Two perceptual units are helpful in the measurement and understanding of loudness, these are the reference unit *phon* and the loudness unit *sone*. One phon is defined as the loudness of a 1 kHz tone at 1 dB SPL. More generally a tone of  $x$  phon has a sound pressure level of  $x$  dB at 1 kHz. One sone is defined as the loudness of a 40 dB SPL tone at 1 kHz. The sone scale is a linear loudness scale, for example a tone of 2 sone is twice as loud as that of 1 sone (i.e. it is twice as loud as a 1 kHz, 40 dB SPL tone). Zero sone and phon represent the threshold of hearing, lower than which sounds are not audible to the statistically average person.

The second most important correlate of loudness is frequency. The relationship between the loudness and frequency of pure tones has been well studied [155], [156]. The ISO 226:2003 equal loudness contours [181] are shown in Figure 6.5 for a frequency range of 20 Hz to 2 kHz. This small frequency interval was chosen since it covers the complete bandwidth of cardiac sounds, as well as the fact that all heart sound collected specifically for this research have been sampled at 4 kHz. Each of the contours represents the sound pressure level that produces an equally loud sound at different frequencies to a "statistically average" person. The phon and sone values are also shown

on either side of 1 kHz frequency. It should be clear from these curves that our auditory systems are increasingly less sensitive to frequencies below around 1kHz.

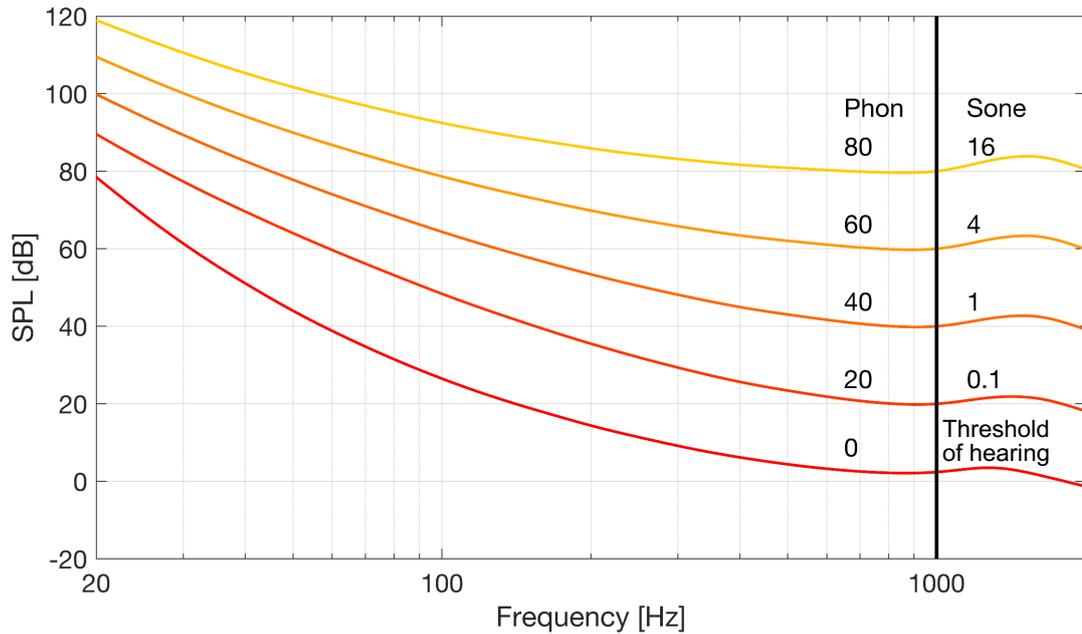


Figure 6.5: ISO 226:2003 equal loudness contours [181]. The sound pressure level (SPL) is plotted against frequency (on a logarithmic scale). Each contour shows the SPL required at the corresponding frequency for the loudness to be considered "equal", i.e. a 20 Hz tone at 100 dB SPL will be equally loud (approximately) as a 1000 Hz tone at 40 dB.

Equal loudness contours are integral in the design of *loudness weighting filters*. Two examples of commonly used weighting filters are shown in Figure 6.6. The A-weighting contour [182] simulates the perceived loudness at low level tones, around 40 phon (1 sone). The ITU-R 468 recommendation is designed to reflect the perceived loudness of not just pure tones but all kinds of noises [183]. The most significant differences between these two contours occurs in the frequency range of 1 kHz to 9 kHz however, and thus the difference between these two standards is minimal regarding this work. We have however opted to use the ITU-R 468 weighting because of the broader bandwidth of the heart sounds compared to pure tones. In this chapter the ITU-R 468 weightings are applied to heart sound recordings before the estimation of any of the psychoacoustic features.

At this point it is necessary to point out that the use of weighting curves (specifically A-weighting curves) are not seen as a sufficient method of measuring loudness by Zwicker and Fastl [156]. Their criticism, however, mainly relates to the use of A-weighting curves for estimating the loudness of high-level noises with varying durations and bandwidths

at high frequencies where the frequency bandwidths of our auditory system becomes wider. Heart sounds are soft (low-level) sounds at relatively low frequencies with limited bandwidths and short durations, thus the weighting curves should prove appropriate for estimating their relative loudness. The applied methodology of magnitude estimation after perceptual weighting has shown to have similar results to the more complex psychoacoustical loudness models [58]. Although beyond the scope of this thesis, further development of methodologies that more strictly conform to psychoacoustic findings and models could potentially provide interesting results in heart sound analysis research.

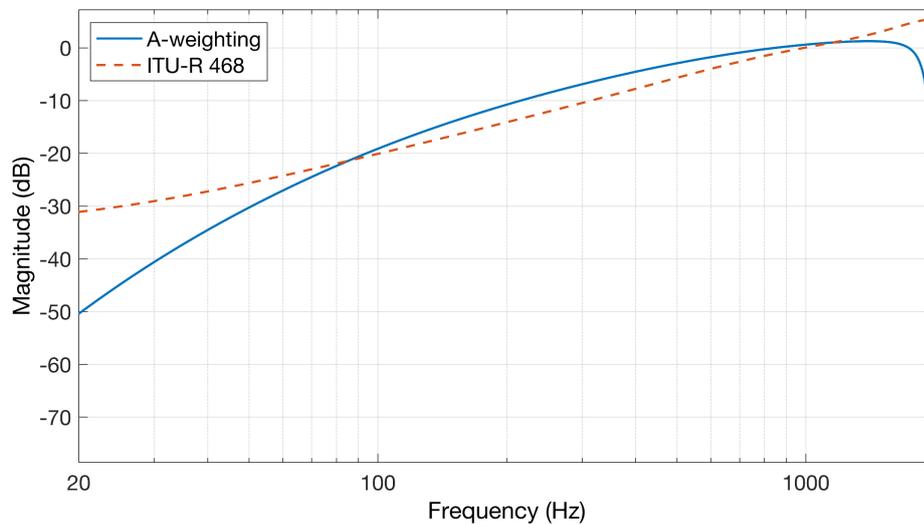


Figure 6.6: Common perceptual weighting filters for the frequency range of 20 to 2000 Hz.

The sound intensity can be calculated as the root mean square (RMS) of the signal amplitude, as described by Lerch [58]. It is not possible, or necessary, to determine the intensity and loudness in physical units, rather these are calculated using the relative changes in sound pressure level as recorded by electrical changes by the stethoscope microphone. Block based processing is employed to determine the intensity envelope (the variation of the RMS/intensity over time). The RMS for a block of T samples can be calculated as

$$v_{RMS}(t) = \sqrt{\frac{1}{T} \sum_{t=T}^{t=1} x(t)^2} \quad (6.1)$$

To estimate the sound intensity level of the audio signal the intensity can be converted to a decibel scale by taking the logarithm of the ratio of the RMS to a reference level, which in this work has been taken as  $v_0 = 1$ , corresponding to the full-scale value. This is computed using

$$v_{dB}(t) = 20 \log_{10} \left( \frac{v_{RMS}(t)}{v_0} \right) \quad (6.2)$$

To calculate an estimate of the loudness of the signal the RMS is calculated for a loudness-filter weighted version of the signal.

$$v_{loudness}(t) = \sqrt{\frac{1}{T} \sum_{t=T}^{t=1} x_l(t)^2} \quad (6.3)$$

where  $x_l$  is the audio recording with loudness weightings applied. The loudness envelope provides a representation of how the audio signal would sound to a listener. This is illustrated in Figure 6.7 using an example signal composed of 1 second intervals of 10, 100, and 1000 Hz with 1 second increasing chirps between them, sampled at 4 kHz. For this example, the intensity and loudness envelopes were calculated using blocks of 100 ms with a hop size of 50 ms. The unweighted magnitude of the signal is practically constant throughout its 5 second duration, corresponding to the equal amplitude of the different sine wave and chirp wave components of the example signal. However, the loudness varies significantly with the frequency of the different components. Frequencies below 20 Hz (the 10Hz component at the start of the signal) are essentially inaudible but produce palpable effects at high enough levels, thus the loudness representation given by ITU-R 468 filtering gives a closer approximation of the sensation associated with listening to this example signal.

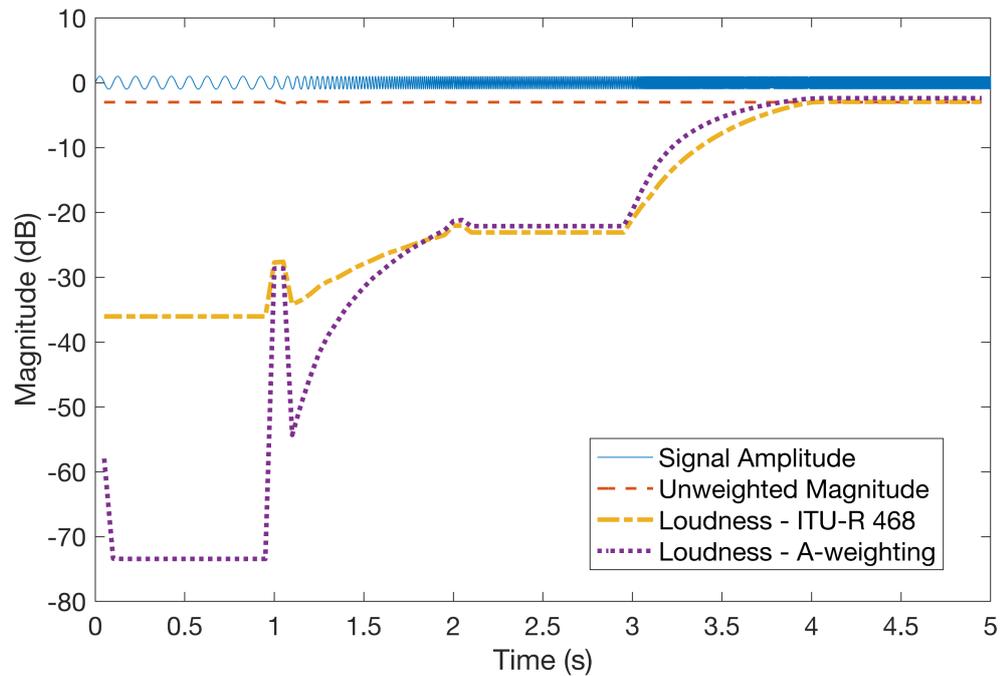


Figure 6.7: Loudness levels for the tone and chirp signal described in Chapter 4, as calculated using two different perceptual weighting curves.

### 6.3.2.1 Methods of calculation

To estimate the loudness of a given murmur the following approach was followed: the heart sound was first segmented using the algorithm described in the previous chapter including a probability that the systolic states contain a murmur. Based on the segmentation results the first heart sound ( $state = S1$ ) and the systolic ( $state = SysMurmur$ ) states were isolated in an ITU-R 468 weighted version of the heart sound recording. An envelope representative of the loudness level was determined for each of these states and magnitude features across the duration of all the examples of the state in the recording were calculated from this envelope. To emulate the Keren et al. grading system the loudness of  $SysMurmur$  state was compared to that of  $S1$ . The arithmetic means across all samples assigned to these two states in a given recording were calculated. The level difference of  $S1$  relative to  $SysMurmur$  was determined and used to classify a murmur as loudness  $I, II, or III$ . Figure 6.8 shows the loudness envelopes and mean loudness levels for an example heart sound.

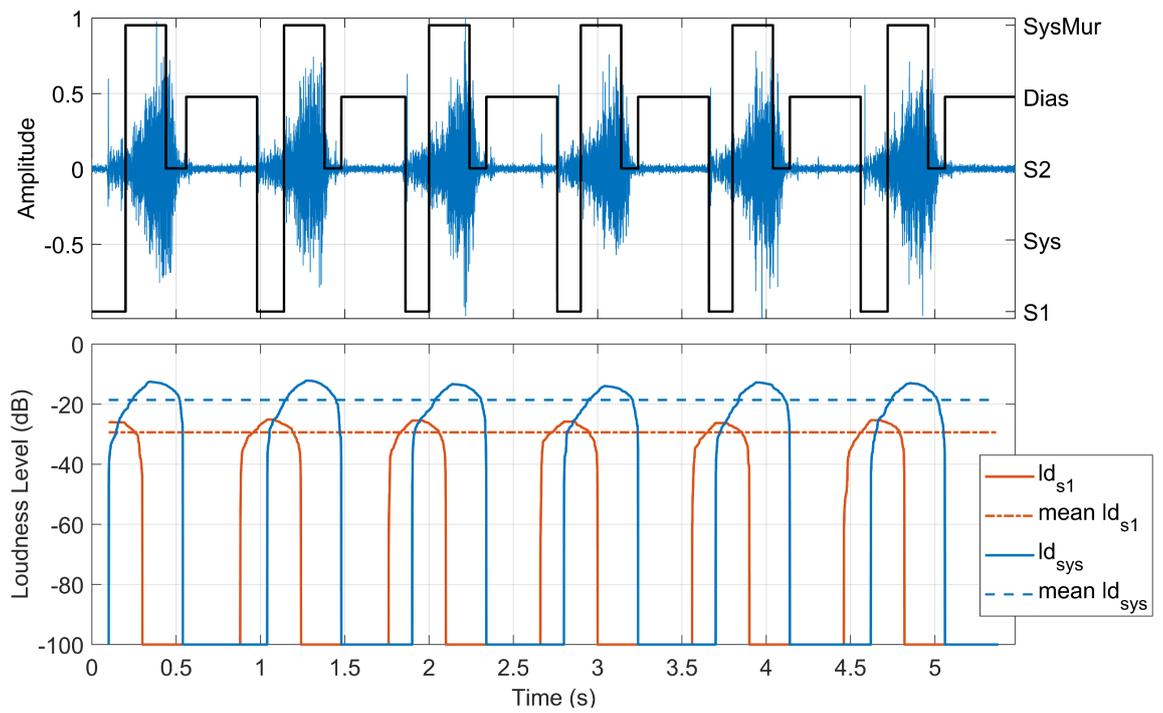


Figure 6.8: An example of the loudness levels of the systolic and first heart sound states across the duration of a heart sound recording. The dashed lines represent the mean loudness level of all of the samples in the recording that have been assigned to that state.

In this example the first heart sound has components with approximately the same amplitude as the murmur components but significantly lower frequencies. Figure 6.9 illustrates the same calculation as Figure 6.8 without applying the loudness weighting to the waveform. The higher frequencies of the murmur lead to it being perceived to be much louder than the first heart sound; this observation is lost in the second analysis.

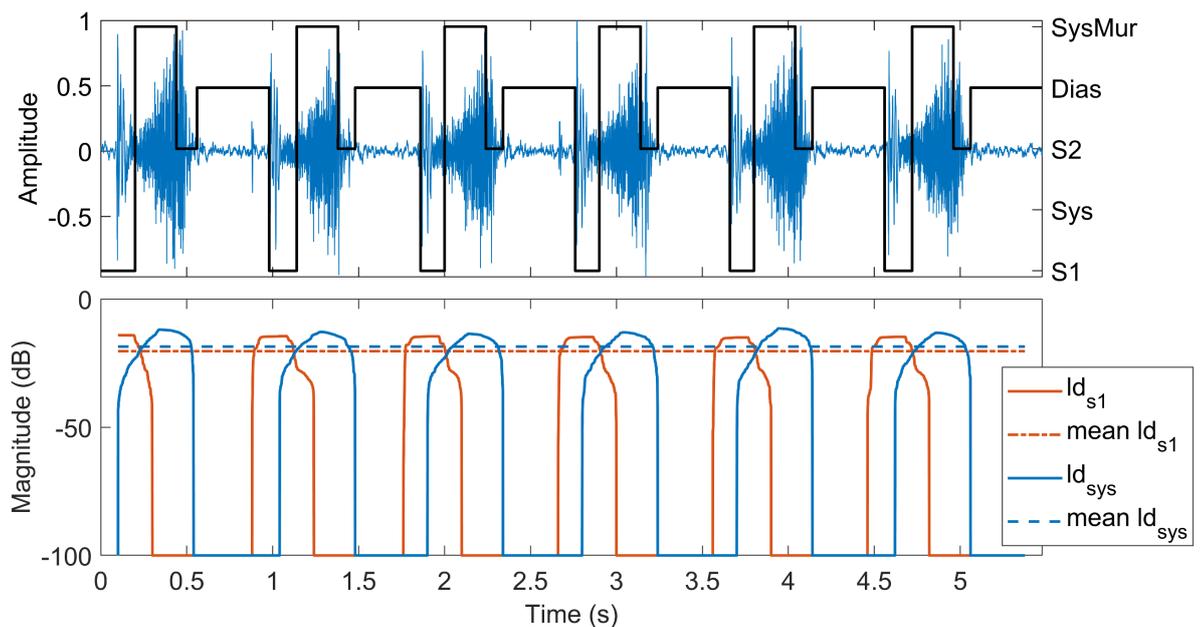


Figure 6.9: The same waveform as Figure 6.8 without the loudness weightings applied.

### 6.3.2.2 Results

The results of the proposed method for examples from the "systolic murmur" subset are shown in Figure 6.10 along with the decision boundaries proposed in this work. The upper boundary (between II and III) represents murmurs with a mean loudness level 1 dB higher than the first heart sound, while the lower boundary represents murmurs that are 10 dB lower in level than the first heart sound in the same recording. These boundaries were determined heuristically, based on the intuition that a murmur just noticeably louder than the fundamental heart sounds would be more readily marked as such, while a murmur would have to be significantly softer than the heart sounds for an observer to class it as I. These boundaries attempt to capture the I, II, and III loudness ratings of both the Levine and Keren et al. scales.

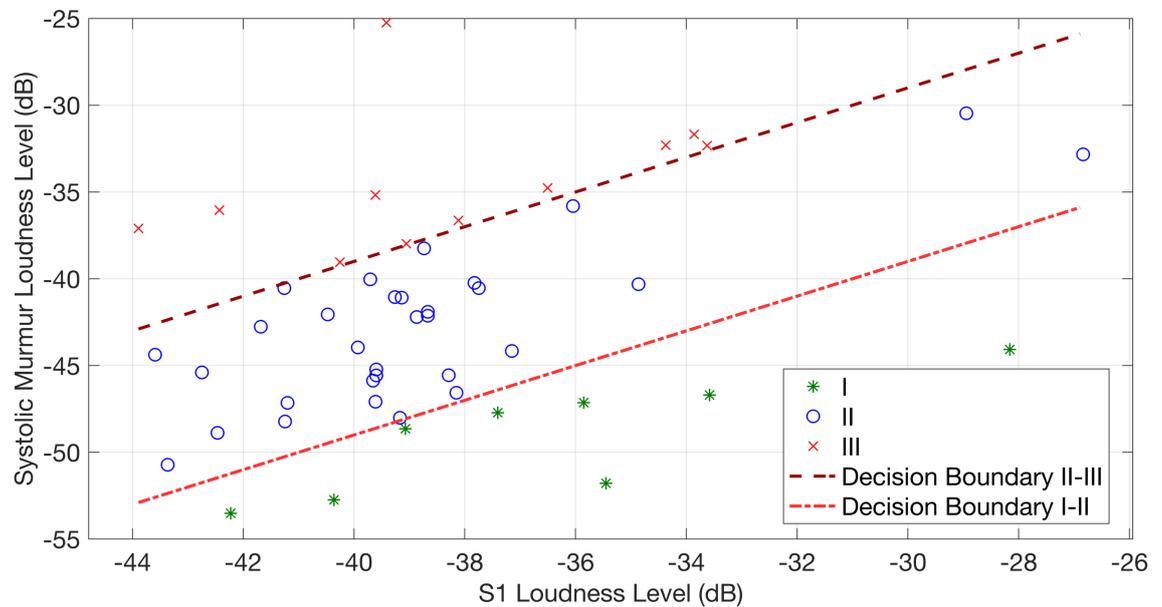


Figure 6.10: The "loudness" decision space populated by the samples from the probabilistic labelling murmur subset. The decision boundaries used in this work are also illustrated. As clearly shown in this figure the decision boundaries represent linear relationships between the S1 and Systolic Murmur loudness levels.

### 6.3.3 Shape

Before continuing with the model derivation for the "shape" feature, a distinction should be made between what can be called objective (or physical) duration and subjective duration, or perhaps we would be better served by referring to a perceptual concept that we can call psychoacoustic duration to specify that we are referring to an "experienced period of time in auditory perception". In any case, data have shown that human perception of the duration of auditory events does not directly correspond to

their objective duration in time in two cases [156], both relevant to this work. The first case is in sounds that have short durations, less than around 100 ms, below this duration sounds tend to be perceived as lasting longer than their "objective" duration. The second discrepancy is that which occurs in the psychoacoustic duration of *pauses*, or brief periods of silence, in between sounds. Using a comparative methodology in which participants were asked to adjust the duration of either a pause or a burst to have the same length, researchers found that the psychoacoustic duration of pauses were around 2 times less than that of bursts for a 200 Hz tone. That is, to have the same psychoacoustic (experienced) duration, a pause would have to last twice as long as a tone of 200 Hz. This effect increases with higher frequencies and can be up to 4 fold for tones around 3 kHz [156]. These two aspects of psychoacoustic duration can then be summarised as: the experienced duration of short sounds is more than their physical duration and the experienced duration of short pauses in between sounds is less than their physical duration.

Psychoacoustic duration is important to consider in heart sound analysis since the heart sounds fit well the description of sounds where psychoacoustic (subjective) duration deviates from objective (physical) duration. The heart sounds have short (objective) durations (see previous chapter) and can easily be described as bursts of sound and pauses, relating to the second case. For the estimation of the timing characteristics of heart murmurs this can be interpreted as the slight "stretching" of the edges of the loudness envelopes of sounds due to pre- and post-masking or equivalently the slight delay of perception as the stimulus decays in the case of pauses.

Zwicker and Fastl [156] model the psychoacoustic duration of bursts by measuring the period in which the sound burst excitation level (a measure of loudness level) exceeds the local minima by 10 dB. Likewise, the (psychoacoustic) duration of pauses can also be estimated by finding the period in which the excitation level is within 10 dB of the local minima. This heuristic also appears to be appropriate for loudness envelopes calculated using the ITU-R 468 weighting and homomorphic envelope (HE) discussed in section 6.3.2. Figure 6.11 shows a reproduction of an example signal used in the previously mentioned psychoacoustic research. In the methodology described by Zwicker and Fastl [156], participants are asked to adjust the duration of the pause in between the two bursts of white noise (at around 1 sec) to match that of the burst of white noise that

occurs at around 200 ms. Interestingly the loudness envelope calculated using the HE exhibits aspects that appear very similar to pre- and post- masking as observed in psychoacoustic studies. This is likely a result of the low pass filtering of the envelope extraction methodology emulating the build-up and decay characteristics of the auditory system, but regardless of the exact cause, this suggests that the HE, combined with ITU-R 468 weighting, as an effective method of estimating the loudness of an audio signal.

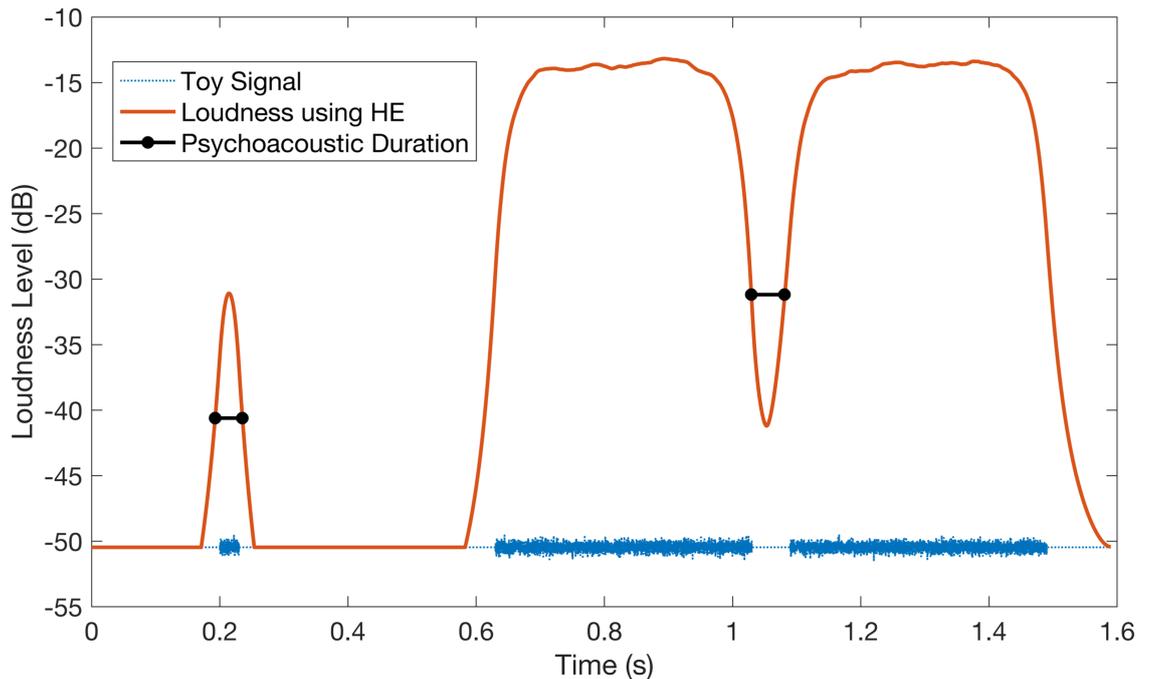


Figure 6.11: The loudness envelope as calculated using HE of an example signal (shown here with an offset mean for visual purposes) that includes a 30 ms burst of white noise followed by two 400 ms bursts of white noise separated by a (perceptually similar in duration to the 30 ms burst) pause of 60 ms. The black lines represent and circles represent 10 dB above the local minima which heuristically represent the subjective duration.

A significant limitation of block-based root-mean-square calculation discussed in section 6.3.2 is the distortion due to the averaging over time and the necessary approximation of new time indices. An alternative methodology is to calculate a loudness measure that naturally possesses time dependency. The homomorphic envelope (HE) of the ITU-R 468 weighted example signal is shown in Figure 6.12. In this case the loudness of the signal is represented by the HE, thus,

$$v_{loudness}(t) = HE_l(t) \quad (6.4)$$

where  $HE_l$  is the homomorphic envelope of the perceptually weighted input signal. Because the HE is calculated from the analytic signal the power of the resulting envelope is doubled, resulting in the approximately 3 dB difference seen in Figure 6.12. While this means that the absolute values of the RMS and HE envelopes are not directly comparable, the effect is a uniform level shift and does not influence the methods discussed further in this work.

The HE envelope has several advantages over the block-based approach. Since the HE is calculated for each of the samples of the original signal the time indices remain unaltered. This means that the HE is simpler to index than a block-based envelope since it does not require separate indices to be calculated for the original signal and its envelope. The RMS approach has the advantage of being a more tried and tested method of loudness calculation and so the results are arguably a more trustworthy representation of the "actual" loudness of an audio signal. The difference between the two methodologies seems to be small, based on the results of the example signal, and the advantages conveyed by the natural time dependencies of the homomorphic envelope make it a more attractive methodology for estimating the change of loudness over time.

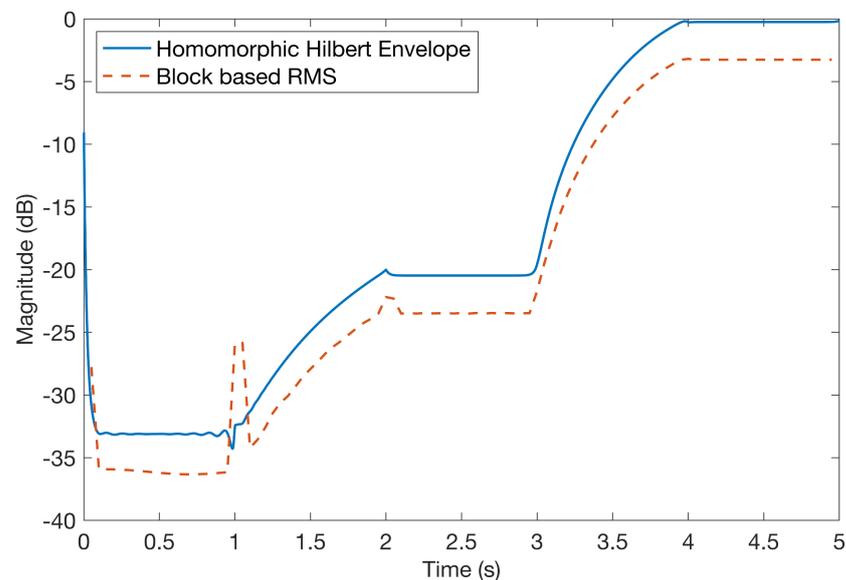


Figure 6.12: Loudness levels as estimated using the homomorphic envelope and RMS of the ITU-R 468 weighted example signal from Figure 6.7.

### 6.3.3.1 Methods of calculation

Determining the shape of a murmur present in the systolic state proved challenging since it represents a complex, time varying relationship. For this reason, a statistical classification framework was adopted. Features related to the variation of the loudness over time were extracted from each of the *SysMurmur* regions labelled by the HSMM. A loudness envelope was estimated by calculating the homomorphic envelope of the ITU-R 468 weighted audio signal. Each of the regions labelled as *SysMurmur* were extracted and divided into 3 parts (each  $\frac{1}{3}$  the length of the region). The following feature extraction was then performed on each of the regions. In the following equations  $i$  represents the index of the part in question,  $S$  represents the shape curve,  $S_i$  the  $i$ th part of the shape curve,  $N_i$  the number of samples in part  $i$ , and  $N_T$  the number of samples in all three parts.

- Firstly, the area under the curve (approximated as the sum of the sample values and named  $AoC$  to clearly distinguish it from the commonly used AUC classification metric) for each of the parts was calculated as a ratio of the total area under the curve for that region.

$$AoC_i = \frac{\sum_{n=1}^{N_i} S_i(n)}{\sum_{n=1}^{N_T} S(n)} \quad (6.5)$$

- The second class of features represent an approximation of the slope in each of the regions and was calculated as the difference between the value of the first and last samples in each of the parts divided by the length of that part.

$$Slope_i = \frac{S_i(N_i) - S_i(1)}{N_i} \quad (6.6)$$

- Lastly, the value of the midpoints of each of the parts rounded down to the nearest sample  $[*]$  were determined.

$$MidPoints_i = S_i\left(\left\lfloor \frac{N_i}{2} \right\rfloor\right) \quad (6.7)$$

The mean of each of these features across all the regions in a given heart sound recording was used to form a representative feature vector for the heart murmur. To

perform classification each of the three classes of features were parameterised by first standardising them across the 3 parts by subtracting the mean and dividing by the standard deviation. This had the effect of mapping all the points in each of the classes onto a circle in the 3-dimensional (represented by the 3 parts that features were extracted from in each systolic region) space. These points were then mapped onto a linear space by arbitrarily choosing a starting vector  $[1, -1, 0]$  on the circle and determining the angle of each of the data points to this vector. This effectively allowed each of the three classes of features to be represented as a single parameter reducing the required feature space from 9 to 3 without any loss of information. These "shape parameters" were then used as the input to an error correcting output code (ECOC) multiclass support vector machine (SVM) classifier.

#### 6.3.3.2 Results

Evaluation of the proposed methodology was performed using data in the systolic murmur *training* subset described previously. The shape of all the murmurs in the dataset were labelled by the researcher according to their loudness envelope and their perceived shape when listened to. These annotations were then used as the "ground truth" in training a Radial Basis Function (RBF) SVM classifier. To compensate for the unbalanced dataset the prior to the RBF-SVM classifier was set to uniform before training. The examples in the systolic murmur *testing* subset were then also annotated by the researcher, and the results of trained model on these samples were compared to their annotations to gauge the ability of the model to estimate the perceptual shape of murmurs (using the researcher's perception as a target). The results shown here are thus an estimate of the ability of an RBF-SVM classifier trained on a dataset of a given perceiver's annotations to reproduce those annotations on an unseen dataset.

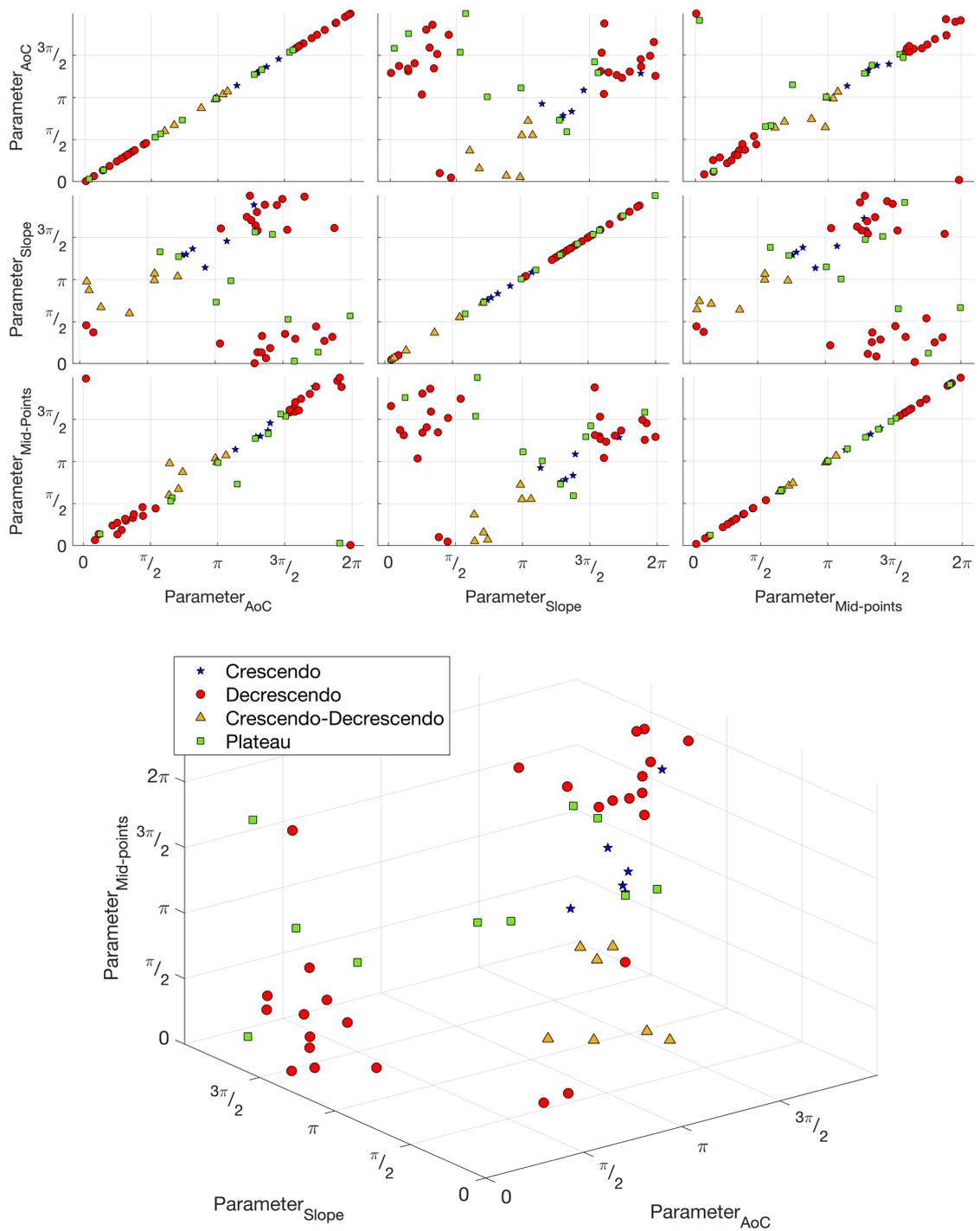


Figure 6.13: The parameterised decision space used to determine the "shape" feature. Each parameter represents the angle between a reference vector and the sample vector mapped onto a circle in the space constructed from the "early", "mid", and "late" time periods.

The shape parameter space for the training data is shown in Figure 6.13 above. Discernible clusters exist for all the "shapes" except "plateau" which shows samples spread over almost the entire decision space. Despite this, the trained model was able to correctly classify the majority of the 23 examples in the testing dataset as can be seen in the confusion matrix shown in Figure 6.14. Unsurprisingly the incorrectly classified

samples were all three either incorrectly classified as "plateau" or examples of "plateau" that were misclassified.

**Confusion Matrix**

<b>Output Class</b>	Crescendo	5 21.7%	0 0.0%	0 0.0%	1 4.3%	83.3% 16.7%
	Decrescendo	0 0.0%	5 21.7%	0 0.0%	0 0.0%	100% 0.0%
	C-D	0 0.0%	0 0.0%	4 17.4%	0 0.0%	100% 0.0%
	Plateau	1 4.3%	1 4.3%	0 0.0%	6 26.1%	75.0% 25.0%
		83.3% 16.7%	83.3% 16.7%	100% 0.0%	85.7% 14.3%	87.0% 13.0%
	Crescendo	Decrescendo	C-D	Plateau		
	<b>Target Class</b>					

Figure 6.14: Results of the proposed "shape" support vector machine on the examples from the probabilistic labelling murmur test subset.

### 6.3.4 Pitch

Human perception of pitch is primarily perception of the frequency of an audio signal. That is, higher frequency signals are perceived as being higher pitched [58]. Although sound pressure level does influence the perception of pitch slightly (less than 3% shift [156]) this effect is neglected in this work as the heart sounds are relatively low sound pressure level signals. Pitch perception of combination of tones is dominated by what is referred to in signal processing literature as the *fundamental frequency*. If a signal is a combination of sinusoidal components with frequencies,  $f_0, 2f_0, 3f_0, \dots$  then pitch perception is dominated by  $f_0$ , even in cases where  $f_0$  has low power. The relationship between the perceived pitch and the fundamental frequency of a tone is positively

correlated but non-linear, although these nonlinearities only become very prominent above 1 kHz they are present at lower frequencies. As an example a tone with frequency 220 Hz will be perceived as approximately half the pitch as that of 440 Hz, but a tone with a frequency of 1.3 kHz will have approximately half the pitch as a tone of 8 kHz [156]. The relationship has been modelled in a variety of ways by different researchers [58].

Examples of the Bark and ERB scales (introduced in section 6.1.1.3) are shown in Figure 6.15 for the frequency range from 20 to 1000 Hz, the range of most interest in heart sound analysis. A logarithmic (base 10) relationship is also shown for comparison. Figure 6.16 shows an example of the magnitude spectrum of a heart sound plotted against the scales shown in Figure 6.15. It should be clear from these figures that the ERB and Bark Scales provide greater resolution at the frequencies of interest to this work. The rest of this section will make use of the Equivalent Rectangular Bandwidth scale (ERBS) since it is a more recent adaptation of Zwicker's Bark scale with a more detailed model of frequencies below 500 Hz [168]. The frequency of a signal can be approximated to an ERBS representation using,

$$ERBS(f) = 21.4 \log_{10}(0.00437f + 1) \quad (6.8)$$

where the frequency  $f$  is in hertz. In practice frequencies were converted to ERBS using the VoiceBox MATLAB package [184] which employs a slightly more complicated approximation. As an illustration using the previously mentioned example of 440 Hz and 8 kHz the relationship in ERBS between 440 Hz and 220 Hz is 0.61 while the relationship between 8 kHz and 4 kHz is 0.84, while the ratio of 1.3 kHz and 8 kHz is 0.55;

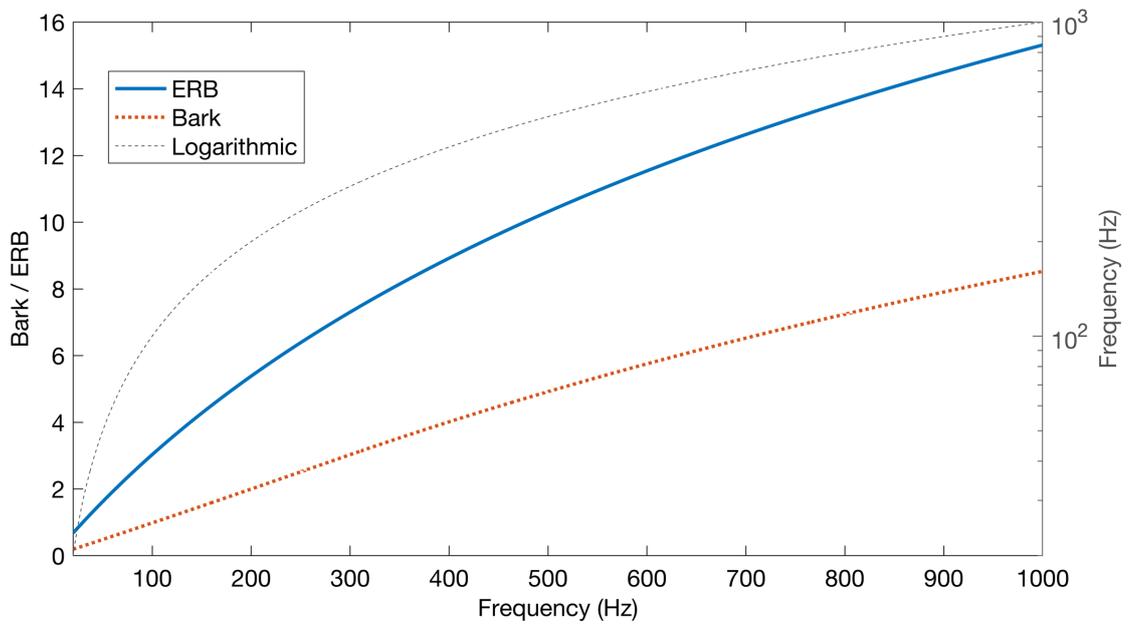


Figure 6.15: The Bark and ERB perceptual pitch scales for the frequency range of 0-1000 Hz. A logarithmic relationship is shown for reference.

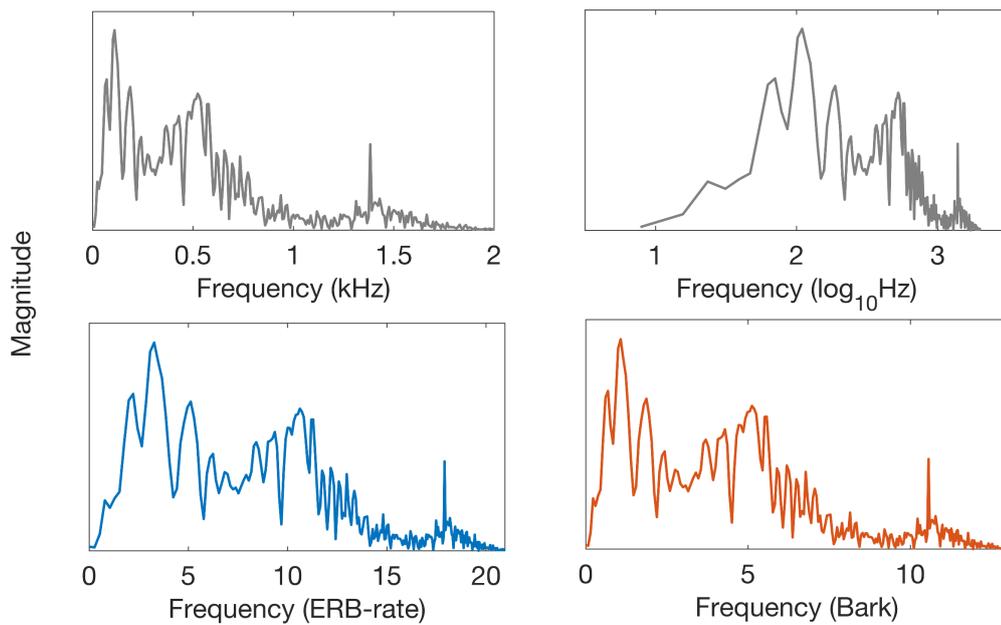


Figure 6.16: Illustration of the magnitude spectrum of a heart cycle as viewed against a linear frequency, logarithmic frequency, ERB, and Bark scale.

In audio content analysis/music information retrieval it is useful to determine the predominant pitch at different times in the music; in this way the musical notation can be determined automatically. The techniques used for pitch detection rely on the signal consisting of musical tones consisting of fundamental frequencies and harmonics. Heart murmurs are more noise-like, they have broader spectra than pure tones, and as such their perceived pitch also differs from that of pure tones. According to Zwicker and Fastl

[156] bandpass noise produces pitch sensations that correspond to the frequencies of the spectral edges. If the bandwidth is narrow enough (roughly less than 200 Hz at less than 1kHz centre frequency) then a single pitch corresponding to the centre frequency of the pass band is perceived. As motivation the frequency spectra for a single trumpet note and a heart murmur resulting from mitral regurgitation are shown in Figure 6.17, both have been weighted to represent human perception (using ITU-R 468 specifications). The trumpet note presents with very clear spectral peaks that occur at spaced intervals (harmonics), while the heart murmur has a much more erratic pattern with energy at a much wider range of frequencies (more noise-like).

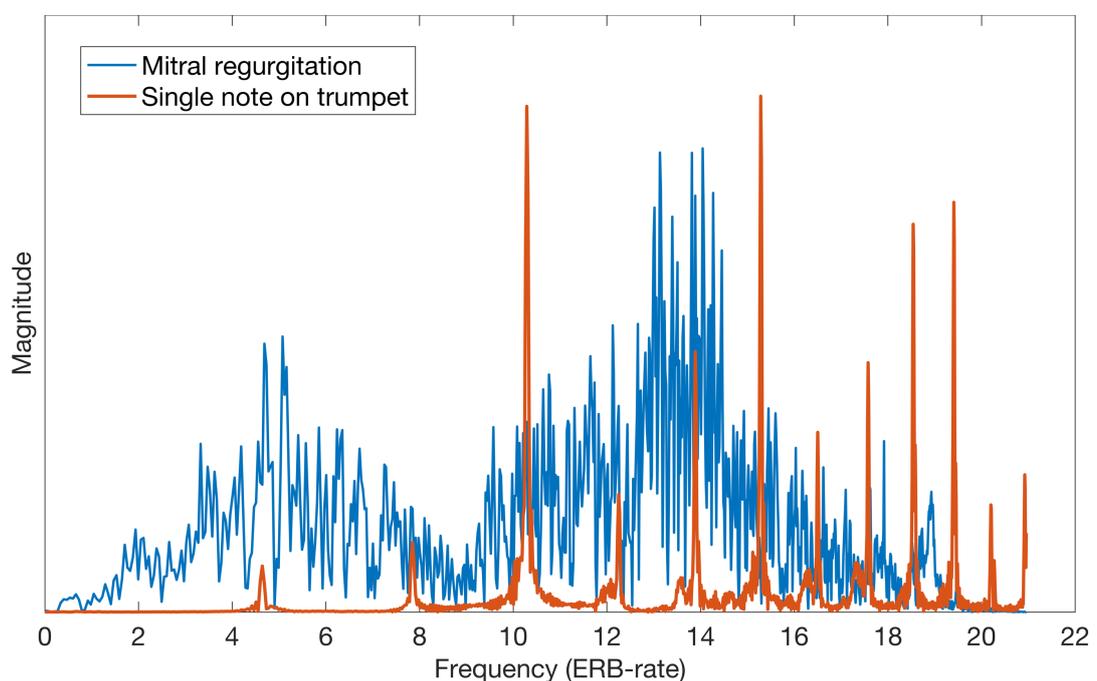


Figure 6.17: Comparison of the magnitude spectra of a trumpet and a murmur containing heart cycle.

#### 6.3.4.1 Methods of calculation

As discussed in the previous section, the frequency spectra of heart sounds, and especially heart murmurs, are broader and more resemble noise than harmonic tones, such as those elicited by many musical instruments. The traditional techniques used in ACA for "pitch-tracking" are thus less applicable to heart murmur analysis. In this work the mean magnitude of the continuous wavelet transform coefficients over the duration of the systolic region has been used to produce a representation of the frequency spectrum of the signal. A comparison of the frequency spectra as calculated using the

CWT and the FT are shown for an example heart sound that contains a distinct mitral regurgitation heart murmur in Figure 6.18.

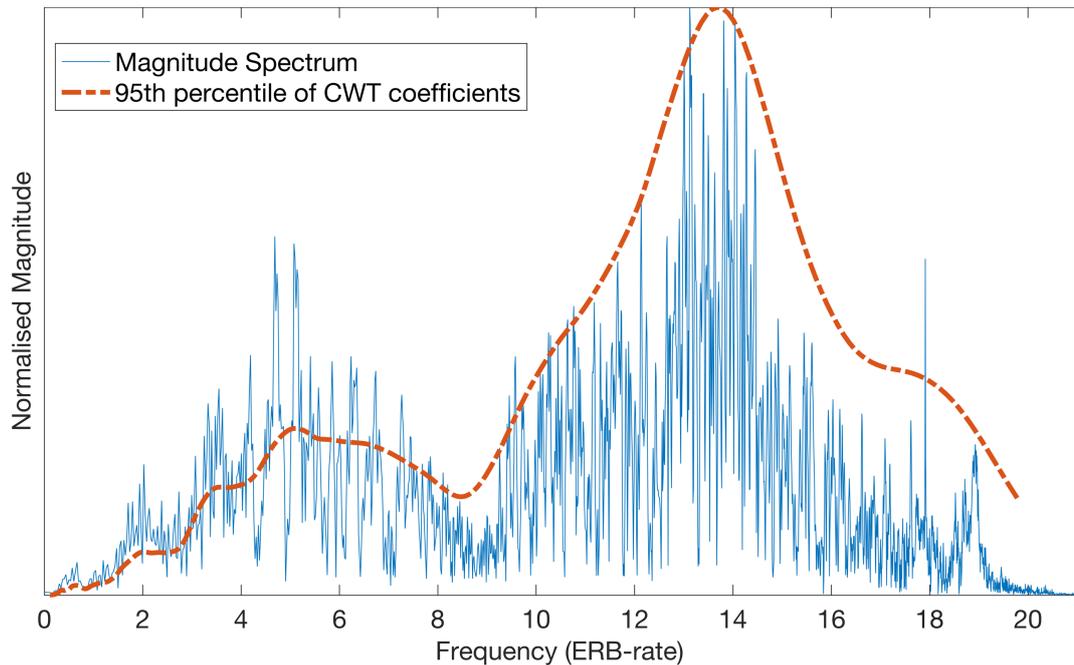


Figure 6.18: Comparison of normalised frequency spectra calculated from the magnitude of Fourier transform and the 95th percentile of the continuous wavelet transform coefficients.

To discretize the pitch of the noise-like murmurs, magnitude spectrum representations of the regions labelled as murmurs were calculated using the CWT. The average of the CWT spectra across all the systolic murmurs was calculated and used as a representative magnitude spectrum for the murmur under analysis. The peak magnitude in 3 frequency bands was then used to discretise the pitch of the murmur as low, medium, or high. The bands were specified in ERB's based on the frequency ranges of the fundamental heart sounds (which range from just below 1 to around 4 ERB). A summary of the chosen ranges is shown in Table 11 and illustrated in Figure 6.19 for examples of low, medium, and high frequency murmurs.

Table 11: ERB and equivalent frequency ranges for Pitch features extraction

Description	Rating	ERB range	Equivalent Frequency range (Hz)
Approximately equal in pitch to FHS	Low	1 - 4	30 - 138
Slightly higher pitched than FHS	Medium	4 - 7.5	138 - 311
Significantly higher pitched than FHS	High	7.5 - 15.5	311 - 1024

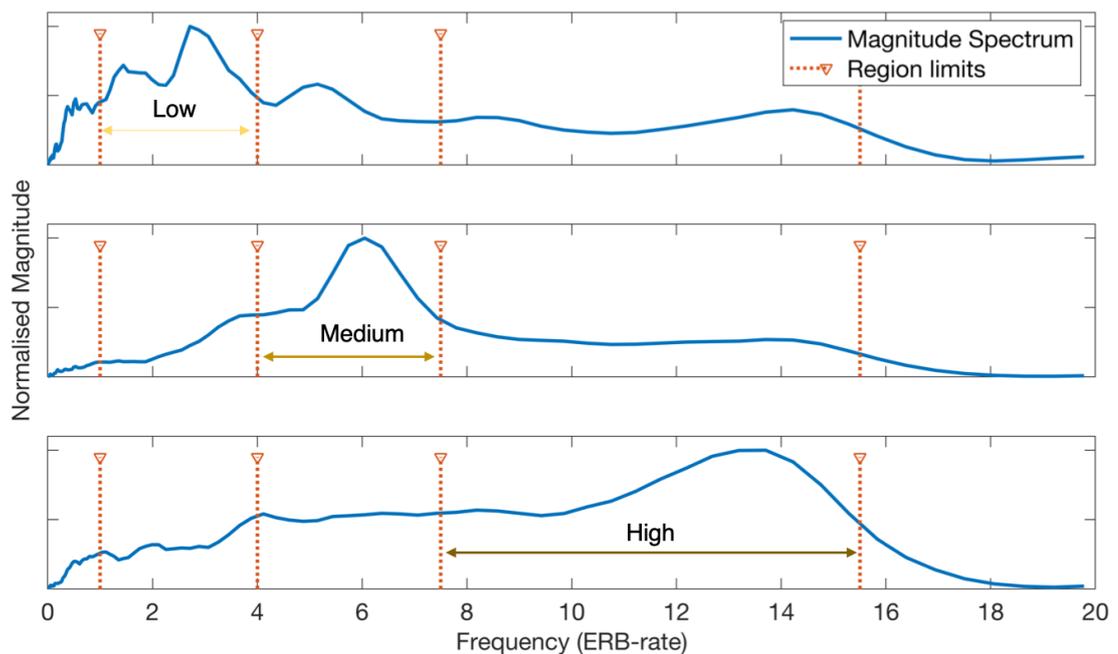


Figure 6.19: The three pitch regions, low, medium, and high, shown with examples of CWT spectra that present with peaks in each of these regions.

#### 6.3.4.2 Results

To provide a classification of the pitch as low, medium, or high, the peak magnitudes of the CWT spectrogram from each of the regions were compared. The pitch of the murmur was then designated according to the region in which the largest peak occurred. Figure 6.20 shows the resulting classifications in a standardised peak space for the 50 examples of the systolic murmur subset. Standardising the normalised peak amplitudes across the low, medium, and high regions and choosing an arbitrary starting vector, a one dimensional pitch parameter was calculated. This pitch parameter is plotted against the heart sound recording indices in Figure 6.21 below.

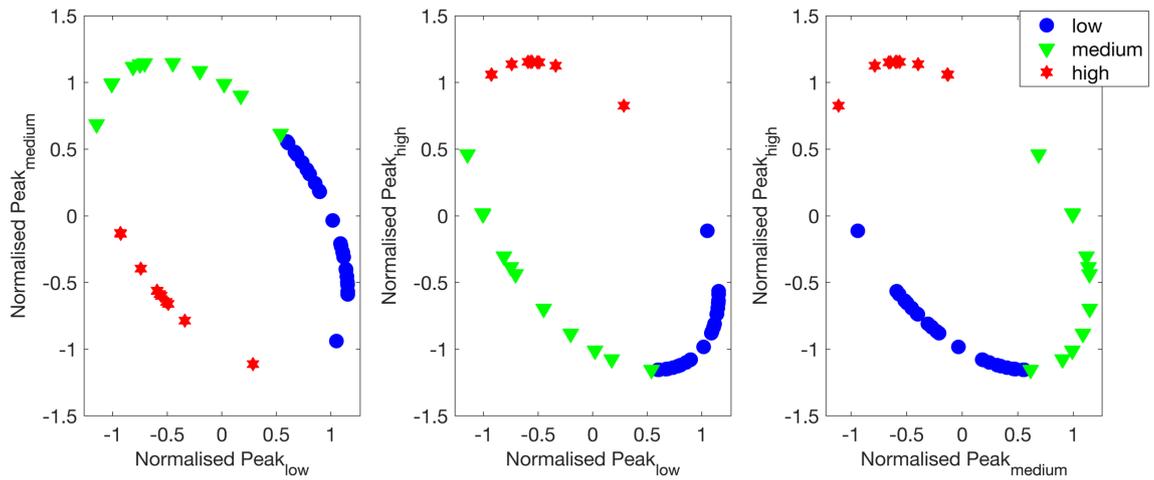


Figure 6.20: The normalised peaks are mapped onto a circle in the three dimensional (low, medium, high) decision space.

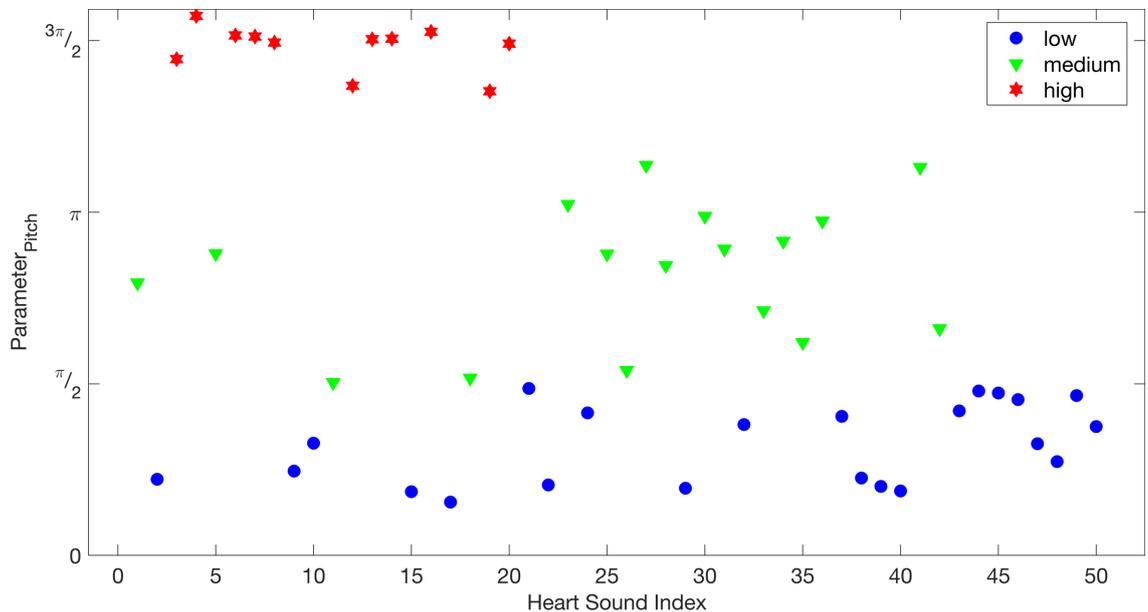


Figure 6.21: By parameterising the peak space we can represent the three dimensional samples using a single dimensional (the angle) without losing any information.

## 6.4 Expert descriptions: Cardiac Auscultation Research Survey (CARS)

A primary hindrance to the proposed methodology is the lack of relevant data. That is, data that are adequately annotated with expert descriptions of murmurs. Since heart sound recordings can be described across numerous dimensions, for example, diagnosis, subject attributes (age, sex, weight, etc.), recording sites, or degree of disease to name but a few, this is not surprising. Collecting detailed descriptions of murmurs also poses serious challenges. Firstly, the inherently subjective nature of such descriptions poses the issue of consensus or disagreement. In other words, when inevitably, there is disagreement between two graders, how do we define the "ground truth". Secondly,

collecting quality ratings from experienced auscultators poses a serious logistical challenge. Since expert auscultators (defined in this work simply as cardiologists) are relatively rare and the task of rating multiple heart sounds can be difficult and time-consuming, recruitment of suitable and sufficiently motivated graders proves challenging. Thirdly, listening to a recording of a heart sound is not the same as cardiac auscultation. Cardiac auscultation is an aspect of physical examination and consists of processes of investigation and deduction in which the stethoscope is used in different ways and at different locations and conscious attention is directed at multiple aspects of the patient, the sound produced by the heart, and the different sensations experienced by the auscultator. Asking for an expert auscultator to simply listen to a heart sound recording ("in vitro") and provide a description of any abnormalities of the sound is akin to providing a baker with a picture of bread and asking about the flavour, while some description can be given the amount of information is severely limited compared to "in vivo" auscultation. This limits the possible ways in which such data can effectively and efficiently be collected. Despite these challenges in this work we have attempted to design a framework for the efficient collection of such data and conducted a study to test the effectiveness of this methodology and collect some preliminary results.

#### 6.4.1 Cardiac Auscultation Research Survey (CARS) design

An online survey was created using the Qualtrics platform [185]. This survey, named the Cardiac Auscultation Research Survey (or CARS), consisted of 20 heart sounds recordings that contained systolic murmurs that participants could listen to as many times as they wished. Participants were asked to classify the loudness, shape, and pitch of the murmur in the recording, as well as the overall quality of the recording. Participants were asked to grade loudness on the ordinal scale proposed by Keren et al. [178] using the fundamental heart sounds as internal reference. The shape/configuration was graded on a nominal scale with four possibilities and participants were asked to select the most relevant option. Finally, the pitch of the murmur was graded on a 3 point ordinal scale (low, medium, high). An example question for a single heart sound recording is shown in Figure 6.22.

Heart Sound 1



Rate the loudness based on the following adaptation of the Levine scale

- I - Clearly softer than the heart sounds
- II - Approximately equal in intensity to the heart sounds
- III - Clearly louder than the heart sounds

Select the most relevant description of the shape or configuration of the murmur

- |                       |                       |                           |                       |
|-----------------------|-----------------------|---------------------------|-----------------------|
| Crescendo             | Decrescendo           | Crescendo-<br>Decrescendo | Plateau               |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> |

Rate the pitch of the murmur

- |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|
| Low                   | Medium                | High                  |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Figure 6.22: An example of the heart sound questions as they were presented to participants. Participants were able to listen to the recordings as many times as they wished and were asked to select the most relevant descriptions out of the choices provided.

The 20 heart sound recordings used in the survey were taken from the MIT heart sound database (MITHSDB) subset of the publicly available Physionet Heart sound Database [138]. The majority (15) of the examples were cases of mitral valve prolapse. The remaining 5 cases consisted of 3 with unspecified diagnoses, 1 benign murmur, and 1 case of aortic disease. All recordings were made either at the apex (mitral site) or the second left intercostal space (pulmonary site). The mean duration of the recordings was  $33 \pm 1.2$  seconds.

Table 12: Summary of information available for survey dataset

Survey ID	Physionet ID	Diagnosis	Site of recording	Duration of recording
1	a0002	Mitral valve prolapse	Pulmonary site	20.8 s
2	a0014	Mitral valve prolapse	Apex	35.7 s
3	a0020	Mitral valve prolapse	Pulmonary site	35.7 s
4	a0022	Mitral valve prolapse	Apex	36.2 s
5	a0023	Mitral valve prolapse	Apex	30.6 s
6	a0024	Mitral valve prolapse	Apex	20.7 s
7	a0031	Mitral valve prolapse	Apex	35.6 s
8	a0033	Mitral valve prolapse	Apex	35.7 s
9	a0036	Mitral valve prolapse	Apex	35.9 s
10	a0042	Unspecified pathology	Apex	36.2 s
11	a0045	Mitral valve prolapse	Apex	35.9 s
12	a0057	Unspecified pathology	Pulmonary site	35.9 s
13	a0059	Mitral valve prolapse	Pulmonary site	35.7 s
14	a0065	Unspecified pathology	Pulmonary site	35.7 s
15	a0066	Mitral valve prolapse	Pulmonary site	35.6 s
16	a0073	Mitral valve prolapse	Apex	35.7 s
17	a0078	Mitral valve prolapse	Apex	35.7 s
18	a0084	Benign	Apex	30.8 s
19	a0089	Aortic disease	Pulmonary site	21.0 s
20	a0101	Mitral valve prolapse	Pulmonary site	35.8 s

Perhaps the most confounding element in the collection of heart murmur descriptions using an online survey is the lack of control over the hardware used to play the sounds. Participants were advised to use in-ear headphones, as these had shown more success in reproducing the low frequency elements of the heart sound recordings. Even so it is not possible to control the sound card or sound drivers that the participant would be using and so to compare the results of different participants a calibration question was added where the participant's perception of low frequency tones was assessed. If a participant gave too low responses to these calibration questions, they were advised that their hardware was likely not suitable for completion of the survey.

## Low Frequency Test

Please play each sound and comment on well it can be heard

	Not at all	Less Clearly	Somewhat Clearly	Very Clearly
30 Hz 	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
40 Hz 	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
60 Hz 	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
100 Hz 	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[← Back](#) [Next →](#)

Figure 6.23: The low-frequency calibration block was designed to assess whether or not the participants' audio playback was suitable for the survey. Participants who reported not being able to hear the example tones were advised that their results might be invalidated.

### 6.4.2 CARS delivery

Ethics approval for the distribution of the CARS was obtained from the Auckland University of Technology Ethics Committee (AUTEC), ethics approval number 17/186. Participants were required to have specialised in cardiology, being either a qualified cardiologist or cardiology fellow. A total of 25 participants were identified and contacted via email. Seven of the contacted participants responded to the survey request and 4 completed the survey in full. Out of the participants that started the survey and did not finish, 2 were unable to hear the low frequency test tones and stopped the survey and 1 participant stopped after only completing half of the survey for unspecified reasons. This translates to a 16% completion rate, only slightly lower than the expected range for a survey distributed via email (around 20-30% [186]–[188]). Unfortunately, one participant who completed the survey reported not being able to hear the low frequency tones presented during calibration at the start of the survey and thus their results could not be included. In total then from the 7 participants who responded to the survey request, data was collected from 43% (3 out of 7) and there was a dropout rate of 57%

(4 out of 7). This small sample size is unlikely to be representative or provide statistically significant reliability measures. This study does however provide an interesting preliminary view of the variability between expert auscultators using an online survey data collection methodology as well as some preliminary data to which to compare the proposed algorithm.

### 6.4.3 CARS results

A total of 4 expert responses were collected for each of the 20 heart sound recordings in the survey. Due to the low rating that Participant 4 gave in response to this calibration block their results have been excluded from the rest of the analysis. Thus, a total of 60 responses were collected for three features from three participants. The results for the low frequency calibration test at the start of the survey are shown in Table 13. The distributions of the responses for each of the heart sounds are shown in Figure 6.24, Figure 6.25, and Figure 6.26 for the loudness, shape, and pitch respectively. These figures show stacked bar graphs for each of the heart sounds used in the survey. Each rating is represented by a different shade and pattern and the area of the bar occupied represents the proportion of experts that gave that rating.

Table 13: Participants' responses to the low frequency calibration block at the start of the survey

	30 Hz	40 Hz	60 Hz	100 Hz	Total
<b>Participant 1</b>	Very Clearly (3)	Very Clearly (3)	Very Clearly (3)	Very Clearly (3)	12
<b>Participant 2</b>	Somewhat Clearly (2)	Somewhat Clearly (2)	Very Clearly (3)	Very Clearly (3)	10
<b>Participant 3</b>	Less Clearly (1)	Somewhat Clearly (2)	Very Clearly (3)	Very Clearly (3)	9
<b>Participant 4</b>	Not at all (0)	Not at all (0)	Not at all (0)	Not at all (0)	0

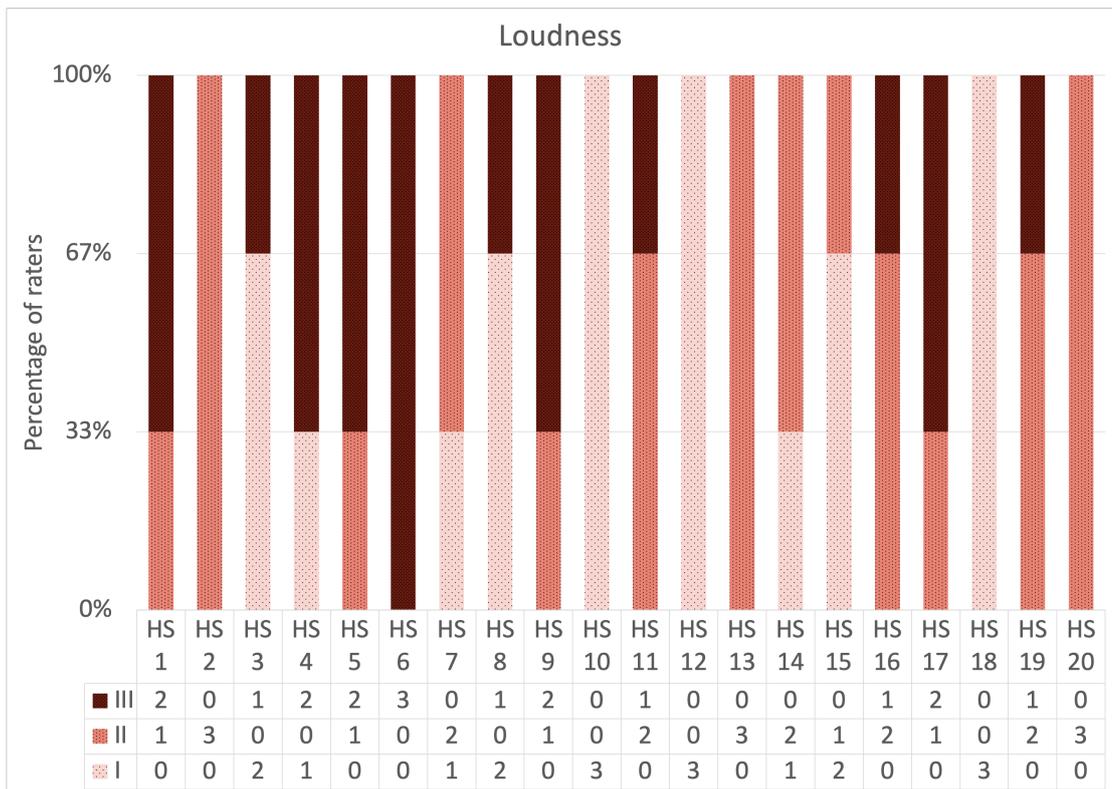


Figure 6.24: Participant results for loudness feature. Uniform color/pattern in a single bar represents high agreement among experts for that heart sound.

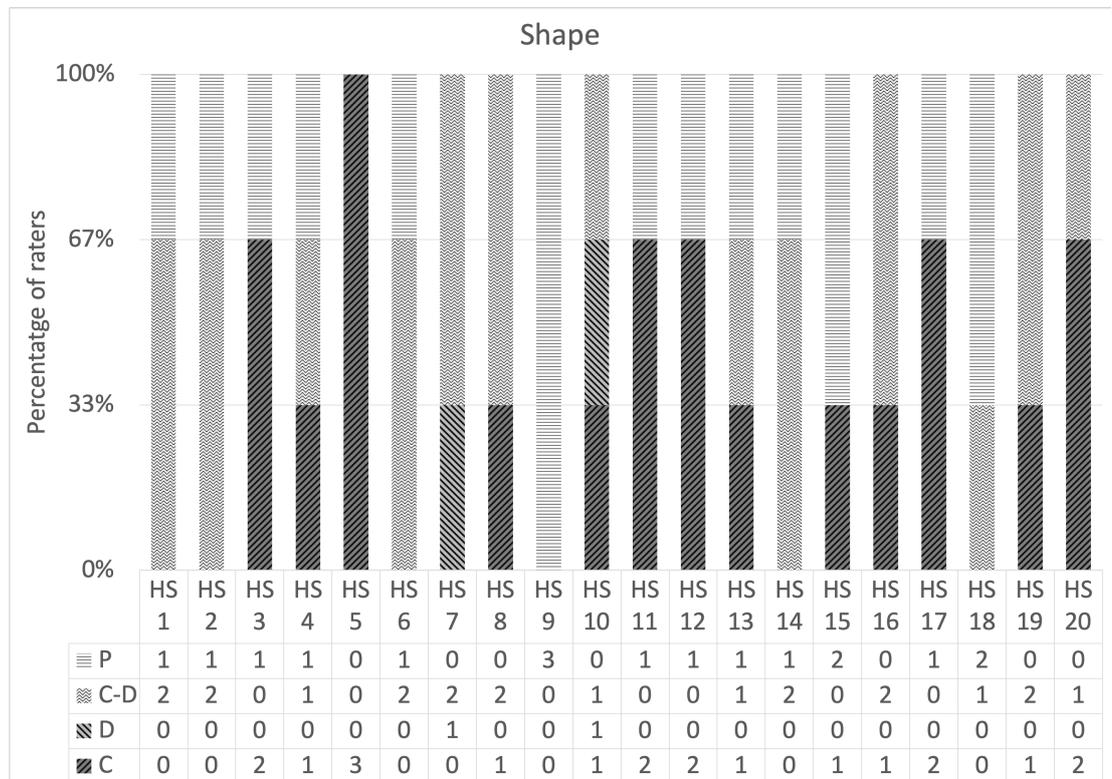


Figure 6.25: Participant results for Shape feature. Uniform pattern in a single bar represents high agreement among experts for that heart sound.

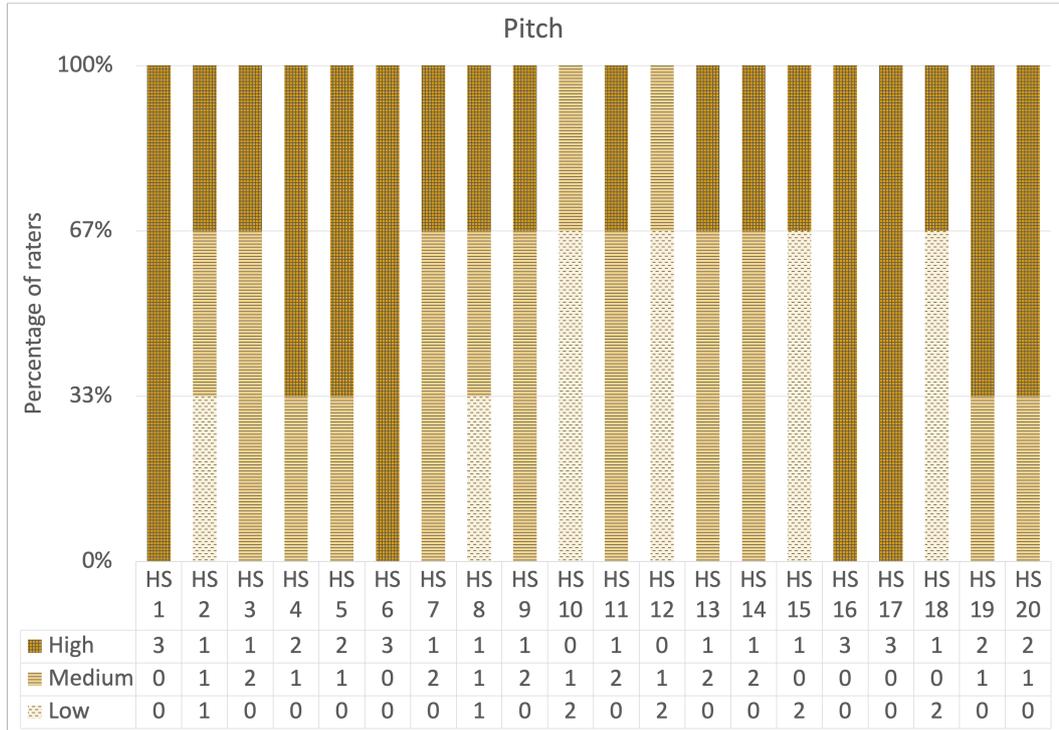


Figure 6.26: Participant results for pitch feature. Uniform color/pattern in a single bar represents high agreement among experts for that heart sound.

As there appears to be little scientific consensus on the statistical measures used to evaluate inter-Participant reliability [189], [190] we have opted, for these preliminary results, to report the most basic interrater-reliability metric of *observed percent agreement* (also called *Accuracy*) with the caveat that this measure does not adjust for "chance" agreement and will overestimate the level of agreement in cases of relatively low inter-Participant agreement [191]. For multiple categories, multiple Participants, and a score weighting scheme this metric can be calculated as,

$$\alpha_0 = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \frac{r_{ik} (\sum_{l=1}^K w_{kl} r_{il} - 1)}{r_i (r_i - 1)} \quad (6.9)$$

with,

$N$  - the total number of items (heart sound recordings) that were rated

$K$  - the total number of categories (e.g. 3 for loudness)

$r_{ab}$  - the number of Participants that assigned item  $a$  to category  $b$

$w_{ab}$  - the weight associated with two Participants assigning an item to categories  $a$  and  $b$

$r_i$  - the number of Participants that assigned item  $i$  to any category

Along with the observed agreement we also report the *specific agreement coefficient* for each of the possible categories of each of the features, this can be thought of as the conditional probability of a randomly selected Participant assigning a particular category given that another randomly selected Participant has assigned the item to that category. The specific agreement is calculated as,

$$\sigma_k = \frac{\sum_{i=1}^N r_{ik}(r_{ik} - 1)}{\sum_{i=1}^N r_{ik}(r_i - 1)} \quad (6.10)$$

with similar notation to Equation (8).

The "mReliability" [192] software package for MATLAB was used to estimate the observed percent agreement and specific agreement coefficient. An "identity" weighting was used for the nominal shape responses; that is, agreement was only scored if Participants gave the same rating. For the ordinal data (loudness and pitch) a "linear" weighting was used, meaning that partial agreement (e.g. low and medium) were scored half the amount that total agreement (e.g. low and low) was. Again, it should be emphasised that the reported metrics are making no assumption about chance agreements, or in other words no correction is made for the reliability of random guessing. The report metrics are therefore, quite naively, ignoring the subjectivity and challenge of the task. As such, these metrics are only suitable to be compared among each other and should be interpreted with care.

Table 14: Inter-Participant agreement metrics for overall and category specific agreement

Feature	Percent Agreement	Specific Agreement			
	$\alpha_0$	I/Low/C	II/Medium/D	III/High/C-D	Plateau
Loudness	0.73	0.75	0.79	0.64	-
Shape	0.35	0.40	0	0.38	0.29
Pitch	0.67	0.55	0.65	0.72	-

#### 6.4.4 CARS discussion

Overall there was little consensus among the Participants. For the ordinal pitch and loudness scales certain heart sounds (e.g. HS 3, 4, and 8 for loudness, and HS 2, 15, and 18 for pitch) are puzzling, since as Participants disagreed at the two extremes of the

scales. Although, considering the scales have only 3 levels this is perhaps not so unexpected. Participants were in complete agreement on both the pitch and loudness levels of HS 6 although a single Participant considered the shape to be "plateau" while the other Participants reported that it was "crescendo-decrescendo". This disparity in describing the shape as "plateau" can be seen in most of the survey heart sounds, suggesting that the difficulties encountered in finding a clear cluster when modelling the "plateau" category of shape is perhaps inherent to the category. The description of "plateau" may represent a sort of "catch-all" category for changes that are too subtle to distinguish and as such contains characteristics of all the other shape categories.

The *percent agreement* and *specific agreement* metrics reported in Table 14 show what can be intuited from the stacked bar graphs: there was much less agreement among experts in regards to the shape than the other 2 categories. No two Participants agreed on the shape category of "decrescendo", although this description was only given twice in all 60 shape responses recorded and most likely represents a statistical artefact due to underrepresentation of the category. Disregarding the specific agreement of the "decrescendo" category, we can also note that the "plateau" category has a lower specific agreement than the other remaining categories. As stated before, these metrics should be interpreted with care and the small sample size of this preliminary study should not be regarded as being statistically significant (or representative of the total population).

## 6.5 Computer Assisted Cardiac Auscultation: Automated descriptions of systolic murmurs

The stage has now been set to compare the results of the proposed model to results obtained from expert auscultators. For this purpose, we have selected the heart sound which showed the most consensus among CARS Participants: heart sound number 6, a prominent example of mitral valve prolapse (MVP) from the MITHSDB. Figure 6.27 shows a five second section of this heart sound recording (perceptually weighted) along with a time-frequency representation and a summary of the psychoacoustic models. Information about the recording along with the results of both the proposed model and the expert auscultators are summarised in Table 15.

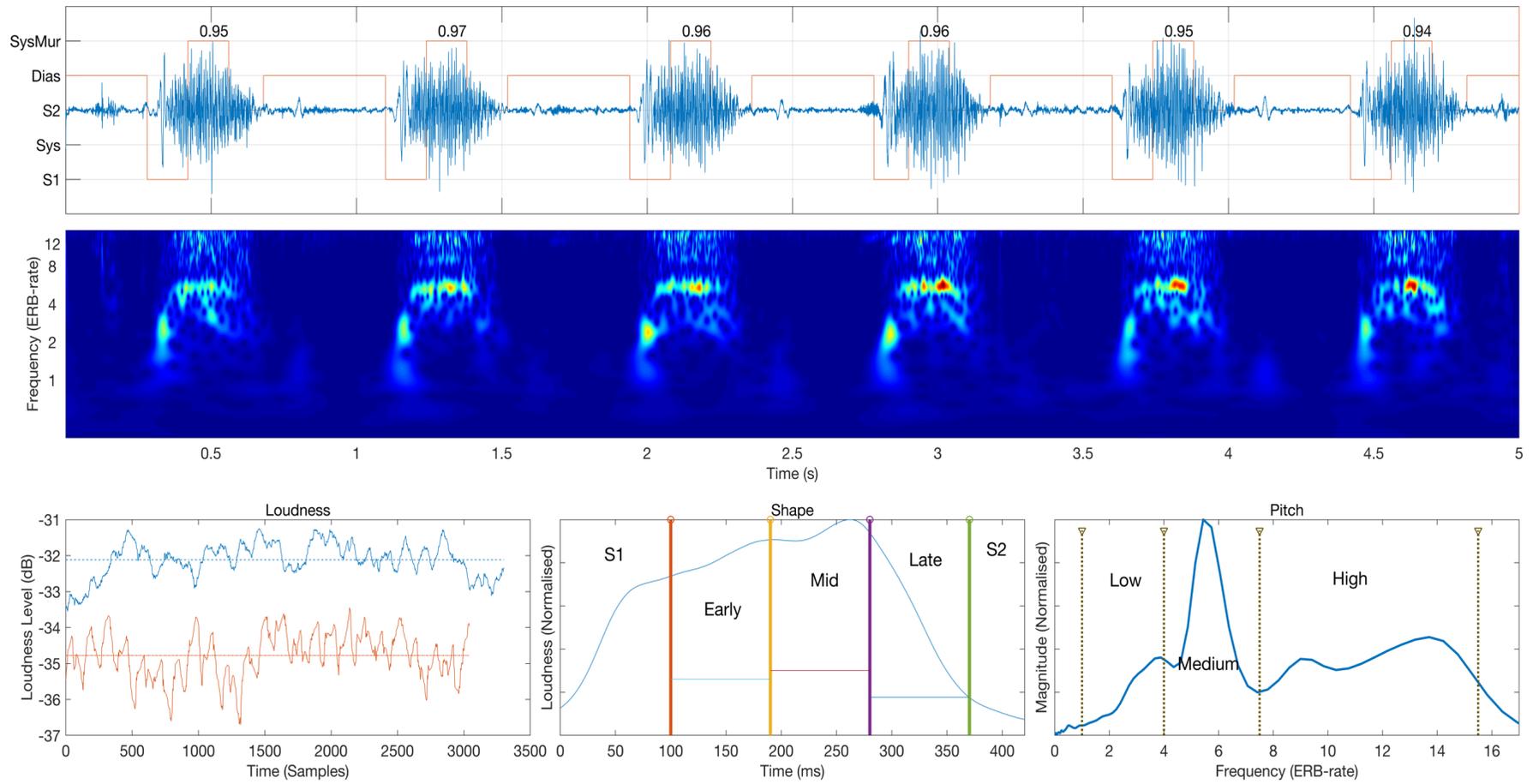


Figure 6.27: Psychoacoustically motivated model of Heart sound 6 from the CARS. Five seconds of the audio recording and the state labels (with murmur probability scores) along with a CWT representation of the same 5 seconds. A third heart sound (S3), consistent with the diagnosis of MVP, is highlighted in the CWT for the sake of interest. The bottom panels show aspects of the models for "loudness", "shape", and "pitch", respectively, for HS6.

Table 15: Summary of information available regarding HS6, including model and CARS psychoacoustic descriptions.

<b>Physionet ID</b>		<b>a0024</b>
<b>CARS ID</b>	HS6	
<b>Reported Diagnosis</b>	Mitral Valve Prolapse	
<b>Location of recording</b>	Apex	
<b>Patient Position</b>	Left Lateral Decubitus	
<b>Estimated Heart Rate</b>	72 bpm	
<b>Length of recording</b>	20.7 s	
<b>Feature Class</b>	Proposed Model results	Expert Annotations
<b>Loudness</b>	III	3x III
<b>Shape</b>	Crescendo-Decrescendo	2x Crescendo-Decrescendo, 1x Plateau
<b>Pitch</b>	Medium	3x High

The results of the proposed model are promising when compared to the expert ratings. The model results agree with the loudness rating of "III" and with the two Participants who assigned "Crescendo-decrescendo" to the shape category. There is however disagreement in the classification of the pitch of the murmurs. The model, which assigns pitch based on the peak spectral energy in specified regions (low, medium, or high), has classified the pitch as "medium" based on the prominent spectral peak occurring at approximately 5 ERB, while the expert auscultators perceived the pitch as "high" presumably based on the relatively large amount of energy present at approximately 12 ERB. Keeping in mind that the waveforms shown have already been weighted towards higher frequencies (according to ITU-R 468 specifications), this suggests that energy at higher frequencies is given preference by observers (listeners), even when comparatively lower than that at lower frequencies. It may also be seen as simply a "difference of opinion" and arguably the murmur is both medium and high pitch, having energy at a broad range of frequencies above 4 ERB (as is clearly visible in Figure 6.27).

The model adds another dimension to the CARS results making the stacked bar graphs inappropriate for visualising a comparison between the survey and model results. Instead we have opted to use pie charts divided first into halves to represent the CARS and model results, and then further divided into 3 parts each to represent the three features classes. These charts can be interpreted as uniformity in pie sections representing agreement amongst experts and symmetry between circle halves agreement between experts and the proposed model. For perfect agreement between

experts we would see uniform colours in the three LH segments; and complete agreement with the model by perfect symmetry between the LH and RH segments. Each of the heart sounds in the CARS survey is represented by a pie chart in Figure 6.28.

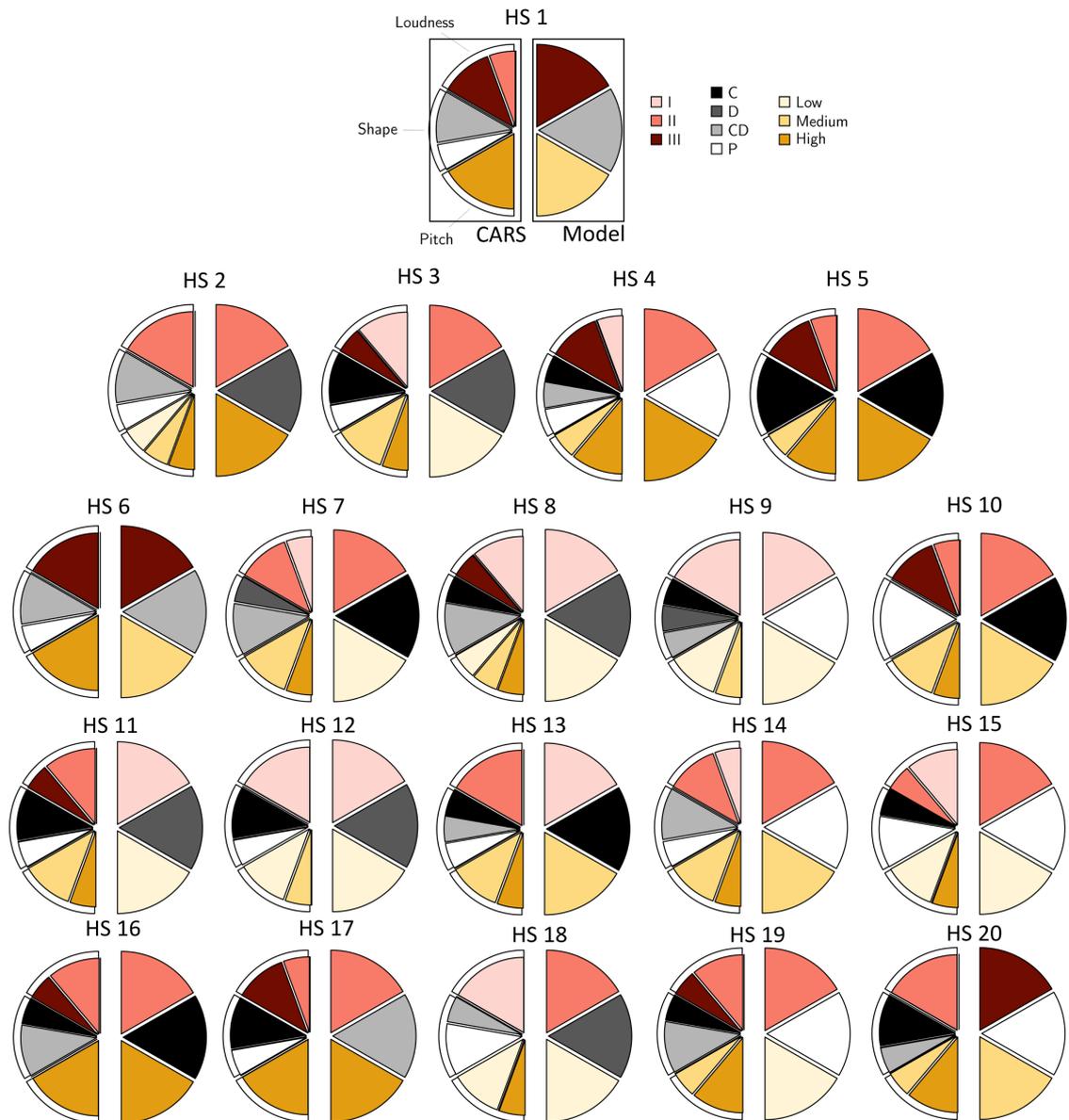


Figure 6.28: A comparison of the results obtained from expert auscultators in the CARS and the descriptions obtained by the proposed model. A uniform coloured slice in the left semi-circle indicates complete agreement between experts, this can then be compared to the findings of the model shown in the right semicircle.

The left hand semicircles of Figure 6.28 reiterates the lack of agreement between the three experts as discussed in section 6.4.4. The loudness feature appears to show the most agreement between experts and the model, although looking for instance at HS 13 and HS 18 we can see that disagreement has occurred in both directions. The shape

feature proved the most difficult for the experts to agree on and the model added another voice of dissent. The model determined the shape of six out of the 20 heart sounds to be "decrescendo", while the "decrescendo" description was only assigned two times out of the possible 60 by experts and not for the same heart sounds as the model or the other experts. The pitch feature again showed a fair amount of agreement between the model and experts and tended to score the pitch as lower than what the expert had in cases of disagreement. As discussed in the following section the amount of data used means that the results as shown in Figure 6.28 are more illustrative than conclusive and so we must take care in drawing generalised conclusions.

### 6.5.1 Limitations

The number of expert opinions we were able to collect by means of the CARS is insufficient to make meaningful inferences. This limitation relates to both the number of experts who completed the survey and the number of heart sound examples in the survey. Furthermore, the methodology used to collect expert opinions is not validated against in-person cardiac auscultation, i.e. we do not know how listening to recordings of heart sounds online compares to actual bedside auscultation. The models described in section 6.3 provide a psychoacoustic framework for deriving the murmur features, but the utility of this framework is dependent on the quality of the data used to tune its parameters. Thus, a larger and more representative dataset would enable us to create a large enough training and testing set to conclusively test and develop the proposed methodologies.

## 6.6 Conclusion

In this chapter we have explored the application of psychoacoustic principles to heart sound analysis. It is expected that by analysing heart sounds from the perspective of the human auditory system the results of the analysis will be more familiar and acceptable to an investigating physician and thus more likely to be useful in their final diagnosis. We have developed signal processing methods that approximate three classes of prevalent features that have been used when describing systolic murmurs in literature, namely "loudness", "shape", and "pitch". These methods have been developed using a variety of available signal processing tools and keeping in mind the wealth of research on human perception of sound and the models used to approximate this. We have

stopped short of the development of more involved psychoacoustic models as this would lead us down the path of researching psychoacoustics itself rather than applying the knowledge of this field to the problem of heart sound analysis. The results suggest that we can model the psychoacoustic features of systolic heart murmurs, and more data would be required to establish the degree of agreement between experts on these features.

We have also attempted to develop a methodology for the efficient collection of psychoacoustic descriptions of heart murmurs and collect preliminary data for a limited dataset of systolic murmurs. This collection presented with several technical and logistic issues, the lack of control in playback quality and the relatively small number of expert auscultators that are available to participate combined with the time-consuming nature and difficulty of the task. The results that were collected suggest that there is some agreement regarding the loudness and, to a lesser degree, the pitch of murmurs, but very little consensus as to the shape of systolic murmurs. Finally, the results of the proposed model were compared with that of the CARS showing both agreement and disagreement between the ratings given by "expert auscultators" to the model proposed in this work.

The results found in this chapter suggest that the development of a "digital expert auscultation system" that describes murmurs in the same way as human auscultators, is a viable and potentially useful enterprise. This system does not attempt to diagnose heart conditions but rather clarifies and accentuates the information in the heart sound. By weighting the frequency information in a way representative of the human auditory system before estimating features the output becomes more representative of what a person listening to the sound would experience. In this way the output of the computational models can be translated to familiar terms and concepts that practitioners can more easily integrate into their diagnoses.



### 7.1 Summary

This thesis is divided into 7 chapters, including the introduction, Chapter 1, in which the problem area was introduced and this concluding chapter in which the study is summarised, and future directions discussed. Chapters 2 introduced the theoretical background of the problem domain. Chapter 3 introduced the signal processing framework used in this work and the aim of the study. Chapters 4 to 6 represent the core contributions of this work in which we applied various signal processing and probabilistic modelling techniques to heart sound data. In this final chapter, Chapter 7, we aim to summarise and conclude this body of work; linking it back to the problem domain to establish if and how the findings of this study have altered our understanding of computer assisted cardiac auscultation.

The overall aim of this work has been the investigation and development of technologies that can add more value to the process of cardiac auscultation, specifically software algorithms that can aid in the analysis of heart sounds. We have investigated *heart sound analysis* from three main perspectives: digital audio signal processing, probabilistic modelling, and psychoacoustic modelling. We have attempted to avoid preemptively coming to conclusions about the diagnostic value of a given sound, instead trying to explain the sounds that are present in the recording. Using these perspectives, we have attempted to find methods that emulate the cardiac auscultation process and provide output similar to what you would expect from an expert auscultator. An expert auscultator that you can carry around in your pocket.

Chapter 4 introduced the paradigm of Bayesian probabilistic modelling. This provided the framework for analysis of time-frequency decompositions of heart sounds. Adapting models developed in the field of machine learning, a probabilistic ICA model employing variational inference for posterior parameter estimation was constructed and applied to time-frequency (CWT) decompositions of normal and several abnormal heart cycles. ICA demonstrated a remarkable ability to separate the CWT coefficients into groups that could be interpreted coherently considering the underlying physiology of the heart and the associated disease conditions present in the recordings.

- This research has provided evidence that probabilistic ICA can be used to decompose heart sounds into components that can be explained by underlying physiology. Notwithstanding this, other explanations of these components may also be possible.

ICA models the data as independent and identically distributed samples ignoring important correlations across time in the time-series data. This limitation was addressed in Chapter 5, in a Markovian manner, using Semi (or duration dependent) Hidden Markov Machines (HSMM). Previous work on heart sound segmentation was adapted and extended to include the identification of systolic murmurs along with the heart cycle states. The added complexity of discerning between normal and abnormal systolic states was dealt with by the development of a more sophisticated emission-probability model. Duration distribution models reported in the literature were updated with data from the heart sound recordings in the Physionet dataset. The proposed labelling model was trained and tested on subsets of the Physionet heart sound database and assessed in its ability to segment heart sounds successfully as well as to differentiate normal systolic regions from those containing murmurs. The model demonstrated a higher level of performance on the murmur testing subset of the Physionet data than previous algorithms and also achieved systolic murmur recognition at levels at or above that reported for the Physionet 2016 heart sound classification challenge, albeit these results are not directly comparable since the algorithm was not attempting to classify examples as normal/abnormal. Finally, the probabilistic output of the model of the model was demonstrated in the form of a "probability score" for each of the regions marked as *Sys/SysMur*; an important feature that increases the transparency of the algorithm and the usefulness of the results.

- Hidden semi-Markov Models have been extended to recognise systolic murmurs. This addition has been shown to improve segmentation performance, particularly in the presence of systolic murmurs.

In Chapter 6, we combined the statistical and signal processing modelling techniques developed in the previous chapters with the field of psychoacoustic modelling and auscultation knowledge and techniques. A literature review was performed in which classes of features to which auscultators pay attention were identified. These classes of

features were analysed in light of psychoacoustic models of the human auditory system in order to create models that are representative not only of the signal contents but of the perceptual stimulus elicited by these, thereby enabling the proposed algorithm to describe a heart sound in a fashion similar to an expert auscultator. Models for the "loudness", "shape", and "pitch" of systolic murmurs were developed using psychoacoustic, statistical, and audio processing techniques. The chapter also discussed the development of a "Cardiac Auscultation Research Survey" (CARS) in which a methodology of collecting expert annotations for heart murmurs was created. The results of a preliminary survey were presented and discussed. In the final part of this chapter the results of the proposed model were compared to that found in the CARS. Although the survey results and the comparison between the survey results and the model results showed considerable variation in responses amongst experts and between the algorithm and experts in some cases, certain examples had high levels of agreement. The main contributions of Chapter 6 include:

- The most important features used by expert auscultators to describe murmurs have been identified from the literature. Psychoacoustic models of segmented systolic murmurs were developed for extraction of these features and successfully demonstrated agreement with subjective observers, although these preliminary results require further validation with larger and more extensively annotated datasets.

## 7.2 Final Conclusions

As discussed in chapter 3, the main shortcomings identified in the literature relate to the lack of wide generalisability and adequate clinical validation. This lack of validation is complicated by the highly variable nature of both heart sounds and methods of auscultation and disagreement in literature about which exact categories computer assisted cardiac auscultation algorithms should attempt to place examples of heart sounds in. This work has taken the viewpoint that a single heart sound recording does not contain the information required to place a patient in a specific disease category, or even in a category of "diseased". Instead we have attempted to address the issues identified in the literature by reducing the need for extensive representative datasets that would be needed for a statistical model able to perform medical diagnosis without

input from a physician. The algorithms and systems developed in this work take the role of biomedical signal clarification or enhancement, increasing the value of the heart sounds by revealing diagnostically relevant information in the heart sound.

As summarised in the research questions, presented in section 3.4, this work was focused on modelling the cardiac auscultation process using digital signal processing and machine learning methodologies. To this end we have explored probabilistic modelling in combination with the field of psychoacoustics to arrive at computer generated descriptions of heart sounds that (in theory) align with what a human observer would report.

The first question: "What are the attributes that expert cardiologists try to ascertain during cardiac auscultation and how can these attributes be converted to mathematical models?" was directly addressed in Chapter 6. A literature review showed that 'loudness', 'shape', and 'pitch' were the most important descriptive features that physicians try to ascertain when listening to heart sounds. Computational models for these features were constructed for systolic murmurs based on approximations of the human auditory system. Agreement with subjective annotations show that these models can successfully capture the perceptual qualities of heart murmurs.

The second question "Can digital signal processing and machine learning techniques be used to extract these features from digital recordings of heart sounds?" was addressed in Chapter 4, 5, and 6. Probabilistic modelling in the guise of independent component analysis and Hidden semi-Markov Model based heart sound segmentation were proposed to this end in Chapter 4 and Chapter 5. The output of the probabilistic heart sound labelling algorithm proposed in Chapter 5, was fundamental in the development of the psychoacoustic models of systolic heart murmurs presented in Chapter 6. Signal processing and machine learning methods were successfully used to ascertain the perceptual features of 'loudness', 'pitch', and 'shape'.

The application of probabilistic ICA, explicit modelling of a systolic murmur state in heart sound segmentation, and models of psychoacoustic features shown in this work, building on the wealth of work already done and being done in the field of computer assisted cardiac auscultation, represent steps in the direction of the development of

tools and methodologies able to assist in both the practice and training of cardiac auscultation and in so doing improving the lives of those with limited access to advanced imaging technologies as well as the burden on those tasked to treat and diagnose cardiac conditions.

### 7.3 Future Work

In the 1960's science fiction series *Star Trek* the medical crew of the Starship Enterprise make use of a small handheld device called a "medical tricorder" to instantly diagnose ailments and diseases. The device uses a non-descript "scanner" to record "life-signs" about the health condition of the patient and can report on internal organ function, physiological processes, and detect infectious organisms. Although the series in which the medical tricorder first appeared is set in the 24<sup>th</sup> century, it is one of the fictional technologies portrayed in the series that is seeming evermore feasible in the 21<sup>st</sup> century. Science fiction writers however have the luxury of using "scanners" to record "life-signs". These abstractions do not have much value for scientists, engineers, and doctors trying to design, build, and use devices that improve the health outcomes and lives of their patients and customers. We must approach the multi-layered and incredibly intricate structure that is the human organism and determine where the signal in the noise lies.

The heart sounds are but one of the multitudes of bio-signals that the body produces. It is fundamental in the sense that it can be detected using our senses, but only with the invention of the stethoscope and the amplification provided by this device did it become truly useful as an aid to diagnosis. We are perhaps again at such a point in time, where technology has advanced enough to reveal further value in a bio-signal and its position within medicine. The vague abstractions of "scanner" and "life-signs" could then be an insightful aspect of the medical tricorder. We can imagine that instead of recording a single signal the device works by integrating many different sources of information obtained from arrays of passive sensors sensitive to different bio-signals, along with probing technologies that transmit and receive different types of signals into the patient's body; one such signal that seems quite obvious to monitor is the mechanical motion of the internal organs and thus even here in this fictional futuristic technology we find room for cardiac auscultation. It is important for future work to continue the

exploration of this bio-signal, the sound waves produced as matter is displaced, and its utility in cardiac assessment and diagnosis.

This work has approached computer assisted cardiac auscultation as a problem in computation but also as one in communication. It is important for the medical devices that we engineer to not only be accurate in their assessment (of course this is critical) but also be able to communicate that information in meaningful and useful ways. It is hoped that this approach has been sufficiently justified in this thesis; if not the exact methodologies then at least the attempt towards translating the outputs of computational models into information that is meaningful to the physician and to the patient being treated. From this researcher's point of view this is an integral step in developing diagnostic aids that are integrated and accepted into our cultures and able to provide real value to doctors and patients.

Less generally, potential future work related to this research should include:

- The extension of the ICA of heart sounds model using state-of-the-art research in this field. Interesting areas include the use of nonlinear ICA [193] which take the temporal structure of the independent components into account. The heart sounds have prominent time correlations and modelling of these would allow the ICA model to more effectively separate events in the heart sound based on when they occur. This could potentially result in an unsupervised heart sound segmentation algorithm.
- The extension of the proposed segmentation algorithm to include more heart states, such as diastolic murmurs or events such as opening snaps and mid-systolic clicks. The addition of more transition states to the current model, or the implementation of ensembles of segmentation models each finding specific events in a recording, would enable a more comprehensive model of the cardiovascular dynamics in the heart and surrounding great vessels.
- The extension of this work beyond cardiac auscultation. Although this work has focused exclusively on the development of methods for CACA, in principle these methods could be adapted for applications in pulmonary and gastric auscultation as well as auscultation of the great blood vessels.

- More representative psychoacoustical models, based on psychoacoustic data obtained directly from expert auscultators rather than the general statistical models used in this work. If the distribution of this psychoacoustic data is significantly different from those collected from the general population this would not only prove interesting in designing automated auscultation algorithms, but also in establishing which areas of the auditory system, e.g. pitch perception, should be trained in order to effectively auscultate. Using data collected from a specialist population would allow us to develop psychoacoustic models of heart murmurs that correspond to the auditory system of a trained auscultator, potentially increasing the diagnostic utility of these models.
- The challenge of creating large databases of both systolic and diastolic murmurs annotated by expert auscultators along with gold standard diagnoses (which should constitute consensus findings based on a variety of information including but not necessarily limited to echocardiography) would provide the most direct value to the models proposed in Chapter 6. Ideally the models should be trained on data that has shown positive outcomes for the patients involved. Instead of aiming to correlate the findings of one diagnostic test with another, future studies should focus on finding correlations between the diagnostic claims of a test and the outcomes of the patients involved.
- Development of teaching solutions that implement the technologies and data discussed in this work. These could include interactive applications that draw on advanced signal processing techniques and computational models to introduce important auscultation concepts and skills piecewise. The visualisations of heart sound data provided by the discussed techniques (e.g. time-frequency decompositions provided by CWT and isolation of interesting areas by ICA, discrimination of specific heart sound states provided by heart sound segmentation, and visual displays of loudness and frequency information) could provide learners with attractive and valuable visual information to enhance their learning.
- Clinical evaluation of the algorithms proposed in this work. Although all the models were tested on data obtained from clinical settings, the integration of the algorithms into forms suitable for embedded solutions and subsequent

widespread clinical evaluation at the point-of-care would provide a true test of their utility.

The suggestions given above are a small sample of all the possible ways in which the work presented in this thesis, and the wider area of computer assisted cardiac auscultation, can and should develop.

## Afterword

Technology has, and will continue to have, a tremendous impact on the way medicine is practiced. A 2016 research report by PricewaterhouseCoopers [194] found that 54% of 12 000 consumers surveyed across Europe, the Middle-East, and Africa were willing to engage with artificial intelligence (AI) and robotics for their healthcare needs. This figure is much higher in developing countries (94% in Nigeria and 85% in Turkey) with 69% of respondents in Nigeria saying they would be willing to undergo *major* surgery performed by a robot. The trend is clear, and most of us would perhaps be surprised to realise how ubiquitous the use of digital tools has already become in our daily lives. The unprecedented connectivity and access to information and processing power that the development of the internet has created is leading to the rapid emergence of new technologies. The responsibility of biomedical engineers is to embrace these emerging technologies and develop them in ways that lead to better outcomes for the patients and doctors using them. My hope is that this work has added value to the field and in turn leads to tools that add value to the lives of the people who use them.

It is safe to say that there remains much work to do and that for all the questions this thesis has answered it has spawned many more. But although the future may at times seem misty and opaque it is always ripe with possibilities.

*No valid plans for the future can be made by those who have no capacity for living now. – Alan Watts*

## Bibliography

- [1] I. Haney, M. Ipp, W. Feldman, and B. W. McCrindle, "Accuracy of clinical assessment of heart murmurs by office based (general practice) paediatricians," *Arch. Dis. Child.*, vol. 81, no. 5, pp. 409–412, Nov. 1999.
- [2] I. Germanakis, E. T. Petridou, G. Varlamis, I. L. Matsoukis, K. Papadopoulou-Legbelou, and M. Kalmanti, "Skills of primary healthcare physicians in paediatric cardiac auscultation," *Acta Paediatr.*, vol. 102, no. 2, pp. e74–e78, Feb. 2013, doi: 10.1111/apa.12062.
- [3] S. Mangione, L. Z. Nieman, E. Gracely, and D. Kaye, "The teaching and practice of cardiac auscultation during internal medicine and cardiology training. A nationwide survey," *Ann. Intern. Med.*, vol. 119, no. 1, pp. 47–54, Jul. 1993.
- [4] "Fact or fiction? Demand for GP appointments is driving the 'crisis' in general practice," *The Nuffield Trust*, Jan. 18, 2017. <https://www.nuffieldtrust.org.uk/news-item/fact-or-fiction-demand-for-gp-appointments-is-driving-the-crisis-in-general-practice> (accessed Feb. 23, 2017).
- [5] L. Bernstein, "Heart doctors are listening for clues to the future of their stethoscopes," *Washington Post*. [https://www.washingtonpost.com/national/health-science/heart-doctors-are-listening-for-clues-to-the-future-of-their-stethoscopes/2016/01/02/bd73b000-a98d-11e5-8058-480b572b4aae\\_story.html](https://www.washingtonpost.com/national/health-science/heart-doctors-are-listening-for-clues-to-the-future-of-their-stethoscopes/2016/01/02/bd73b000-a98d-11e5-8058-480b572b4aae_story.html) (accessed Feb. 23, 2017).
- [6] R. D. Conn and J. H. O'Keefe, "Cardiac physical diagnosis in the digital age: an important but increasingly neglected skill (from stethoscopes to microchips)," *Am. J. Cardiol.*, vol. 104, no. 4, pp. 590–595, Aug. 2009, doi: 10.1016/j.amjcard.2009.04.030.
- [7] Tavel Morton E., "Cardiac auscultation. A glorious past--but does it have a future?," *Circulation*, vol. 93, no. 6, pp. 1250–1253, Mar. 1996, doi: 10.1161/01.CIR.93.6.1250.
- [8] S. Jeffries, "Take a deep breath: the stethoscope is dying," *The Guardian*, Jan. 24, 2014.
- [9] L. Bernstein, "Has the stethoscope had its day?," *The Guardian*, Jan. 09, 2016.
- [10] S. Kaul, "VIEWS FROM THE MASTERS: Pocket ultrasound devices: time to discard the stethoscope?," *Echo Res. Pract.*, vol. 1, no. 2, pp. E7–E8, Dec. 2014, doi: 10.1530/ERP-14-0062.
- [11] M. Mehta *et al.*, "Handheld Ultrasound Versus Physical Examination in Patients Referred for Transthoracic Echocardiography for a Suspected Cardiac Condition," *JACC Cardiovasc. Imaging*, vol. 7, no. 10, pp. 983–990, Oct. 2014, doi: 10.1016/j.jcmg.2014.05.011.
- [12] W. H. Frishman, "Is the Stethoscope Becoming an Outdated Diagnostic Tool?," *Am. J. Med.*, vol. 128, no. 7, pp. 668–669, Jul. 2015, doi: 10.1016/j.amjmed.2015.01.042.
- [13] V. F. Panoulas *et al.*, "Pocket-size hand-held cardiac ultrasound as an adjunct to clinical examination in the hands of medical students and junior doctors," *Eur. Heart J. - Cardiovasc. Imaging*, vol. 14, no. 4, pp. 323–330, Apr. 2013, doi: 10.1093/ehjci/jes140.
- [14] V. Fuster, "The Stethoscope's Prognosis Very Much Alive and Very Necessary," *J. Am. Coll. Cardiol.*, vol. 67, no. 9, pp. 1118–1119, Mar. 2016, doi: 10.1016/j.jacc.2016.01.005.

- [15] W. R. Thompson, "In defence of auscultation: a glorious future?," *Heart Asia*, vol. 9, no. 1, pp. 44–47, Feb. 2017, doi: 10.1136/heartasia-2016-010796.
- [16] E. R. Edelman and B. N. Weber, "Tenuous Tether," *N. Engl. J. Med.*, vol. 373, no. 23, pp. 2199–2201, Dec. 2015, doi: 10.1056/NEJMp1509625.
- [17] M. E. Tavel, "Cardiac auscultation: a glorious past--and it does have a future!," *Circulation*, vol. 113, no. 9, pp. 1255–1259, Mar. 2006, doi: 10.1161/CIRCULATIONAHA.105.591149.
- [18] M. E. Legget, M. Toh, A. Meintjes, S. Fitzsimons, G. Gamble, and R. N. Doughty, "Digital devices for teaching cardiac auscultation - a randomized pilot study," *Med. Educ. Online*, vol. 23, no. 1, p. 1524688, Jan. 2018, doi: 10.1080/10872981.2018.1524688.
- [19] M. J. Barrett, A. S. Mackie, and J. P. Finley, "Cardiac Auscultation in the Modern Era: Premature Requiem or Phoenix Rising?," *Cardiol. Rev.*, vol. 25, no. 5, pp. 205–210, Oct. 2017, doi: 10.1097/CRD.000000000000145.
- [20] J. P. Finley, R. Caissie, P. Nicol, and B. Hoyt, "International trial of online auditory training programme for distinguishing innocent and pathological murmurs," *J. Paediatr. Child Health*, vol. 51, no. 8, pp. 815–819, Aug. 2015, doi: 10.1111/jpc.12839.
- [21] H. Dalen, B. O. Haugen, and T. Graven, "Feasibility and clinical implementation of hand-held echocardiography," *Expert Rev. Cardiovasc. Ther.*, vol. 11, no. 1, pp. 49–54, Jan. 2013, doi: 10.1586/erc.12.165.
- [22] S. D. Solomon and F. Saldana, "Point-of-Care Ultrasound in Medical Education — Stop Listening and Look," *N. Engl. J. Med.*, vol. 370, no. 12, pp. 1083–1085, Mar. 2014, doi: 10.1056/NEJMp1311944.
- [23] "Medical Radiation Technologist - How to enter the job." <http://www.careers.govt.nz/jobs-database/health-and-community/health/medical-radiation-technologist/how-to-enter-the-job> (accessed Aug. 23, 2016).
- [24] John Hopkins Medicine, "Valvular Heart Disease." [http://www.hopkinsmedicine.org/heart\\_vascular\\_institute/conditions\\_treatments/conditions/valvular\\_heart\\_disease.html](http://www.hopkinsmedicine.org/heart_vascular_institute/conditions_treatments/conditions/valvular_heart_disease.html) (accessed Aug. 23, 2016).
- [25] S. Leng, R. S. Tan, K. T. C. Chai, C. Wang, D. Ghista, and L. Zhong, "The electronic stethoscope," *Biomed. Eng. OnLine*, vol. 14, Jul. 2015, doi: 10.1186/s12938-015-0056-y.
- [26] P. C. Benias *et al.*, "Structure and Distribution of an Unrecognized Interstitium in Human Tissues," *Sci. Rep.*, vol. 8, no. 1, p. 4947, Mar. 2018, doi: 10.1038/s41598-018-23062-6.
- [27] Wapcaplet, *Diagram of the human heart, created by Wapcaplet in Sodipodi. Cropped by Yaddah to remove white space (this cropping is not the same as Wapcaplet's original crop)*. 2006.
- [28] OpenStax College, *Heart Valves*. 2011.
- [29] "Types of Valve Disease | Cleveland Clinic." <http://my.clevelandclinic.org/services/heart/disorders/heart-valve-disease/valve-disease-types> (accessed Aug. 26, 2016).
- [30] "Heart Valve Problems," *Patient Education Center*. <http://www.patienteducationcenter.org/articles/heart-valve-problems/> (accessed Aug. 26, 2016).

- [31] Heart Foundation NZ, "Curious what a heart murmur means?," *Heart Foundation NZ*. <https://www.heartfoundation.org.nz/your-heart/heart-conditions/heart-valve-disease> (accessed Jun. 13, 2019).
- [32] R. E. Schneider, *Anatomy & physiology revealed*. McGraw-Hill Higher Education, 2007.
- [33] "Univeristy of Michigan Medical School | Professional Skill Builder." [http://www.med.umich.edu/lrc/psb\\_open/html/menu/index.html](http://www.med.umich.edu/lrc/psb_open/html/menu/index.html) (accessed May 02, 2017).
- [34] M. E. Silverman, "The Third Heart Sound," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Boston: Butterworths, 1990.
- [35] E. S. Williams, "The Fourth Heart Sound," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Boston: Butterworths, 1990.
- [36] M. A. Alpert, "Systolic Murmurs," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Boston: Butterworths, 1990.
- [37] H. K. Walker, W. D. Hall, and J. W. Hurst, Eds., *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed. Boston: Butterworths, 1990.
- [38] Madhero88, *English: Phonocardiograms from normal and abnormal heart sounds*. [object HTMLTableCellElement].
- [39] R. Hajar, "The Art of Listening," *Heart Views Off. J. Gulf Heart Assoc.*, vol. 13, no. 1, pp. 24–25, 2012, doi: 10.4103/1995-705X.96668.
- [40] I. R. Hanna and M. E. Silverman, "A history of cardiac auscultation and some of its contributors," *Am. J. Cardiol.*, vol. 90, no. 3, pp. 259–267, Aug. 2002.
- [41] A. C. Guyton and J. E. Hall, "Heart Valves and Heart Sounds," in *Textbook of medical physiology*, Philadelphia: Elsevier Saunders, 2006, pp. 269–271.
- [42] "Heart murmurs and other sounds: MedlinePlus Medical Encyclopedia." <https://www.nlm.nih.gov/medlineplus/ency/article/003266.htm> (accessed Mar. 22, 2016).
- [43] L. S. KING, "THE MEANING OF MEDICAL DIAGNOSIS," *ETC Rev. Gen. Semant.*, vol. 8, no. 3, pp. 202–211, 1951.
- [44] R. S. Ledley and L. B. Lusted, "Reasoning Foundations of Medical Diagnosis," *Science*, vol. 130, no. 3366, pp. 9–21, 1959.
- [45] C. Spiers, "Cardiac auscultation.," *Br. J. Card. Nurs.*, vol. 6, no. 10, pp. 482-486 5p, Oct. 2011.
- [46] OpenStax College, *Illustration from Anatomy & Physiology, Connexions Web site*. <http://cnx.org/content/col11496/1.6/>, Jun 19, 2013. [object HTMLTableCellElement].
- [47] C. Spiers, "Advanced auscultation: aortic stenosis and mitral regurgitation.," *Br. J. Card. Nurs.*, vol. 7, no. 9, pp. 415-418 4p, Sep. 2012.
- [48] M. A. Chizner, "Cardiac auscultation: rediscovering the lost art," *Curr. Probl. Cardiol.*, vol. 33, no. 7, pp. 326–408, Jul. 2008, doi: 10.1016/j.cpcardiol.2008.03.003.
- [49] "CV Physiology: Valsalva." <http://www.cvphysiology.com/Hemodynamics/H014.htm> (accessed Sep. 07, 2016).

- [50] M. J. Barrett, C. S. Lacey, A. E. Sekara, E. A. Linden, and E. J. Gracely, "Mastering cardiac murmurs: the power of repetition," *Chest*, vol. 126, no. 2, pp. 470–5, Aug. 2004, doi: 10.1378/chest.126.2.470.
- [51] Science Museum London, *Laennec's Stethoscope*. .
- [52] American Diagnostic Corporation, "Adscope 609." <https://www.adctoday.com/products/609> (accessed Jun. 13, 2019).
- [53] T. Loubani, "Glia Free Medical hardware," *GliaX Repo*. <https://github.com/GliaX> (accessed Aug. 15, 2016).
- [54] C. Ralph and R. E. Irons, "Acoustic and electronic stethoscope," US3247324A, Apr. 19, 1966.
- [55] M.-C. Grenier, K. Gagnon, J. Genest, J. Durand, and L.-G. Durand, "Clinical Comparison of Acoustic and Electronic Stethoscopes and Design of a New Electronic Stethoscope," *Am. J. Cardiol.*, vol. 81, no. 5, pp. 653–656, Mar. 1998, doi: 10.1016/S0002-9149(97)00977-6.
- [56] H. Høyte, T. Jensen, and K. Gjesdal, "Cardiac auscultation training of medical students: a comparison of electronic sensor-based and acoustic stethoscopes," *BMC Med. Educ.*, vol. 5, p. 14, May 2005, doi: 10.1186/1472-6920-5-14.
- [57] E. Kalinauskienė, H. Razvadauskas, D. J. Morse, G. E. Maxey, and A. Naudžiūnas, "A Comparison of Electronic and Traditional Stethoscopes in the Heart Auscultation of Obese Patients," *Medicina (Mex.)*, vol. 55, no. 4, Apr. 2019, doi: 10.3390/medicina55040094.
- [58] A. Lerch, *An Introduction to Audio Content Analysis*. Hoboken, NJ: Wiley, 2012.
- [59] G. E. Pfander, *Sampling Theory, a Renaissance : Compressive Sensing and Other Developments*. Cham: Birkhäuser, 2015.
- [60] A. J. B. J. Butterfield and J. S. Szymanski, "Nyquist rate," in *A Dictionary of Electronics and Electrical Engineering*, A. J. Butterfield and J. Szymanski, Eds. Oxford University Press, 2018.
- [61] M. E. Silverman, P. R. Fleming, A. Hollman, D. G. Julian, and D. M. Krikler, *British Cardiology in the 20th Century*. Springer Science & Business Media, 2012.
- [62] B. A. Sheno, *Introduction to digital signal processing and filter design*. Wiley, 2005.
- [63] C. Ahlström, "Processing of the phonocardiographic signal: methods for the intelligent stethoscope," Department of Biomedical Engineering, Linköpings universitet, Linköping, 2006.
- [64] O. Poisson, P. Rioual, and M. Meunier, "New signal processing tools applied to power quality analysis," *IEEE Trans. Power Deliv.*, vol. 14, no. 2, pp. 561–566, Apr. 1999, doi: 10.1109/61.754104.
- [65] P. S. Addison, *The illustrated wavelet transform handbook : introductory theory and applications in science, engineering, medicine and finance*. Bristol : Institute of Physics Publishing, [2002], 2002.
- [66] D. Gill, N. Gavrieli, and N. Intrator, "Detection and identification of heart sounds using homomorphic envelopogram and self-organizing probabilistic model," 2005, pp. 957–960, doi: 10.1109/CIC.2005.1588267.
- [67] I. Rezek and S. Roberts, *Envelope Extraction via Complex Homomorphic Filtering*. 1998.
- [68] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, Apr. 2010, doi: 10.1088/0967-3334/31/4/004.

- [69] A. L. Goldberger *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.
- [70] D. Springer, L. Tarassenko, and G. Clifford, “Logistic Regression-HSMM-based Heart Sound Segmentation,” *IEEE Trans. Biomed. Eng.*, pp. 1–1, 2015, doi: 10.1109/TBME.2015.2475278.
- [71] D. S. Gerbarg, F. W. Holcomb, J. J. Hofler, C. E. Bading, G. L. Schultz, and R. E. Sears, “Analysis of phonocardiogram by a digital computer,” *Circ. Res.*, vol. 11, pp. 569–576, Sep. 1962.
- [72] F. de Lima Hedayioglu, M. T. Coimbra, and S. da Silva Mattos, “A Survey of Audio Processing Algorithms for Digital Stethoscopes,” in *HEALTHINF*, 2009, pp. 425–429, Accessed: Feb. 05, 2015. [Online]. Available: [http://www.dcc.fc.up.pt/~fheday/Papers/A\\_SURVEY\\_OF\\_AUDIO\\_PROCESSING\\_ALGORITHMS\\_FOR\\_DIGITAL\\_STETHOSCOPIES.pdf](http://www.dcc.fc.up.pt/~fheday/Papers/A_SURVEY_OF_AUDIO_PROCESSING_ALGORITHMS_FOR_DIGITAL_STETHOSCOPIES.pdf).
- [73] T. H. Falk and Wai-Yip Chan, “Modulation filtering for heart and lung sound separation from breath sound recordings,” Aug. 2008, pp. 1859–1862, doi: 10.1109/IEMBS.2008.4649547.
- [74] J. P. Ramos, P. Carvalho, R. P. Paiva, and J. Henriques, “Modulation filtering for noise detection in heart sound signals,” in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 6013–6016, Accessed: Feb. 05, 2015. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6091486](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6091486).
- [75] D. Kumar, P. Carvalho, M. Antunes, R. P. Paiva, and J. Henriques, “Noise detection during heart sound recording using periodicity signatures,” *Physiol. Meas.*, vol. 32, no. 5, pp. 599–618, May 2011, doi: 10.1088/0967-3334/32/5/008.
- [76] Z. Syed, D. Leeds, D. Curtis, F. Nesta, R. A. Levine, and J. Guttag, “A Framework for the Analysis of Acoustical Cardiac Signals,” *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 651–662, Apr. 2007, doi: 10.1109/TBME.2006.889189.
- [77] Z. Syed and J. Guttag, “Prototypical Biological Signals,” 2007, pp. I-397–I-400, doi: 10.1109/ICASSP.2007.366700.
- [78] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sörnmo, R. Sepponen, and E. Pesonen, “Computer-based detection and analysis of heart sound and murmur,” *Ann. Biomed. Eng.*, vol. 33, no. 7, pp. 937–942, 2005.
- [79] C. Ahlstrom, T. Länne, P. Ask, and A. Johansson, “A method for accurate localization of the first heart sound and possible applications,” *Physiol. Meas.*, vol. 29, no. 3, pp. 417–428, Mar. 2008, doi: 10.1088/0967-3334/29/3/011.
- [80] H. Liang, S. Lukkarinen, and I. Hartimo, “Heart sound segmentation algorithm based on heart sound envelopogram,” in *Computers in Cardiology 1997*, 1997, pp. 105–108, Accessed: Jan. 07, 2015. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=647841](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=647841).
- [81] C. Barabasa, M. Jafari, and M. D. Plumbley, “A robust method for S1/S2 heart sounds detection without ecg reference based on music beat tracking,” in *Electronics and Telecommunications (ISETC), 2012 10th International Symposium on*, 2012, pp. 307–310, Accessed: Jan. 07, 2015. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6408110](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6408110).
- [82] “Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016.” <https://www.physionet.org/challenge/2016/> (accessed Sep. 05, 2016).

- [83] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, Apr. 2010, doi: 10.1088/0967-3334/31/4/004.
- [84] F. Schoonjans, "Logistic regression," *MedCalc*. [https://www.medcalc.org/manual/logistic\\_regression.php](https://www.medcalc.org/manual/logistic_regression.php) (accessed Sep. 06, 2016).
- [85] F. A. Sonnenberg and J. R. Beck, "Markov models in medical decision making: a practical guide," *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.*, vol. 13, no. 4, pp. 322–338, Dec. 1993.
- [86] S. E. Schmidt, E. Toft, C. Holst-Hansen, C. Graff, and J. J. Struijk, "Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov Model," Sep. 2008, pp. 345–348, doi: 10.1109/CIC.2008.4749049.
- [87] S.-W. Deng and J.-Q. Han, "Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps," *Future Gener. Comput. Syst.*, vol. 60, pp. 13–21, Jul. 2016, doi: 10.1016/j.future.2016.01.010.
- [88] S. Yuenyong, A. Nishihara, W. Kongprawechnon, and K. Tungpimolrut, "A framework for automatic heart sound analysis without segmentation," *Biomed. Eng. Online*, vol. 10, no. 13, pp. 10–13, 2011.
- [89] G. E. Güraksin and H. Uguz, "Classification of heart sounds based on the least squares support vector machine," 2011, Accessed: Feb. 04, 2015. [Online]. Available: <http://www.ijcic.org/ijcic-10-08003.pdf>.
- [90] C. N. Gupta, R. Palaniappan, S. Swaminathan, and S. M. Krishnan, "Neural network classification of homomorphic segmented heart sounds," *Appl. Soft Comput.*, vol. 7, no. 1, pp. 286–297, Jan. 2007, doi: 10.1016/j.asoc.2005.06.006.
- [91] C. Ahlstrom *et al.*, "Feature Extraction for Systolic Heart Murmur Classification," *Ann. Biomed. Eng.*, vol. 34, no. 11, pp. 1666–1677, Nov. 2006, doi: 10.1007/s10439-006-9187-4.
- [92] T. Ölmez and Z. Dokur, "Classification of heart sounds using an artificial neural network," *Pattern Recognit. Lett.*, vol. 24, no. 1–3, pp. 617–629, Jan. 2003, doi: 10.1016/S0167-8655(02)00281-7.
- [93] A. Gelman, "Objections to Bayesian statistics," *Bayesian Anal.*, vol. 3, no. 3, pp. 445–449, Sep. 2008, doi: 10.1214/08-BA318.
- [94] C. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006.
- [95] S. E. Page, "Why 'Many-Model Thinkers' Make Better Decisions," *Harv. Bus. Rev. Digit. Artic.*, pp. 48–54, Nov. 2018.
- [96] R. Choudrey, "Variational Methods for Bayesian Independent Component Analysis," University of Oxford, 2002.
- [97] P. Comon, C. Jutten, and J. Herault, "Blind separation of sources, part II: Problems statement," *Signal Process.*, vol. 24, no. 1, pp. 11–20, Jul. 1991, doi: 10.1016/0165-1684(91)90080-3.
- [98] P. Comon, "Independent component analysis, A new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994, doi: 10.1016/0165-1684(94)90029-9.
- [99] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Netw.*, vol. 13, no. 4, pp. 411–430, Jun. 2000, doi: 10.1016/S0893-6080(00)00026-5.
- [100] J.-F. Cardoso, "High-Order Contrasts for Independent Component Analysis," *Neural Comput.*, vol. 11, pp. 157–192, 1999, doi: 10.1162/089976699300016863.

- [101] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995, doi: 10.1162/neco.1995.7.6.1129.
- [102] N. D. Lawrence and C. M. Bishop, "Variational Bayesian Independent Component Analysis," 1999.
- [103] N. D. Lawrence and C. M. Bishop, "Variational Bayesian Independent Component Analysis," *Univ. Camb. Tech Rep.*, p. 11, 2000.
- [104] P. A. d F. R. Højen-Sørensen, O. Winther, and L. K. Hansen, "Mean-field approaches to independent component analysis," *Neural Comput.*, vol. 14, no. 4, pp. 889–918, Apr. 2002, doi: 10.1162/089976602317319009.
- [105] O. Winther and K. B. Petersen, "Bayesian independent component analysis: Variational methods and non-negative decompositions," *Digit. Signal Process.*, vol. 17, no. 5, pp. 858–872, Sep. 2007, doi: 10.1016/j.dsp.2007.01.003.
- [106] H. J. Park and T. W. Lee, "Modeling Nonlinear Dependencies in Natural Images using Mixture of Laplacian Distribution," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. MIT Press, 2005, pp. 1041–1048.
- [107] "3M™ Littmann® Electronic Stethoscope Model 3200." [https://www.3mnz.co.nz/3M/en\\_NZ/company-nz/all-3m-products/~3M-Littmann-Electronic-Stethoscope-Model-3200/?N=5002385+3293188392&rt=rud](https://www.3mnz.co.nz/3M/en_NZ/company-nz/all-3m-products/~3M-Littmann-Electronic-Stethoscope-Model-3200/?N=5002385+3293188392&rt=rud) (accessed Jul. 09, 2019).
- [108] S. E. Schmidt, E. Toft, C. Holst-Hansen, C. Graff, and J. J. Struijk, "Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov Model," Sep. 2008, pp. 345–348, doi: 10.1109/CIC.2008.4749049.
- [109] S. C. Olhede and A. T. Walden, "Generalized Morse wavelets," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2661–2670, Nov. 2002, doi: 10.1109/TSP.2002.804066.
- [110] J. M. Lilly and S. C. Olhede, "Higher-Order Properties of Analytic Wavelets," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 146–160, Jan. 2009, doi: 10.1109/TSP.2008.2007607.
- [111] D. Springer, L. Tarassenko, and G. Clifford, "Logistic Regression-HSMM-based Heart Sound Segmentation," *IEEE Trans. Biomed. Eng.*, pp. 1–1, 2015, doi: 10.1109/TBME.2015.2475278.
- [112] J. M. Felner, "The Second Heart Sound," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Boston: Butterworths, 1990.
- [113] J. M. Felner, "The First Heart Sound," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Boston: Butterworths, 1990.
- [114] T. E. Meyer, "Auscultation of cardiac murmurs in adults," in *UpToDate*, W. Zoghbi and S. B. Yeon, Eds. Waltham, MA: Wolters Kluwer, 2019.
- [115] T. Biancianiello, "Innocent Murmurs," *Circulation*, vol. 111, no. 3, pp. e20–e22, Jan. 2005, doi: 10.1161/01.CIR.0000153388.41229.CB.
- [116] I. S. Crawley, "Diastolic Murmurs," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Boston: Butterworths, 1990.
- [117] T. E. Meyer, "Auscultation of heart sounds," in *UpToDate*, S. B. Yeon, Ed. Waltham, MA: Wolters Kluwer, 2019.

- [118] "PhysioToolkit Software Index." <https://physionet.org/physiotools/software-index.shtml> (accessed Feb. 19, 2019).
- [119] H. Ali, T. J. Ahmed, and K. Shoab, "Heart Sound Signal Modeling and Segmentation based on Improved Shannon Energy Envelopogram using Adaptive Windows," in *Asialink International Conference on Biomedical Engineering & Technology, 2007*, Accessed: Feb. 04, 2015. [Online]. Available: [http://www.utdallas.edu/~hxa098020/publications/AsiaLinkConference\\_Jakarta\\_2007.pdf](http://www.utdallas.edu/~hxa098020/publications/AsiaLinkConference_Jakarta_2007.pdf).
- [120] S. Sanei, M. Ghodsi, and H. Hassani, "An adaptive singular spectrum analysis approach to murmur detection from heart sounds," *Med. Eng. Phys.*, vol. 33, no. 3, pp. 362–367, Apr. 2011, doi: 10.1016/j.medengphy.2010.11.004.
- [121] D. Kumar, P. Carvalho, M. Antunes, R. P. Paiva, and J. Henriques, "An adaptive approach to abnormal heart sound segmentation," May 2011, pp. 661–664, doi: 10.1109/ICASSP.2011.5946490.
- [122] V. Nigam and R. Priemer, "Accessing heart dynamics to estimate durations of heart sounds," *Physiol. Meas.*, vol. 26, no. 6, pp. 1005–1018, Dec. 2005, doi: 10.1088/0967-3334/26/6/010.
- [123] L. G. Gamero and R. Watrous, "Detection of the first and second heart sound using probabilistic models," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, Sep. 2003, vol. 3, pp. 2877-2880 Vol.3, doi: 10.1109/IEMBS.2003.1280519.
- [124] A. D. Rieke, R. J. Povinelli, and M. T. Johnson, "Automatic segmentation of heart sound signals using hidden markov models," in *Computers in Cardiology, 2005*, Sep. 2005, pp. 953–956, doi: 10.1109/CIC.2005.1588266.
- [125] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989, doi: 10.1109/5.18626.
- [126] S.-Z. Yu, "Hidden semi-Markov models," *Artif. Intell.*, vol. 174, no. 2, pp. 215–243, Feb. 2010, doi: 10.1016/j.artint.2009.11.011.
- [127] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967, doi: 10.1109/TIT.1967.1054010.
- [128] C. Liu *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, Dec. 2016, doi: 10.1088/0967-3334/37/12/2181.
- [129] N. Atanasov and T. Ning, "Isolation of Systolic Heart Murmurs Using Wavelet Transform and Energy Index," 2008, pp. 216–220, doi: 10.1109/CISP.2008.758.
- [130] C. Bunluechokchai and W. Ussawawongaraya, "A Wavelet-based Factor for Classification of Heart Sounds with Mitral Regurgitation," *Int. J. Appl. Biomed. Eng.*, vol. 2, no. 1, pp. 44–48, 2009.
- [131] A. Castro, T. T. Vinhoza, S. S. Mattos, and M. T. Coimbra, "Heart sound segmentation of pediatric auscultations using wavelet analysis," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, 2013*, pp. 3909–3912, Accessed: Jan. 15, 2015. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6610399](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6610399).
- [132] C. C. Balili, M. C. C. Sobrepena, and P. C. Naval, "Classification of heart sounds using discrete and continuous wavelet transform and random forests," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Nov. 2015, pp. 655–659, doi: 10.1109/ACPR.2015.7486584.

- [133] Jung Jun Lee, Sang Min Lee, In Young Kim, Hong Ki Min, and Seung Hong Hong, "Comparison between short time Fourier and wavelet transform for feature extraction of heart sound," 1999, vol. 2, pp. 1547–1550, doi: 10.1109/TENCON.1999.818731.
- [134] W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and partial least squares regression," *Biomed. Signal Process. Control*, vol. 32, pp. 20–28, Feb. 2017, doi: 10.1016/j.bspc.2016.10.004.
- [135] P. Rakovic, E. Sejdic, L. J. Stankovic, and J. Jiang, "Time-frequency signal processing approaches with applications to heart sound analysis," *Comput. Cardiol.*, vol. 33, pp. 197–200, 2006.
- [136] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*. Cambridge ; New York : Cambridge University Press, 2000., 2000.
- [137] "Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016." <https://www.physionet.org/challenge/2016/> (accessed Jan. 23, 2018).
- [138] G. Clifford *et al.*, "Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016," presented at the 2016 Computing in Cardiology Conference, Sep. 2016, doi: 10.22489/CinC.2016.179-154.
- [139] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, and R. Getz, "Classifying Heart Sounds Challenge," 2012. <http://www.peterjbentley.com/heartchallenge/#background> (accessed Feb. 26, 2019).
- [140] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of Feature:based and Deep learning:based Classifiers for Detection of Abnormal Heart Sounds," presented at the 2016 Computing in Cardiology Conference, Sep. 2016, doi: 10.22489/CinC.2016.182-399.
- [141] M. Zabihi, A. Bahrami Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, "Heart Sound Anomaly and Quality Detection using Ensemble of Neural Networks without Segmentation," presented at the 2016 Computing in Cardiology Conference, Sep. 2016, doi: 10.22489/CinC.2016.180-213.
- [142] E. Kay and A. Agarwal, "DropConnected neural networks trained on time-frequency and inter-beat features for classifying heart sounds," *Physiol. Meas.*, vol. 38, no. 8, pp. 1645–1657, Jul. 2017, doi: 10.1088/1361-6579/aa6a3d.
- [143] I. J. D. Bobillo, "A tensor approach to heart sound classification," in *2016 Computing in Cardiology Conference (CinC)*, Sep. 2016, pp. 629–632, doi: 10.23919/CIC.2016.7868821.
- [144] M. Nabhan Homsy *et al.*, "Automatic Heart Sound Recording Classification using a Nested Set of Ensemble Algorithms," presented at the 2016 Computing in Cardiology Conference, Sep. 2016, doi: 10.22489/CinC.2016.237-325.
- [145] S. A. Pavlopoulos, A. C. Stasis, and E. N. Loukis, "A decision tree – based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds," *Biomed. Eng. OnLine*, vol. 3, p. 21, 2004, doi: 10.1186/1475-925X-3-21.
- [146] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers," *J Mach Learn Res*, vol. 1, pp. 113–141, Sep. 2001, doi: 10.1162/15324430152733133.

- [147] “Random decision forests,” in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Aug. 1995, vol. 1, pp. 278–282 vol.1, doi: 10.1109/ICDAR.1995.598994.
- [148] H. He and E. A. Garcia, “Learning from Imbalanced Data,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.
- [149] M. Sokolova, N. Japkowicz, and S. Szpakowicz, *Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation*, vol. Vol. 4304. 2006.
- [150] S. Boughorbel, F. Jarray, and M. El-Anbari, “Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric,” *PLoS ONE*, vol. 12, no. 6, Jun. 2017, doi: 10.1371/journal.pone.0177678.
- [151] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, USA, 2012*, pp. 1097–1105, [Online]. Available: <http://dl.acm.org/citation.cfm?id=2999134.2999257>.
- [152] F. Renna, J. H. Oliveira, and M. T. Coimbra, “Deep Convolutional Neural Networks for Heart Sound Segmentation,” *IEEE J. Biomed. Health Inform.*, Jan. 2019, doi: 10.1109/JBHI.2019.2894222.
- [153] A. Meintjes, A. Lowe, and M. Legget, “Fundamental Heart Sound Classification using the Continuous Wavelet Transform and Convolutional Neural Networks,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2018, pp. 409–412, 2018, doi: 10.1109/EMBC.2018.8512284.
- [154] C. Liu, D. Springer, and G. D. Clifford, “Performance of an open-source heart sound segmentation algorithm on eight independent databases,” *Physiol. Meas.*, vol. 38, no. 8, pp. 1730–1745, Aug. 2017, doi: 10.1088/1361-6579/aa6e9f.
- [155] M. Florentine, A. N. Popper, and R. R. Fay, *Loudness*. New York: Springer, 2011.
- [156] H. ( 1 ) Fastl and E. ( 2 ) Zwicker, *Psychoacoustics: Facts and models*. Springer Berlin Heidelberg, 2007.
- [157] K. K. Patil, B. S. Nagabhushan, and B. P. Vijay Kumar, “Psychoacoustic Models for Heart Sounds,” in *Advances in Networks and Communications*, vol. 132, N. Meghanathan, B. K. Kaushik, and D. Nagamalai, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 556–563.
- [158] I.-S. Lee, H. Preissl, and P. Enck, “How to Perform and Interpret Functional Magnetic Resonance Imaging Studies in Functional Gastrointestinal Disorders,” *J. Neurogastroenterol. Motil.*, vol. 23, no. 2, pp. 197–207, Apr. 2017, doi: 10.5056/jnm16196.
- [159] A. Eklund, T. E. Nichols, and H. Knutsson, “Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 28, pp. 7900–7905, 12 2016, doi: 10.1073/pnas.1602413113.
- [160] “Much of what we know about the brain may be wrong: The problem with fMRI,” *ideas.ted.com*, Aug. 30, 2016. <https://ideas.ted.com/much-of-what-we-know-about-the-brain-may-be-wrong-the-problem-with-fmri/> (accessed Feb. 06, 2019).
- [161] L. Chittka and A. Brockmann, “Perception Space—The Final Frontier,” *PLOS Biol.*, vol. 3, no. 4, p. e137, Apr. 2005, doi: 10.1371/journal.pbio.0030137.
- [162] OpenStax, “Anatomy and Physiology,” May 03, 2019. <http://cnx.org/contents/14fb4ad7-39a1-4eee-ab6e-3ef2482e3e22@15.2>.

- [163] The Open University, "Hearing," *OpenLearn*, 2019. <https://www.open.edu/openlearn/science-maths-technology/science/biology/hearing/content-section-0> (accessed May 26, 2019).
- [164] "Creative Commons — Attribution-NonCommercial-ShareAlike 4.0 International — CC BY-NC-SA 4.0." [https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en\\_GB](https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en_GB) (accessed May 26, 2019).
- [165] OpenStax, *1406 Cochlea*. 2016.
- [166] OpenStax, *1408 Frequency Coding in The Cochlea*. 2016.
- [167] B. C. J. Moore, "Development and Current Status of the 'Cambridge' Loudness Models," *Trends Hear.*, vol. 18, Sep. 2014, doi: 10.1177/2331216514550620.
- [168] B. C. J. Moore and B. R. Glasberg, "A revision of Zwicker's loudness model," *Acust. United Acta Acust. J. Eur. Acoust. Assoc.*, vol. 82, no. 2, pp. 335–345, 1996.
- [169] N. A. Cyphers, C. G. Mest, and M. E. Doyle-Tadduni, "Effect of Psychoacoustic Learning on Cardiac Auscultation Proficiency in Nurse Practitioner Students:," *Nurse Educ.*, vol. 44, no. 2, pp. 79–83, 2019, doi: 10.1097/NNE.0000000000000585.
- [170] K. Tremblay, "Training-Related Changes in the Brain: Evidence from Human Auditory-Evoked Potentials," 2007, doi: 10.1055/s-2007-973438.
- [171] C. Micheyl, K. Delhommeau, X. Perrot, and A. J. Oxenham, "Influence of musical and psychoacoustical training on pitch discrimination," *Hear. Res.*, vol. 219, no. 1–2, pp. 36–47, Sep. 2006, doi: 10.1016/j.heares.2006.05.004.
- [172] K. L. Tremblay, A. J. Shahin, T. Picton, and B. Ross, "Auditory training alters the physiological detection of stimulus-specific cues in humans," *Clin. Neurophysiol.*, vol. 120, no. 1, pp. 128–135, Jan. 2009, doi: 10.1016/j.clinph.2008.10.005.
- [173] M. J. Barrett, C. S. Lacey, A. E. Sekara, E. A. Linden, and E. J. Gracely, "Mastering cardiac murmurs: the power of repetition," *Chest*, vol. 126, no. 2, pp. 470–475, Aug. 2004, doi: 10.1378/chest.126.2.470.
- [174] QSR, *NVivo 12*. Victoria, Australia: QSR International Pty Ltd.
- [175] W. P. Harvey, "Cardiac pearls," *Dis.--Mon. DM*, vol. 40, no. 2, pp. 41–113, Feb. 1994.
- [176] S. A. Levine, "THE SYSTOLIC MURMUR: ITS CLINICAL SIGNIFICANCE," *J. Am. Med. Assoc.*, vol. 101, no. 6, pp. 436–438, Aug. 1933, doi: 10.1001/jama.1933.02740310020005.
- [177] M. E. Silverman and C. F. Wooley, "Samuel A. Levine and the History of Grading Systolic Murmurs," *Am. J. Cardiol.*, vol. 102, no. 8, pp. 1107–1110, Oct. 2008, doi: 10.1016/j.amjcard.2008.06.027.
- [178] R. Keren, M. Tereschuk, and X. Luan, "Evaluation of a Novel Method for Grading Heart Murmur Intensity," *Arch. Pediatr. Adolesc. Med.*, vol. 159, no. 4, pp. 329–334, Apr. 2005, doi: 10.1001/archpedi.159.4.329.
- [179] A. J. Taylor, *Learning cardiac auscultation: from essentials to expert clinical interpretation*. London : Springer, [2015], 2015.
- [180] J. Constant, "How to differentiate ejection murmurs from systolic regurgitant murmurs," *Keio J. Med.*, vol. 44, no. 3, pp. 85–87, Sep. 1995.
- [181] International Organization for Standardization, *ISO 226:2003 Acoustics. Normal equal-loudness contours*. Geneva: International Organization for Standardization, 2003.
- [182] American National Standards Institute, *Design Response Of Weighting Networks For Acoustical Measurement*. New York: American National Standards Institute, 2016.

- [183] International Telecommunication Union, "RECOMMENDATION ITU-R BS.468-4 - Measurement of audio-frequency noise voltage level in sound broadcasting," p. 7, Jul. 1986.
- [184] M. Brookes, *Voicebox: Speech Processing Toolbox for Matlab*. 2011.
- [185] Qualtrics, *Qualtrics*. Provo, Utah, USA: Qualtrics, 2019.
- [186] "What's the average survey response rate? [2018 benchmark]," *Survey Anyplace*, Apr. 05, 2018. <https://surveyanyplace.com/average-survey-response-rate/> (accessed May 21, 2019).
- [187] "What's a Good Survey Response Rate?," *SurveyGizmo*, Jul. 27, 2015. <https://www.surveygizmo.com/resources/blog/survey-response-rates/> (accessed May 21, 2019).
- [188] A. Ramshaw, "The Complete Guide to Acceptable Survey Response Rates," *Genroe*, Jan. 14, 2019. <https://www.genroe.com/blog/acceptable-survey-response-rate-2/11504> (accessed May 21, 2019).
- [189] A. F. Hayes and K. Krippendorff, "Answering the Call for a Standard Reliability Measure for Coding Data," *Commun. Methods Meas.*, vol. 1, no. 1, pp. 77–89, Apr. 2007, doi: 10.1080/19312450709336664.
- [190] X. Zhao, G. C. Feng, J. song Liu, and K. Deng, "We agreed to measure agreement - Redefining reliability de-justifies Krippendorff's alpha," 2018.
- [191] X. Zhao, J. song Liu, and K. Deng, "Assumptions behind Intercoder Reliability Indices," 2013, doi: 10.1080/23808985.2013.11679142.
- [192] J. Girard, *MATLAB Functions for Computing Inter-Observer Reliability: jmgirard/mReliability*. 2019.
- [193] A. Hyvarinen, H. Sasaki, and R. E. Turner, "Nonlinear ICA Using Auxiliary Variables and Generalized Contrastive Learning," *ArXiv180508651 Cs Stat*, May 2018, Accessed: Jul. 29, 2019. [Online]. Available: <http://arxiv.org/abs/1805.08651>.
- [194] PricewaterhouseCoopers, "What doctor? Why AI and robotics will define New Health," *PwC*, Nov. 2016. <https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health.html> (accessed Jun. 17, 2019).