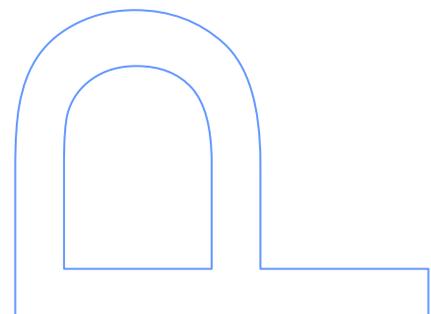
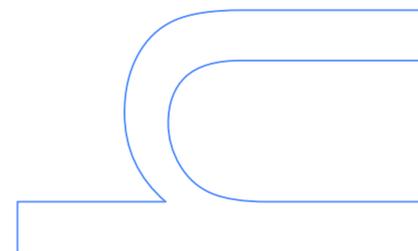
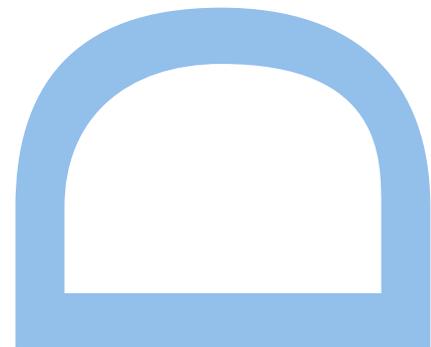




Subject-driven supervised and unsupervised Hidden Markov Models for heart sound segmentation in real noisy environments



Jorge Henrique Santos Oliveira

Doutoramento em Informática
Departamento de Ciências dos Computadores
2018

Orientador

Professor Auxiliar Miguel Coimbra, Faculdade de Ciências da Universidade do Porto

Contents

1	Introduction	2
1.1	Auscultation	2
1.1.1	Why we need to auscultate	3
1.1.2	Electronic stethoscopes can help in CAD systems	4
1.1.3	Signal processing pipeline for heart sounds	5
1.2	HMM's as a promising avenue	6
1.3	Objectives	6
1.4	Contributions	7
1.5	Thesis structure	8
2	Heart Sound Physiology	9
2.1	Heart Anatomy and Sound Generation	10
2.1.1	Heart Cycle	11
2.1.2	Heart Sound Auscultation History	12
2.1.3	Auscultation Procedure	13
2.2	Rhythmical Excitation of the Heart	15
2.2.1	The Electrocardiogram	16
2.3	Electrocardiography	17
2.3.1	Electrocardiography History	19
3	State-of-art and Background	20
3.1	Heart Sound Modeling Perspectives	20
3.1.1	Fluid Dynamics Perspective	21
3.1.2	Electrical Analogy Perspective	23
3.1.3	Signal Processing Perspective	24
3.2	Background on Signal Processing for Heart Sounds	26
3.2.1	Denosing and Signal Enhancement	26
3.2.2	Segmentation - Introduction	27
3.2.3	Segmentation - Single Channel System	27

3.2.4	Segmentation - Multi Channel System	29
3.2.5	Feature Extraction	30
3.2.6	Classification	31
3.3	Heart Sound Segmentation using Statistical Models	32
3.3.1	Hidden Markov Models	33
3.3.2	Semi-Hidden Markov Models	33
3.3.3	Coupled Hidden Markov Models	34
3.4	Background	34
3.4.1	Nomenclature	34
3.4.2	Hidden Markov Models	35
3.4.3	Semi-Hidden Markov Models	36
3.4.4	Coupled Hidden Markov Models	38
4	Heart Sound Segmentation using a Subject Dependent Approach	40
4.1	Introduction	41
4.1.1	Why implement a subject dependent approach?	41
4.1.2	Objectives	41
4.2	Methodology	42
4.2.1	Model Architecture	42
4.2.2	HSMM and HMM distributions	43
4.2.3	Experimental methodology	46
4.2.4	Pre-processing and feature extraction	47
4.2.5	Training HMM and HSMM distributions	47
4.2.6	Initialization of HMM and HSMM distributions	48
4.2.7	Tuning HMM and HSMM distributions	48
4.2.8	Decoding	50
4.3	Materials	50
4.3.1	DigiScope dataset	50
4.3.2	Metrics of performance	51
4.4	Results	52
4.4.1	Why should you model time on a HMM?	52
4.4.2	How should you model time on a HSMM?	54
4.4.3	Where should you model time on a HSMM?	57
4.5	Discussion	59
5	Heart Sound Segmentation using a Subject Independent Approach	60
5.1	Introduction	60
5.1.1	Why implement a subject independent approach?	61

5.1.2	Motivation and objectives	61
5.2	Methodology	62
5.2.1	Model Architecture	62
5.2.2	HSMM distributions	62
5.2.3	Experimental methodology	63
5.2.4	Pre-processing and feature extraction	64
5.2.5	Training HSMM distributions	64
5.2.6	Initialization of HSMM distributions	64
5.2.7	Tuning sojourn time distributions	65
5.2.8	E-step	67
5.2.9	M-step	68
5.2.10	Decoding	69
5.3	Materials	69
5.3.1	Physionet dataset	69
5.3.2	Pascal dataset	70
5.3.3	DigiScope dataset	70
5.3.4	Metrics of performance	70
5.4	Results	71
5.5	Discussion	73
6	Heart Sound Segmentation using an Unsupervised Approach	75
6.1	Introduction	76
6.1.1	Towards unsupervised heart sound segmentation	76
6.1.2	Motivation and objectives	76
6.2	Methodology	77
6.2.1	Model Architecture	77
6.2.2	HSMM distributions	77
6.2.3	Experimental methodology	78
6.2.4	Pre-processing and feature extraction	79
6.2.5	Training HSMM distributions	79
6.2.6	Initialization of HSMM distributions	79
6.2.7	Tuning GMM parameters on a HSMM	81
6.2.8	Decoding	86
6.3	Materials	86
6.3.1	PhysionNet dataset	86
6.3.2	Metrics of performance	86
6.4	Results	87
6.5	Discussion	90

7	Electrocardiogram and Phonocardiogram Segmentation in a Multi-Channel System	91
7.1	Introduction	92
7.1.1	Current limitations using a single PCG channel	92
7.1.2	Advantages of also using an ECG channel	92
7.1.3	Motivation	93
7.1.4	Objectives	93
7.2	Methodology	94
7.2.1	Model architecture	94
7.2.2	Fully connected model	94
7.2.3	Partially connected model	95
7.2.4	CHMM distributions	95
7.2.5	Experimental methodology	96
7.2.6	Pre-processing and feature extraction	96
7.2.7	Training and initialization of CHMM distributions	96
7.2.8	Tuning CHMM distributions	98
7.2.9	EM algorithm	98
7.2.10	sEM algorithm	100
7.2.11	Decoding	100
7.3	Materials	100
7.3.1	HeartSafe dataset	100
7.3.2	Metrics of performance	101
7.4	Results	101
7.5	Discussion	103
8	Conclusions	104
	Bibliography	106

List of Figures

1.2	Diagram of the standard signal processing pipeline.	5
2.1	A diagram of longitudinal wave propagating through a medium. . .	9
2.2	Diagram of a normal heart.	10
2.4	A normal heart, the white arrows show the blood flow direction . .	13
2.8	(a) Conventional arrangement of electrodes (Lead I, II and III) and (b) the precordial leads, picture adapted from [5].	18
3.5	Four state HMM for a cycle of a normal heart sound signal (adapted from [51]).	35
3.6	Four state HSMM for a cycle of a normal heart sound signal. . . .	37
3.7	The CHMM scheme for a particular case when two channels are completely connected. The first layer is composed by the nodes (A,B,C) and the second layer is composed by the nodes (D,E,F). Each node is a hidden state and the state transition probabilities are denoted by arrows [61]).	39
4.2	A PCG signal and its corresponding elements.	43
4.3	The probabilistic density function of a Geometric distribution . . .	44
4.4	The probabilistic density function of a Poisson distribution	44
4.5	The probabilistic density function of a Gaussian distribution	45
4.6	The probabilistic density function of a Gamma distribution	45
4.7	Non-parametric mass density function	46
4.8	Prototype of the DigiScope Collector system, composed by a tablet and an electronic stethoscope Littmann 3200.	51
4.9	A Picture of a normalized heart sound signal and the corresponding expert annotations made using the Audacity software.	51
4.10	Classification results of heart sound recordings from a normal sub- ject. The dashed lines present the states classified by an expert, HMM, HSMM; and the solid lines present the observation input to the model.	54

4.11	Subject dependent results. Average positive predictability (a) per sample P_{Sample}^+ and (b) per state P_{State}^+ for the tested HMM's, HSMM's over the DigiScope dataset.	55
4.12	The conditional distribution P_r generated by the HSMM using the Poisson distribution.	57
4.13	Relationship of the conditional distribution P_r with positive predictability P_{Sample}^+, RP^+, EP^+ and confidence P_c in our dataset. . .	58
4.14	Relationship of the conditional distribution P_r and the subset positive predictability which is obtained using a high threshold in our dataset.	59
5.2	A two-dimensional logistic regression function, where $w_i = \{-2, -2\}$.	63
5.3	Precision results: A) When using the 10 fold cross validation over the PhysioNet dataset. B) When using the PhysioNet dataset to train and the DigiScope dataset to test. C) When using the PhysioNet dataset to train and the Pascal dataset to test.	72
5.6	Accuracy results: A) When using the 10 fold cross validation over the PhysioNet dataset. B) When using the PhysioNet dataset to train and the DigiScope dataset to test. C) When using the PhysioNet dataset to train and the Pascal dataset to test.	72
5.4	Recall results: A) When using the 10 fold cross validation over the PhysioNet dataset. B) When using the PhysioNet dataset to train and the DigiScope dataset to test. C) When using the PhysioNet dataset to train and the Pascal dataset to test.	73
5.5	F-Score results: A) When using the 10 fold cross validation over the PhysioNet. B) When using the PhysioNet dataset to train and the DigiScope dataset to test. C) When using the PhysioNet dataset to train and the Pascal dataset to test.	74
6.2	A Gaussian mixture model distribution, with $\mu_i = \{0.2, 0.5\}$, $\sigma_i = \{0.1, 0.1\}$ and $\Pi_i = \{0.5, 0.5\}$	78
6.3	Segmentation performance obtained by two supervised segmentation algorithms with the PhysioNet dataset: HSMM with GMM emission distributions (blue thick lines); HSMM with logistic regression function emission distributions [34] (purple thin lines). . . .	87

6.4	Segmentation performance obtained when using all 792 sounds from the PhysioNet dataset. Supervised HSMM-GMM algorithm (blue thick lines), unsupervised HSMM-GMM algorithm when using only the heuristics proposed in Section 6.2.6 (red lines), unsupervised HSMM-GMM algorithm when using the tuning routines presented in Section 6.2.7 (gray thin lines).	88
6.5	Segmentation performance obtained using a subset of the PhysioNet dataset composed by 314 sounds with a duration longer than 10 seconds. Supervised HSMM-GMM algorithm (blue thick lines), unsupervised HSMM-GMM algorithm when using only the heuristics proposed in Section 6.2.6 (red lines), unsupervised HSMM-GMM algorithm when using the tuning routines presented in Section 6.2.7 (gray thin lines).	89
7.2	An example CHMM state sequence for a particular case when two channels are fully connected.	95
7.3	An example CHMM state sequence for a particular case when two channels are partially connected.	95
7.4	Positive predictability per sample (P_{sample}^+) results as function of Ω in the (A) ECG channel and (B) PCG channel. The CHMM parameters (Θ) are re-estimated using the sEM algorithm.	102
7.5	Positive predictability per sample (P_{sample}^+) results as function of Ω in the (A) ECG channel and (B) PCG channel. The state transition parameters are re-estimated using the sEM algorithm, the emission Gaussian parameters and the initial state parameters are re-estimated using the standard EM algorithm ($\Omega = 0$).	102

List of Tables

3.1	A summary of the standard heart sound segmentation algorithms. . .	29
4.1	Performance of HSMM's when using a Poisson sojourn time distribution in detecting $S1$ and $S2$	56

Acknowledgements

I would like to thank my supervisor Prof. Miguel Coimbra for the helpful guidance and for giving me the opportunity to work in such an important topic, it has been an interesting and enriching experience. Finally, I thank my family and my wife for the ongoing support and understanding.

Abstract

The analysis of heart sounds is a challenging task, due to the quick temporal onset between successive events and the fact that an important fraction of the information carried by phonocardiogram (PCG) signals lies in the inaudible part of the human spectrum. For these reasons, computer-aided analysis of the PCG can improve dramatically the quantity of information recovered from such signals. In this thesis, a family of hidden Markov models are used to automatically segment PCG signals in both supervised and unsupervised approaches, resulting in four distinct scientific contributions to the field. The presented algorithms have shown to be effective in decoding the true state sequence of events in real noisy PCG signals. The first two main contributions, consist in several advances to tune the emission and the sojourn time distributions from the training data to the tested subject, with a special highlight to the semi-hidden Markov models (HSMMs). Among the contributions made perhaps the most important one is a novel subject driven unsupervised heart sound segmentation algorithm, where the emission probability distributions are tuned to the tested subject regardless of the training done. Perhaps surprisingly, our method outperformed the state-of-the-art supervised approaches, when provided with a sufficiently long unlabeled heart sounds. Finally, the last contribution is an automatic electrocardiogram (ECG) and PCG segmentation algorithm for a multichannel system.

Chapter 1

Introduction

In this chapter, we are going to give a quick introduction to the topic of heart sound analysis. We start by motivating the reader, about the role of auscultation and how electronic stethoscopes might be used in computer assisted decision (CAD) systems. Furthermore, we explain our choice for hidden Markov models instead of other statistical models or algorithms in the literature. Finally, the objectives, contributions and thesis structure are explained in the last sections.

1.1 Auscultation

From the rumbling in the belly to a pumping heart, the human body is a strange symphony of sounds. Most of them are perfectly normal but others are not and the simplest and cheapest medical exam to assess them is auscultation. Auscultation is the act of listening the internal body sounds, usually using a stethoscope. Auscultation is typically used to evaluate (assess) the circulatory and the respiratory systems (heart and breath sounds) as well as the gastrointestinal system (bowel sounds).



Figure 1.1: A researcher doing an auscultation to a child. Picture taken from the ‘Caravana do Coração’ event in Paraiba, Brazil.

1.1.1 Why we need to auscultate

Cardiovascular diseases are the leading cause of death in developed and developing countries and one of the major causes of hospitalization. Cardiovascular diseases are highly constrained and they can occur in any age, although senior citizens are more susceptible than any other age group. In Europe, the proportion of the population aged > 65 years is projected to increase 25% in 2025 and globally by 14% until 2040 [1]. These diseases can be largely avoided through healthy lifestyle or/and by an active (close) relationship with healthcare agents (e.g: seeing your doctor once a year for a physical examination) and also by doing periodically (monthly, yearly) specific screening exams (e.g: heart sound auscultation, echocardiography, echo-doppler etc.) [2]. Nowadays, effective preventive healthcare strategies require robust and inexpensive solutions for early detection of heart diseases. In this context, heart sound auscultation is a very important and interesting solution for several reasons:

- It is a non-invasive method to assess the heart status.
- When used properly, it is a very effective medical exam to screen heart diseases.
- It is a simple and quick medical exam.
- It is cheaper than for example, echocardiography.

1.1.2 Electronic stethoscopes can help in CAD systems

On the other hand, cardiac auscultation is a difficult skill to master since it requires an extensive and continuous training in order to distinguish the heart sound nuances. This is due to the fact that heart sounds have a very quick temporal onset between events and an important fraction of the information carried by a phonocardiogram (PCG) signal lies in the inaudible part of the human spectrum. Usually a medical student or a physician needs to listen around 500 repetitions of each type of murmur in order to learn how to identify them properly [3] and only 20% of graduate students and physicians are able to accurately detect abnormal heart sounds [4]. Furthermore, physicians with poor auscultation skills are likely to fail in detecting cardiac diseases, which might result in disastrous consequences for the patient. This is a very problematic issue where emerging technologies might play an important role in the nearby future. Some developed or under development examples are:

- The is4Learning ¹ technology is an affordable virtual patient simulator that enables the teaching and training of the three fundamental skills required for an effective cardiopulmonary auscultation: positioning, gesture, and listening.
- The DigiScope2 ² prototype is a computer assisted decision system for auscultation, integrating heart sound algorithms and machine learning techniques to recognize hidden patterns of the signal.

Auscultation is still the quickest and less expensive screening tool available for heart diseases in developed and developing countries, and possibly the only one available in some isolated communities around the world.

¹Developed by the is4Health company.

²Under development by the is4Health company.

1.1.3 Signal processing pipeline for heart sounds

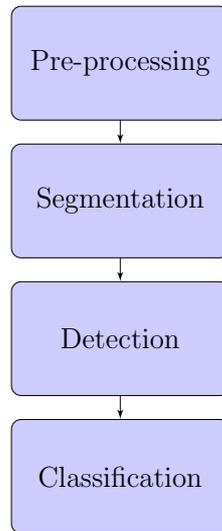


Figure 1.2: Diagram of the standard signal processing pipeline.

In order to recognize and distinguish normal from abnormal heart sounds we need to follow the steps in Figure 1.2 :

- pre-process the signal: the noise is removed or attenuated and the signal components (heart sounds) are enhanced. This is usually achieved by removing some undesired frequencies or frequency bands in the signal, a process known as filtering.
- segment the signal: the heart sound signal is splitted into heart cycles. Each heart cycle is mainly composed by the first heart sound (S1), the systolic period (siSys), the second heart sound (S2), and the diastolic period (siDias). In the literature, there are several possible approaches to segment the PCG signals. The simplest ones try to identify the time instant and duration of the S1 and S2 sounds, by using some sort of a peak-picking algorithm. On the other hand, the most advanced algorithms apply temporal statistical models to search for the most likely hidden state sequence according to a set of observations.
- recognize the presence of abnormal sounds: the heart sounds are classified according to the presence or not of abnormal heart sounds (e.g: third and fourth heart sounds as well as heart murmurs and ejection clicks that may be associated to cardiac pathologies). This is usually done by applying standard pattern recognition techniques: 1) usually features are extracted from

the signal, 2) these are fed into a classifier, (e.g: artificial neural networks (aNN), k-nearest neighbors(k-NN), support vector machine (SVM), etc) and 3) finally, the algorithm infers the presence or not of abnormal heart sounds.

- classify the signal: abnormal heart sounds are ranked according to specific criteria. For example, murmurs are further classified according to the timing, duration, quality (pitch, change in intensity), intensity (loudness) and point of maximum intensity and radiation. These features are used by the physicians to categorize the murmur and also to predict its source. For the signal processing community, this is typically achieved by applying the same aforementioned rationale, although sadly it has been neglected or forgotten by the scientific community.

1.2 HMM's as a promising avenue

Heart sound segmentation is still a very challenging task and an unsolved problem in several applications. For example, sounds from very noisy environments, healthy neonates, children and in senior citizens (although for distinct reasons) and of course in the presence of abnormal sounds. To address this problem and among several possibilities, we chosen by hidden Markov models since:

- they offer a more natural modeling approach than other methods. In a hidden Markov model, each Markov state represents a stage of the cardiac cycle and the Markovian state transitions obey physiological time constraints in the cardiac muscle. This give us a more natural insight and understanding of the cardiac system that we are trying to model.
- These models are mathematically convenient to predict future or missing heart beat sequences.
- the rapidly emerging success over the past years in several similar applications (e.g: ECG and EEG segmentation).

1.3 Objectives

In this thesis, we aim to design efficient heart sound segmentation algorithms, which are capable of decoding the 'true' state sequence of events in very noisy PCG signals. This is indeed a very challenging task and to do so, we divided our main objective into sub objectives. Our first objective is to segment uncompleted

annotated heart sounds. In this approach, we have made the assumption that the first heart beats of each tested subject are known. This annotated data is afterwards used to properly initialize the HMM's for the segmentation of the last heart beats of each signal. Our second objective is a more demanding one since we aim to segment heart sound signals for unannotated test subjects. To do so, generic HMM's parameters are trained using an independent training dataset. These parameters might be further adjusted to the tested subject, using Expectation Maximization (EM) approaches. In our third objective, we are going even further and nothing is assumed to be known about the signal and the system that we are trying to segment, i.e we segment heart sounds in the absence of a representative annotated heart sound dataset. Finally, in our last objective, we segment heart sounds in the presence of synchronous auxiliary signals, such as the electrocardiogram. For this contribution, we have made the assumption that the first heart beats of each tested subject are known.

1.4 Contributions

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

- We proved that semi-hidden Markov models (HSMM's) are more efficient than hidden Markov models (HMM's) to decode the 'true' state sequence of events in a PCG signal.
- We propose to model the sojourn time distribution in a HSMM by a Poisson distribution instead of a Gaussian distribution.
- We propose to compute the probability that a sample classified by our model is correct.
- We presented re-estimation routines to tune the sojourn time parameters extracted from the training dataset to the tested subject in a HSMM.
- We suggested to model the emission probability distribution by a Gaussian mixture model (GMM), instead of the current state-of-art logistic regression function in a HSMM.
- We proposed re-estimation routines to tune the emission parameters to the tested subject regardless the training or not done in a HSMM.

- To be best of our knowledge, we designed the first coupled hidden Markov model for an ECG and PCG multi-channel system.

1.5 Thesis structure

The thesis is organized as follows:

- **Chapter 2** Introduces a brief summary of the physiology of the heart. The history of auscultation and electrocardiography is also presented.
- **Chapter 3** Presents the current state-of-art algorithms for the different schools of heart sound modeling. The background about HMM's, HSMM's and CHMM's are also explained.
- **Chapter 4** Answers the following questions *Why, How and Where should you model time when using a hidden Markov model?*
- **Chapter 5** Explains the limitations of the current state-of-algorithms when trying to model the sojourn time distribution in a HSMM. As a solution, an EM algorithm to tune the sojourn time distribution in a HSMM from the training dataset to the tested subject is proposed.
- **Chapter 6** Explains the limitations of the current state-of-algorithms when trying to model the emission distribution in a HSMM. In this chapter, an EM algorithm to tune the emission distribution in a HSMM to the tested subject is proposed.
- **Chapter 7** Presents the first CHMM for an ECG and PCG multi-channel system. Two different CHMM architectures are presented and compared among them in two different scenarios.
- **Chapter 8** Presents the main conclusions of our work, discuss some possible future research lines.

Chapter 2

Heart Sound Physiology

A sound is roughly speaking a wave traveling over a medium. A sound is generated when a particle (or a set of particles) starts to oscillate around a resting position and as it moves other nearby particles are pushed and pulled. As a result, an air pressure gradient is generated (see Figure 2.1).

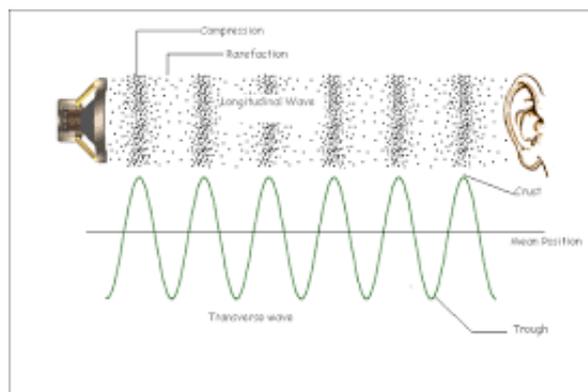


Figure 2.1: A longitudinal wave propagating through a medium, picture adapted from ¹.

As the sound wave travels, it reflects off objects, creating further disturbances in the surrounding medium. When the sound waves reach the eardrum, nerve signals are sent to the brain and are further perceived as sounds. In this thesis, our focus is on the sounds generated by our bodies, more specifically heart sounds.

¹website: <https://physics818.wordpress.com/2015/03/26/waves/> on 11/10/2017.

2.1 Heart Anatomy and Sound Generation

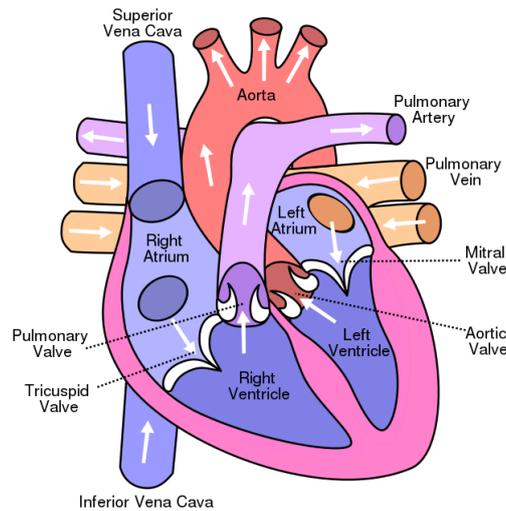


Figure 2.2: Diagram of a normal heart. The white arrows show the blood flow direction, picture adapted from ².

In this section, we are going to explain very briefly the heart anatomy in order to understand the mechanisms and the hemodynamic events concerning heart sounds. The heart is a muscular organ and it works as a pump in the circulatory system forcing the blood to circulate throughout the human body. The heart has two main functions:

- collect oxygen-rich blood from the lungs and send it to all the tissues of the body, then oxygen and carbon dioxide are exchanged from the blood to the tissue cells.
- collect the blood rich in carbon dioxide from the tissues and send it to the lungs, then carbon dioxide and oxygen are exchanged from the alveolar blood to the alveolar air.

The heart is divided into four chambers separated from each other by cardiac muscle:

- The upper left atria, receives oxygenated blood from the left and right pulmonary veins and pumps it to the left ventricle through the mitral valve.
- The upper right atria, receives deoxygenated blood from the superior vena cava, inferior vena cava, anterior cardiac veins and smallest cardiac veins and

²website: <https://en.wikipedia.org/wiki/Heart> on 11/10/2017.

the coronary sinus and it pumps to the right ventricle through the tricuspid valve.

- lower left ventricle, receives oxygenated blood from the left atria and pumps it out to the aorta artery through the aortic valve.
- lower right ventricle, receives deoxygenated blood from the right atria and sends it out to the pulmonary artery through the pulmonary valve.

The internal heart structure and components compel the blood to flow in one-way only. The atrioventricular valves (tricuspid and bicuspid) allow blood to flow only from the atriums to the ventricles. The semilunar valves (pulmonary and aortic) allow blood to flow out of the heart from the ventricles to the great arteries, as depicted in Figure 2.1.

2.1.1 Heart Cycle

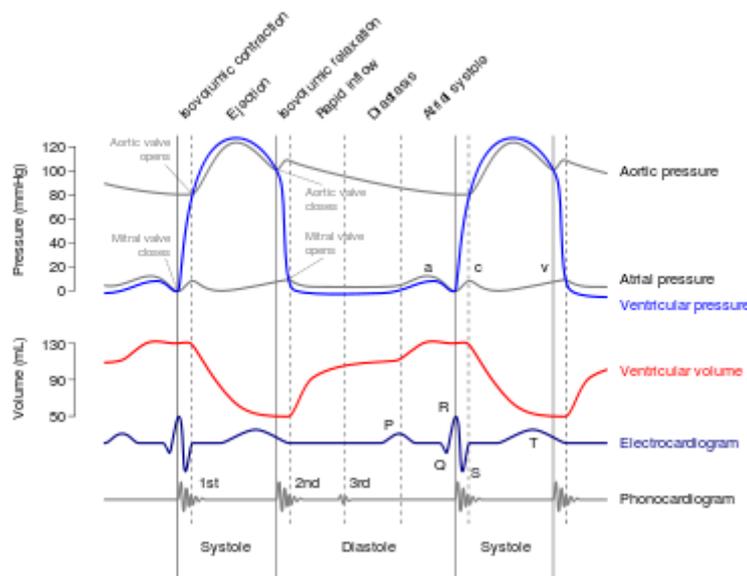


Figure 2.3: The Wiggers diagram, including the ECG and the PCG at bottom (adapted from [5]).

During the systole, large amounts of blood are stored in the atriums since the atrioventricular valves are closed. At the end of the systole, the atrioventricular valves open suddenly due to the increasing pressure in the atriums and the decreasing pressure in the ventricles, a period known as the rapid filling of the ventricles.

This period of rapid filling corresponds to 2/3 of the diastolic time and the last 1/3 corresponds to the atrial contraction [6]. This rapid filling results in a

rising pressure in the ventricles, causing at the end of the diastole the closing once again of the atrioventricular valves, the resulting vibration is low in pitch and relatively long-lasting and it is known as the first heart sound (S1). On the other hand, the semilunar valves do not open immediately and it takes around 0.02 to 0.03 seconds to do it, during this period, contraction is occurring in the ventricles, but there is no emptying a period known as isometric contraction [6]. When the left ventricular pressure rises slightly above 80 mm Hg (and the right ventricular pressure slightly above 8 mm Hg), the semilunar valves open and blood is ejected outside of the ventricles, this period of ejection corresponds to the systole, see Figure 2.3. At the end of this period, ventricular relaxation begins suddenly, allowing both the right and left intraventricular pressures to decrease rapidly, in contrast the pressure in large arteries are very high since they have been just filled with blood from the contracted ventricles, at the end of this period, some expected blood flows back to the ventricles, forcing the aortic and pulmonary valves to close resulting in a rapid snap sound called the second heart sound (S2). The ventricle muscle continues to relax (isometric relaxation) and the intraventricular pressures decrease rapidly compelling once again atrioventricular valves to open, therefore marking the beginning of a new heart cycle [6].

2.1.2 Heart Sound Auscultation History

Heart sound auscultation as a diagnostic (screening) tool goes back to the Ancient Greece. Hippocrates documented the first auscultation in history, by directly applying the ear to a patient's chest and abdomen. This immediate auscultation is very uncomfortable for both physicians and patients. In hospitals, this method is impractical due to the great corporal contact and the significantly risk of infection. From Hippocratic Greece to the 17th century, little knowledge was added to auscultation until 1628 when William Harvey attempted to describe the heart sound phenomenon [7]. Only later in the 19th century, Laennec created a device which resembles to the nowadays stethoscope although very rudimentary (see Figure 2.4). In his own words, *"It consists simply of a cylinder of wood, perforated in its center longitudinally, by a bore three lines in diameter, and formed so as to come apart in the middle, for the benefit of being more easily carried.... The complete instrument, that is, with the funnel-shaped plug in fixed, is used in exploring the signs obtained through the medium of the voice and the action of the heart, the other modification or with the stopper removed, is for examining the sounds communicated by respiration."* [7]. Laennec also made the first physical examination although he failed in explaining the heart sound phenomenon and the relationship between cardiac

sounds and murmurs to specific heart diseases. Only in 1831, Hope provided clinical and pathological findings for different type of heart diseases and the evidence of valvular origin for the second heart sound. Finally in 1834, Rouanet explained and proved experimentally that heart sounds were generated by valvular motion. [7].



Figure 2.4: Laennec stethoscope invented in 1816, picture adapted from ³.

From Rouanet to the 20th century, several other researchers contributed significantly not only in terms of technology (device) but also in our understanding of cardiac sounds and murmurs or in the proper scientific methodology in the heart sound auscultation. Einthoven, the inventor of the modern electrocardiogram, was also the first to record heart sounds successfully in 1907. Later on Rappaport and Sprague showed how the stethoscope and chest modified the frequencies perceived by the human ear. Their work improved the bedside understanding of auscultation and led to the design of a stethoscope with a bell and diaphragm combination [7] (see Figure 2.5).

From Laennec's discovery of the stethoscope in 1816 to nowadays, heart sound auscultation remained as the standard medical exam to screen the presence of heart diseases. It is a cheap and non-invasive medical procedure (exam) to assess the internal functional and hemodynamic behaviors of the heart. It is also an ancient art, which over the centuries has been defining the relationship between physicians and patients.

2.1.3 Auscultation Procedure

The stethoscope is an ingenious device, very well designed to listen to heart sounds. The diaphragm of the stethoscope can detect high frequency sounds, such as the systolic ejection murmurs, whereas the bell of the stethoscope can detect low frequency sounds, such as S3 and S4 gallops or the diastolic rumble of mitral stenosis

³website: <https://tokbox.com/blog/telehealth-can-you-hear-my-heart-now/> on 11/10/2017.

(see Figure 2.5).



Figure 2.5: A traditional stethoscope, picture adapted from ⁴.

Auscultation is typically done in a quiet environment, in order to minimize the external noise. First, the patient should be instructed about the procedure in order to decrease stress and anxiety. The patient must stand still and be comfortable during the exam. Finally at the end, the patient must be informed about the results and also further clarified.

Up to our knowledge, there isn't a standard methodology to collect heart sounds, although two systematic approaches are well accepted: the physician should first auscultate the right upper sternal border, followed by the left upper sternal border. Afterwards the down left sternal and finally the apex is also auscultated. The other way around is also acceptable as long the sequence is kept (see Figure 2.6). In each spot, the frequencies listened are dominated by a unique heart valve, enabling us to uniquely assess the mechanical properties of a specific heart valve. Good examples are: the murmur of aortic stenosis is heard best at the right second inter space; the murmur of pulmonic stenosis is heard best at the left second interspace; the murmur of tricuspid stenosis is heard best along the lower left sternal border; and the murmur of mitral stenosis is heard best at the apex.

⁴website: <http://www.tildee.com/LHki1k> on 11/10/2017.

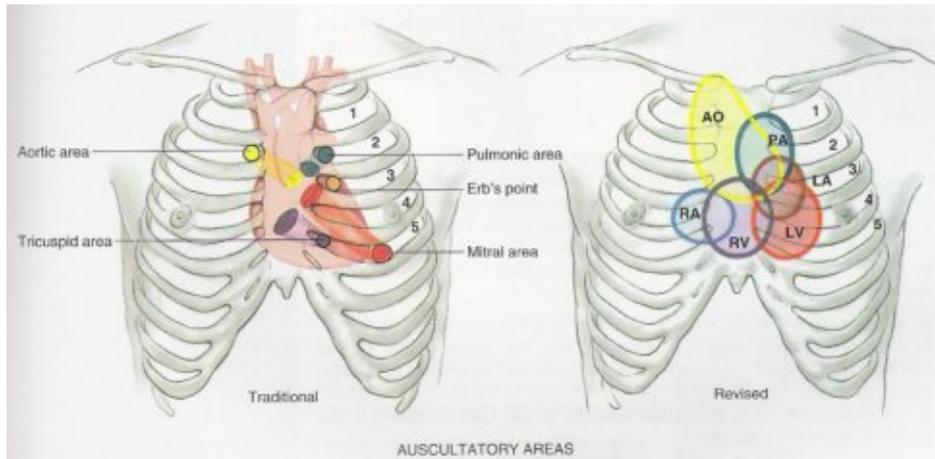


Figure 2.6: Auscultation spots, picture adapted from ⁵.

Some murmurs can radiate far from the source [5]. Examples are pulmonary valve stenosis that sometimes radiates to the left clavicle, aortic valve stenosis that radiates to the carotid arteries, and mitral regurgitation to the axilla or precordium (see Figure 2.6).

2.2 Rhythmical Excitation of the Heart

The heart is composed by cells called cardiac myocytes or cardiomyocytes. These cells are by nature polarized, i.e negatively charged when compared to the outside. The electrical potential is created by the presence of a delicate K^+ and Na^+ equilibrium from inside and outside of a cell. When an action potential is generated in the sinus node (termed as SA node) ⁶, an impulse wave travels through electrical conduction system of the heart to each cardiac cell (see Figure 2.7). As a net result, the Na^+ and Ca^{2+} channels open, allowing large amounts of Na^+ and Ca^{2+} ions to move inside of a cell [5].

⁵website: www.topregisterednurse.com/apical-pulse-definition-process-measurement/ on 11/10/2017.

⁶located at the top of the heart's upper-right chamber (the right atrium)

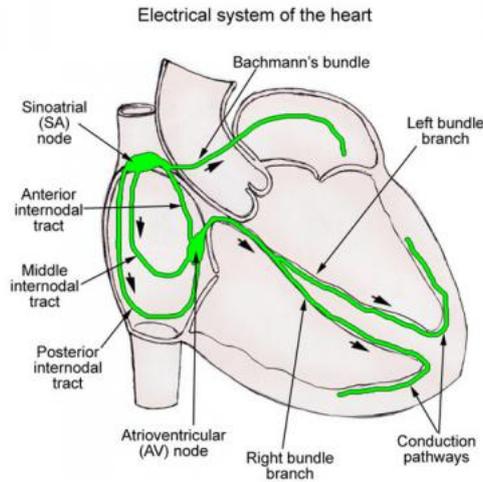


Figure 2.7: The electrical signal first stimulates the upper chambers of the heart (the atria) to contract. This signal then travels through the atrioventricular (AV) node, where the signal is delayed and then sent to the muscle fibers of the lower chambers (the ventricles), picture adapted from ⁷.

These Ca^{2+} join to the troponin, which triggers the cross-bridge binding that leads to the sliding of actin filaments past myosin filaments. The sliding of the filaments produces cell contraction. After the Na^+ channels have been opened for some milliseconds, the Na^+ gate is the first to close and later on the Ca^{2+} . In contrast, the K^+ channels opens and K^+ diffuses out of the cardiac cell. These events restore the cardiac cell to its original polarization, except that the positions of K^+ and Na^+ on each side of the membrane are reversed. After a refractory period follows, where K^+ and Na^+ are actively restored to their appropriate places through Na^+/K^+ pumps. During this period, the cardiac cells do not contract until Na^+ and K^+ levels have been restored to their original stages. The refractory period of a cardiac muscle is dramatically longer than that of skeletal muscle so there is enough time to refill the heart chambers with blood before the next contraction [5].

2.2.1 The Electrocardiogram

When the cardiac impulse spreads through the heart, electromagnetic waves also spread from the heart into the adjacent tissues surrounding the heart. These are detected and recorded by placing electrodes on opposite sides of the heart. This recording is known as an electrocardiogram (ECG). The normal electrocardiogram (see Figure 2.3) is composed by a P, QRS complex, and a T waves. The P wave

⁷website: <https://emedicine.medscape.com/article/1922987-overview> on 11/10/2017.

is caused by electrical potentials generated when the atria depolarize before the atrial contraction begins. The QRS complex is caused by potentials generated when the ventricles depolarize. The T wave is caused by potentials generated as the ventricles recover from the state of depolarization. This process normally occurs in ventricular muscle from 0.25 to 0.35 seconds after depolarization [5].

2.3 Electrocardiography

Nowadays, modern electrocardiography uses computer-based systems, while traditional ones use a direct pen recorder that writes the electrocardiogram with a pen directly on a moving sheet of paper [5]. Regardless of the instrument used, two electrodes and a meter are at least needed⁸. Usually, the electrodes are disposed in the limbs and in opposite sides of the heart and therefore closing a circuit between the body and the electrocardiograph [5]. Three distinct electrode arrangements are used:

- Lead I: The negative terminal of the electrocardiograph is connected to the right arm and the positive terminal to the left arm.
- Lead II: The negative terminal of the electrocardiograph is connected to the right arm and the positive terminal to the left leg.
- Lead III: The negative terminal of the electrocardiograph is connected to the left arm and the positive terminal to the left leg.

When someone wants to diagnose a damage in the ventricular or atrial muscle or in the Purkinje conducting system, it is important to choose very carefully the recording lead since abnormalities in cardiac muscle contraction or cardiac impulse conduction do change their patterns in the electrocardiograms markedly in some leads, and yet may not affect other leads [5]. In contrast, it does not matter greatly which lead is recorded when someone wants to diagnose cardiac arrhythmias, because diagnosis of arrhythmias depends mainly on the time relations between the different waves of the cardiac cycle.

Another standard layout is to place an electrode on the anterior surface of the chest directly over the heart. This electrode is connected to the positive terminal of the electrocardiograph, and the negative electrode, called the indifferent electrode, is connected through equal electrical resistances to the right arm, left arm, and left leg all at the same time.

⁸When a meter is connected with its negative terminal on a depolarized area and its positive terminal on one of the still-polarized areas, it reads positive otherwise negative.

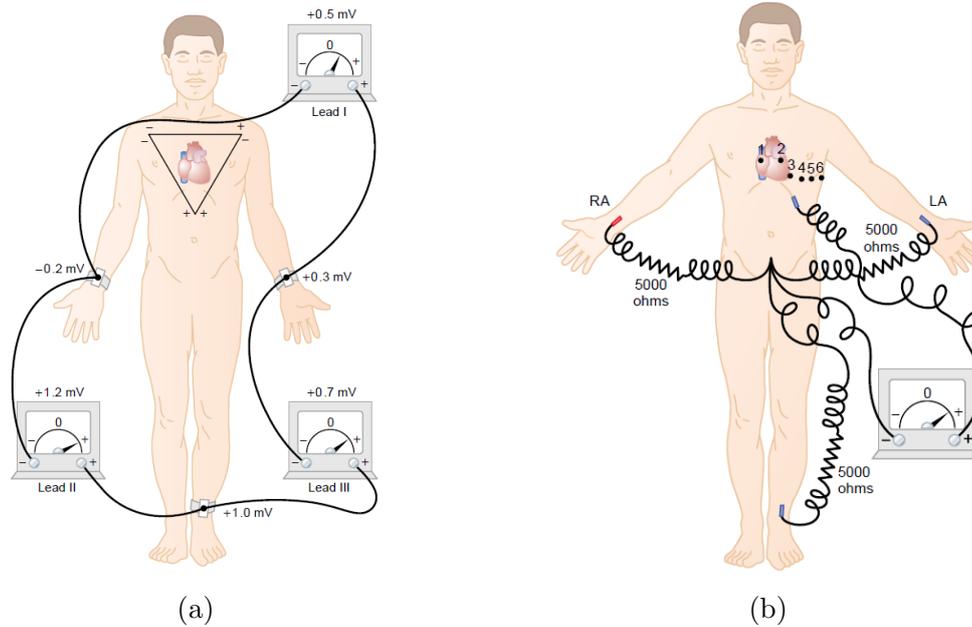


Figure 2.8: (a) Conventional arrangement of electrodes (Lead I, II and III) and (b) the precordial leads, picture adapted from [5].

Usually six standard chest leads are displaced sequentially on the anterior chest wall and recorded one at a time. Because the heart surfaces are close to the chest wall, each chest lead records mainly the electrical potential of the cardiac musculature immediately beneath the electrode. Therefore, relatively minute abnormalities in the ventricles, particularly in the anterior ventricular wall, can cause marked changes in the electrocardiograms recorded from individual chest leads [5].

2.3.1 Electrocardiography History

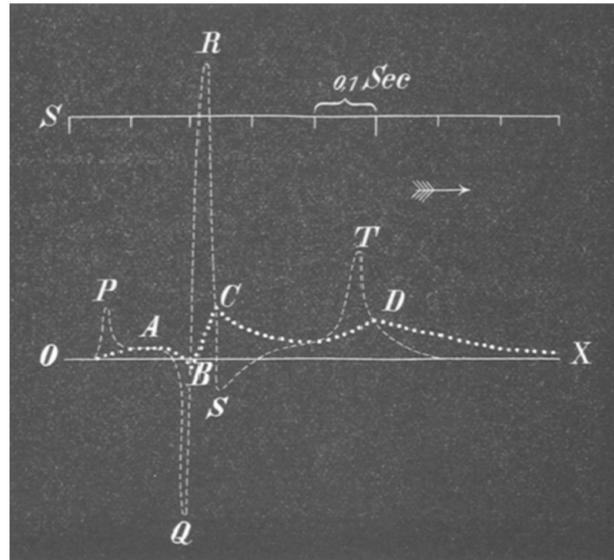


Figure 2.9: Two superimposed ECGs. Uncorrected curve is labeled ABCD. This tracing was made with refined Lippmann capillary electrometer, picture adapted from [8].

The interest in electromagnetic phenomena dates back to the ancient Greeks although it remained as an intellectual curiosity for millennia. Several researchers tried to explain and extract the fundamentals of electrostatics. Benjamin Franklin proved that lightning is an electric phenomenon, when he performed the famous kite experiment, but only in the 18th century, that static electric phenomena are explained by Coulomb's work. In 1786, Galvani first noted that electrical current could be recorded from skeletal muscles. In 1842, Matteucci demonstrated that electrical current accompanies every heart beat in a frog and thirty-five years later, Waller published the first human electrocardiogram using a capillary electrometer and electrodes placed on the chest and back of a human. Einthoven demonstrated the existence of the five deflection waves using a refinement of the Waller capillary electrometer (see Figure 2.9).

Chapter 3

State-of-art and Background

In this chapter, we will start by explaining the different modeling perspectives of the cardiac system, with a special emphasis for the signal processing perspective. Afterwards, the standard signal processing pipeline for heart sounds are explained deeply and always followed by appropriate references. We will give a special focus on Hidden Markov Model algorithms for heart sound segmentation tasks for a single and multi-channel system, as it is suggested by our thesis title. Furthermore, the nomenclature is presented and a simple explanation of the different Hidden Markov Model families are presented and discussed finally.

3.1 Heart Sound Modeling Perspectives

In this section, we are going to present the signal processing, fluid dynamic and electrical analogy perspectives of the cardiac system.

The signal processing school infers from the signal itself several characteristics (statistical, morphological, spectral and fractal) in order to identify similarity and dissimilarity patterns among healthy and non-healthy individuals. The fluid dynamic perspective tries to understand the precise mechanisms about the generation of heart sounds in the different pathological conditions. By comparing the expected output of the model with the real observations, someone might infer if it is a normal or abnormal heart sound, although under some predefined initial conditions and settings. In the electrical analogy perspective, the mechanical events in the heart are explained by establishing some analogies with a real physical electrical circuit. The abnormalities are simulated over different settings and over a controlled environment. The generated synthetic data is compared with real data in order to withdraw conclusions.

3.1.1 Fluid Dynamics Perspective

The first pioneers on this topic, tried to mathematically model a vessel's distention and contraction over the cardiac cycle. Streeter et al. [9] modeled the flow through distensible vessels by solving simultaneously the continuity and momentum equation when applied to a small vessel segment. The model runs over several approximations:

- only a one dimensional flow is considered.
- the vessel walls are incompressible.
- constant blood density ρ .
- the pressure pulse speed a in a vessel is $\frac{hY}{\rho D}$, where Y is the elastic modulus of the vessel wall, h is the wall thickness and D is the vessel diameter.

They predicted that internal pressure P in a vessel and D are correlated:

$$\frac{A}{A_0} = \frac{1}{1 - \frac{PD_0}{h_0Y}}, \quad (3.1)$$

where A is the vessel cross section area and the subscripts denote the variables magnitude at time equal to zero. This equation is discontinuous at $PD_0 = h_0Y$, although for normal pressure waves and assuming a linear elastic domain, this limit is never reached. Finally, they also showed that a and D are also correlated:

$$D = \frac{D_0 a_0}{a}, \quad (3.2)$$

Using equation 3.2, they concluded that reflections in the pressure pulse wave happen when the diameter of vessels changes.

Sikarskie et al. [10] proposed a one-dimensional mathematical model for the aortic valve vibration. The aorta artery is tapered and circular with an elastic valve at one end (proximal condition). The driving force across the valve results from a pressure gradient between the aorta artery and the left ventricle. The proposed model runs over several assumptions such as:

- The pressures in the aorta artery are assumed to be known (distal condition).
- It is only applied at the beginning of each diastole.
- The pressure in the ventricles is known.
- The valve has parabolic shape.

- The distal pressure is assumed to be constant.
- The ventricular pressure is assumed to be:

$$P_v = -\Delta P_v - \frac{\partial P_v}{\partial t}t, \quad (3.3)$$

where ΔP_v and $\frac{\partial P_v}{\partial t}$ are the initial pressure drop and the slope respectively.

In their work [10], the aortic valve vibrates according to the following differential equation:

$$M \frac{\partial V}{\partial t} + f(U) = A_v \rho g (P - P_v) \quad t > 0, \quad (3.4)$$

where P_v is the pressure in the ventricles, M is the valve mass, $V = V(z, t)$ is the average velocity, g is the acceleration of the gravity and $U = U(t) = \int_0^t V(0, t) dt$ is the valve displacement. From the presented model, several parameters can be extracted as a function of time: the valve displacement and pressure, velocity and pressure downstream in the aorta artery. The authors concluded that stiffness and effective mass (valve mass plus the fluid trapped in the valve leaflets) play a major role in the aortic valve vibration [10].

Blick et al [11] also proposed a one-dimensional model for the aortic valve vibration. In their model, the valve motion is described by the following differential equation:

$$m\ddot{x} + d\dot{x} + kx = \Delta p \pi a^2, \quad (3.5)$$

where m is the effective mass of vibration, d is the damping factor, k is the elastic stiffness and Δp is the pressure gradient across the valve. They experimentally measured during catheterization the instantaneous pressure gradient across the semi-lunar valve during a diastole. They observed that the pressure gradient increases linearly with time until a time t_1 is reached, and afterwards it remains essentially constant until the end of the diastole. Using such knowledge, they solved the differential equation (3.5), by approximating Δp with a linear function t for time $t < t_1$ and by a constant function for time $t > t_1$. The solutions show a good agreement with experimental measurements of the vibrations of a normal stent-mounted porcine valve incorporated in a hydraulic chamber in the cardiovascular system. They concluded that the amplitude of the sound pressure generated P_{amp} by the aortic valve is directly proportional to its velocity and it

can be expressed as:

$$P_{amp} = \frac{G\pi\omega\rho x_0 a^2}{R} \quad (3.6)$$

where R is the distance from the valve to the recording spot, G proportionality constant, x_0 the velocity of centerline deflection, a valve radius and ω frequency of the valve vibration.

3.1.2 Electrical Analogy Perspective

In the later 19th century, Frank Windkessel [12] described the heart and the systemic arterial system as a closed hydraulic circuit. In his analogy, the circuit contained a water pump connected to a chamber, filled with water except for a pocket of air. As it's pumped, the water compresses the air, which in turn pushes the water out of the chamber. This analogy resembles the mechanics of the heart and is known as Windkessel model [12]. The simplest one entitled *Two Element Model* is represented in Figure 3.1. This model takes into account during a cardiac cycle, the effect of arterial compliance and the total peripheral resistance. In the electrical analogy, the arterial compliance (C) is represented as a capacitor, peripheral resistance of the systemic arterial system (R) is represented by a resistor. The blood flow is analogous to a current flowing in a circuit and the blood pressure in the aorta ($P(t)$) is modeled as a time-varying electric potential [13].

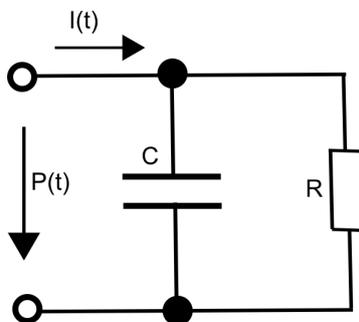


Figure 3.1: The 2D Windkessel model of the arterial system (adapted from [13]).

The system is mathematically governed by the following differential equation:

$$I(t) = \frac{P(t)}{R} + C \frac{\partial P(t)}{\partial t} \quad (3.7)$$

Although this model is a good approximation of the arterial system's behavior, it has some obvious limitations when used to model peripheral arterial blood pressure. One of its major weaknesses is that it does not account for the propagation

effects through the vessels and assumes that the pressure rise occurs simultaneously in the entire arterial tree [13].

More advanced models, study the relationship among different cardiac compartments, as the one described in Figure 3.2

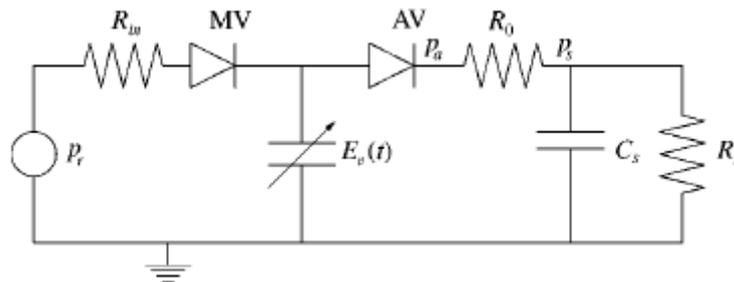


Figure 3.2: A three-element modified Windkessel model (adapted from [14])

This electrical analogy describes the interaction between the left atrium, left ventricle and the arterial tree. In these models, the mitral and the aortic valves (indicated by diodes) allow only bloody flow in one direction. When the ventricular pressure exceeds the aortic pressure, the aortic valve (AV) opens and blood flows into the arterial tree. It closes when ventricular pressure falls below the aortic pressure. The mitral valve (MV) behaves in a similar fashion.

Although it is an excellent model and far more accurate than the previous $2D$ Windkessel model, it still fails to predict some physiological phenomenons such as the isovolumetric contraction of the ventricles, ventricular injection acceleration or deceleration phases and finally an insensitivity to different loading conditions in the ventricles.

3.1.3 Signal Processing Perspective

The signal processing perspective looks to the data extracted directly from the subject in order to infer and extract knowledge of the cardiac system. This is indeed the perspective chosen by us and we are going to devote the remaining of the current section in explaining it step-by-step.

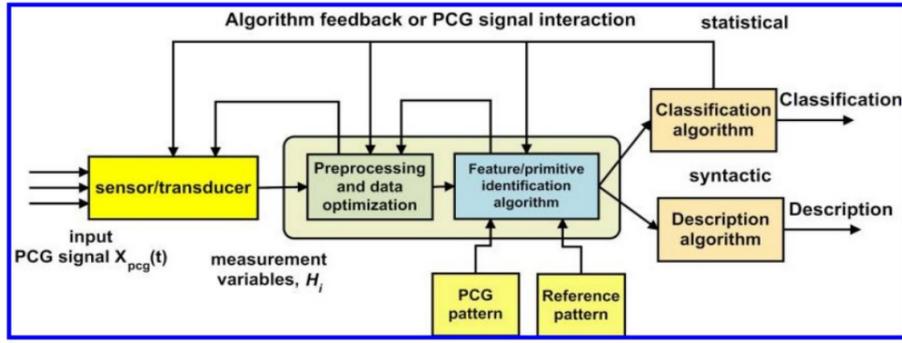


Figure 3.3: PCG pattern recognition system and signal flow diagram, which represent the principal strategy for PCG classification, adapted from [15]

Usually, standard heart sound segmentation algorithms start by converting an analog signal to a digital one, this job is done by the AC/DC converter. To do so, a continuous signal is sampled using an impulse train with a decay frequency (among impulses), which should be at least two times faster than the frequency of the observed phenomenon, therefore obeying the Nyquist rate constraint, see Figure 3.3. The next step, involves filtering a signal with a filter, the aim is to remove undesired sources or components from the original raw signal, see Figure 3.3. The following step, involves segmenting the filtered signal. To do so, several techniques have been proposed in the literature, the simplest ones involve computing an envelopegram on a specific domain, see Section 3.2.2. More advanced techniques apply statistical models, like the well known hidden Markov model, in order to infer hidden states under the original signal, we will devote the entire Section 3.3 to explain the most important contributions, see Figure 3.3. The next step is usually the extraction of features from the segmented signal, see Figure 3.3. Usually, researchers look for characteristics that are very discriminative, what we mean, for a transformation where samples from different classes are furthest way as they can be, see Section 3.2.5. Finally, in the last step a classifier is going to make a decision (based on the features extracted in the previous step) about the most likely class for an input signal, see Figure 3.3. Different classifiers are going to be presented in Section 3.2.6.

3.2 Background on Signal Processing for Heart Sounds

3.2.1 Denoising and Signal Enhancement

Heart sound recordings are very often corrupted by external and noisy sources such as: respiration sounds, chest movements, noise from the contact between the recording device and the skin ("shear noises"), acoustic damping through the bones and tissues external sounds from the environment, etc.

The existing methods usually apply digital band pass filters (most commonly IIR or FIR filters) as a simple denoising method [16]. The cut-off frequencies of the filters are determined empirically, although several statistical results showed that the major concentration of energy, for both first heart sound (S1) and second heart sound (S2), is below 150 Hertz (Hz). We have also confirmed this in the DigiScope dataset [17], in which the frequency content of the *S1* and *S2* heart sounds is around 30 – 80 Hz. The S1 peak is around 50Hz and the S2 peak is around 60Hz (see Figure 3.4). Similar results can be found in [18].

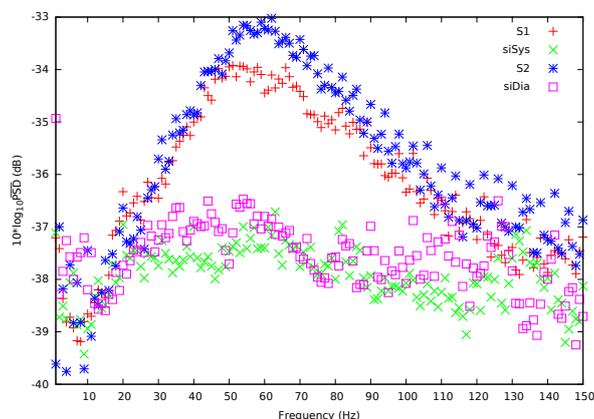


Figure 3.4: Average power spectral density (PSD) for each state over the frequency range [0 – 150]Hz. The PSD is computed using the annotated events and through a short time Fourier transform (STFT). The STFT is computed using a Hamming window of 0.25 seconds length with 95% of overlapping.

Other alternative methods use wavelet reconstruction to denoise the PCG signals. Several authors searched for the optimal wavelet decomposition which gives the highest signal-to-noise ratio and the corresponding wavelet reconstruction algorithm, regardless of the PCG signal. Messer et al. [20] stated that the wavelet coefficients resulting from noisy sources are smaller in amplitude when compared to heart sound sources, over different time scales. By only retaining large coeffi-

cients for further reconstruction and neglecting the others, the noisy components are effectively suppressed.

3.2.2 Segmentation - Introduction

Heart sounds segmentation is a fundamental step in our pipeline. Our goal is to extract the heart sound components in a PCG signal. These algorithms are divided into two main approaches: those that use a synchronized ECG reference signal and those that do not. It is typically preceded by a pre-processing step, and followed by feature extraction and classification steps in a traditional pattern recognition approach (see Figure 3.3).

3.2.3 Segmentation - Single Channel System

Using a single channel to segment the PCG signal is a demanding task for reasons aforementioned explained. On the other hand, such algorithms do not require extra hardware or clumsy wiring arrangement for data acquisition. It is still the most used setting in clinical environments and therefore many researches still try to identify $S1$ and $S2$ sounds by several means of signal processing and statistics without using ECG as a reference. Many methods of heart sound segmentation have been studied over the past few decades. The standard algorithms extract envelopograms from the original signals and further temporal criteria in order to detect and classify heart sounds. Liang et al [21] were the first to compute the Shannon energy envelopogram from the original signal. This non-linear transformation emphasizes medium signal intensities, which are highly correlated with heart sound components and attenuates low and high signal intensities, which are highly correlated with noisy signal segments.

They applied a peak-picking algorithm over the Shannon energy envelopogram in order to select a set of heart sound candidates. A subset of candidates are retained according to statistical (mean and variance of the peak intervals) and morphological (amplitude of the normalized Shannon energy envelopogram) criteria. Finally, the remaining candidates are identified as $S1$ or $S2$ according to some temporal criteria, e.g: the diastolic period is always longer than the systolic one. Moukadem et al [22] extracted the Shannon energy envelopogram not from the original signal but from the frequency spectrum calculated using the S-transform. In this way, the authors expect to improve the algorithm performance by emphasizing straightforward the medium frequencies in the local spectrum, which are assumed to be correlated with heart sounds. Finally, the $S1$ and $S2$ are classified using

the temporal criteria presented by Kumar [23]. Other envelopograms can be extracted using the Hilbert transform. Sun et al [24] extracted an envelopogram from a decimated signal by using the real part of a complex analytic signal. The instantaneous frequency is the derivative of the imaginary part of complex analytic signal. More advanced information theory techniques use complexity signatures in order to identify and distinguish heart sounds from noisy segments, such as in Nigam et al [25].

Oliveira et al. [26] proposed a creative solution using information theory techniques too. In their work, the signal is wrapped around a cylinder and divided in two sections. In the first section, delay vectors are collected and stored in an embedding matrix from the original signal. In the second section, the data is discarded and not post-processed. When this imaginary cylinder spins, delay vectors are displaced from one section to the other one, and as a net result an entropy gradient is generated in the two extremes of the cylinder. This cylinder responds in a unique way to heart sound components and noisy segments. The post-processed signal is further used to segment the PCG signal. Another set of algorithms claim that wavelet transformations (a time-scale representation) are a more efficient signal representation than the previous transformations. These algorithms search for an optimal wavelet decomposition where the heart sounds are more easily distinguishable in some signal sub-levels. Castro et al. [27] proposed a segmentation algorithm based on the time and scale characteristics of the signal. The heart sounds are classified by using first the heuristic rules proposed by Liang [28]. Later these are refined using the relative signal energy distribution of the Daubechie wavelet in the sub-band details 3 and 4. They concluded that in these sub-bands, the $S2$ component exhibits higher energy when compared to the $S1$, although no explanation is given. Wang et al. [29] proposed a tracking algorithm based on wavelet transform to separate the $S1$ and $S2$ from other extra-sounds such as murmurs or clicks. In the first stage, the fundamental components (sounds) are separated from the noisy environment by using an adaptive sub-level tracking algorithm. In the second stage, the Shannon energy envelopogram is computed from the post-processed signal. Finally, $S1$ and $S2$ heart sounds are classified according to some temporal criteria, similar to Liang and Hartimo [28].

Kumar et al. [23] extracted frequency markers through the fast wavelet decomposition in order to fragment the signal into heart cycles. These frequency markers are physiologically motivated by an accentuated pressure difference found across heart valves, both in native and prosthetic valves. The $S1$ and $S2$ are classified in each heart cycle according to some temporal criteria. Later on, Kumar

Author	Year	Dataset Description	Transformation
Liang [21]	1997	37 normal and abnormal heart sounds.	Shannon energy envelopogram computed over the pre-processed signal.
Moukadem [22]	2013	40 normal and 40 pathological cardiac sounds.	Shannon energy envelopogram computed over the S-Transform.
Sun [24]	2014	Michigan Heart Sound database ^a .	Envelopogram computed from the complex analytical signal.
Oliveira [26]	2014	Pascal Challenge dataset ^b .	Envelopogram computed from entropy gradient.
Castro [27]	2013	Pascal Challenge dataset ^c	Shannon energy envelopogram computed over a Daubechie sublevel 4 decomposition.
Wang [29]	2005	30 normal and abnormal heart sounds.	Shannon energy envelopogram computed over a processed Daubechie sublevel 4 decomposition.
Kumar [23]	2006	55 patients with different prosthetic valve implants.	Shannon energy envelopogram computed over a Daubechie sublevel 5 decomposition.
Kumar [30]	2007	5 normal and abnormal heart sounds with mitral regurgitation.	Simplicity envelopogram computed from the Daubechie sublevel 6 decomposition

Table 3.1: A summary of the standard heart sound segmentation algorithms.

^ahttp://www.med.umich.edu/lrc/psb_open/html/repo/primer_heartsound/primer_heartsound.html

^b<http://www.peterjbentley.com/heartchallenge/>

^c<http://www.peterjbentley.com/heartchallenge/>

[30] used a simplicity filter to separate and detect $S1$, $S2$ and $S3$ sounds from background noise. The third heart sound is detected by setting a threshold on the low frequency content of the Kumar filter.

3.2.4 Segmentation - Multi Channel System

In a multi-channel PCG system, the segmentation task is relatively easier when compared to a single channel system since we have more data at our disposal to process. If one channel is corrupted, there might be a chance that the others are not (e.g: shear noises) or even if they are, one can apply noise cancellation algorithms to remove it. Assuming that noisy sources are uncorrelated over the different channels. In 2009, Li [31] developed a multi-channel acquisition system

using five different auscultation spots to remove the delay between sensors and to synchronize the signals with respect to the different cardiac and breathing cycles. Later on Liu [32] used such a system to segment heart sounds. The homomorphic envelopogram is extracted from the original signal at different chest spots through the proposed frequency alterable homomorphic filtering method. Finally, heart sounds are detected and classified by using a simple tracking algorithm.

Another possibility is to use the QRS complex and the T-wave from the ECG channel in order to localize the S1 and S2 waves. In this case, the signal processing ‘complexity’ is transferred from the PCG to an ECG analysis. This is due to the fact that the ECG signal is a recording of an electromagnetic phenomenon and in resting conditions the interference from external sources are easily controlled. In contrast, the PCG is a recording of an acoustic phenomenon and therefore interference problems from external sources are more problematic and not so easy to control. Chen et al [33] used an ECG analysis method in order to identify the heart sounds. Springer et al. [34] generated synthetic PCG annotations using the agreement of four R-peak and four end-T-wave detectors over a ECG channel. Although interesting, the aforementioned solutions do not try to modulate mathematically both ECG and PCG channels simultaneously but only one channel is analyzed, while ignoring the other. This is a waste of data and moreover an oversimplification from the current models, since these do not take into account the important and evident interactions among the two systems. Therefore it remains important and necessary to design a multi-channel algorithm which does not neglect the interaction between the electrical (ECG) and the mechanical (PCG) components of the cardiac system.

3.2.5 Feature Extraction

The extraction of features is the following step. The selected features are expected to discriminate normal from abnormal classes, to be uncorrelated among themselves (no redundancy) and also in a reduced number, if possible (no curse of dimensionality). The PCG features vary according to the domain and the application, from the signal itself acquired through a stethoscope to more sophisticated ones such as: frequency spectrum, phase delay or the energy profile of a signal [15].

Bentley et al. [35] used Choi Williams Distribution (CWD) features to classify normal and abnormal heart sounds. They argue that CWD is a more discriminate than other time-frequency (T-F) representations. According to them, a simple description of the T-F distribution allows the analysis of the heart valves over

different conditions. However, they highlight the need for a more comprehensive evaluation (e.g: a larger population to test).

Other standard approaches do not extract a single but multiple features from the PCG signal. The rationale follows the fact that the distance among the classes is expected to increase as the feature space increases. On the other hand, it is very common that the extracted features are correlated spatially or temporally, therefore a feature selection algorithm is necessary in order to remove these dependencies. Liang [36] used several features in order to classify normal and abnormal heart sounds. The feature vector included 95 elements extracted from the original signal, three detail sub-bands (3rd, 4th and 5th) and one average sub-band (6th) signal obtained from the discrete wavelet transform. Finally, using Principal Component Analysis (PCA), a subset of features is retained and further used in a neural network classifier.

3.2.6 Classification

The artificial neural network (ANN) is one of the most widely used machine learning-based approaches for heart sound classification, although relatively little work has been done in optimizing the network architecture. Typically, the features used as input are: time-scale, time, frequency, time-frequency and complexity based features. Akay et al [37] combined time-scale (wavelet) and morphological features with an ANN for the automatic detection of coronary heart diseases (CAD). These are combined with patient-based features (gender, age, weight, height) and fed into a fuzzy neural network. Sepehri et al. [38] is an example of frequency-based features. They have identified five frequency bands where the spectral energy is significantly different from normal and pathological patients. These are further used as input of an ANN.

Another successful machine learning algorithm applied to heart sound classification are the support vector machine (SVM). Since SVMs are another form of supervised machine learning, the features chosen are rather similar to those based on ANN approaches. Zheng et al [39] decomposed heart sounds using wavelet decomposition methods, afterwards the total energy and the sample entropy of each sub-level are used as input to a SVM. In contrast, Maglogiannis et al. [40] used information and frequency based features. He used Shannon energy and frequency features from four frequency bands (50-250, 100-300, 150-350, 200-400 Hz) to develop an automated diagnosis system for the identification of heart valve diseases based on an SVM classifier.

A number of researchers use k-nearest neighbour (k-NN) algorithms to classify

abnormal heart sounds. Like the aforementioned strategies, there is not a consensus about the most suitable feature or set of features to be used in a K-NN algorithm, although several attempts have been made, such as Delgado-Trejos [41] used a K-NN classifier to detect cardiac murmurs. They compared and concluded that fractal features outperformed significantly the perceptual and spectral features. According to the authors, this might be explained by the existence of long range (fractal) correlation among the distinct classes. Oliveira et al. [42] expanded Trejos's [41] work, by proposing new fractal features, which are based on the distinct signatures of complexity and self-similarity registered from the normal and pathogenic cases. On the other hand, Bentley et al. [43] claim that discrete wavelet transform features outperform morphological features (time and frequency features from S1 and S2) when performing heart sound classification using a k-NN algorithm.

Finally, hidden Markov models (HMM) are also applied to heart sound classification problems. In such cases, the posterior probability of the heart sound signal when given a set of trained HMM parameters is used to differentiate between healthy and pathological recordings. For example: Wang et al. [44] used HMM's to classify abnormal heart sounds (murmurs). They used time-domain features, short-time Fourier transforms and mel-frequency cepstral coefficients (MFCCs) features as evidence. Later on, they also used MFCCs to extract representative features and developed a HMM-based method for heart sound classification [45].

3.3 Heart Sound Segmentation using Statistical Models

In the last decade, several approaches have been proposed to assign features extracted from the PCG to the different segments/states, e.g. ANN [46], Deep Neural Networks (DNN) [47], k-NN [48], SVM [49] and HMM.

Among these, HMM's and their variations have the advantage of naturally modeling the sequential nature of heart sound signals. Hidden Markov models (HMM) emerged in the 1970s as a very important statistical framework with application in several pattern recognition applications, such as speech and handwriting recognition.

3.3.1 Hidden Markov Models

Recently, HMM's have shown to be effective when modeling heart sound signals: in Gill et al. [50], the signal is pre-processed and a subset of candidates (peaks) are extracted from the homomorphic envelopogram. These candidates are further classified using a discrete-time HMM. In Chung[51], heart sounds are detected and classified using first a left-right HMM model (the first state is assumed to be known) and later a fully-connected HMM. The emission probability distribution in each state is modeled by a multivariate Gaussian mixture model distribution.

3.3.2 Semi-Hidden Markov Models

In the 1980s, Hidden semi-Markov models (HSMM's) with a nonparametric state occupancy distributions were first proposed in speech recognition by Ferguson [52] and later on other several researchers proposed statistical inference methods. Guedon et al. [53] analyzed branching and flowering patterns in plants using a HSMM. They proposed a non-parametric EM algorithm in order to infer emission and sojourn time parameters using forward and backward quantities.

Finally, Sansom et al. [54] modeled the emission and sojourn time probability distributions using several distinct approximations. They also derived the corresponding re-estimation equations.

Keiichiro et al. [55] incorporated re-estimation routine for the emission probability distribution on a HSMM. They designed a Japanese speech recognition system based on HSMM, in which the emission probability distributions are also approximated with a GMM. Keiichiro et al. [55] proposed re-estimation equations for both emission and sojourn time probability distribution but without providing any mathematical proof or any strong experimental validation.

Schmidt et al.[56] implemented a hidden semi-Markov model (HSMM) [57] using the homomorphic filtering envelopogram as an observation to the system. This model extends the traditional HMM, as state duration distributions are no longer strictly approximated by a geometric distribution. The state duration distribution function is modeled by a Gaussian distribution, where the systolic and diastolic duration parameters are estimated through autocorrelation analysis of the homomorphic filtering envelopogram. Springer et al. extended Schmidt's algorithm mainly by studying the use of two different emission probability distributions, SVMs [58] and logistic regression functions [34], thus obtaining state-of-art performance for PCG segmentation.

Recently, Johnson et al. [59] proposed an explicit-duration hierarchical Di-

richlet process hidden semi Markov model (HDP-HSMM) and also a sampling algorithm for efficient posterior inference. He observed that HDP-HSMM outperformed the typical HMM's and HSMM's in real and synthetic experiments.

3.3.3 Coupled Hidden Markov Models

Coupled hidden Markov models (CHMM) have been proposed by Brand in [60] as a generalization of the HMM for a multi-channel system. He observed that CHMM outperformed HMM when classifying visual tasks (two-handed actions). Montazeri [61] presented a CHMM, where the state inter-dependencies are modeled using a stochastic matrix. They proposed a novel apnea-bradycardia detection method for preterm infants, integrating a phase of multivariate feature extraction from the ECG, and a phase of time-series characterization through the proposed CHMM. Despite, being quite a recent method, CHMM have been successfully applied in several distinct fields: forensic electronics [62], genetics [63], audio-visual speech recognition system [64] and target tracking [65].

3.4 Background

After seeing the state-of-art we decided by hidden Markov models since HMMs and their variations have the advantage of naturally modeling the sequential and periodic nature of the heart sounds. In a HMM, each state represents a real stage of the cardiac cycle and transitions among states are explicitly modeled. Such a flexibility allows the HMM to easily obey some fundamental state transition constraints that govern each cardiac cycle, e.g.: it is forbidden to move from an S1 state directly to an S2 state.

3.4.1 Nomenclature

To denote observation or state sequences of length l , starting at a time instant t_1 , we use the following notation $X_{t_1:t_1+l}$ and $S_{t_1:t_1+l}$, respectively. We denote s_t and x_t the state and the observation at time instant t respectively. We further denote $s(r)$ as the r^{th} visited state, $v(r)$ the corresponding time instant when such transition happened and $u(r)$ the sojourn time of the r^{th} visited state. Finally, we denote R as the total number of state transitions in a signal of length T .

3.4.2 Hidden Markov Models

The PCG signal is most likely a non-stationary sequential time series, meaning that the data is generated by a distribution which evolves over the time. Although for simplicity and convenience, we are going to assume a stationary regime in order to satisfy the Markov requirements (assumptions). This is a serious drawback in arrhythmic sounds, where the waiting time distributions are constantly and rapidly changing over the time. HMM's are probabilistic models, where the observation sequence $X = x_1, x_2, \dots, x_T$ depends on the underlying hidden state sequence $S = s_1, s_2, \dots, s_T$ and the unobserved Markov process [66]. In such models, the latent variables (hidden states) are discrete, describing which component of the mixture is activated when the data x_t was generated at time t . A homogeneous hidden Markov model assumes that the state transition probability distribution A is constant over time. In this case, the (i^{th}, j^{th}) element is:

$$a_{ij} = Pr(s_t = j | s_{t-1} = i), \quad \sum_{i \in S} a_{ij} = 1 \quad \text{with} \quad a_{ij} \geq 0, \quad (3.8)$$

is the probability of being in state j knowing that the previous state was i , and such probability is independent of the current evaluation time t [51]. The specification of a HMM is completed by defining the emission probability distribution E . Its elements are $e_k(x_t)$, the probability of having observed x_t knowing that the state at time t is k .

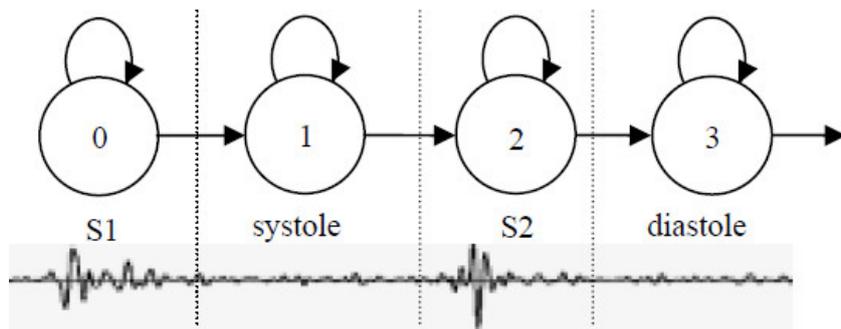


Figure 3.5: Four state HMM for a cycle of a normal heart sound signal (adapted from [51]).

The initial state distribution π is discrete:

$$\pi_k = P(s_1 = k) \quad \text{with} \quad \sum_{k \in S} \pi_k = 1. \quad (3.9)$$

The π_k represents the probability that a sequence starts at state k .

The complete likelihood [66] of a HMM given S , X and Θ is:

$$P(X, S, \Theta) = \pi_{s_1} \left\{ \prod_{t=2}^T A_{s_{t-1}s_t} \right\} \prod_{t=1}^T e_{s_t}(x_t), \quad (3.10)$$

where $\Theta = \{\pi, A, E\}$ denotes the set of HMM parameters. Then given a HMM, heart sound segmentation is obtained by looking at the state sequence S which maximizes the likelihood in (3.10).

3.4.3 Semi-Hidden Markov Models

One of the major limitations when using a simple and standard HMM's to segment heart sounds, is based on the fact that the Markovian hypothesis imposes restrictions on the distribution of the sojourn time in a state, which should be geometrically distributed (in discrete case). This is an unrealistic assumption or very unlikely for example in heart sound signals because: a) the geometric distribution monotonically decreases over the time and as a result the most likely sojourn time duration is always a few time steps and b) the state transition probabilities are constantly changing over time (e.g children arrhythmias) [58]. A more natural choice is to smooth the Markovian assumption in order to:

- Allow any arbitrary distribution of the sojourn time in any state.
- Allow the Markovian hypothesis to exist although the conditional independency between past and future visited states when given the current state is not measured in the usual time scale (sample by sample) but in the jump scale of the system. In other words, if we know the past visited states, (denoted by $s(0) \dots s(n-1)$) and the corresponding time instants when such transitions happened (denoted by $v(0) \dots v(n)$) as well as its present state (denoted by $s(n)$), the future visited state (denoted by $s(n+1)$) and the sojourn time in the present state (denoted by $u(n)$) depend only on the current state $s(n)$, as it is showed in equation 3.11.

$$\begin{aligned} P(s(n+1) = j, u(n) = k | s(0), \dots, s(n); v(0) \dots v(n)) \\ = P(s(n+1) = j, u(n) = k | s(n)) \end{aligned} \quad (3.11)$$

HSMM's [66] are statistical models that do follow this more flexible Markovian assumption. The major advantage of such models, is based on the fact that the sojourn time is explicitly and intrinsically modeled inside of a HMM's. To do so,

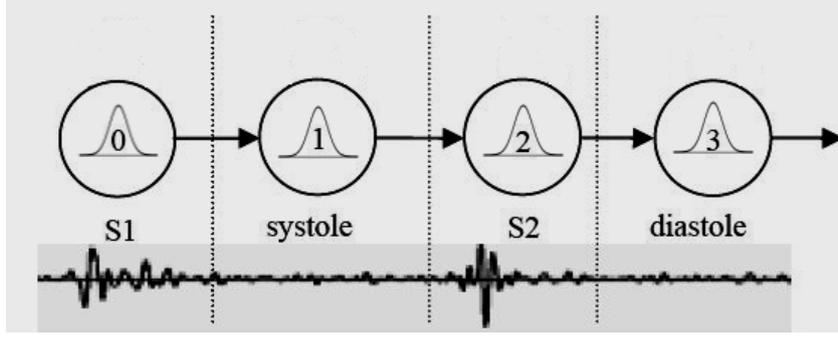


Figure 3.6: Four state HSM for a cycle of a normal heart sound signal.

we start by denoting D , the sojourn time probability distribution. Its elements are $d_k(u)$, the probability of spending u units of time in the state k , as it is shown in equation 3.12.

$$d_k(u) = P(s(n+1) \neq k, v(n) = u | s(n) = k). \quad (3.12)$$

We further define $d_k^*(u)$ as the survivor function of the sojourn time.

$$d_k^*(u) = \sum_{v \geq u} d_k(v). \quad (3.13)$$

In the HSM's, the state transition probability distribution Γ is constant over the time. In this case, the (i^{th} , j^{th}) element is:

$$\gamma_{ij} = Pr(s(r) = j | s(r-1) = i), \quad \sum_{i \in S} \gamma_{ij} = 1 \quad \text{with} \quad \gamma_{ij} \geq 0, \quad (3.14)$$

the probability of the r^{th} visited state being j knowing that the $r-1^{\text{th}}$ visited state was i [53] (see Figure 3.6). Since we are assuming an homogeneous semi-Markov chain, the $d_k(u)$ and γ_{ij} quantities are invariant to the evaluation time, regardless the time scaled used. The complete likelihood of a state sequence S given a set of observations X and a model Θ is expressed as:

$$\begin{aligned} L(X, S, \Theta) &= p(X, S | \Theta) \\ &= \pi_{s_1} d_{s_1}^*(u(1)) \left\{ \prod_{r=2}^{R-1} \gamma_{s(r-1), s(r)} \times d_{s(r)}(u(r)) \right\} \times \\ &\quad \times \gamma_{s(R-1), s(R)} \times d_{s(R)}^*(u(R)) \times \prod_{t=1}^T e_{s_t}(x_t), \end{aligned} \quad (3.15)$$

where $\Theta = \pi, \Gamma, E, D$ denotes the set of HSM parameters. Then given a HSM,

heart sound segmentation is obtained by looking at the state sequence S which maximizes the likelihood in (3.15).

3.4.4 Coupled Hidden Markov Models

The coupled hidden Markov models (CHMM's) are an extension of the typical hidden Markov model to a multichannel system, where each channel, represents the evolution of an underlying generation process and the state transition probabilities depends on the current state on all the channels [61]. In order to overcome, the exponential growth in complexity as the number of channels increases, it is advisable to use the following simplification:

$$P(s_t^{c'} | s_{t-1}^1, \dots, s_{t-1}^C) = \prod_{c=1}^C P(s_t^{c'} | s_{t-1}^c), \quad (3.16)$$

where C is the number of independent channels in our CHMM. Using such a simplification, the state transition probability A , and its elements are further defined as:

$$a_{ij}^{c'c} = P(s_t^c = j | s_{t-1}^{c'} = i), \quad \sum_{i \in M(c')} a_{ij}^{c'c} = 1 \quad a_{ij}^{c'c} \geq 0, \quad (3.17)$$

the probability of being state j channel c knowing that in the previous time instant, the system was in state i channel c' , where $M(c')$ denotes the set of states in the channel c' . Since we are assuming an homogeneous coupled Markov chain, the state transition quantities $a_{ij}^{c'c}$ are invariant regarding the evaluation time t , a schematic diagram of a CHMM's is presented in Figure 3.7.

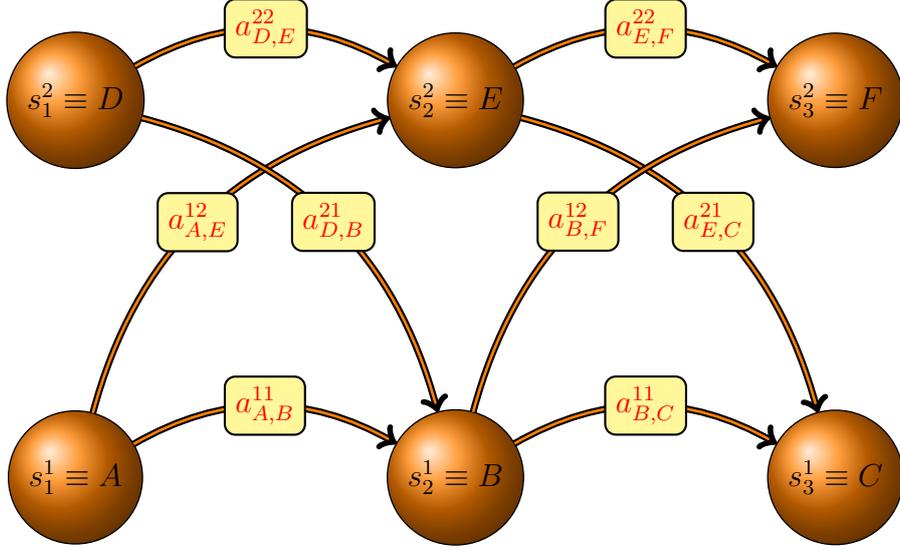


Figure 3.7: The CHMM scheme for a particular case when two channels are completely connected. The first layer is composed by the nodes (A,B,C) and the second layer is composed by the nodes (D,E,F). Each node is a hidden state and the state transition probabilities are denoted by arrows [61]).

The emission probability distribution is denoted by E and its elements are:

$$e_j^c(x_t^c) = P(x_t^c | s_t^c = j), \quad (3.18)$$

the probability of observing x_t^c in the state j channel c at time t . Finally π is the initial state probability distribution, and its elements are:

$$\pi_j^c = P(s_1^c = j) \quad \sum_{j \in M(c)} \pi_j^c = 1 \quad \pi_j \geq 0, \quad (3.19)$$

the probability that a process in the channel c starts at the state j . The complete likelihood of a state sequence S given a set of observations X and a model Θ is expressed as:

$$L(X, S, \Theta) = \prod_{c=1}^C \pi_{s_1^c}^c e_{s_1^c}^c(x_1^c) \times \prod_{t=2}^T \prod_{c'=1}^C a_{s_{t-1}^c s_t^{c'}}^{c'c} e_{s_t^{c'}}^c(x_t^c), \quad (3.20)$$

where $\Theta = \pi, A, E$, denotes the set of CHMM parameters. Then given a CHMM, heart sound segmentation is obtained by looking at the state sequence S which maximizes the likelihood in 3.20.

Chapter 4

Heart Sound Segmentation using a Subject Dependent Approach

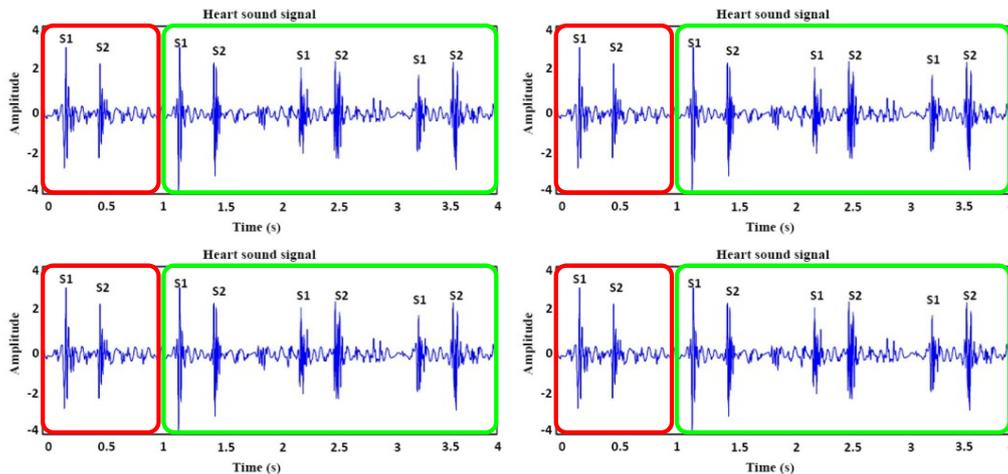


Figure 4.1: A diagram of the proposed subject-dependent approach when using the Digiscope dataset. The red-box represents the annotated data used to train our HMM's, the green-box represents the data used to test our HMM's.

This chapter is based on the following contributions:

- J. Oliveira, T. Mantadelis, and M. Coimbra, "Why should you model time when you use Markov Models for analysing heart sounds", in Proc. IEEE EMBC 2016, Orlando, USA, Aug 2016.
- J. Oliveira, T. Mantadelis, F. Renna, P. Gomes and M. Coimbra, "On modifying the temporal modeling of HSMM's for pediatric heart sound segmentation", in Proc. of IEEE International Workshop on Signal Processing Systems, SiPS 2017, Lorient, France, Oct 2017.

4.1 Introduction

4.1.1 Why implement a subject dependent approach?

In this chapter, we are going to address the problem of heart sound segmentation in a subject dependent approach, i.e: we split the signal from each subject into two parts. The first part is used to train our HMM's and HSMM's and the second part of the signal is used to evaluate our HMM's and HSMM's. We choose this strategy for several reasons:

- The parameters extracted during the training phase are close to the optimal ones needed to set up properly our HMM's or HSMM's, since we are using data from the subject itself.
- Using generic parameters extracted from an independent training data, might result in a poor model initialization, since the optimal HMM and HSMM parameters diverge from subject to subject and therefore the algorithm might not converge to the optimal solution.
- The re-estimation routines for the HMM's and HSMM's are more likely to converge to the optimal solution since the initial HMM parameters (inferred from the training set) are expected to be close enough to the optimal set.

In standard HMM's, the sojourn time (waiting time) is geometrically distributed over all states. This is an unrealistic assumption in heart sound signals: a) the geometric distribution monotonically decreases over the time and as a result the most likely sojourn time duration is a few-time steps and b) the state transition probabilities are constantly changing over time (e.g children arrhythmias) [58]. Nowadays, the paradigm is shifting and the sojourn time distributions are explicitly and intrinsically modeled in the HMM, leading to a new class of statistical models known as HSMM. Recently, these models have been proposed as an alternative to the typical and standard HMM and several studies have been made in the topic [58]. Mostly are related to the signal representation and the optimal distribution used as an approximation for the emission probability. We do expand and raise other important questions concerning heart sound segmentation problems when using HSMM's.

4.1.2 Objectives

The goal of this chapter, is to answer three fundamental questions:

1. *Why should you model time?* We answer this question by using an exhaustive cross-validation to compare 5 different models, over a real auscultation dataset.
2. *How should you model time?* To answer this question, we compare the HSMM performance over three different parametric distributions and one non-parametric probability density mass function [67] as an approximation for the sojourn time distribution.
3. *Where should you model time?* We answer this question by proposing a confidence metric function based on the conditional distribution of one observation given all the rest. Finally, by setting a threshold over this metric function we are able to exclude highly uncertain sample classifications from our model.

4.2 Methodology

4.2.1 Model Architecture

In this section, each state i of a HMM or HSMM corresponds to an element of the heart sound signal $\mathcal{S} = \{S1, siSys, S2, siDia\}$ because the signal characteristics are thought to be homogeneous. For simplicity, our HMM's and HSMM's will ignore $S3$, $S4$ and murmur sounds thus implementing a four state HMM or HSMM respectively, see 3.5 and 3.6.

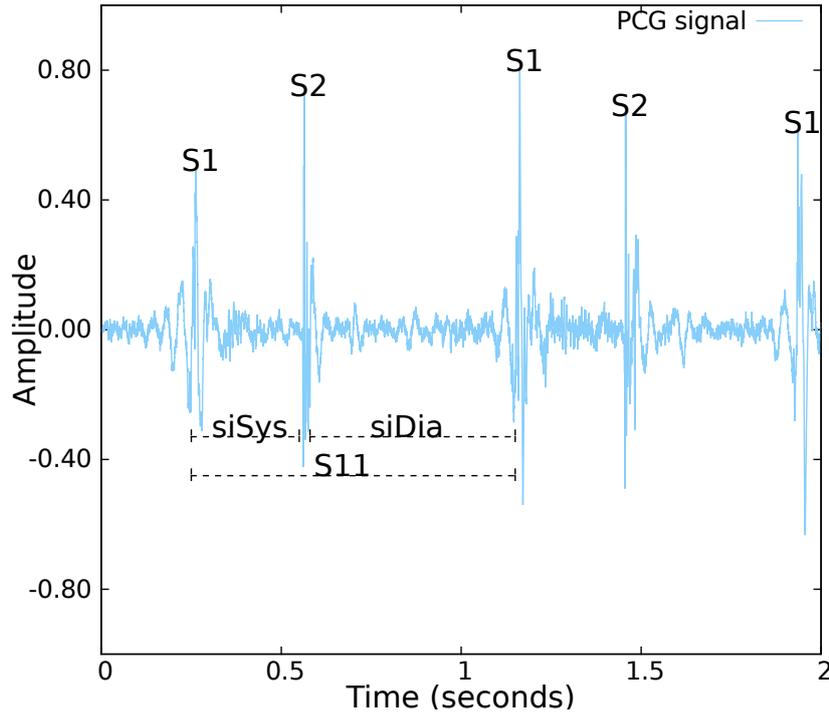


Figure 4.2: A PCG signal and its corresponding elements.

4.2.2 HSMM and HMM distributions

The emission probability distribution E is assumed to be a continuous probability Gaussian function:

$$p(x_t | \mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(x_t - \mu_i)^2 / (2\sigma_i^2)}, \quad (4.1)$$

with μ_i, σ_i^2 being the expected and variance emission of the state i respectively [66]. The sojourn time distribution D is approximated using five different approaches. Four of them are parametric distributions, whereas the last one is a non-parametric probability density mass function:

- Parametric sojourn time distributions:

- Geometric:

$$d_i(u) = (1 - \vartheta_i)^{u-1} \vartheta_i, \quad (4.2)$$

where ϑ_i is the probability of leaving the state i in the next time step, meaning $\vartheta_i = \sum_{j \neq i} P(s_{t+1} = j | s_t = i)$ regardless the evaluation time t ¹.

This is the approximation used in the standard HMM's, see Figure 4.3.

¹we are considering a homogeneous hidden Markov model.

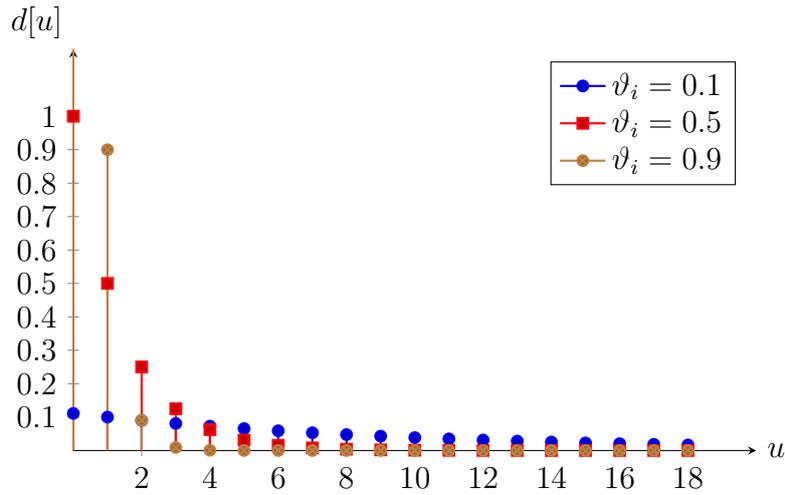


Figure 4.3: The probabilistic density function of a Geometric distribution

- Poisson:

$$d_i(u|\lambda_i) = \frac{e^{-\lambda_i} \lambda_i^u}{u!}, \quad (4.3)$$

where λ_i is the expected sojourn time in the state i , see Figure 4.4. Note that the Poisson distribution is a strictly positive definite distribution, see Figure 4.4.

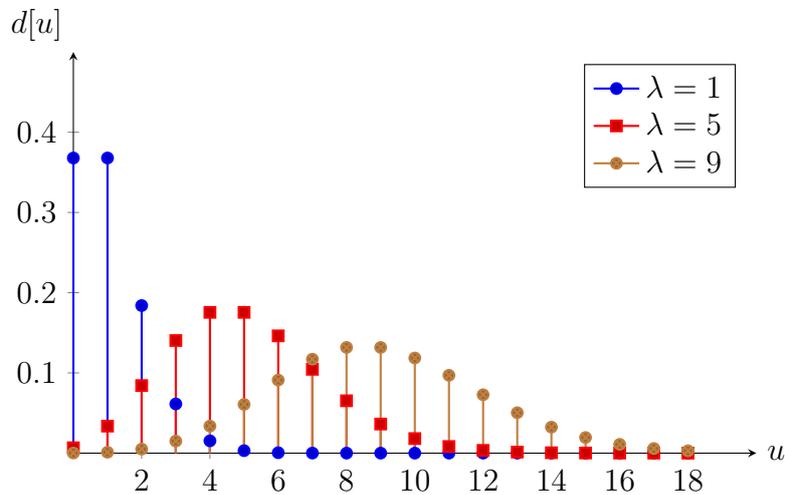


Figure 4.4: The probabilistic density function of a Poisson distribution

- Gaussian:

$$d_i(u|\lambda_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \cdot e^{-\frac{(u-\lambda_i)^2}{2\sigma_i^2}}, \quad (4.4)$$

where λ_i is the expected sojourn time in the state i and σ_i^2 is the variance of the sojourn time in the state i . Note that λ_i and σ_i^2 are chosen so that the

probability that the corresponding sojourn time is negative is negligible, see Figure 4.5.

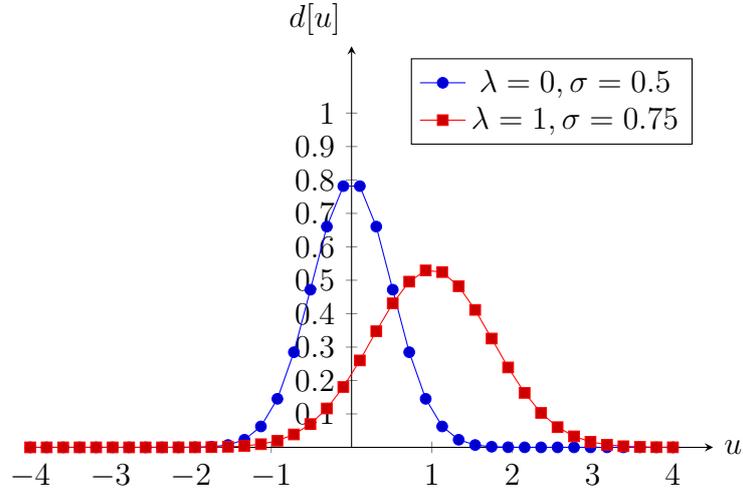


Figure 4.5: The probabilistic density function of a Gaussian distribution

- Gamma:

$$d_i(u|a_i, b_i) = \frac{1}{\Gamma(a_i)b_i^{a_i}} \cdot u^{(a_i-1)} \cdot e^{-\frac{u}{b_i}}, \quad (4.5)$$

where $u, a_i, b_i > 0$. The a_i, b_i are the shape and scale sojourn time parameters for the state i , respectively, see Figure 4.6.

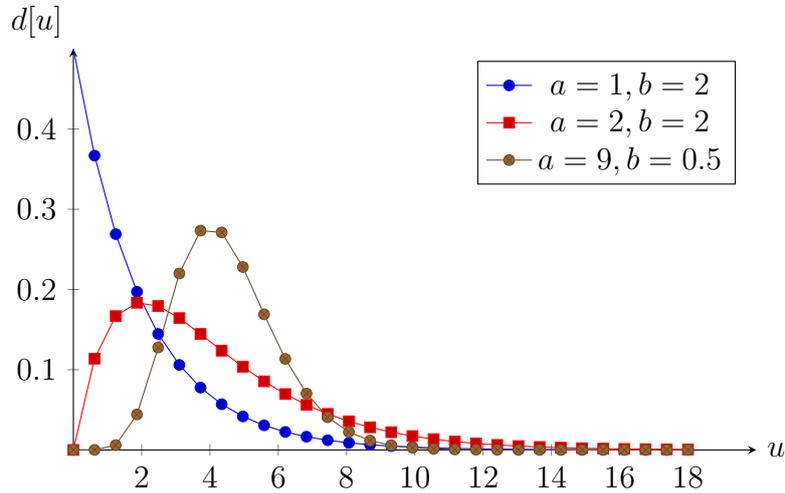


Figure 4.6: The probabilistic density function of a Gamma distribution

- The non-parametric probability density mass function:

$$d_i(u|c_i, f_i) = \begin{cases} \frac{1}{f_i - c_i}, & c_i \leq u < f_i \\ 0, & \text{otherwise} \end{cases}, \quad (4.6)$$

where c_i is a constant equal to the minimum acceptable sojourn time in the state i and f_i is a constant equal to the maximum acceptable sojourn time in the state i , see Figure 4.7.

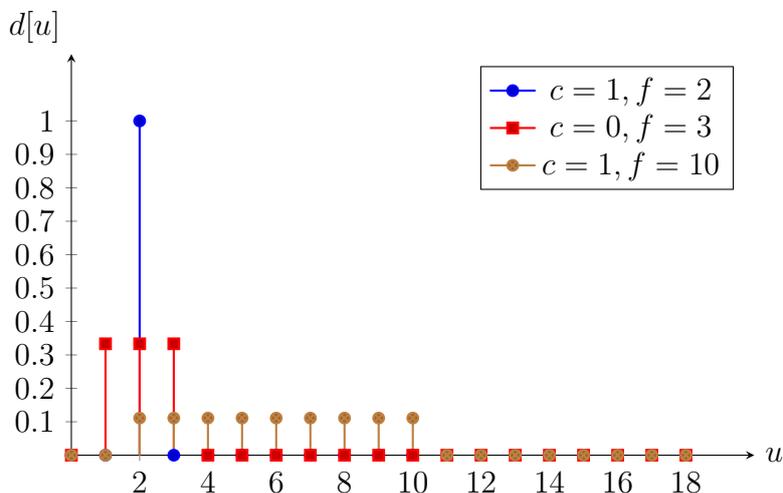


Figure 4.7: Non-parametric mass density function

4.2.3 Experimental methodology

In this subsection, we are going to explain the experimental setup used in this chapter. Furthermore, we use only the Digiscope dataset in our experiments ². In order to train and evaluate our HMM's and HSMM's in a subject dependent approach, we use an exhaustive cross-validation strategy. The models are trained using from 1 to 7 heart beats and the remaining ones are used for testing purposes only. An important restriction is that, the selected training sets must allow for continuous testing sets. For example, when we have a signal with four heart beats and we only need two heart beats to train, we can use:

- The first two heart beats to train and the last two to test.
- The last two heart beats to train and the first two to test.
- the first and last heart beats to train and the second and third heart beats to test.

but we are forbidden (according to our rules) to use the first and the third heart beats to train and the second and the fourth heart beats to test.

²For a full explanation of the dataset, see subsection 4.3.1

4.2.4 Pre-processing and feature extraction

Following previous literature [68, 50, 69], we first normalize the signal, making it vary between 0 and 1, by subtraction followed by scaling. The normalized signal is filtered using a Butterworth lowpass filter of order 10 with a cutoff frequency of $100Hz$, since the majority of the frequency content of the S1 and S2 (for the Di-giScope dataset) is over $30 - 80Hz$ as it is shown in Figure 3.4. Similar results can be found in [34]. From the filtered signal, the homomorphic envelopogram is computed. In this transformation, the signal is viewed as a product of slowly varying components (heart sounds) with fast oscillatory components (noise). These fast components are rejected by applying a non-linear transformation and is computed as in [50]. The pre-processing and feature extraction steps are the same for both training and testing phases.

4.2.5 Training HMM and HSMM distributions

The emission parameters are trained using a hierarchical clustering algorithm [70] over the annotated segments corresponding to each state $i \in S$. The sojourn time parameters are trained using temporal statistics extracted from the annotated segments as follows:

- Parametric distributions
 - Poisson: we use the average annotated time lapse between the beginning and the end of each corresponding state $i \in S$.
 - Gaussian: we use the average and standard deviation of the annotated time lapse between the beginning and the end of each corresponding state $i \in S$ respectively.
 - Gamma: we use the maximum likelihood estimation (MLE) proposed by Choi over each annotated state $i \in S$ [71].
- Non-parametric probability density mass functions
 - Non-parametric probability density mass function is initialized uniformly for each state $i \in S$ as $U(a_i, b_i)$, where a_i is the minimum allowed sojourn time in the state i and b_i is the maximum sojourn time annotated in the state i respectively.

4.2.6 Initialization of HMM and HSMM distributions

The initial state probability distribution (π) is initialized with equal starting probabilities for each state $i \in S$. The state transition probability distribution A in a HMM is given by:

$$A = \begin{array}{cc} & \begin{array}{cccc} S1 & siSys & S2 & siDia \end{array} \\ \begin{array}{c} S1 \\ siSys \\ S2 \\ siDia \end{array} & \begin{array}{cccc} 0.80 & 0.20 & 0 & 0 \\ 0 & 0.80 & 0.20 & 0 \\ 0 & 0 & 0.80 & 0.20 \\ 0.20 & 0 & 0 & 0.80 \end{array} \end{array} . \quad (4.7)$$

The state transition probability distribution Γ in HSMM is given by:

$$\Gamma = \begin{array}{cc} & \begin{array}{cccc} S1 & siSys & S2 & siDia \end{array} \\ \begin{array}{c} S1 \\ siSys \\ S2 \\ siDia \end{array} & \begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{array} \end{array} , \quad (4.8)$$

since in a normal cardiac system the state sequence $\{S1 \rightarrow siSys \rightarrow S2 \rightarrow siDia \rightarrow S1\}$ is fixed, as it is shown in Figure 3.6. The emission and the sojourn time probability distributions are initialized using the corresponding parameters extracted during the training phase.

4.2.7 Tuning HMM and HSMM distributions

Regardless of the sojourn time probability distribution used in the expectation step, we need to calculate the η_{iv}^k quantities over each k^{th} iteration of the E-M algorithm, i.e: the expected number of times that a model remains in a state i for v units of time.

- Gaussian probability distribution: In the maximization step, the Gaussian sojourn time parameters and also the corresponding shifted versions ζ are updated as:

$$\lambda_{i,\zeta}^{k+1} = \sum_{v=1}^{T_{max}} \frac{\eta_{iv}^k}{\eta_i^k} (v - \zeta), \quad (4.9)$$

and

$$(\sigma_{i,\zeta}^{k+1})^2 = \sum_{v=1}^{T_{max}} \frac{\eta_{iv}^k}{\eta_i^k} (\lambda_{i,\zeta}^{k+1} - v)^2, \quad (4.10)$$

where ζ 's are:

$$\zeta = \min(p : \eta_{i,p}^k > 0). \quad (4.11)$$

Finally, the shifted Gaussian parameter set, which gives the maximum likelihood (3.15) is retained. Note that the shifted ζ versions are very useful for very short heart sound signals, where inference is more problematic [53].

- Poisson probability distribution: In the maximization step, the Poisson sojourn time parameters and also the corresponding shifted versions ζ are updated as:

$$\lambda_{i,\zeta}^{k+1} = \sum_{v=1}^{T_{max}} \frac{\eta_{iv}^k}{\eta_i^k} (v - \zeta). \quad (4.12)$$

Finally, the shifted Poisson parameter set, which gives the maximum likelihood (3.15) is retained.

- Non-parametric probability density mass function: In a non-parametric probability density mass function each entry is re-estimated as:

$$d_i^{k+1}(u) = \frac{\eta_{iu}^k}{\sum_{v=1}^{T_{max}} \eta_{iv}^k}. \quad (4.13)$$

- Gamma probability distribution: In a Gamma sojourn time probability distribution, the shape a_i^{k+1} parameter is obtained by maximizing the following equation:

$$\log(a_i^{k+1}) - \psi(a_i^{k+1}) = \log(\bar{u}_i) - \overline{\log(u_i)}, \quad (4.14)$$

where $\bar{u}_i = \frac{\sum_v v \times \eta_{iv}^k}{\sum_v \eta_{iv}^k}$ and $\overline{\log(u_i)} = \frac{\sum_v \log(v) \times \eta_{iv}^k}{\sum_v \eta_{iv}^k}$. This equation can be solved using the Newton's method (see [72]). The scale parameter is obtained as $b_i^{k+1} = \frac{\bar{u}_i}{a_i^{k+1}}$

In the expectation step of the EM algorithm, two important quantities are computed:

- Forward quantities α [53] defined as:

$$\alpha_i^k(t) = P(s_{t+1} \neq i, s_t = i | X_{1:t}, \Theta^k), \quad (4.15)$$

- Backward quantities β [53] defined as:

$$\beta_i^k(t) = \frac{P(s_{t+1} \neq i, s_t = i | X_{1:T})}{P(s_{t+1} \neq i, s_t = i | X_{1:t}, \Theta^k)}, \quad (4.16)$$

In the maximization step, the Gaussian emission parameters for each state $i \in S$ are re-estimated as:

$$\mu_i^{k+1} = \frac{\sum_{t=1}^T \alpha_i^k(t) x_t}{\sum_{t=1}^T \alpha_i^k(t)}, \quad (4.17)$$

and

$$(\sigma_i^{k+1})^2 = \frac{\sum_{t=1}^T \alpha_i^k(t) (x_t - \mu_i^{k+1})^2}{\sum_{t=1}^T \alpha_i^k(t)}, \quad (4.18)$$

assuming that our observation variable is unidimensional. For a more complete explanation please see [53].

4.2.8 Decoding

In this section, we chose the Viterbi algorithm to predict future heart beat sequences [73]. The Viterbi algorithm does not try to classify every single point separately, but instead, it tries to find the most likely hidden state sequence that maximizes the likelihood equations 3.10, 3.15 for a HMM and a HSMM, respectively.

4.3 Materials

4.3.1 DigiScope dataset

The DigiScope dataset is composed of samples from 29 different healthy individuals, ranging in age from six months to 17 years old. The recordings have a minimum, maximum and average duration of $\approx 2, 20$ and 8 seconds, respectively. This is a very challenging dataset given the highly varying heart rates of individuals in this age range. A dataset with healthy adults is potentially easier to process, given their heart rate stability and the full maturity of the heart. Heart sounds have been collected in Real Hospital Português (Recife, Brasil) using a Littmann 3200 stethoscope embedded within the DigiScope Collector [17] technology, see Figure 4.8. The sounds are recorded at 4 kHz and they have all been collected from the mitral spot using the following methodology: 1) search for the best possible heart sound; 2) hold the head of the stethoscope as firmly as possible; 3) start recording, holding the position for a minimum time; 4) stop the recording. This methodology tries to minimize external noise and it is used to collect sounds in telemedicine scenarios [74]. These sounds were then manually annotated by cardiopulmonologists using the Audacity software. The annotations contain information about the beginning and the ending stages of S1 and S2 during a variable number of heart cycles, see Figure 4.9.



Figure 4.8: Prototype of the DigiScope Collector system, composed by a tablet and an electronic stethoscope Littmann 3200.

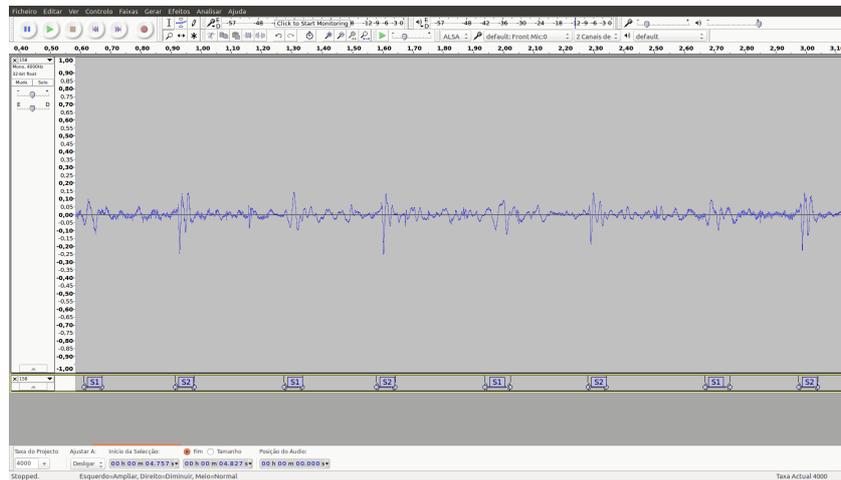


Figure 4.9: A normalized heart sound signal and the corresponding expert annotations made using the Audacity software ³.

4.3.2 Metrics of performance

The typical standard performance metrics measure the model’s capability in detecting the precise position of the principal heart sounds $S1$ and $S2$. In this case true and false positives, and true and false negatives are computed by comparing the average time instant annotated by the expert and by the model, when the event occurred. For example, a true positive happens when, given the average time instant of an $S1$ ($S2$) sound in the output sequence, the closest sound in the annotation state sequence is also associated to an $S1$ ($S2$) sound. We report

³www.audacityteam.org.

four standard performance metrics concerning the detection of the principal heart sounds: precision (Prec), recall (Rec), accuracy (Acc) and F-1 score (F-1).

$$\text{Precision} = \frac{TP_{State}}{TP_{State} + FP_{State}}. \quad (4.19)$$

$$\text{Recall} = \frac{TP_{State}}{TP_{State} + FN_{State}}. \quad (4.20)$$

$$\text{F-measure} = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4.21)$$

$$\text{Accuracy} = \frac{TP_{State} + TN_{State}}{TP_{State} + TN_{State} + FP_{State} + FN_{State}}. \quad (4.22)$$

Other standard metrics are more concerned in measuring a model’s capability in recreating the state sequence annotated by the expert. In this case, true and false positives are computed by comparing the predicted and the annotated state sample. For example, a sample at time t is a true positive when the predicted state sample and the annotated state sample are the same, otherwise it is a false positive. We compute the positive predictability per sample (P_{Sample}^+) as:

$$P_{Sample}^+ = \frac{TP_{Sample}}{TP_{Sample} + FP_{Sample}}, \quad (4.23)$$

where TP_{Sample} is the sum of all positive samples and FP_{Sample} is the sum of all negative samples respectively. Furthermore, we define P_{high}^+ as:

$$P_{high}^+ = \frac{TP_{Sample}^{high}}{TP_{Sample}^{high} + FP_{Sample}^{high}}, \quad (4.24)$$

where $TP_{Samples}^{high}$, $FP_{Samples}^{high}$ are the correctly and wrongly classified samples respectively, which are above of a pre-defined threshold condition.

4.4 Results

4.4.1 Why should you model time on a HMM?

The state transitions in PCG signals are rare events for two reasons:

1. Because PCG signals usually are sampled at a high frequency. The sampling rate of our signals is $f = 4kHz$ for the DigiScope dataset. One could downsample the signals but would still need to respect the Nyquist-Shannon sampling criterion, which imposes that the sampling rate f must be at least

two times greater than the source rate f_h , ($f \geq 2 \times f_h$) [75]. The majority of the heart sounds frequencies f_h lie between $30 - 80Hz$.

2. Because of physiological time constraints that exist in the cardiac cycle. As an example, the cardiac muscle, like any excitable tissue, exhibits a refractory period to re-stimulation. During this time interval, normal cardiac impulse cannot re-excite an already excited area of the cardiac muscle [5]. The normal refractory period of the ventricle is 0.25 to 0.30 seconds. Even the heart sounds, which are produced when the heart valves close are not an instantaneous phenomenon and their sojourn time appear to be proportional to the quantity of ejected blood during the atrial or ventricular deflation [5].

In order to answer our first question (Why should you model time?) we experimented both HMM's and HSMM's. The HMM is not capable to detect the right sequence of events and not even the state duration in each state as it can be seen in Figure 4.10(a). On the other hand, the HSMM correctly classifies the signal as is shown in Figure 4.10(b). This is even more emphasized in Figures 4.11(a) and 4.11(b), where the HMM's average positive predictability per sample P_{Sample}^+ and per state P_{State}^+ is much lower compared to the P_{Sample}^+ and P_{State}^+ of any HSMM we tested respectively.

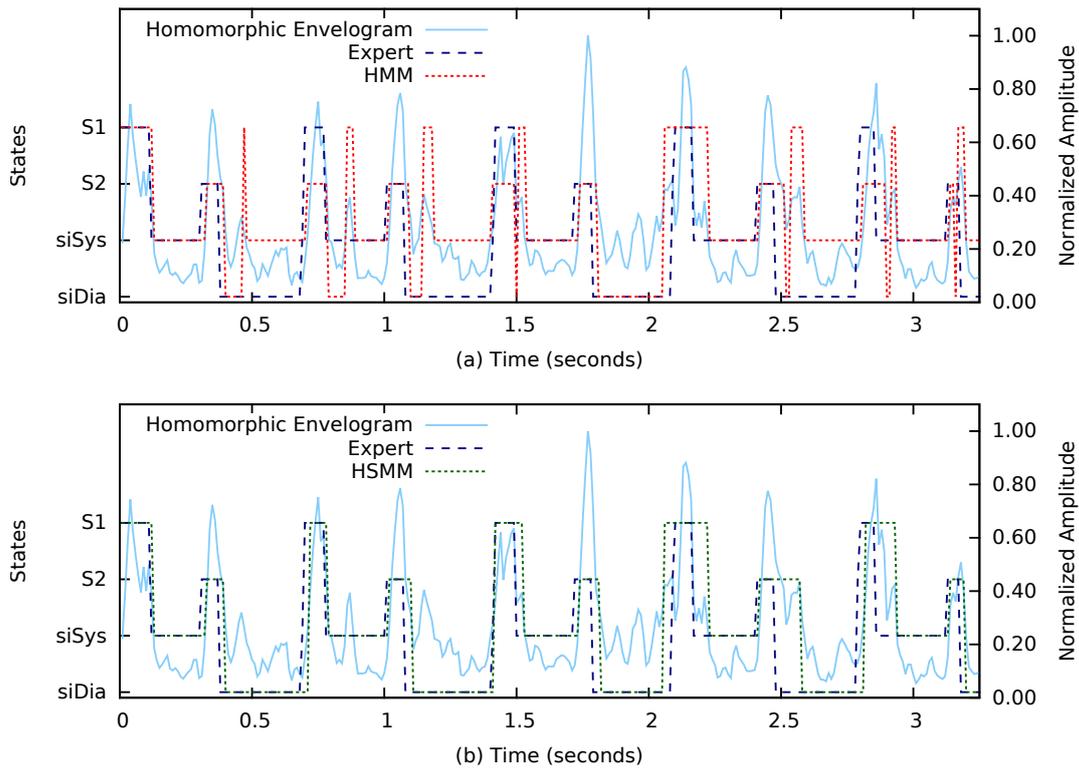


Figure 4.10: Classification results of heart sound recordings from a normal subject. The dashed lines present the states classified by an expert, HMM, HSMM; and the solid lines present the observation input to the model.

A non-parametric paired sampled test (two sampled Kolmogorov Smirnov (KS) test) is performed in order to compare the HMM's and HSMM's performance over our dataset. A p-value lower than 0.05 was considered statistically significant. We have tested with different HMM's configurations, although the results did not change significantly. The HMM's low performance might be a consequence of using a static state transition matrix. In the HSMM's, the assumption that the Markov chain is homogeneous is dropped. Instead, it is assumed that the state transition matrix is dependent on time. This ultimately, leads to a better model capable of describing the non-stationary events in the heart sound signal than the standard HMM's as depicted in Figure 4.10(b).

4.4.2 How should you model time on a HSMM?

In the current state-of-the-art for heart sound segmentation when using HSMM's [56, 58], only the Gaussian distribution is examined as an approximation for the sojourn time distribution. But the standard Gaussian distribution is not strictly positively defined and therefore it is not the most advisable distribution, since the sojourn

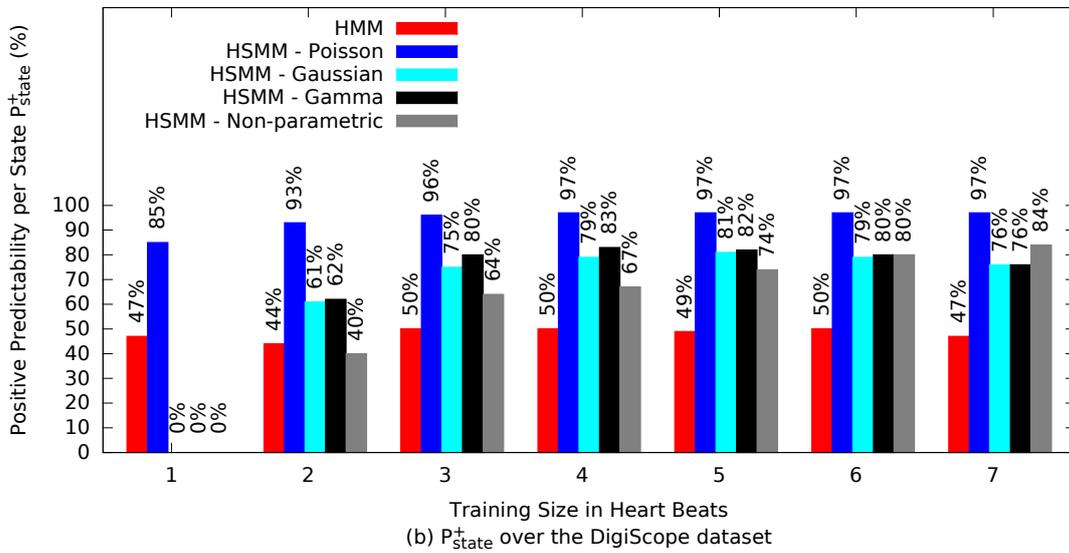
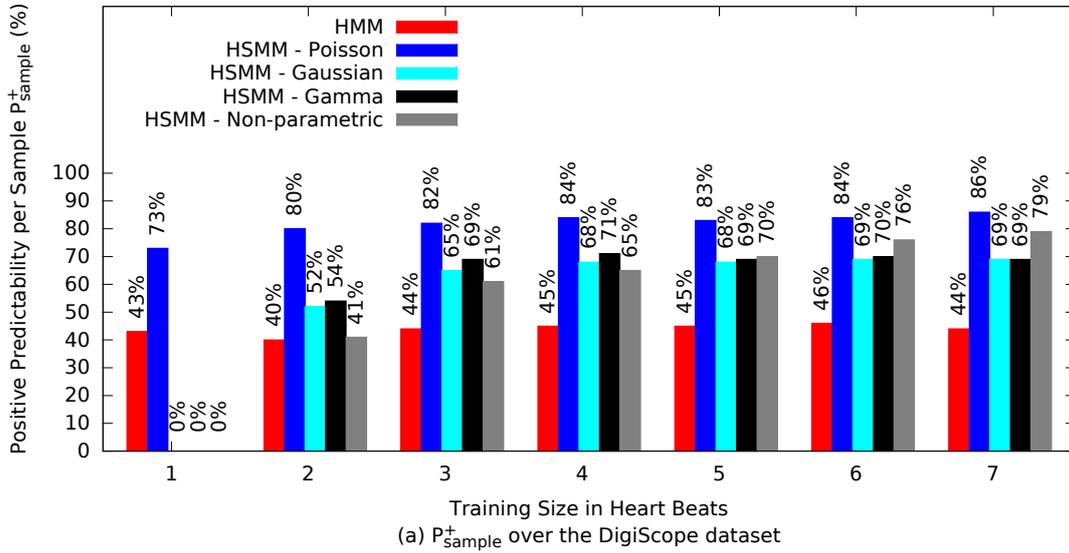


Figure 4.11: Subject dependent results. Average positive predictability (a) per sample P_{Sample}^+ and (b) per state P_{State}^+ for the tested HMM's, HSMM's over the DigiScope dataset.

times are by nature strictly positive. Furthermore, in terms of complexity, the AIC (Akaike Information Criterion) increases for the Gaussian distribution when compared to simpler distributions like the Poisson [76] (when keeping the other parameters constant). Therefore, the Gaussian distribution should not be our first choice as an approximation for the sojourn time in an one-channel or in a multi-channel system. To address our second question (How should you model time?) we compared HSMM using Poisson, Gaussian, gamma distributions and a non-parametric probability density mass function in terms of their ability to recreate the "true" state sequence. Using a subject dependent approach, the model using the Poisson distribution outperformed significantly the Gaussian, gamma and the non-parametric probability density mass function. The p-value lower than 0.05 was also considered significant in the KS test, see Figure 4.11(a). In order to gain deeper insight in our results, we also computed the positive predictability per state P_{State}^+ , see Figure 4.11(b). The same conclusions were withdrawn. Furthermore, we can see that the HSMM using a non-parametric probability density mass function starts with weak performance, but it improves significantly as the size of the training set increases. We also tested the logarithmic distribution which performed worse than the other HSMM's; for brevity we excluded it from our thesis. The KS test showed that the gamma distribution is not statistical significant (p-value > 0.05) from the Gaussian distribution. We suspect that the weak performance of the gamma and Gaussian distribution (when compared to the Poisson distribution) could be a result of a poor parameter initialization. This happens mainly because the size of the training set is relatively small, especially when someone uses only two heartbeats to initialize the model.

Table 4.1: Performance of HSMM's when using a Poisson sojourn time distribution in detecting $S1$ and $S2$.

	Training Size in Heart Beats						
	1	2	3	4	5	6	7
Prec	80%	91%	95%	96%	97%	97%	97%
Rec	96%	98%	99%	99%	99%	99%	99%
F-1	87%	94%	97%	98%	98%	98%	98%
Acc	88%	95%	97%	98%	98%	98%	98%

Table 4.1 reports the performance obtained by the proposed HSMM when using the Poisson sojourn time distribution in detecting the principal heart sounds. The results suggest that three heart beats are enough to capture the relevant information associated to a given recording. For the remaining of this chapter, we chose to use the Poisson distribution as an approximation of the sojourn time.

4.4.3 Where should you model time on a HSMM?

Not all sample classifications have the same degree of confidence. For example, samples near the transition between states are difficult to classify, since identifying the exact location where one state ends and another begins, is a hard task. Similarly, high level noise could be easily misinterpreted as heart sounds because of its high amplitude in the homomorphic envelopogram. These samples should have inherently low confidence in their classifications. On the other hand, samples in the middle of heart sound states ($S1$ and $S2$ sounds) have very high amplitude and are easier to classify, providing to their classifications a higher confidence. Could these observations be an indicator that the model cannot reliably classify specific parts of the time series? Should we reject state transitions and noisy segments? In order to answer where you should model time, we propose a measure of confidence based on the conditional probability distribution P_r ⁴. The probability distribution of x_t conditioned on all the remaining observations $X_{\setminus t} = (X_1, \dots, X_{t-1}, X_{t+1}, \dots, X_T)$ is given by:

$$P_r = P(X_t = x_t | X_{\setminus t} = x_{\setminus t}) = \sum_{i \in S_k} \ln(\epsilon_i(t)) \cdot e^{\zeta_i(t)}, \quad (4.25)$$

where $\epsilon_i(t) = \frac{e^{-(\alpha_i(t) \times \Gamma + \beta_i(t))}}{\sum_{j \in S} e^{-(\alpha_j(t) \times \Gamma + \beta_j(t))}}$, $\zeta_i(t) = \frac{-\ln(e_i(x_t))}{\sum_{k=1}^T \ln(e_i(x_k))}$. The exponentiation of α , β and the logarithms of $e_i(x_t)$ are used in order to reduce the chance of underflow and overflow respectively [77]. Figure 4.12 shows an example of the proposed confidence metric function as an overlay over the homomorphic envelopogram. One can notice that the conditional distribution exhibits sudden low peaks around the transitions between different states. Furthermore, we notice that noise has lower probabilities compared to waveforms corresponding to heart sound segments.

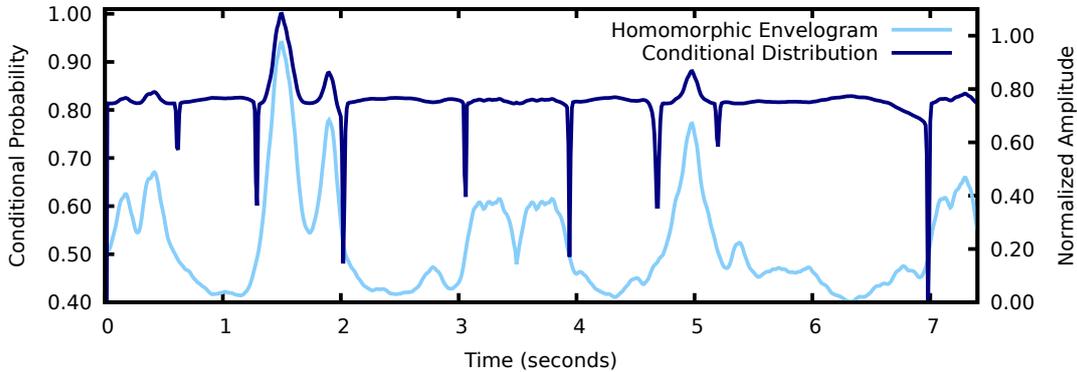


Figure 4.12: The conditional distribution P_r generated by the HSMM using the Poisson distribution.

⁴For the rest of the paper we shorthand $P(X_t = x_t | X_{\setminus t} = x_{\setminus t})$ as P_r .

Motivated by the above observations, we used the samples from all Poisson cross-fold iterations in order to compute P_{Sample}^+ as a discrete function of P_r . These results are presented in Figure 4.13. Furthermore, one can notice from this plot that the majority of the samples follow a linear trend. In Figure 4.13, the circles are centered around P_{Sample}^+ and the color intensity (from low to high) indicates the number of samples (from few to many, respectively). Furthermore, using the nonlinear least-squares (NLLS) Marquardt-Levenberg algorithm with a first degree polynomial, we get the following regression function:

$$RP^+(x_t) = 0.98 \cdot P_r - 0.10, \quad (4.26)$$

with a weighted Pearson product-moment correlation coefficient (WPCC) of 0.93. The regression line is presented in Figure 4.13. For a $P_r \lesssim 0.50$, we do not have enough data to withdraw any conclusions, although we can safely argue, that in our dataset, the conditional distribution $P_r \gtrsim 0.50$ gives a good estimate of our P_{Sample}^+ . Furthermore, we compute P_{high}^+ by setting our threshold according to the

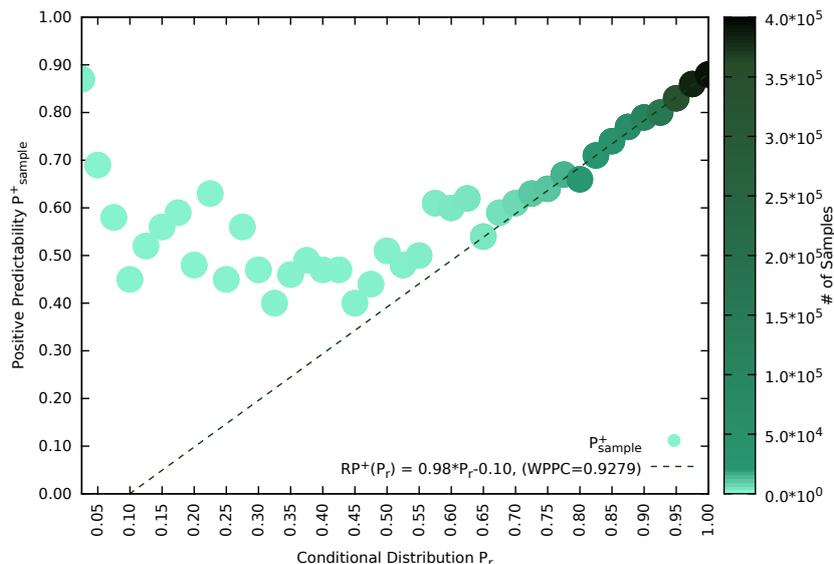


Figure 4.13: Relationship of the conditional distribution P_r with positive predictability P_{Sample}^+ , RP^+ , EP^+ and confidence P_c in our dataset.

conditional probability P_r , as depicted in Figure 4.14.

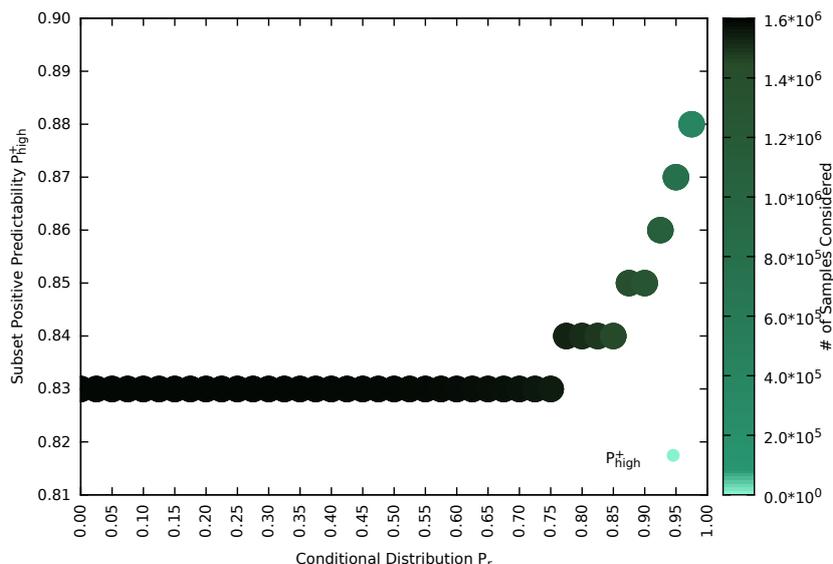


Figure 4.14: Relationship of the conditional distribution P_r and the subset positive predictability which is obtained using a high threshold in our dataset.

For a high threshold of $P_r \geq 0$ we consider all data to our computation and we computed a $P_{high}^+ \simeq 0.83$. By setting thresholds, we observe that we can increase our positive predictability to almost 90%. From Figure 4.14 we concluded, that by setting a high threshold in P_r , we can still select the majority of the sampling points, and at the same time, be more selective and confident with respect to the classification set.

4.5 Discussion

In this chapter, we saw the importance of time modelling on a HMM performance. The assumption of a geometric sojourn time distribution is ‘too strict’ to describe the non-stationary and dynamic events in the heart sound signal. Therefore, different sojourn time distributions are tested and compared in a subject dependent approach. In such a setting, the Poisson distribution outperformed any other distribution tested, which might be explained due to the amount of data available to train our HSMM models. In the next chapter, HSMM models are going to be designed and tested in a subject independent approach. Finally, in order to weight the probability of the samples classified by our HSMM are correct, a confidence metric is proposed. Using such a metric unreliable data points are discarded while keeping the majority of the sampling points.

Chapter 5

Heart Sound Segmentation using a Subject Independent Approach

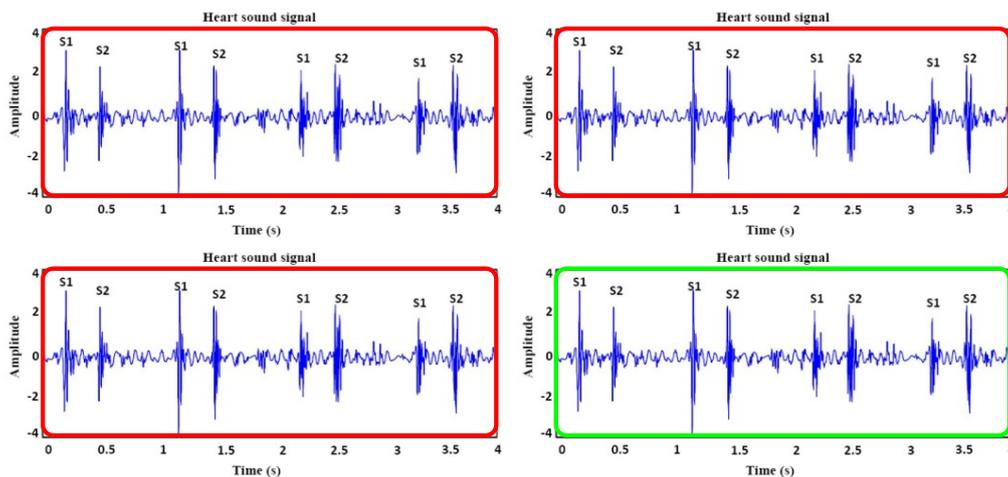


Figure 5.1: A diagram of the proposed subject-independent approach when using the Pascal dataset. The red-box represents the annotated data used to train our HMM's, the green-box represents the data used to test our HMM's.

5.1 Introduction

This chapter is based on the following contribution:

- J. Oliveira, F. Renna, Theofrastos Mantadelis and M. Coimbra, "Adaptative Sojourn Time HSMM for Heart Sound Segmentation", submitted in IEEE Journal of Biomedical and Health Informatics, 2017.

5.1.1 Why implement a subject independent approach?

In the previous chapter, we have studied HMM and HSMM segmentation algorithms for heart sounds using a subject dependent approach. In reality algorithms based on subject independent approaches are more easily deployed in real clinical environments such as hospitals and more easily accepted by physicians, since they do not need any additional action from the user (more specifically, the manual annotation of the first heart beats). On the other hand, this is indeed a more challenging task, since our algorithms must learn unknown HMM or HSMM parameters, which are specific for each tested subject and even in the same subject, it changes dramatically according to the subject physical activity, i.e: the PCG trace is completely different from someone who is doing exercise to someone who is sleeping. In this chapter, we aim to design algorithms that do not need annotated data from the subject in study. Using a subject independent approach means:

- We need an independent training dataset to train our HSMM's models. The emission HSMM parameters are trained in such way. We allow that, the sojourn time parameters are estimated using the data from the tested subject but without any annotations (unsupervised estimation).
- We might need re-estimation routines when the testing data is statistically significant different from the training data.

5.1.2 Motivation and objectives

In current state-of-the-art methods, the HSMM parameters are usually learned from training data. Then, in the testing phase, such parameters are used to recreate the underlying hidden state sequence of a PCG signal. However, this approach does not fully take into account the severe inter and intra subject variability present in heart sound signals, and can therefore lead to poor performance when there is a considerable mismatch between the training and testing data. This is precisely what will happen quite often when this type of systems are used in the future to support clinical decisions, in which real subjects will not have any annotated data. A first attempt to solve this problem is represented by the heuristics proposed by [56], where the estimated sojourn time distribution parameters are computed from the heart rate by assuming, for example, that the diastolic period is always longer than the systolic one. However, a more robust way is needed in order to adapt HSMM parameters to an unknown heart sound with the objective of maximizing

the probability of a successful segmentation. Motivated by this problem, we pose the following research question:

How should we optimally tune HSMM sojourn time parameters to improve heart sound segmentation performance?

In order to address this research question, we present the following contributions in this chapter:

1. Development of an algorithm, that searches for the most likely sojourn time parameters of a HSMM for each individual subject.
2. Testing and comparison of the performance of the state-of-art algorithm with the proposed algorithm over different datasets.
3. Testing of the effectiveness of the proposed algorithm in transfer learning tasks, i.e., considering the case when training samples and testing samples are taken from different datasets.

5.2 Methodology

5.2.1 Model Architecture

In this section, each state i of a HSMM's corresponds to an element of the heart sound signal $\mathcal{S} = \{S1, siSys, S2, siDia\}$ because the signal characteristics are thought to be homogeneous. For simplicity, HSMM's will ignore $S3$, $S4$ and murmur sounds thus implementing a four state HSMM, see 3.6.

5.2.2 HSMM distributions

The sojourn time distributions D are approximated using a Gaussian probability distribution:

$$d_i(u_i|\lambda_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \cdot e^{-\frac{(u_i-\lambda_i)^2}{2\sigma_i^2}}, \quad (5.1)$$

where λ_i is the expected sojourn time in the state i and σ_i^2 is the variance of the sojourn time in the state i . Note that λ_i and σ_i^2 are chosen so that the probability that the corresponding sojourn time is negative is negligible. The emission probability distributions E are modeled using a logistic regression function, where we have used the following Bayes rule to express the probability of observing the

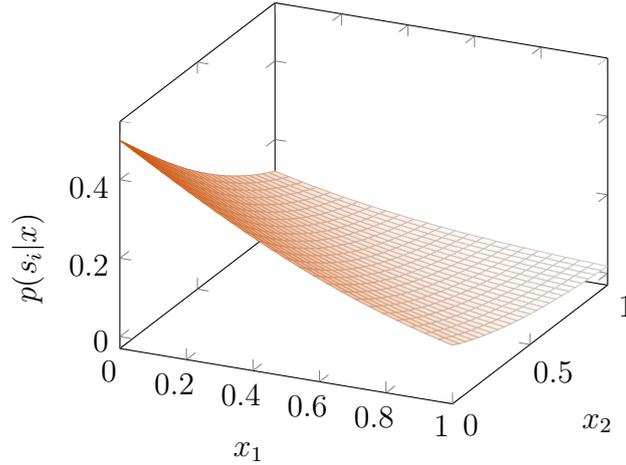


Figure 5.2: A two-dimensional logistic regression function, where $w_i = \{-2, -2\}$.

emission x_t conditioned on being in the state i :

$$p(x_t|s_i) = \frac{p(s_i|x_t)p(x_t)}{p(s_i)}, \quad (5.2)$$

the $p(s_i|x_t)$ is computed through a logistic regression function:

$$p(s_i|x_t) = \frac{1}{1 + \exp(-w_i^T x_t)}, \quad (5.3)$$

where w_i^T is the transpose of the weight vector of the state i and x_t is the observation vector. For more details see [34].

5.2.3 Experimental methodology

We tested the performance of the proposed approach in recreating the hidden "true" state sequence of a PCG signal in two distinct cases:

- Training and testing using the same dataset.
- Training and testing using different datasets.

In the first case, we split the PhysioNet dataset into 10 subsets (10-fold cross-validation), in order to get statistically significant results. Each subset has different subject records from another subset. Records of the same subject are in a single subset. Each subset is tested separately and the remaining subsets are used for training, therefore avoiding overfitting problems. In the second case, we want to assess the performance of our models in a transfer learning scenario. In order to achieve this, we use the entire PhysioNet dataset for training and the Digiscope

and Pascal datasets for testing. Note that the PhysioNet dataset contains heart sounds from healthy and unhealthy subjects from different ages. In contrast, the Digiscope and Pascal datasets contain only healthy children and teenagers. Since the Pascal and Digiscope datasets are only composed by children and teenagers, the heart rhythm variability is expected to be higher than in the PhysioNet dataset. As a result, we expect that the fine tuning of sojourn time distributions will play a fundamental role in these cases.

5.2.4 Pre-processing and feature extraction

Following previous literature [68, 50], the system first normalizes the signal into the range $[0, 1]$. Then, the signal is filtered using Butterworth lowpass and highpass filters of order 4, with cutoff frequencies at 400 Hz and 25 Hz, respectively. Since signals from different datasets are sampled at different frequency rates, each signal is downsampled to 1 kHz. Four different features are extracted from the filtered signal, as in [34]: homomorphic envelogram, Hilbert envelogram, wavelet-based features and power spectral density features.

5.2.5 Training HSMM distributions

During the training phase, annotated training data is used to determine the emission probability distribution for every state $i \in \mathcal{S}$. During the training, all samples belonging to the state k are collected. Then, using a leave-one-out strategy, the weights w_i are computed by an iterative re-weight least-squares algorithm [66] (see details in [34]). The algorithm requires samples from each state i and complementary samples equally distributed over the remaining states to compute w_i .

5.2.6 Initialization of HSMM distributions

In order to initialize the parameters of the Gaussian sojourn time distributions, we compute an autocorrelation function over the homomorphic envelogram. From this, we use the heuristics proposed by Schmidt *et al.* [56] to estimate the heart rate and the systolic sojourn time. Moreover, $S1$ is assumed to have an average duration of 122 ms and a standard deviation of 32 ms as in [56]. $S2$ is assumed to have an average duration of 92 ms and a standard deviation of 28 ms as in [56]. The average diastolic duration is inferred from the heart rate and from the others state duration . The initial state probability distribution (π) is initialized with equal starting probabilities for every state $i \in S$. The state transition probability

distribution Γ in HSMM is given by:

$$\Gamma = \begin{matrix} & S1 & siSys & S2 & siDia \\ S1 & 0 & 1 & 0 & 0 \\ siSys & 0 & 0 & 1 & 0 \\ S2 & 0 & 0 & 0 & 1 \\ siDia & 1 & 0 & 0 & 0 \end{matrix} . \quad (5.4)$$

5.2.7 Tuning sojourn time distributions

Next, we describe the proposed method to fine tune the sojourn time distribution parameters. We assume that during the testing phase the signal is pre-processed and features are extracted as described in Section 5.2.4. Our goal is to find a set of sojourn time parameters that maximize the incomplete likelihood for a given X [53]:

$$L(X, \Theta) = \sum_{\forall s_1 \dots s_T} L(S, X, \Theta), \quad (5.5)$$

where $\sum_{\forall s_1 \dots s_T}$ denotes the sum over all possible state sequences of length T . To do so, we use the expectation maximization (EM) algorithm [66].

Algorithm 5.1 Expectation Maximization algorithm devised to tune sojourn time parameters.

Data: $d^0 \rightarrow$ initial sojourn time parameters

Result: $d^{opt} \rightarrow$ optimal sojourn time parameters

$L^{opt} \rightarrow$ optimal likelihood

```

1 begin
2    $k \leftarrow 0$ 
    $D^k \leftarrow$  ComputeSojournTimeDistribution( $d^k$ )
    $[\alpha^k, \beta^k, L1^k, N^k] \leftarrow$  ExpectationStep( $D^k$ )
    $L^k \leftarrow$  ComputeLikelihood( $N^k$ )
   while true do
3      $d^{k+1} \leftarrow$  MaximizationStep( $\alpha^k, \beta^k, L1^k, D^k$ )
      $D^{k+1} \leftarrow$  ComputeSojournTimeDistribution( $d^{k+1}$ )
      $[\alpha^{k+1}, \beta^{k+1}, L1^{k+1}, N^{k+1}] \leftarrow$  ExpectationStep( $D^{k+1}$ )
      $L^{k+1} \leftarrow$  ComputeLikelihood( $N^{k+1}$ )
     if  $L^{k+1} < L^k$  then
4        $d^{opt} \leftarrow d^k$ 
        $L^{opt} \leftarrow L^k$ 
       return  $[d^{opt}, L^{opt}]$ 
5     else
6        $k \leftarrow k+1$ 
7     end
8   end
9 end
```

The EM algorithm is an iterative algorithm that tries to maximize the following quantity:

$$Q(\Theta|\Theta^k) = E[\log(L(S, X, \Theta)|X_1^T, \Theta^k)], \quad (5.6)$$

where Θ^k are the HSMM parameters at iteration k . This conditional expectation quantity $Q(\Theta|\Theta^k)$ is decomposed as a sum of terms (see [53]):

$$\begin{aligned} Q(\Theta|\Theta^k) &= \sum_{j \in \mathcal{S}} Q_\pi(\{\pi_j\}|\Theta^k, X) + \sum_{i \in \mathcal{S}} \sum_{j \neq i} Q_\Gamma(\{\gamma_{ij}\}|\Theta^k, X) \\ &+ \sum_{j \in \mathcal{S}} \sum_{\mathbf{u}} \mathbf{Q}_d(\mathbf{d}_j(\mathbf{u})|\Theta^k, \mathbf{X}) + \sum_{j \in \mathcal{S}} \sum_{t=1}^T Q_e(\{e_t(x_t)\}|\Theta^k, X). \end{aligned} \quad (5.7)$$

In this section, we will try to maximize only the 3rd term in equation (5.7), which is related to the sojourn time parameter set:

$$\begin{aligned} \sum_{j \in \mathcal{S}} \sum_u Q_d(d_j(u)|\Theta^k, X) &= \sum_{j \in \mathcal{S}} \sum_u \left\{ \sum_{t=0}^{T-1-u} P(s_{t+u+1} \neq j, \right. \\ &\quad \left. s_{t+u-v} = j, v = 0 \dots u-1, s_t \neq j | X, \Theta^k) \right. \\ &\quad \left. + P(s_{u+1} \neq j, s_{u+1-v} = j, v = 1 \dots u | X, \Theta^k) \right\} \log(d_j(u)). \end{aligned} \quad (5.8)$$

To do so, we perform the following operations iteratively:

1. Assign state posterior probability distributions, namely: forward α , backward β and smoothed state transition $L1$ quantities (E-step). These are defined as:

$$\alpha_j(t) = P(s_{t+1} \neq j, s_t = j | X_1^t), \quad \forall j \in \mathcal{S}. \quad (5.9)$$

$$\beta_j(t) = \frac{P(s_{t+1} \neq j, s_t = j | X_1^T)}{P(s_{t+1} \neq j, s_t = j | X_1^t)}, \quad \forall j \in \mathcal{S}. \quad (5.10)$$

$$L1_j(t) = P(s_{t+1} \neq j, s_t = j | X_1^T), \quad \forall j \in \mathcal{S}. \quad (5.11)$$

2. Select the sojourn time parameter set that maximizes the function in (5.8), while keeping the E, Γ, π parameters constant (M-step).
3. Check if the stop conditions are satisfied, e.g the likelihood did not increase.

In the following section, we describe in more detail what are the operations that are implemented over the different stages of the EM algorithm in each iteration.

5.2.8 E-step

In this step, the posterior state probability distributions are estimated using the α , β and $L1$ quantities, which are computed similarly to [53]. The key difference with respect to [53], concerns on both censoring sojourn time at the first and last visited states. Then we need to calculate the expected number of times $\eta_{i,u}$ that the model remains in state i for u time steps:

$$\begin{aligned} \eta_{i,u} = & P(s_{u+1} \neq i, s_{u+1-v} = i, v = 1, \dots, u | X, \Theta) + \\ & \sum_{t=1}^{T-1-u} P(s_{t+u+1} \neq i, s_{t+u-v} = i, \\ & v = 0, \dots, u-1, s_t \neq i | X, \Theta). \end{aligned} \quad (5.12)$$

The first term is expressed for $v \leq T$ as:

$$\frac{L1_j(v)}{\alpha_j(v)} \left\{ \prod_{u=1}^{v-1} \frac{e_j(x_{v-u})}{N_{v-u}} \right\} d_j^*(v) \pi_j, \quad (5.13)$$

where $N_t = P(x_t | x_1^{t-1})$ can be written as:

$$\begin{aligned} N_t = \sum_j e_j(x_t) \left\{ \sum_{v=1}^{t-1} \left\{ \prod_{u=1}^{v-1} \frac{e_j(x_{t-u})}{N_{t-u}} \right\} d_j(v+1) \sum_{i \neq j} \gamma_{i,j} \alpha_i(t-v) \right. \\ \left. + \left\{ \prod_{u=1}^{t-1} \frac{e_j(x_{t-u})}{N_{t-u}} \right\} d_j^*(t) \pi_j \right\}, \end{aligned} \quad (5.14)$$

and for $v > T$:

$$\left\{ \prod_{u=1}^{T-1} \frac{e_j(x_{T-u})}{N_{T-u}} \right\} d_j^*(v) \pi_j. \quad (5.15)$$

The general term in (5.12) is expressed for $v \leq T-1-t$ as:

$$\frac{L1_j(t+1+v)}{\alpha_j(t+1+v)} \left\{ \prod_{u=1}^v \frac{e_j(x_{t+1+v-u})}{N_{t+1+v-u}} \right\} d_j(v) \sum_{i \neq j} \gamma_{i,j} \alpha_i(t), \quad (5.16)$$

and for $v > T-1-t$ as:

$$\left\{ \prod_{u=0}^{T-1-t} \frac{e_j(x_{T-u})}{N_{T-u}} \right\} d_j^*(v) \sum_{i \neq j} \gamma_{i,j} \alpha_i(t). \quad (5.17)$$

The η_i , the expected number of times that the model will stay in the state i , is given by:

$$\eta_i = \sum_{u=0}^{T_{max}} \eta_{i,u}, \quad (5.18)$$

where T_{max} is the maximum sojourn time allowed in state i . In order to speed up the computation, we set T_{max} equal to:

$$T_{max} = \frac{60}{H_R} * F_s, \quad (5.19)$$

where H_R is the estimated heart rate (see [56]) and F_s is the sampling frequency.

5.2.9 M-step

In the maximization step, the Gaussian sojourn time parameters and the corresponding shifted versions ζ are updated as:

$$\hat{\lambda}_{i,\zeta} = \sum_{v=1}^n \frac{\eta_{iv}}{\eta_i} (v - \zeta), \quad (5.20)$$

and

$$\hat{\sigma}_{i,\zeta}^2 = \sum_{v=1}^n \frac{\eta_{iv}}{\eta_i} (\lambda_{i,\zeta} - v)^2, \quad (5.21)$$

where ζ 's are:

$$\zeta = \min(u : \eta_{i,u} > 0). \quad (5.22)$$

Finally, the shifted Gaussian parameter set, which gives the maximum likelihood (5.8) is retained. The shifted ζ versions are very useful for very short heart sound signals, where inference is more problematic [53]. It is also advisable and useful in practice to consider shifted parameter arrangements when searching for the optimal parameter set.

The convergence to a local optimal solution is guaranteed, although it might not be the global maximum [66]. Therefore, the success of the optimization procedure might rely on parameter initialization choices. In order to surpass this limitation, we randomly generate seeds. Each seed is a perturbation of the initial sojourn time parameter set (d^0), which is obtained via the heuristics described in Section 5.2.6. Finally, we select the optimal parameter set (d^{opt}), corresponding to the seed for which the EM procedure achieves the largest value of the incomplete likelihood in (5.8) (see Algorithm 5.2).

Algorithm 5.2 Random initializations of the EM algorithm to avoid finding local optimal sojourn time parameters.

Data: $d^0 \rightarrow$ initial sojourn time parameters

Result: $d^{opt} \rightarrow$ optimal sojourn time parameters

```
10 begin
11    $N \leftarrow$  Total Number of Seeds
      for  $n=1 \rightarrow N$  do
12      $\tilde{d} \leftarrow$  random_seed( $d^0$ )
       $[d^n, L^n] \leftarrow$  EM( $\tilde{d}$ )
13   end
14    $d^{opt} \leftarrow S[\text{argmax}(L)]$ 
15 end
```

5.2.10 Decoding

In this section, we choose the Viterbi algorithm to determine the hidden state sequence [73]. The algorithm does not try to classify every single point separately, but instead, it searches for the most likely hidden state sequence as the one that maximizes the likelihood function (3.15), given a set of parameters Θ and a set of observations X .

5.3 Materials

5.3.1 Physionet dataset

In 2016, the Physionet/CinC Challenge released a large dataset of heart sounds. The data was collected from different research groups and recorded in different clinical and non-clinical environments. In this work, we consider 792 heart sounds recordings of 135 patients extracted from the Physionet dataset.¹ From those, 406 sounds are collected from patients with pathological heart damage (most commonly mitral valve prolapse), as assessed by echocardiography. The remaining 386 sounds are collected from healthy patients. Sound recordings have variable duration in the range from 1 to 35.5 seconds and they are sampled at 1 kHz. They are collected from several spots over the chest and they are possibly corrupted by different sources and noise levels. The annotations provided with the dataset are computed via the analysis of synchronous ECG recordings, based on the agreement between five different automatic R-peak and end-T-wave detectors [78].

¹The sounds are available online at <https://physionet.org/physiotools/hss>.

5.3.2 Pascal dataset

The Pascal dataset includes PCGs from pediatric patients, although their ages are not specified. The corresponding S1 and S2 positions were manually annotated by certificated cardiopulmonologists. In this dataset, 90 healthy heart sounds are collected, corresponding to a total of 1415 annotations examples of S1 and S2, within a time range between 1.2 and 14.7 seconds². The recorded sounds are sampled at 4 kHz.

5.3.3 DigiScope dataset

The DigiScope dataset is composed of samples from 29 different healthy individuals, ranging in age from six months to 17 years old. The recordings have a minimum, maximum and average duration of $\approx 2, 20$ and 8 seconds, respectively. This is a very challenging dataset given the highly varying heart rates of individuals in this age range. A dataset with healthy adults is potentially easier to process, given its heart rate stability and the full maturity of the heart. Heart sounds have been collected in Real Hospital Português (Recife, Brasil) using a Littmann 3200 stethoscope embedded with the DigiScope Collector [17] technology. The sounds are recorded at 4 kHz and they have all been collected from the mitral spot using the following methodology: 1) look for the best possible heart sound; 2) hold the head of the stethoscope as firmly as possible; 3) start recording, holding the position for a minimum time; 4) stop the recording. This methodology tries to minimize external noise and it is used to collect sounds in telemedicine scenarios [74]. These sounds were then manually annotated by cardiopulmonologists using the Audacity software³. The annotations contain information about the beginning and the ending stages of S1 and S2 during a variable number of heart cycles.

5.3.4 Metrics of performance

The typical standard performance metrics measure the system’s capability in detecting the precise position of the principal heart sounds $S1$ and $S2$. In this case true, false positives and true, false negatives are computed by comparing the average time instant annotated by the expert and by the model, when the event occurred. For example, a true positive happens when, given the average time instant of an $S1$ ($S2$) sound in the output sequence, the closest sound in the annotation state sequence is also associated to an $S1$ ($S2$) sound. We report

²The sounds are available online at <http://www.peterjbentley.com/heartchallenge/>.

³www.audacityteam.org

four standard performance metrics concerning the detection of the principal heart sounds: precision (Prec), recall (Rec), accuracy (Acc) and F-1 score (F-1).

5.4 Results

In this section, a subject independent heart sound segmentation algorithm is presented, where the emission probability distributions are modeled using a logistic regression function and the sojourn time distributions by a Gaussian one. In contrast, with the current state-of-art algorithm presented by Springer *et al.* [34], the sojourn time distributions are tuned to the tested subject through the tuning routines presented in this chapter. As a result, the algorithm manages to infer sojourn time occupancy in each state more accurately than the solutions presented in [34] and [56].

- The current state-of-the-art algorithm presented in [34], where the sojourn time parameters are extracted using heuristic decision rules from the autocorrelation function of the homomorphic envelopogram. The results are displayed in Figures 5.3, 5.4, 5.5 and 5.6 as a thin red line.
- Our proposed solution, where the sojourn time parameters are initialized using the heuristics proposed in [34] and then re-estimated using the EM routines presented in Section 5.2.7. The results are displayed in Figures 5.3, 5.4, 5.5 and 5.6 as a blue thick line.

The algorithm by Springer *et al.* [34] achieves a similar performance in the PhysioNet dataset when compared to our proposed approach. On the other hand, our algorithm clearly outperformed the algorithm proposed by Springer *et al.* [34], when applied over a testing dataset which is remarkably different from the training dataset.

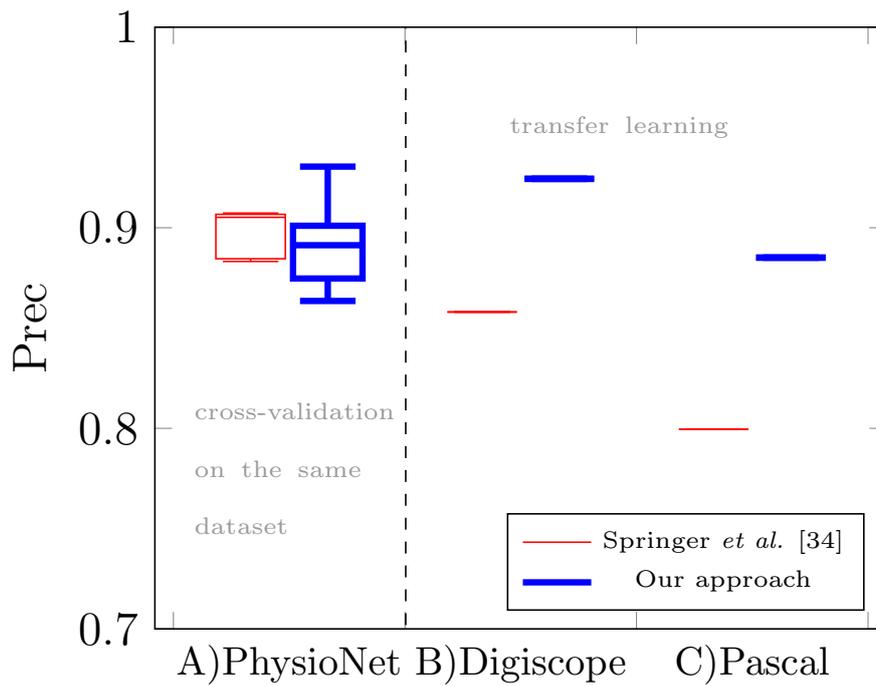


Figure 5.3: Precision results: A) When using the 10 fold cross validation over the PhysioNet dataset. B) When using the PhysioNet dataset to train and the Digiscope dataset to test. C) When using the PhysioNet dataset to train and the Pascal dataset to test.

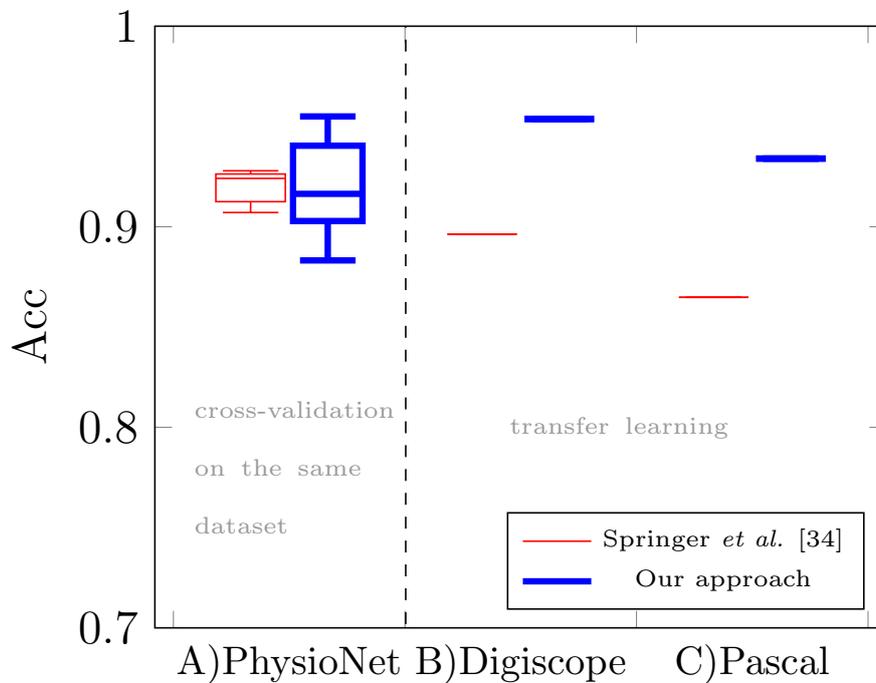


Figure 5.6: Accuracy results: A) When using the 10 fold cross validation over the PhysioNet dataset. B) When using the PhysioNet dataset to train and the Digiscope dataset to test. C) When using the PhysioNet dataset to train and the Pascal dataset to test.

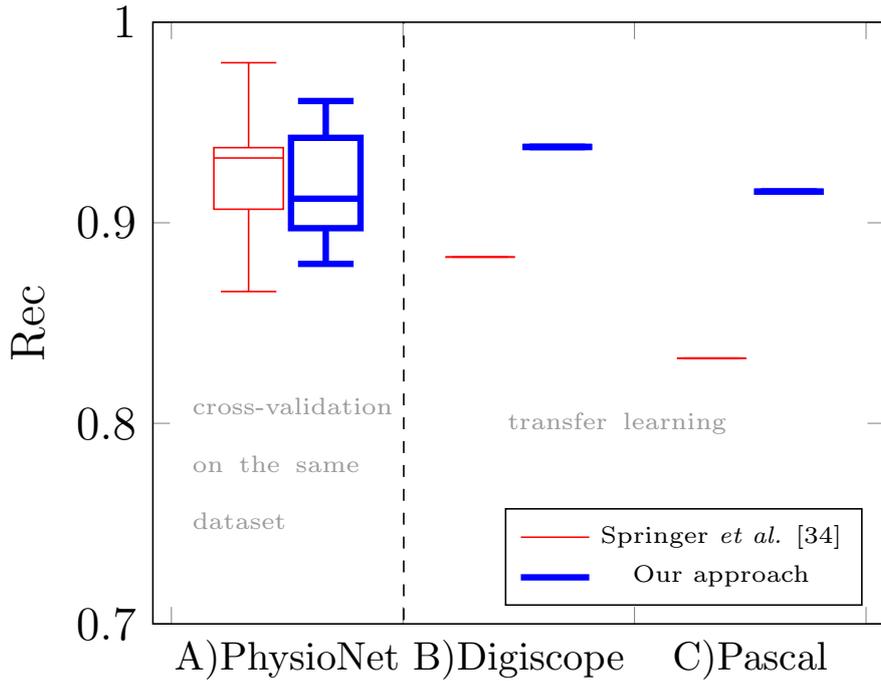


Figure 5.4: Recall results: A) When using the 10 fold cross validation over the PhysioNet dataset. B) When using the PhysioNet dataset to train and the Digiscope dataset to test. C) When using the PhysioNet dataset to train and the Pascal dataset to test.

This might be due to the naive heuristics rules used to initialize the sojourn time distributions on a HSMM. E.g: in Schmidt *et al.* [56], the systolic duration is defined from the auto-correlation function over the homomorphic envelopogram as the time from lag zero to the highest peak in the interval between 200 ms and half of the heart cycle duration. This is not always the case, e.g: in neonates and children, it is normal to have diastolic duration times shorter than systolic ones. Therefore, we can conclude that the naive sojourn time parameters extracted using the heuristic rules proposed in [56] and used later in [34] are not reliable enough when applied to datasets where the sojourn time distribution per state are largely variable (e.g: pediatric datasets, such as the Digiscope and the Pascal datasets). This observation confirms the necessity of tuning routines such as the EM algorithm presented in Section 5.2.7 in order to re-estimate a more reliable sojourn time parameter set.

5.5 Discussion

In this chapter, we presented a subject independent heart sound segmentation algorithm. The algorithm is tested and compared with the current state-of-art

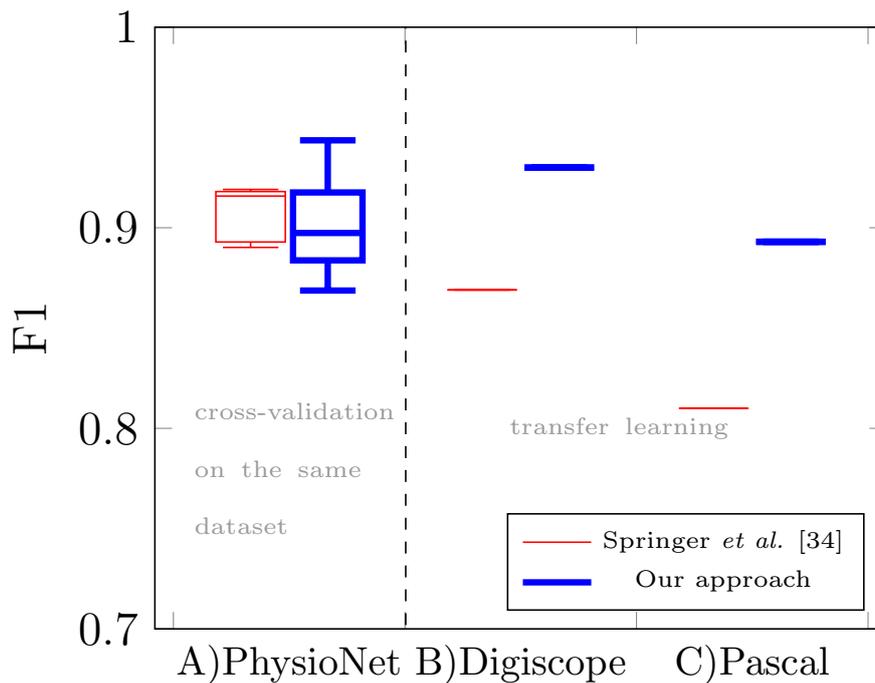


Figure 5.5: F-Score results: A) When using the 10 fold cross validation over the PhysioNet. B) When using the PhysioNet dataset to train and the Digiscope dataset to test. C) When using the PhysioNet dataset to train and the Pascal dataset to test.

algorithm presented by Springer *et al.* [34] over real data. The key difference in respect to the algorithm presented by Springer concerns the EM tuning routines for the sojourn time distribution. The EM routines discussed in Section 5.2.7, search for a more likelihood sojourn times parameters than the ones proposed initially by [56]. On the other hand, EM routines do incur in an increased computational complexity with respect to method presented in [34] and [56]. In the next chapter, we will switch our attention to the modelling of the emission probability distribution on a HSMM and the corresponding EM routines to tune the emission parameters for each subject tested.

Chapter 6

Heart Sound Segmentation using an Unsupervised Approach

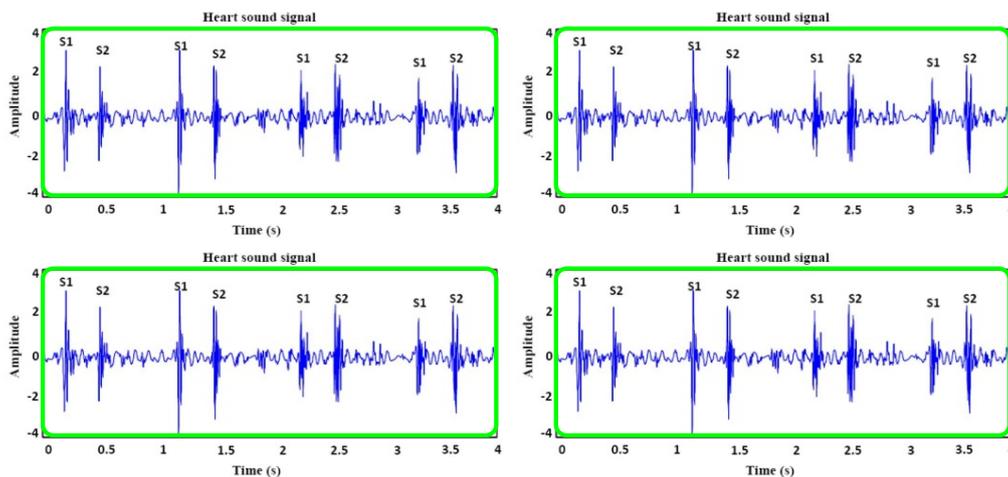


Figure 6.1: A diagram of the proposed unsupervised approach when using the Pascal dataset. The green-box represents the data used to test our HMM's.

This chapter is based on the following contributions:

- F. Renna, J. Oliveira and M. Coimbra, "A Data-Driven Feature Extraction Method for Enhanced Phonocardiogram Segmentation", in Proc. of Computing in Cardiology, CInC 2017, Rennes, France, Sep 2017.
- J. Oliveira, F. Renna and M. Coimbra, "A subject-driven automatic HSMM-GMM for heart sound segmentation", submitted in IEEE Journal of Biomedical and Health Informatics, 2017.

6.1 Introduction

6.1.1 Towards unsupervised heart sound segmentation

In this section, we are going even further in our topic and we present the first unsupervised algorithm for heart sound segmentation when using HSMM's. This is indeed a very ambitious goal and it is far more difficult to design and to implement than any other method presented before. This is easy to understand, since we do not possess any information about the signal that we are trying to model. However there is an increasing interest for such algorithms in different clinical and non-clinical environments. Mainly, when the training dataset, if it exists at all, is not representative of the population tested. In such cases, the training done leads to very poor translation to clinical reality. On the other hand, re-estimation routines which are capable of automatically adjusting the HSMM parameters to each subject tested are expected to play a major role.

6.1.2 Motivation and objectives

In the current state-of-the-art, the emission distribution parameters are learned from the training dataset. In the testing phase, these global parameters are always used to initialize our models regardless of the heart sound tested and without any tuning. Although this is the standard pipeline, it might result in weak model performances if the test dataset is statistically different from the training set. A simple example is when HSMM's are trained using pediatric data and tested on adult data. This is mostly related to the heart maturity and growing necessities of a child, which are not present in an adult. Another example is when HSMM's are trained using healthy subjects and tested over unhealthy subjects. This is mainly explained due to dysfunctional mechanisms not recorded in the training phase. But the major limitation relates to the subject-to-subject variability over the testing data. This is even more evident in heart sounds considered as outliers in respect to the remaining test population. Therefore, re-estimation parameter routines adapted to each subject might play a major role in heart sound segmentation. In this chapter, the problem of how to manage inter-patient variability when using HSMM's for heart sound segmentation is addressed. In particular, this section focuses on the study of how to model emission distributions associated to a HSMM when analyzing a PCG signal and it provides the following contributions:

- *Propose an efficient approximation for the emission probability distribution.*

To do so, we compare the current state-of-art logistic regression [34] function

with a Gaussian mixture model distribution when:

- we train HSMM's emission probability distributions on a dataset and test the algorithm with other heart sounds from the same dataset.
- *We describe an algorithm that searches for the most likely emission GMM parameters for a given heart sound signal.*
- *We propose an efficient unsupervised heart sound segmentation algorithm and compare its performance with state-of-the-art supervised approaches, thus emphasizing the operational regimes where the proposed unsupervised solution can outperform other methods described in the literature.*

6.2 Methodology

6.2.1 Model Architecture

In this section, each state i of a HSMM's corresponds to an element of the heart sound signal $\mathcal{S} = \{S1, siSys, S2, siDia\}$ because the signal characteristics are thought to be homogeneous. For simplicity, HSMM's will ignore $S3$, $S4$ and murmur sounds thus implementing a four state HSMM, see 3.6.

6.2.2 HSMM distributions

The sojourn time distributions D are approximated using a Gaussian probability distribution:

$$d_i(u_i|\lambda_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \cdot e^{-\frac{(u_i-\lambda_i)^2}{2\sigma_i^2}}, \quad (6.1)$$

where λ_i is the expected sojourn time in the state i and σ_i^2 is the variance of the sojourn time in the state i . Note that λ_i and σ_i^2 are chosen so that the probability that the corresponding sojourn time is negative is negligible.

The emission probability distributions E are approximated using:

- **Logistic Regression:** This is done by using the approach described in [34], where the Bayes' rule is used to express the probability of observing the emission vector x_t conditioned on being in state i at time t , as follows:

$$e_i(x_t) = p(x_t|s_t = i) = \frac{p(s_t = i|x_t)p(x_t)}{p(s_t = i)}. \quad (6.2)$$

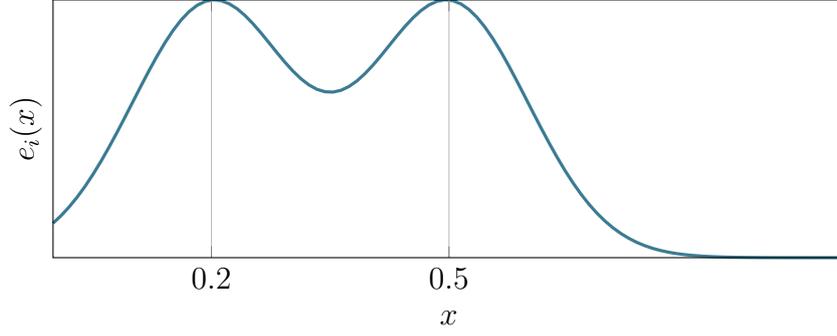


Figure 6.2: A Gaussian mixture model distribution, with $\mu_i = \{0.2, 0.5\}$, $\sigma_i = \{0.1, 0.1\}$ and $\Pi_i = \{0.5, 0.5\}$

Then, $p(s_t = i|x_t)$ is computed through a logistic regression function:

$$p(s_t = i|x_t) = \frac{1}{1 + \exp(-w_i^T x_t)}, \quad (6.3)$$

where w_i are the weights associated to the state $i \in \mathcal{S}$.

- **Gaussian Mixture Model:** The entries of E are defined as:

$$e_i(x_t|\Pi_i, \mu_i, \Sigma_i) = \sum_{m=0}^M N(x_t|\mu_{im}, \Sigma_{im}) \times \Pi_{im}, \quad (6.4)$$

where $N(\mu_{im}, \Sigma_{im})$ represents the multivariate Gaussian pdf with mean μ_{im} and covariance matrix Σ_{im} . The Π_{im} is the probability of the m^{th} component being activated in the state i , so that:

$$\sum_{m=1}^M \Pi_{im} = 1, \quad 1 \geq \Pi_{im} \geq 0 \quad \forall i \in \mathcal{S}, \quad (6.5)$$

where M is the number of Gaussian mixing components in each state-dependent multivariate GMM distribution.

The initial state probability distribution is defined as π , where $\sum_{i \in \mathcal{S}} \pi_i = 1$ and $1 \geq \pi_i \geq 0$.

6.2.3 Experimental methodology

In order to showcase the performance of the proposed approach, we test its capability in recreating the hidden true state sequence of a PCG signal in two distinct cases:

1. we train emission distributions on a dataset and test the algorithm with other heart sounds from the same dataset;
2. we consider an unsupervised approach where we test the proposed algorithm without any training phase.

In the first case, we split the PhysioNet dataset into 10 subsets (10 fold cross validation). In order to prevent overfitting, records from the same subject are not shared among different subsets. Each subset is tested separately and the remaining subsets are used for training. The emission parameters are extracted from the training data but the sojourn time parameters are extracted from the testing data directly. In the second case, each of the 10 subsets extracted from the PhysioNet dataset is tested separately but the remaining ones are not used at all. In this case, both sojourn time and emission parameters are inferred directly from the testing data.

6.2.4 Pre-processing and feature extraction

Following previous literature [68, 50], the system first normalizes each signal by subtracting its minimum and scaling it properly. The signal is filtered using a Butterworth lowpass and highpass filter order 4 with a cutoff frequency at 400 Hz and 25 Hz respectively. From the filtered signal, several features are extracted: homomorphic envelopogram, Hilbert envelopogram, wavelet-based features and PSD-based features, as in [34]. The pre-processing and feature extraction steps are identical for both training and testing phases, regardless of the methodology used.

6.2.5 Training HSMM distributions

- **Supervised Approach** In standard approaches in the literature [34] [58], the emission parameters are learned from the training dataset and afterwards used to initialize and test our HSMM's without any tuning to the tested subject. In the present case, the GMM parameters $(\mu_{jm}, \Sigma_{jm}, \Pi_{jm})$ are trained using the EM algorithm for GMM distributions as in [79].
- **Unsupervised Approach** No training is performed.

6.2.6 Initialization of HSMM distributions

- **Supervised Approach** The GMM distributions are initialized using the parameters learned during the training phase.

- **Unsupervised Approach** As an alternative to supervised methods, we propose to initialize the GMM parameters using the following heuristics:

$$\mu_{jm}^0[n] = \begin{cases} \max(X[n, :]), & j \in \{S1, S2\} \\ \min(X[n, :]), & j \in \{siSys, siDia\} \end{cases}$$

$$\forall m \in \{1, \dots, M\}, \forall n \in \{1, \dots, N\}, \quad (6.6)$$

$$\Sigma_{jm}^0 = I_{N \times N}, \quad \Pi_{jm}^0 = \frac{1}{M}, \quad \forall m \in \{1, \dots, M\}, \quad \forall j \in \mathcal{S}^1, \quad (6.7)$$

where $X[n, :]$ is the n -th row vector of a $N \times T$ observation feature matrix X and $I_{N \times N}$ is a $N \times N$ identity matrix. The initial GMM parameters are indeed the $k = 0$ step of our undermentioned EM algorithm. The proposed simple heuristics are based on the following facts:

1. In the feature space, heart sounds ($S1, S2$) correspond to high amplitude peaks, whereas $siSys$ and $siDia$ are silent states corresponding to low amplitude peaks immersed in the background noise.
2. Without any prior knowledge of the tested subject, $S1$ features are remarkably similar to the features extracted from $S2$ sounds.
3. Without any prior knowledge of the tested subject, it is reasonable to assume that the different mixture components in every state $j \in \mathcal{S}$ have equal and isotropic covariance matrices. Note that the emission distributions obtained with the proposed simple heuristics are in fact simple Gaussian distributions rather than GMM. However, the outputs of the fine tuning emission distribution procedure described in Section 6.2.7 are in general GMM distributions.

The initial state probability distribution (π) is initialized with equal starting probabilities for every state $j \in \mathcal{S}$. The state transition probability distribution Γ in HSMM's is given by:

$$\Gamma = \begin{array}{cc} & \begin{array}{cccc} S1 & siSys & S2 & siDia \end{array} \\ \begin{array}{c} S1 \\ siSys \\ S2 \\ siDia \end{array} & \begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{array} \end{array} . \quad (6.8)$$

To estimate the parameters of a Gaussian sojourn time distribution, we compute an auto-correlation function over the homomorphic envelopogram. From this, we use the heuristics proposed by Schmidt et al. [56] and later by Springer et al. [34],

which are heart rate dependent. These parameters are not further re-estimated or tuned to the subject, since we aim to isolate the effect of the proposed re-estimation equations for the emission probability distribution.

6.2.7 Tuning GMM parameters on a HSMM

In the following subsection, we describe the proposed method to fine tune the Gaussian mixture model parameters. Our goal is to find a set of emission parameters that maximizes the incomplete likelihood for a given X [53]:

$$L(X, \Theta) = \sum_{\forall s_1 \dots s_T} L(S, X, \Theta), \quad (6.9)$$

where $\sum_{\forall s_1 \dots s_T}$ denotes the sum over all possible state sequences of length T . To do so, we use the expectation maximization (EM) algorithm [66]. The EM algorithm is an iterative algorithm that tries to maximize the following quantity:

$$Q(\Theta|\Theta^k) = E[\log(L(S, X, \Theta)|X, \Theta^k)],^2 \quad (6.10)$$

where Θ^k are the HSMM parameters at iteration k . In other words:

$$Q(\Theta|\Theta^k) = \sum_{s_1, \dots, s_T} \sum_{x_1, \dots, x_T} \log(P(S, X, \Theta)) \times P(S, X, \Theta|X = x_1^T, \Theta^k) \quad (6.11)$$

$$= \sum_{s_1, \dots, s_T} \log(P(S, X = x_1^T, \Theta)) \times P(S|X = x_1^T, \Theta^k). \quad (6.12)$$

Combining equations 3.15 and 6.12, we get the following result:

$$\begin{aligned} &= \sum_{s_1, \dots, s_T} \left\{ \log(\pi_{s_1}) + \sum_{r=1}^R \log(d_{s(r)}(u(r))) + \sum_{r=1}^R \log(\gamma_{s(r-1), s(r)}) \right. \\ &\quad \left. + \sum_{t=1}^T \log(e_{s_t}(x_t)) \right\} \times P(S|X = x_1^T, \Theta^k). \end{aligned} \quad (6.13)$$

Rearranging some terms and after some algebraic manipulation, it is easy to see that each term inside of the bracket corresponds to a separate maximization of a

²For mathematical convenience, we maximize the expected logarithm of the complete likelihood instead of the expected complete likelihood.

particular HSMM distribution:

$$\begin{aligned}
Q(\Theta|\Theta^k) &= \sum_{j \in \mathcal{S}} Q_\pi(\{\pi_j\}|\Theta^k, X) + \\
&+ \sum_{j \in \mathcal{S}} \sum_u Q_d(d_j(u)|\Theta^k, X) + \sum_{i \in \mathcal{S}} \sum_{j \neq i} Q_\Gamma(\{\gamma_{ij}\}|\Theta^k, X) + \sum_{j \in \mathcal{S}} \sum_{t=1}^{\mathbf{T}} \mathbf{Q}_e(\{\mathbf{e}_t(\mathbf{x}_t)\}|\Theta^k, \mathbf{X}).
\end{aligned} \tag{6.14}$$

For the sake of simplicity, let us use some shorthand notation for each term in equation 6.14.

$$Q(\Theta|\Theta^k) = Q_\pi + Q_D + Q_\Gamma + \mathbf{Q}_E. \tag{6.15}$$

In this subsection, we will explore only the fourth term, corresponding to the likelihood parcel of the emission probability distribution, mathematical derivations for the remaining terms can be found in [53]. For the sake of brevity, we will not display the state and observation variables but only their values.

$$Q_E = \sum_{s_1, \dots, s_T} \left\{ \sum_{t=1}^T \log(e_{s_t}(x_t)) \right\} \times P(s_1^T | x_1^T, \Theta^k). \tag{6.16}$$

Note that we can rewrite Q_E by inverting the order of summations as:

$$Q_E = \sum_{t=1}^T \sum_{s_1, \dots, s_T} \{ \log(e_{s_t}(x_t)) \} \times P(s_1^T | x_1^T, \Theta^k). \tag{6.17}$$

Then, we can split the second summation into two different summations, by considering separately the value assumed by the state sequence at time t as:

$$\begin{aligned}
Q_E &= \sum_{t=1}^T \sum_{j \in \mathcal{S}} \sum_{\substack{s_1, \dots, s_T \\ s_t = j}} \{ \log(e_j(x_t)) \} \times P(s_1^T | x_1^T, \Theta^k) \\
&= \sum_{t=1}^T \sum_{j \in \mathcal{S}} \log(e_j(x_t)) \sum_{\substack{s_1, \dots, s_T \\ s_t = j}} P(s_1^T | x_1^T, \Theta^k).
\end{aligned} \tag{6.18}$$

Then, on noting that

$$\sum_{\substack{s_1, \dots, s_T \\ s_t = j}} P(s_1^T | x_1^T, \Theta^k) = P(s_t = j | x_1^T, \Theta^k), \tag{6.19}$$

we obtain:

$$Q_E = \sum_{j \in S} \sum_{t=1}^T L_j^k(t) \times \log(e_j(x_t)), \quad (6.20)$$

where $L_j^k(t) = P(s_t = j | x_1^T, \Theta^k)$ are the smoothed state probability quantities. If we approximate the emission probability distribution using a Gaussian mixture model in equation 6.4, we get the following result:

$$Q_E = \sum_{j \in S} \sum_{t=1}^T L_j^k(t) \times \left\{ \log \left(\sum_{m=0}^M N(x_t | \mu_{jm}^k, \Sigma_{jm}^k) \times \Pi_{jm}^k \right) \right\}. \quad (6.21)$$

Before moving on, and for the sake of simplicity let us define the following two quantities:

- The relative responsibility quantities:

$$\alpha_{jm}^k(t) = \frac{N(x_t | \mu_{jm}^k, \Sigma_{jm}^k) \times \Pi_{jm}^k}{\sum_{l=0}^L N(x_t | \mu_{jl}^k, \Sigma_{jl}^k) \times \Pi_{jl}^k}. \quad (6.22)$$

- The absolute responsibility quantities:

$$q_{jm}^k(t) = N(x_t | \mu_{jm}^k, \Sigma_{jm}^k). \quad (6.23)$$

Using the Jensen's inequality, we get the following result:

$$\begin{aligned} Q_E &= \sum_{j \in S} \sum_{t=1}^T L_j^k(t) \times \left\{ \log \left(\sum_{m=0}^M q_{jm}^k(t) \right) \right\} \\ &\geq \sum_{j \in S} \sum_{t=1}^T L_j^k(t) \times \left\{ \sum_{m=0}^M \alpha_{jm}^k(t) \times \log \left\{ \frac{q_{jm}^k(t)}{\alpha_{jm}^k(t)} \right\} \right\} = R_E. \end{aligned} \quad (6.24)$$

This quantity R_E , represents a lower bound to our Q_E . For mathematical convenience, we will try to maximize iteratively the lower bound R_E using the standard EM techniques, instead of maximizing straightforward Q_E .

$$R_E = \sum_{j \in S} \sum_{t=1}^T L_j^k(t) \times \left\{ \sum_{m=0}^M \alpha_{jm}^k(t) \times \left\{ \log(q_{jm}^k(t)) - \log(\alpha_{jm}^k(t)) \right\} \right\}. \quad (6.25)$$

The second log term, can be considered as a constant and not important for our maximization task. Therefore, our equations are even more simplified:

$$R_E = \sum_{j \in S} \sum_{t=1}^T L_j^k(t) \times \left\{ \sum_{m=0}^M \alpha_{jm}^k(t) \times \log(q_{jm}^k(t)) \right\}. \quad (6.26)$$

Since we must take into account the equation 6.5 constraint, the final equation to be maximized looks like:

$$R_E = \sum_{j \in S} \sum_{t=1}^T L_j^k(t) \times \left\{ \sum_{m=0}^M \alpha_{jm}^k(t) \times \log(q_{jm}^k(t)) \right\} + \lambda_j \left(\sum_{m=0}^M \Pi_{jm} - 1 \right), \quad (6.27)$$

where λ_j is the Lagrange multiplier associated with the j state. Computing the partial derivative of R_E in respect to μ_{jm} and setting equal to zero, we get:

$$\sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t) \times \Sigma_{jm}^{-1} \{x_t - \mu_{jm}^{k+1}\} = 0. \quad (6.28)$$

If we rearrange some terms, we can easily demonstrate that:

$$\mu_{jm}^{k+1} = \frac{\sum_{t=0}^{\tau-1} L_j^k(t) \times \alpha_{jm}^k(t) \times x_t}{\sum_{t=0}^{\tau-1} L_j^k(t) \times \alpha_{jm}^k(t)}. \quad (6.29)$$

The resulting update μ_{jm}^{k+1} can now be replaced in equation 6.27. Computing the partial derivative of R_E in respect to Σ_{jm} and setting equal to zero, we get:

$$\sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t) \left\{ -\frac{1}{2} \Sigma_{jm}^{-1} + \frac{1}{2} \Sigma_{jm}^{-1} S \Sigma_{jm}^{-1} \right\} = 0. \quad (6.30)$$

where S is defined as:

$$S = (x_t - \mu_{jm}^{k+1}) \times (x_t - \mu_{jm}^{k+1})^T, \quad (6.31)$$

and $(\cdot)^T$ indicates the transpose operator. Following the same line, and imposing symmetry and positive definiteness constraints, we get:

$$\sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t) \times \frac{1}{2} \Sigma_{jm}^{-1} S \Sigma_{jm}^{-1} = \sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t) \times \frac{1}{2} \Sigma_{jm}^{-1}. \quad (6.32)$$

Multiplying in both sides of equation 6.32 by Σ_{jm} on the left and by Σ_{jm}^T on the right, we get:

$$\sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t) \times \frac{1}{2} \Sigma_{jm} \Sigma_{jm}^{-1} S \Sigma_{jm}^{-1} \Sigma_{jm}^T = \sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t) \times \frac{1}{2} \Sigma_{jm} \Sigma_{jm}^{-1} \Sigma_{jm}^T, \quad (6.33)$$

if, we rearrange some terms, we can easily demonstrate that:

$$\Sigma_{jm}^{k+1} = \frac{\sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t) \times (x_t - \mu_{jm}^{k+1})^T \times (x_t - \mu_{jm}^{k+1})}{\sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t)}. \quad (6.34)$$

Finally, computing the partial derivative of R_E in respect to Π_{jm} and setting equal to zero, we found:

$$\sum_{t=1}^T \frac{L_j^k(t) \times \alpha_{jm}^k(t)}{\Pi_{jm}} + \lambda_j = 0. \quad (6.35)$$

If we rearrange some terms, and afterwards summing over all possible m components for every state $j \in S$ on both sides of equation 6.35, we get:

$$-\lambda_j \sum_{m=0}^M \Pi_{jm} = \sum_{m=0}^M \sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t). \quad (6.36)$$

Using the constraint equation 6.5, we can rewrite equation 6.36 as:

$$-\lambda_j = \sum_{m=0}^M \sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t). \quad (6.37)$$

Finally, using equation 6.37, we can eliminate λ_j from equation 6.35 as:

$$\sum_{t=1}^T \frac{L_j^k(t) \times \alpha_{jm}^k(t)}{\Pi_{jm}} = \sum_{m=0}^M \sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t), \quad (6.38)$$

this equation is further simplified and after some algebraic manipulations, we can easily demonstrate that:

$$\Pi_{jm}^{k+1} = \frac{\sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t)}{\sum_{m=0}^M \sum_{t=1}^T L_j^k(t) \times \alpha_{jm}^k(t)}. \quad (6.39)$$

6.2.8 Decoding

In this section, we choose the Viterbi algorithm to determine the hidden state sequence [73]. The algorithm does not try to classify every single point separately, but instead, it searches for the most likely hidden state sequence as the one that maximizes the likelihood function (3.15), given a set of parameters Θ and a set of observations X .

6.3 Materials

6.3.1 PhysionNet dataset

In 2016, the Physionet/CinC Challenge released a large dataset of heart sounds. The data was collected from different research groups and recorded in different clinical and non-clinical environments. In this work, we consider 792 heart sounds recordings of 135 patients extracted from the Physionet dataset.⁴ From those, 406 sounds are collected from patients with pathological heart damage (most commonly mitral valve prolapse), as assessed by echocardiography. The remaining 386 sounds are collected from healthy patients. Sound recordings have variable duration in the range from 1 to 35.5 seconds and they are sampled at 1 kHz. They are collected from several spots over the chest and they are possibly corrupted by different sources and noise levels. The annotations provided with the dataset are computed via the analysis of synchronous ECG recordings, based on the agreement between five different automatic R-peak and end-T-wave detectors [78].

6.3.2 Metrics of performance

The typical standard performance metrics measure the system’s capability in detecting the precise position of the principal heart sounds $S1$ and $S2$. In this case true, false positives and true, false negatives are computed by comparing the average time instant annotated by the expert and by the model, when the event occurred. For example, a true positive happens when, given the average time instant of an $S1$ ($S2$) sound in the output sequence, the closest sound in the annotation state sequence is also associated to an $S1$ ($S2$) sound. We report four standard performance metrics concerning the detection of the principal heart sounds: precision (Prec), recall (Rec), accuracy (Acc) and F-1 score (F-1).

⁴The sounds are available online at <https://physionet.org/physiotools/hss>.

6.4 Results

In order to validate the choice of GMMs to model emission distributions, a comparison of the performance obtained with two different supervised segmentation approaches is presented: the first one uses an HSMM with GMM emission distributions (with $M = 4$ components, $\forall j \in \mathcal{S}$) that are estimated from training data via the standard EM algorithm [66]; the second one is the state-of-the-art approach proposed by Springer *et al.* [34], which uses an HSMM and models the emission distributions via the logistic regression function. In Figure 6.3, box-plot results are presented, which show that GMM distributions can outperform logistic regression emission distribution functions. This seems to hint to an enhanced capability of multi-feature GMM priors in describing signals from the different states and discriminating among them with respect to logistic regression functions.

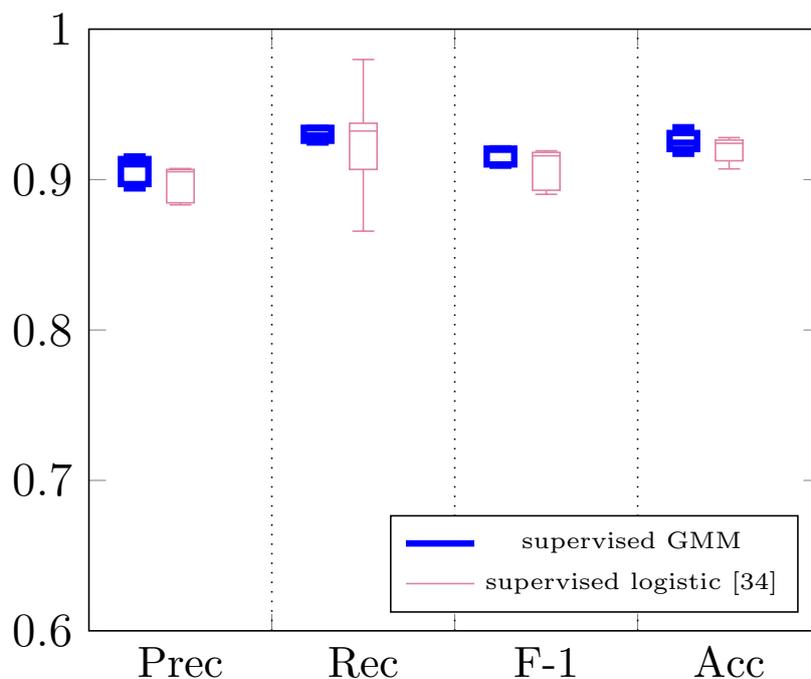


Figure 6.3: Segmentation performance obtained by two supervised segmentation algorithms with the PhysioNet dataset: HSMM with GMM emission distributions (blue thick lines); HSMM with logistic regression function emission distributions [34] (purple thin lines).

The proposed unsupervised segmentation algorithm is compared with its analog supervised segmentation algorithm, based on the use of HSMM's with GMM emission distributions, which achieved the best performance reported in Figure 6.3. When considering the unsupervised algorithm, both the results for the case when emissions are simply modeled via the heuristics described in Section 6.2.6,

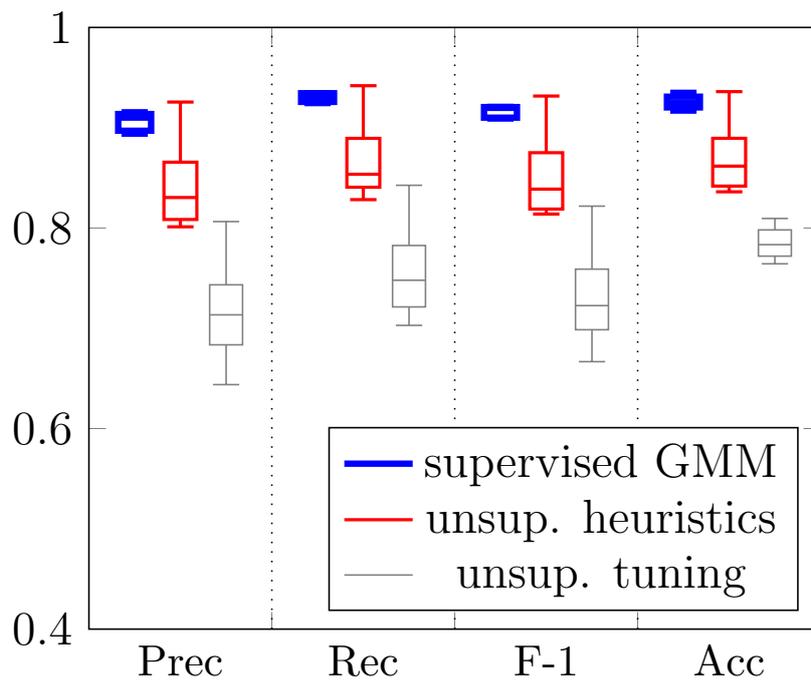


Figure 6.4: Segmentation performance obtained when using all 792 sounds from the PhysioNet dataset. Supervised HSMM-GMM algorithm (blue thick lines), unsupervised HSMM-GMM algorithm when using only the heuristics proposed in Section 6.2.6 (red lines), unsupervised HSMM-GMM algorithm when using the tuning routines presented in Section 6.2.7 (gray thin lines).

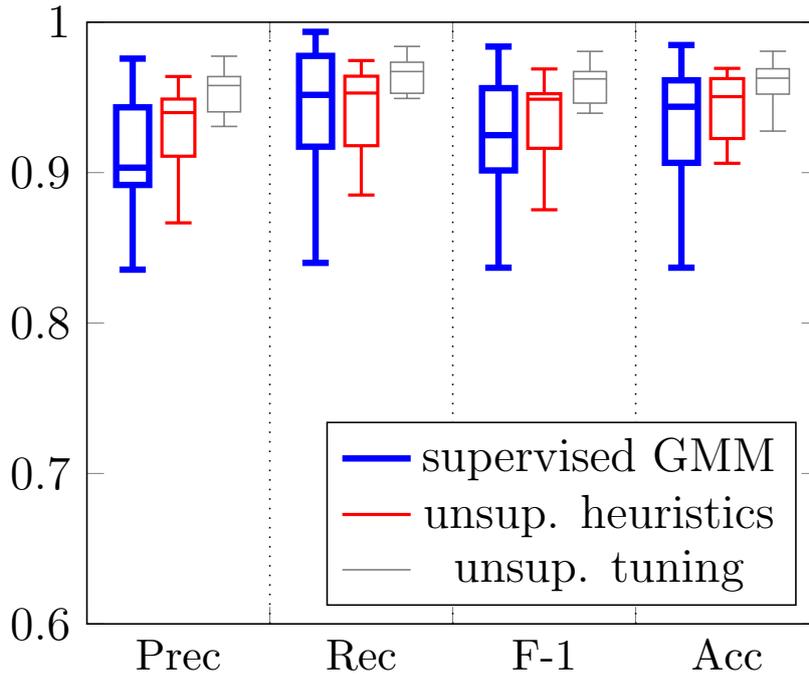


Figure 6.5: Segmentation performance obtained using a subset of the PhysioNet dataset composed by 314 sounds with a duration longer than 10 seconds. Supervised HSMM-GMM algorithm (blue thick lines), unsupervised HSMM-GMM algorithm when using only the heuristics proposed in Section 6.2.6 (red lines), unsupervised HSMM-GMM algorithm when using the tuning routines presented in Section 6.2.7 (gray thin lines).

as well as the case when emission distributions are obtained via the tuning algorithm described in Section 6.2.7, are reported. Moreover, two different scenarios are considered: in the first one, numerical results are obtained by testing the different segmentation algorithms with all the 792 heart sounds from the PhysioNet dataset. In the second case, only heart sounds longer than 10 second are considered, thus retaining 314 heart sounds from the dataset. Figure 6.4 reports the results obtained when considering all the sounds from the PhysioNet dataset. It is possible to observe that, in this case, a supervised approach where emissions distributions are inferred from training data outperforms significantly the proposed unsupervised approaches. In particular, the fine tuning method described in Section 6.2.7 deteriorates the performance when compared to the proposed heuristics for emission probability distribution. This behavior is mainly explained by observing that short heart sounds do not provide enough data to properly tune the emission distributions to the specific patient.

In fact, it is possible to verify the impact of short heart sounds on the performance of the proposed unsupervised segmentation methods by observing the results

reported in Figure 6.5, which are obtained by considering only heart sounds longer than 10 seconds. In this case, perhaps surprisingly, the proposed unsupervised method outperforms the supervised state-of-the-art algorithm which infers the emission distribution parameters from training data. In fact, the simple heuristics proposed in Section 6.2.6 are already providing results in line with those obtained by supervised segmentation. On the other hand, the proposed tuning method guarantees further performance gains. This behavior is explained by noting that, when a sufficient amount of testing data is available, automatic model adaptation to fit the tested signal characteristics can significantly improve the discrimination between heart sound features belonging to different states.

6.5 Discussion

In this chapter, we have discussed the advantages and disadvantages of unsupervised heart sound segmentation algorithms in respect to supervised ones, and we have presented (up to our knowledge) the first unsupervised HSMM heart sound segmentation algorithm. In its core, the emission probability distributions are modeled using a GMM instead of the logistic regression function proposed by Springer *et al.*. This choice allowed us to design and implement EM routines to tune GMM parameters for each subject tested without the support of any annotated data. The proposed unsupervised method outperforms the supervised state-of-the-art algorithm which infers the emission distribution parameters from training data, although when only provided with sufficient enough data from each subject tested. On the other hand, the presented method represents an increase of the computational complexity in respect to simpler solutions [34] and [56].

The heart sound segmentation algorithms presented and discussed in the previous chapters do not incorporate any side information provided from other external sources, such as the ECG signal. In the next chapter, we will address this thematic by presenting a HMM for a multi-channel system entitled CHMM, which was originally proposed by [61].

Chapter 7

Electrocardiogram and Phonocardiogram Segmentation in a Multi-Channel System

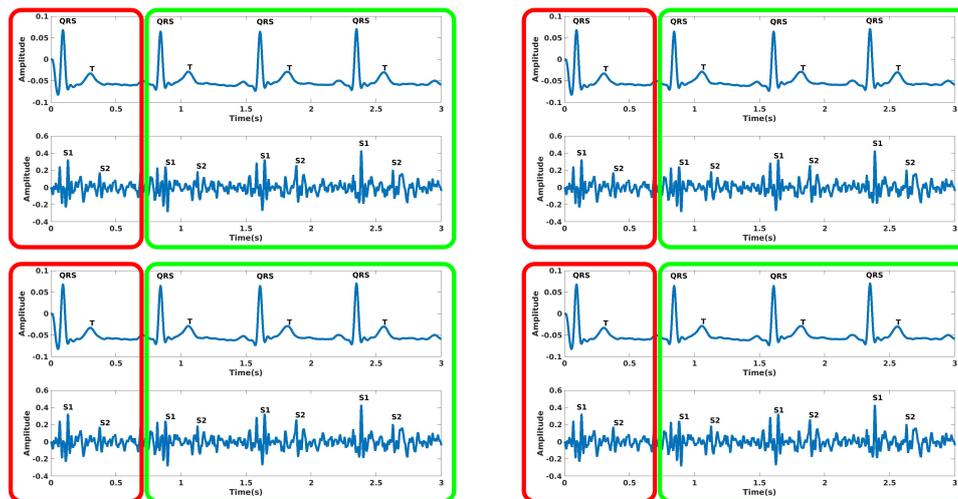


Figure 7.1: A diagram of the proposed subject-dependent approach when using the HeartSafe dataset. The red-box represents the annotated data used to train our CHMM's, the green-box represents the data used to test our CHMM's.

This chapter is based on the following contributions:

- J. Oliveira, C. Sousa, M. Coimbra, "Coupled Hidden Markov Model for Automatic ECG and PCG Segmentation", in Proc. IEEE ICASSP 2017, New Orleans, USA, Mar 2017.

7.1 Introduction

7.1.1 Current limitations using a single PCG channel

In the previous sections, we designed and implemented HMM's and HSMM's using a single PCG data stream. Although we got interesting results, there are several restrictions and limitations when using such kind of systems:

- Our results are constrained to the signal quality. Usually when our data (from a single stream) is too noisy, we discard it regardless of our goal.
- Abnormal heart sounds are in general always difficult to segment even when the signal quality is good. This happens, for example: when abnormal heart sounds (e.g: murmurs) overlap the main heart sounds for a long period.
- Arrhythmic heart sounds are also in general difficult to segment even when the signal quality is good. In such cases, the heart beats are irregular, an example the first heart beats are normal and consistent and suddenly it starts to slow down (bradycardia) or to speed up (tachycardia). In such cases, it is difficult to decode the hidden states when using the statistical models explained in the previous sections.

7.1.2 Advantages of also using an ECG channel

Recent advances in microelectronics and sensors allow the simultaneous recording of the electrocardiogram (ECG) and the phonocardiogram (PCG) signals in a single device, an example is the Coala system ¹. This is indeed a new opportunity to improve our PCG segmentation algorithms since:

- The ECG sensor is a more mature sensor when compared to the PCG sensor. The ECG is a recording of an electromagnetic phenomenon generated by the heart itself (depolarization and repolarization of the cardiac cells), where the interference from external sources (depolarization and repolarization of other muscular cells, e.g: arms, legs, etc) can be easily controlled when the patient is at the resting position. On the other hand, the PCG is a recording of an acoustic phenomenon generated by the heart (closing of the heart valves), where the interference from external sources are more difficult to control even when the patient is at the resting position.

¹<https://www.coalalife.com/english/>

- The ECG and PCG signals are complementary. The ECG signal records the heart electrical activity via electrodes placed over the skin, mapping ‘commands’ for the myocardium to contract and relax [5]. The PCG signal, gives us the heart physiological ‘response’ to the aforementioned ‘commands’. Together these signals can give us a general portrait over the different phases of the cardiac cycle.
- The ECG segmentation algorithms are more popular and solid than PCG segmentation algorithms. In the past decade, some of these algorithms have been successfully embedded in the so-called smart wearable devices in order to extract physiological measurements (e.g: heart rate).
- The ECG and the PCG signals are at least temporally correlated. For example, the first heart sound (S1) happens shortly after the beginning of the QRS complex and the second heart sound (S2) happens slightly after the end of the T-wave.

7.1.3 Motivation

From our past experience [69], we know that a simple HMM on a single PCG channel is not enough accurate to decode the ‘true’ state sequence of events in a PCG signal, see our results in subsection 4.4.1. The model is too static and fails to predict future and past heart beat sequences. But what happens when someone has access not only to one PCG channel but also to one or more ECG channels?. *Is it possible to enhance the performance of our PCG segmentation algorithms using data from the ECG channel?. Is there an efficient statistical model, where the two systems are modeled simultaneously?* after a deep literature review, the coupled HMM seems to be an interesting solution. Since it assumes that the state sequences happen sequentially, and both ECG and PCG channels are co-dependent through their past states and observations [61]. On the other hand, these models have been recently successfully applied in several different scientific fields, such as: forensic electronics [62], audio-visual speech recognition systems [64], target tracking [65], and more recently apnea-bradycardia detection [61].

7.1.4 Objectives

The objectives of this chapter are:

- Implement and evaluate Montazeri CHMM’s ability to recreate the ‘true’ state sequence of events on both coupled ECG and PCG signals.

- Test and select the most efficient CHMM architecture for both ECG and PCG segmentation over different CHMM parameter re-estimation strategies.

7.2 Methodology

7.2.1 Model architecture

In this section, we are going to present two distinct CHMM architectures. In the first model, both ECG and PCG channels mutually interact over time as it is shown in Figure 7.2, while in the second model, the links from the PCG to the ECG channel are removed and therefore only the ECG channel interacts with the PCG channel as it is shown in Figure 7.3. It is important to point out that in all presented CHMM architectures, the ECG channel does not leverage instantaneously the PCG channel (no vertical links in Figure 7.2 and 7.3), and there is a certain time lag (diagonal links in Figure 7.2 and 7.3) when the information is carried from the ECG to the PCG channel (the opposite is also true). This is due to the fact, that current state distributions in a CHMM are only conditioned on past visited states. In both model architectures, the ECG signal is modeled using 4 single states $P = \{QRS, ST, T, TQ\}$, where QRS is the QRS complex wave, ST denotes the set of events between the QRS complex and the T wave, T is the T wave and TQ is the set of events between the T wave and the QRS complex wave. The PCG signal is modeled using 4 single states $S = \{S1, siSys, S2, siDias\}$. The $S1$ is the first heart sound, $siSys$ is the systole, $S2$ is the second heart sound and $siDias$ is the diastole.

7.2.2 Fully connected model

This is the most realistic CHMM architecture for healthy subjects, since it is reasonable to expect that the mechanical (manifested in the PCG signal) and the electrical (manifested in the ECG signal) perspectives of the cardiac system are mutually correlated. This model is the most complete (the state transition probability distribution alone contains 48 independent parameters). This model architecture is expect to perform well when ECG and PCG signals are not corrupted and the information inflow and outflow among the 2 channels is reliable (see Figure 7.2).

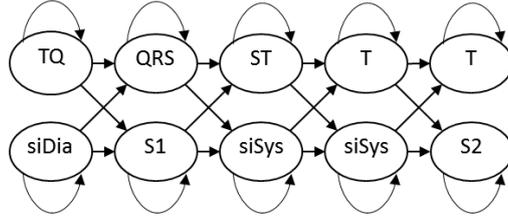


Figure 7.2: An example CHMM state sequence for a particular case when two channels are fully connected.

7.2.3 Partially connected model

This model is a simplified version of our first model (the state transition probability distribution contains 36 independent parameters) and yet it can be considered quite realistic since the ECG signal maps the ‘commands’ for the myocardium to contract and relax [5], whereas the PCG signal, gives us the heart physiologic ‘response’ to such ‘commands’. This model architecture is expected to perform well when PCG channel is corrupted and ECG channel is not, therefore only the information outflow from the ECG channel is reliable (Figure 7.3).

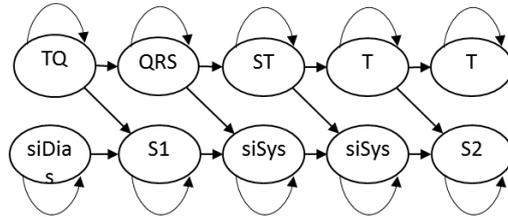


Figure 7.3: An example CHMM state sequence for a particular case when two channels are partially connected.

7.2.4 CHMM distributions

Regardless of the channel or the state analyzed, the emission probability distributions E are modeled using a continuous probability Gaussian function:

$$e_k^c(x_t) = p(x_t | \mu_k^c, \sigma_k^c) = \frac{1}{\sigma_k^c \sqrt{2\pi}} \frac{e^{-(x_t - \mu_k^c)^2}}{2(\sigma_k^c)^2}, \quad (7.1)$$

where σ_k^c is the standard deviation and μ_k^c the expected emission for the state k channel c [66].

7.2.5 Experimental methodology

In our experiments, we used the HeartSafe dataset presented in the subsection 7.3.1. In order to train and to evaluate our models, we implemented a subject dependent approach, where we have used the first half part of each signal to train and the second half to test our CHMM. In both training and testing phases, the signal is pre-processed and features are extracted in the same way.

7.2.6 Pre-processing and feature extraction

In the ECG channel, we choose to implement the popular Pan and Tompkins algorithm [80]. In the pre-processing step, several filters are applied in order to attenuate the noisy segments and to enhance the QRS complex wave in the ECG signal. Finally, the signal is properly normalized, by subtracting its minimum and scaling it properly. In the PCG channel, we implemented the entropy gradient algorithm. This measures the state predictability by looking to the total entropy fluctuation in the ‘expanded region’ as the original time series is shifted in a circular motion and it is computed as in [26]. Finally, the signal is also properly scaled.

7.2.7 Training and initialization of CHMM distributions

The emission CHMM parameters are trained using a hierarchical clustering algorithm [70] over the annotated segment corresponding to each ECG and PCG states. These parameters are further used to initialize the emission probability distributions (E) in our CHMM’s. The initial state probability distribution (π) is initialized with equal starting probabilities for all the states in the PCG and ECG signals. The state transition probability distribution (A) is generally defined as:

$$A = \left[\begin{array}{c|c} A^{ECG,ECG} & A^{ECG,PCG} \\ \hline A^{PCG,ECG} & A^{PCG,PCG} \end{array} \right], \quad (7.2)$$

where the i^{th} row and j^{th} column entry defines the state transition probability distribution from the channel i to the channel j . For a fully connected CHMM, the state transition probability distribution (A) is uniquely defined as:

$$A = \left[\begin{array}{c|cccc} & QRS & ST & T & TQ \\ \hline QRS & .9 & .1 & .0 & .0 \\ ST & .0 & .9 & .1 & .0 \\ T & .0 & .0 & .9 & .1 \\ TQ & .1 & .0 & .0 & .9 \\ \hline & QRS & ST & T & TQ \\ \hline S1 & .9 & .1 & .0 & .0 \\ siSys & .0 & .9 & .1 & .0 \\ S2 & .0 & .0 & .9 & .1 \\ siDias & .1 & .0 & .0 & .9 \end{array} \right] \quad (7.3)$$

On the other hand, in a partially connected CHMM, the state transition probability distribution (A) is uniquely defined as:

$$A = \left[\begin{array}{c|cccc} & QRS & ST & T & TQ \\ \hline QRS & .9 & .1 & .0 & .0 \\ ST & .0 & .9 & .1 & .0 \\ T & .0 & .0 & .9 & .1 \\ TQ & .1 & .0 & .0 & .9 \\ \hline & QRS & ST & T & TQ \\ \hline S1 & n.d & n.d & n.d & n.d \\ siSys & n.d & n.d & n.d & n.d \\ S2 & n.d & n.d & n.d & n.d \\ siDias & n.d & n.d & n.d & n.d \end{array} \right] \quad (7.4)$$

where n.d is a not defined state transition probability. In both model architectures, $A_{i,j}^{cc'} = a_{i,j}^{cc'}$ represents the state transition probability from the channel c state i to channel c' state j . The state transition probability distribution (A) for both fully and partially connected CHMM (equations 7.3 and 7.4) are initialized assuming that the S1 and the QRS-complex happen around the same time (although S1 occurs shortly after the beginning of the QRS-complex) and the S2 and T-wave also happen around the same time too (although the S2 occurs slightly after the end of the T-wave). We have empirically set our state transition probability entries assuming that transitions among states are rare events in a PCG and ECG signals, which should be true when the sampling rate is very high.

7.2.8 Tuning CHMM distributions

In the following subsection, we describe the proposed method to fine tune the CHMM parameters. Our goal is to find a Θ that maximizes the incomplete likelihood for a given X [53]:

$$L(X, \Theta) = \sum_{\forall s_1 \dots s_T} L(S, X, \Theta), \quad (7.5)$$

where $\sum_{\forall s_1 \dots s_T}$ denotes the sum over all possible state sequences in C channels of length T . The $L(X, S, \Theta)$ is the complete likelihood of a state sequence S given a set of observations X and a model Θ and it is expressed as:

$$L(X, S, \Theta) = \prod_{c=1}^C \pi_{s_1}^c e_{s_1}^c(x_1^c) \times \prod_{t=2}^T \prod_{c'=1}^C a_{s_{t-1}s_t}^{c'c} e_{s_t}^c(x_t^c). \quad (7.6)$$

To do so, we use the expectation maximization (EM) algorithm presented by [61].

7.2.9 EM algorithm

The EM algorithm is an iterative algorithm that tries to maximize the following quantity:

$$Q(\Theta|\Theta^k) = E[\log(L(S, X, \Theta)|X_1^T, \Theta^k)], \quad (7.7)$$

where Θ^k are the CHMM parameters at iteration k . As in previous sections, for mathematical convenience, we will try to maximize the expected logarithm of the complete likelihood instead of the expected complete likelihood, constrained to:

$$\sum_{m=1}^{M(c)} a_{m'm}^{c'c} = 1, \quad \text{where } c', c \text{ is a } C \subset (1 \rightarrow C) \text{ and } \forall m' \in M(c') \quad (7.8)$$

$$\sum_{m=1}^{M(c)} \pi_m^c = 1, \quad \text{where } c \text{ is a } C \subset (1 \rightarrow C). \quad (7.9)$$

These constraints are integrated in the maximization equation through the Lagrange multiplier methods yielding:

$$Q(\Theta|\Theta^k) = \sum_{s_1, \dots, s_T} \log(P(S, X = x_1^T, \Theta)) \times P(S|X = x_1^T, \Theta^k) \quad (7.10)$$

$$+ \sum_{c=1}^C \sum_{c'=1}^C \sum_{m=1}^{M(c)} \xi_{cc'm} \left(\sum_{m'=1}^{M(c')} a_{m'm}^{c'c} - 1 \right) + \sum_{c=1}^C \varsigma_c \left(\sum_{m=1}^{M(c)} \pi_m^c - 1 \right), \quad (7.11)$$

where $\xi_{cc'm}$ and ς_c are the Lagrange parameters. In the following paragraphs, we describe in more detail the operations implemented over the different stages of each iteration of the EM algorithm. For a more complete explanation see [61].

- **E-step**

In this step, the $\alpha_{t|T}^c(m) = P(s_t^c = m | X_{1:p})$ $p = t-1, t, T$ predicted, filtered and smoothed probability distributions are computed for each channel c , state m and time t . The recursive α and β are derived as in [61]. The backward quantities $\beta_t^c(m)$ are also computed as:

$$\beta_t^c(m) = \frac{P(s_t^c = m | X_{1:T})}{P(s_t^c = m | X_{1:t-1})}, \quad (7.12)$$

for every channel c , state m and time t .

- **M-step**

In this step, the emission parameters are re-estimated using the aforementioned quantities as :

$$\mu_m^c = \frac{\sum_{t=1}^T \alpha_{t|T}^c(m) x_t^c}{\sum_{t=1}^T \sum_{m'}^{M(c)} \alpha_{t|T}^c(m')}, \quad (7.13)$$

$$(\sigma_m^c)^2 = \frac{\sum_{t=1}^T \alpha_{t|T}^c(m) (x_t^c - \mu_m^c)^2}{\sum_{t=1}^T \sum_{m'}^{M(c)} \alpha_{t|T}^c(m')}. \quad (7.14)$$

For simplicity, we have removed the iteration index k in our variables. The initial state probability distributions are re-estimated as:

$$\pi_m^c = \frac{\alpha_{1|T}^c(m)}{\sum_{m'}^{M(c)} \alpha_{1|T}^c(m')}. \quad (7.15)$$

Finally, the state transition probabilities distribution are re-estimated as:

$$a_{m'm}^{c'} = \frac{\sum_{t=1}^T \Delta_t^{c'}(m', m)}{\sum_{t=1}^T \sum_{m''=1}^{M(c)} \Delta_t^{c'}(m', m'')}, \quad (7.16)$$

where $\Delta_t^{c'}(m', m) = P(s_{t-1}^{c'} = m', s_t^c = m | X)$ is equal to:

$$\Delta_t^{c'}(m', m) = \beta_t^c(m) a_{m'm}^{c'} \alpha_{t-1|t-2}^{c'}(m') e_{m'}^{c'}(x_{t-1}^c). \quad (7.17)$$

The EM algorithm iterates until some stopping criteria is satisfied, e.g a local or a global maximum of equation 7.11 is achieved or a predefined maximum number

of iteration of the EM algorithm is reached.

7.2.10 sEM algorithm

One interesting variation of the *EM* algorithm is called stepwise *EM* algorithm (*sEM*). In this algorithm, the updated CHMM parameters (Θ_{sEM}^{k+1}) are a linear interpolation between the previous (Θ_{sEM}^k) and the current *EM* estimation (Θ_{EM}^{k+1}) through a stepsize $\Omega \in [0, 1]$ as in:

$$\Theta_{sEM}^{k+1} = (1 - \Omega) \times \Theta_{EM}^{k+1} + \Omega \times \Theta_{sEM}^k. \quad (7.18)$$

The algorithm keeps only one step memory from the EM preceding iterations. The algorithm takes smaller steps in the direction of a local or a global maximum when compared to the EM algorithm. The *sEM* algorithm starts with an initial guess (Θ^0) and iterates until some stop criteria is satisfied as in 7.2.9.

7.2.11 Decoding

In this section, we choose a local decoder algorithm to determine the most likely hidden state sequence. In the last iteration of our EM algorithm, the smoothed probability distribution quantities $\alpha_{t|T}^c(m) = P(s_t^c = m | X_{1:T})$ are computed using the *EM* or *sEM* algorithms are kept. These quantities are afterwards used locally to compute the most likely state at time t in channel c conditioned in our set of observations, as in:

$$\rho_t^c = \underset{m}{\operatorname{argmax}} P(s_t^c = m | X_{1:T}) = \underset{m}{\operatorname{argmax}} \alpha_{t|T}^c(m). \quad (7.19)$$

Finally, to reconstruct the ‘true’ ECG and PCG state sequences ρ_t^c are computed for all time instants t in both channels.

7.3 Materials

7.3.1 HeartSafe dataset

The HeartSafe dataset is composed by synchronous PCG and ECG signals from 16 healthy male adults (the average age is 30). The data acquisition was performed in a quiet and relaxed environment, and under the supervision of a clinical technician. The PCG and the ECG signals were recorded at 44100Hz sampling rate, during at least 6 complete heartbeats. The PCG is recorded in the pulmonic spot and the

ECG is measured in the Einthoven-II lead. One cardiopulmonologist manually annotated the beginning and ending of each ECG and PCG state, using adequate software [81].

7.3.2 Metrics of performance

In this section, the system’s capability in recreating the ‘true’ state sequence annotated by the expert is analyzed. In this case, true, false positives are computed by comparing the predicted and the annotated state sample. For example, a sample at time t is a true positive when the predicted state sample and the annotated state sample are the same, otherwise it is a false positive. We compute the positive predictability per sample (P_{Sample}^+) as:

$$P_{Sample}^+ = \frac{TP_{Sample}}{TP_{sample} + FP_{Sample}}, \quad (7.20)$$

where TP_{Sample} is the sum of all positive samples and FP_{Sample} is the sum of all negative samples respectively.

7.4 Results

In this section, we present our experimental results for both fully and partially connected CHMM’s. Depending on the selected model architecture, an ECG or PCG channel might give a negative feedback to the PCG or ECG channel respectively, e.g: corrupted signal, arrhythmia, undetected state, etc. Therefore hurting our ability to decode the ‘true’ state sequence of events in both ECG and PCG channels. The fully connected CHMM model is a very realistic model and it is expected to outperform the partially connected CHMM. In reality, this does not happen. The EM algorithm fails in converging to an interesting solution. We suspect that a greater amount of data from the tested subject is needed in order to do an efficient search in the CHMM parameter space. Therefore, in short heart sound signals, the proposed inference routines are more likely to be damaged leading to a poor CHMM parameter re-estimation. In contrast, for a partially connected CHMM, the EM algorithm succeeded in converging to an interesting solution. The searching parameter space is indeed smaller when compared to the fully connected CHMM and it possibly fits our requirements.

In Figure 7.4, the fully connected CHMM achieves a maximum P_{sample}^+ , when we do not re-estimate ($\Omega = 1$) CHMM parameters. The partially connected CHMM

seems to behave differently and the maximum P_{sample}^+ is achieved when using the standard EM algorithm to re-estimate ($\Omega = 0$) CHMM parameters.

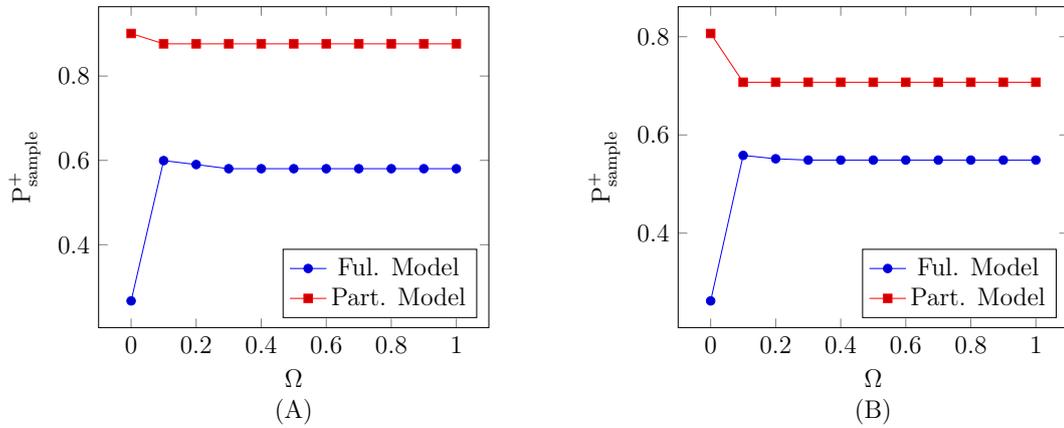


Figure 7.4: Positive predictability per sample (P_{sample}^+) results as function of Ω in the (A) ECG channel and (B) PCG channel. The CHMM parameters (Θ) are re-estimated using the sEM algorithm.

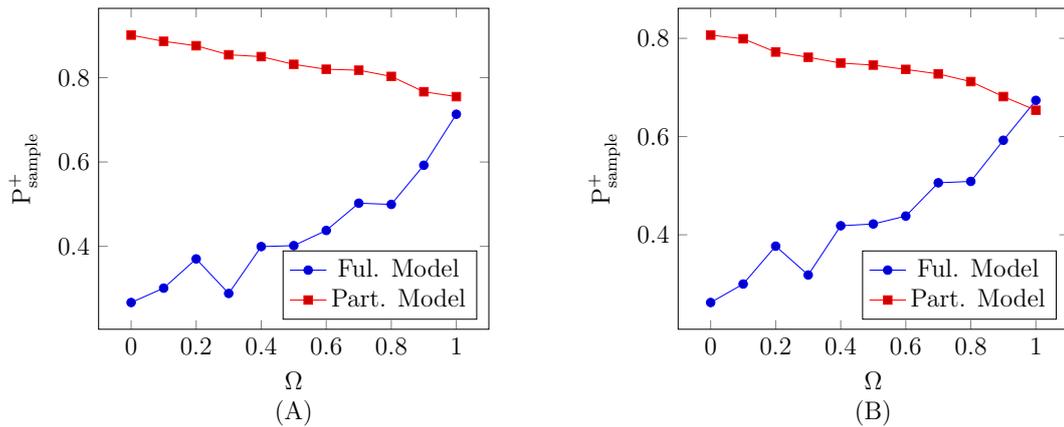


Figure 7.5: Positive predictability per sample (P_{sample}^+) results as function of Ω in the (A) ECG channel and (B) PCG channel. The state transition parameters are re-estimated using the sEM algorithm, the emission Gaussian parameters and the initial state parameters are re-estimated using the standard EM algorithm ($\Omega = 0$).

Another interesting experiment is to study the effect of re-estimating the state transition parameters (A) using the sEM algorithm and the emission (E) and initial state (π) parameters using the EM algorithm ($\Omega = 0$), the results are displayed in Figure 7.5 for both ECG and PCG channels. In the fully connected CHMM, it is mandatory to choose a large Ω , and the maximum P_{sample}^+ is achieved when we do not re-estimate the state transition parameters (A). The partially connected CHMM, again does not respond efficiently to the sEM and the maximum P_{sample}^+

is achieved when we use the standard EM algorithm ($\Omega = 0$) to re-estimate the state transition parameters (A).

7.5 Discussion

In this chapter, we have presented the first CHMM operating over an ECG and PCG multi-channel system. Two distinct models have been devised in order to segment heart sounds: the fully connected and the partially connect model. Both models are tested over real life data, and EM routines are applied in order to re-estimate more likelihood CHMM parameters than the ones extracted during the training phase. As it was expected, the EM algorithm is very inefficient when not provided with a sufficient amount of data, and as a result, the fully connected model is more damaged and inefficient when compared to simpler solutions, such as the partially connected model. In the next chapter, we are going to summarize the different HMM algorithms presented over this thesis, the main contributions and limitations of each one of them. Finally, we are going to suggest future research lines.

Chapter 8

Conclusions

During this thesis, we tried to decode the ‘true’ state sequence of events in a phonocardiogram signal through statistical models such as the hidden Markov models (HMM’s). These models have been successfully applied over four different clinical and non-clinical scenarios. The first main contribution was to suggest an algorithm for heart sound segmentation using a hidden Markov model in a subject dependent approach. We experimentally observed, that a simple HMM does not efficiently decode the hidden state sequence of events in a phonocardiogram (PCG) signal and it underperforms significantly when using models where the sojourn time is explicitly and intrinsically modelled in a HMM, such as the semi-hidden Markov models (HSMM’s). This is simply explained by the fact that the sojourn time state distributions in a PCG signal do not follow a simple geometric distribution, where as the short events are always more likely to happen than the longest ones. We concluded that using information concerning the sojourn time distribution in each state is a compulsory step when modeling heart sound signals.

Due to this ineptitude, a set of probability distributions are used as an approximation for the sojourn time in a HSMM’s. From the tested parametric and one non-parametric probability mass function, the Poisson clearly outperformed the Gaussian, Gamma distributions and the non-parametric probability mass function in our pediatric DigiScope dataset. This is due to the fact that a great amount of data is needed to train a Gaussian and a Gamma distribution when compared to simpler distributions such as the Poisson. The same conclusions are drawn for the non-parametric probability mass function. For a Poisson probability distribution, only four heart beats are needed to properly train a HSMM. For future work, we would like to test the influence of the heart rate variability in the initialization step. Furthermore, regardless of the approximations used for the emission and the sojourn time distributions in a HSMM’s, not every sample classified by our models

is correct. In this thesis, we presented a novel way (based on conditional probabilities) to compute a probability that a certain sample classified by our model is correct. Using the proposed metric, suspicious data points are discarded while keeping the most believable ones. For future work, we would like to propose another confidence metric, where the conditional probability that sample is correctly classified is based not only on its the past and future observations but also on observation samples from the training dataset.

Our second main contribution was to suggest an algorithm for heart sound segmentation using a HSMM in a subject independent approach. To do so, we started by studying the current state-of-art algorithm proposed by Springer *et al* [34]. In this algorithm, the sojourn time distributions are estimated from the tested subject through some heuristics rules extracted from the autocorrelation function of the homomorphic envelopgram. These rules are clearly unsuitable in many applications, since it assumes for example that the diastolic times are always longer than the systolic ones, which is not always true in neonates and children. In order to overcome this limitation, an algorithm that searches for the most likely sojourn time state distribution in a HSMM's was proposed. Our solution is significantly more reliable and effective than other approaches in the literature, when the testing and training datasets are statistically different. Therefore, the proposed approach is foreseen to have significant impact in the application of heart sound segmentation algorithms to real-world scenarios characterized by high variability between training and testing data.

On the other hand, in the current state-of-art algorithms [34] [56], the HSMM emission parameters are trained using a training dataset and further used to initialize the emission probability distributions in the testing data. This procedure might result in a poor HSMM initialization when the testing data is statistically different from the training data. In order to overcome this limitation, we proposed to model the emission probability distribution in a HSMM by a Gaussian mixture model (GMM) priors instead of the current logistic regression function [34]. Our results show that such distributions can outperform the current state-of-art algorithm in PCG segmentation and maybe more importantly, it gives us the mathematical tools needed to design an algorithm that searches for the most likely emission probability distribution regardless of the training done and driven by tested subject itself. This allowed us to propose a completely unsupervised method, where the GMM emission parameters are directly extracted from the testing data and without any annotated data. Perhaps surprisingly, the proposed approach guarantees a better segmentation performance than the standard su-

pervised approaches in the PhysioNet dataset, but only for heart sounds longer than 10 seconds. We can then conclude that the proposed unsupervised algorithm is able to better adapt to inter-patients variability than standard supervised approaches when provided with sufficiently long unlabeled heart sounds.

Finally, in our last contribution, the first multi-channel system for both ECG and PCG segmentation when using a hidden Markov model was proposed. Such HMM is also known as coupled hidden Markov models (CHMM's). This is indeed a very demanding and far more difficult model to design than the previous ones, since we must take into account the dynamics and the interaction among the two ECG and PCG channels. We proposed two distinct CHMM model architectures: the fully connected CHMM's where both channels are free to interact with each other and the partially connected CHMM where only the ECG channel interacts with the PCG channel. In our experiments the fully connected CHMM did not perform as efficiently as expected, not because of the model design but because the EM algorithm gets very often stuck in a local maximum (this happens more often as the parameter searching space increases). The partially connected CHMM (a simplified alternative to the fully connected CHMM) outperformed the fully connected one, mainly because, the EM algorithm succeeded in converging to an interesting solution, since the parameter searching space is indeed smaller. The partially connected CHMM performance is strictly dependent on the QRS and T-wave signatures. The QRS event is synchronized with the S1 sound and the T-wave is also synchronized with the S2 sound. Therefore, when the ECG is a trustful signal, the PCG uses the QRS and the T wave to synchronize beat-by-beat. For future work, we would like to overcome the current limitations and also to propose the coupled semi-hidden Markov model, where the sojourn time is explicitly modeled in both ECG and PCG channels.

Bibliography

- [1] F. Braunschweig, M. R. Cowie, and A. Auricchio, “What are the costs of heart failure,” *EP Europace*, vol. 13, no. 2, pp. ii13–ii17, 2011.
- [2] J. L. Durstine, B. Gordon, Z. Wang, and X. Luo, “Chronic disease and the link to physical activity,” *Journal of Sport and Health Science*, vol. 2, no. 1, pp. 3 – 11, 2013, children’s Physical Activity and Health: Chronic disease in children and young adults.
- [3] 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, 2004.
- [4] S. Mangione, “Cardiac auscultatory skills of physicians-in-training: a comparison of three english-speaking countries,” *The American Journal of Medicine*, vol. 110, no. 3, pp. 210 – 216, 2001.
- [5] J. E. Hall and A. C. Guyton, *Textbook of medical physiology*, 12th ed. Philadelphia, Pa.: Saunders/Elsevier, 2011.
- [6] J. E. Hall, *Guyton and Hall Textbook of Medical Physiology: Enhanced E-book*. Elsevier Health Sciences, 2010.
- [7] I. R. Hanna and M. E. Silverman, “A history of cardiac auscultation and some of its contributors,” *The American Journal of Cardiology*, vol. 90, no. 3, pp. 259 – 267, 2002.
- [8] M. AlGhatrif and J. Lindsay, “A brief review: history to understand fundamentals of electrocardiography,” vol. 2, 04 2012.
- [9] V. L. Streeter, W. F. Keitzer, and D. F. Bohr, “Pulsatile pressure and flow through distensible vessels,” *Circulation Research*, vol. 13, no. 1, pp. 3–20, 1963.

- [10] D. L. Sikarskie, P. D. Stein, and M. Vable, “A mathematical model of aortic valve vibration,” *Journal of Biomechanics*, vol. 17, pp. 831–837, 1984.
- [11] E. F. Blick, H. N. Sabbah, and P. D. Stein, “One-dimensional model of diastolic semilunar valve vibrations productive of heart sounds,” *Journal of Biomechanics*, vol. 12, no. 3, pp. 223 – 227, 1979.
- [12] O. Frank, “Die grundform des arteriellen pulses,” *Erste Abhandlung Mathematische Analyse Z Biol*, vol. 37, pp. 483–526, 1899.
- [13] N. Westerhof, J.-W. Lankhaar, and B. E. Westerhof, “The arterial windkessel,” *Medical and Biological Engineering and Computing*, vol. 47, no. 2, pp. 131–141, Feb 2009.
- [14] M. Danielsen and J. Ottensen, “Describing the pumping heart as a pressure source,” *Journal of Theoretical Biology*, vol. 212, no. 1, pp. 71 – 81, 2001.
- [15] B. Wang, *Medical Equipment Maintenance: Management and Oversight*, ser. Synthesis Lectures on Biomedical Engineering. Morgan and Claypool Publishers, 2012.
- [16] S. B. Patel, T. F. Callahan, M. G. Callahan, J. T. Jones, G. P. Graber, K. S. Foster, K. Glifort, and G. R. Wodicka, “An adaptive noise reduction stethoscope for auscultation in high noise environments,” vol. 103, pp. 2483–91, 06 1998.
- [17] D. Pereira, F. Hedayioglu, R. Correia, T. Silva, I. Dutra, F. Almeida, S. Matos, and M. Coimbra, “DigiScope - Unobtrusive collection and annotating of auscultations in real hospital environments,” in *Engineering in Medicine and Biology Society, IEEE Conference*, 2011, pp. 1193–1196.
- [18] S. M Debbal and F. bereksi reguig, “Frequency analysis of the heartbeat sounds,” vol. 13, pp. 85–90, 01 2008.
- [19] S. R. Messer, J. Agzarian, and D. Abbott, “Optimal wavelet denoising for phonocardiograms,” *Microelectronics Journal*, vol. 32, no. 12, pp. 931 – 941, 2001.
- [20] S. R Messer, J. Agzarian, and D. Abbott, “Optimum wavelet denoising for phonocardiograms,” vol. 32, pp. 931–941, 04 2001.

- [21] H. Liang, S. Lukkarinen, and I. Hartimo, “Heart sound segmentation algorithm based on heart sound envelopogram,” *Computers in Cardiology*, vol. 24, 1997.
- [22] A. Moukadem, A. Dieterlen, N. Hueber, and C. Brandt, “A robust heart sounds segmentation module based on s-transform,” *Biomedical Signal Processing and Control*, 2013.
- [23] D. Kumar, P. Carvalho, M. Antunes, J. Henriques, L. Eugenio, R. Schmidt, and J. Habetha, “Detection of s1 and s2 heart sounds by high frequency signatures,” pp. 1410–1416, Aug 2006.
- [24] S. Sun, Z. Jiang, H. Wang, and Y. Fang, “Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified hilbert transform,” *Computer Methods and Programs in Biomedicine*, vol. 114, no. 3, pp. 219 – 230, 2014.
- [25] V. Nigam and R. Priemer, “Simplicity based gating of heart sounds,” pp. 1298 – 1301 Vol. 2, 09 2005.
- [26] J. Oliveira, A. Castro, and M. Coimbra, “Exploring embedding matrices and the entropy gradient for the segmentation of heart sounds in real noisy environments,” in *2014 36th IEEE-EMBC Conference*, Aug 2014, pp. 3244–3247.
- [27] A. Castro, T. Vinhoza, S. Mattos, and M. Coimbra, “Heart sound segmentation of pediatric auscultations using wavelet analysis,” vol. 2013, pp. 3909–3912, 07 2013.
- [28] S. L. H. Liang and I. Hartimo, “A heart sound segmentation algorithm using wavelet decomposition and reconstruction,” in *19th International Conference - IEEE/EMBS*, Chicago, IL, USA, Oct., Nov. 1997.
- [29] P. Wang, Y. Kim, L. H. Ling, and C. B. Soh, “First heart sound detection for phonocardiogram segmentation,” pp. 5519–5522, Jan 2005.
- [30] D. Kumar, P. Carvalho, M. Antunes, J. Henriques, A. S. e. Melo, R. Schmidt, and J. Habetha, “Third heart sound detection using wavelet transform-simplicity filter,” pp. 1277–1281, Aug 2007.
- [31] Y. Li, X. Wang, C. Sun, and C. Liu, “Identifying heart sound sources through multi-channel acquisition and a 3-d model,” in *2009 2nd International Conference on Biomedical Engineering and Informatics*, Oct 2009, pp. 1–4.

- [32] Y.Li, X.Wang, and C.Sun, “A novel envelope extraction method for multichannel heart sounds signal detection,” June 2011.
- [33] T. Chen, K. Kuan, L. A. Celi, and G. D. Clifford, “Intelligent heartsound diagnostics on a cellphone using a hands-free kit.” 01 2010.
- [34] D. B. Springer, L. Tarassenko, and G. D. Clifford, “Logistic regression-hsmm-based heart sound segmentation.” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2016.
- [35] P. M. Bentley, J. T. E. McDonnell, and P. M. Grant, “Classification of native heart valve sounds using the choi-williams time-frequency distribution,” in *Proceedings of 17th International Conference - IEEE/EMBS*, vol. 2, Sept 1995, pp. 1083–1084 vol.2.
- [36] H. Liang and I. Hartimo, “A heart sound feature extraction algorithm based on wavelet decomposition and reconstruction,” in *Proc. IEEE EMBS*, vol. 20, no. 3, 1998.
- [37] M. Akay, W. Welkowitz, J. L. Semmlow, Y. M. Akay, and J. Kostis, “Non-invasive acoustical detection of coronary artery disease using the adaptive line enhancer method,” *Medical and Biological Engineering and Computing*, vol. 30, no. 2, pp. 147–154, Mar 1992.
- [38] A. A. Sepehri, J. Hancq, T. Dutoit, A. Gharehbaghi, A. Kocharian, and A. Kiani, “Computerized screening of children congenital heart diseases,” *Computer Methods and Programs in Biomedicine*, vol. 92, no. 2, pp. 186 – 192, 2008.
- [39] Y. Zheng, X. Guo, and X. Ding, “A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification,” *Expert Systems with Applications*, vol. 42, no. 5, pp. 2710 – 2721, 2015.
- [40] I. Maglogiannis, E. Loukis, E. Zafiroopoulos, and A. Stasis, “Support vectors machine-based identification of heart valve diseases using heart sounds,” *Computer Methods and Programs in Biomedicine*, vol. 95, no. 1, pp. 47 – 61, 2009.
- [41] E. Delgado-Trejos, A. Quiceno-Manrique, J. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, “Digital auscultation analysis for heart murmur detection,” *Annals of Biomedical Engineering*, vol. 37, no. 2, pp. 337–353, Feb 2009.

- [42] J. Oliveira, C. Oliveira, B. Cardoso, M. S. Sultan, and M. T. Coimbra, “A multi-spot exploration of the topological structures of the reconstructed phase-space for the detection of cardiac murmurs,” in *2015 37th IEEE-EMBC Conference*, Aug 2015, pp. 4194–4197.
- [43] P. M. Bentley, P. M. Grant, and J. T. E. McDonnell, “Time-frequency and time-scale techniques for the classification of native and bioprosthetic heart valve sounds,” *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, pp. 125–128, Jan 1998.
- [44] P. Wang, C. S. Lim, S. Chauhan, J. Y. A. Foo, and V. Anantharaman, “Phonocardiographic signal analysis method using a modified hidden markov model,” *Annals of Biomedical Engineering*, vol. 35, no. 3, pp. 367–374, Mar 2007.
- [45] S. Chauhan, P. Wang, C. S. Lim, and V. Anantharaman, “A computer-aided mfcc-based hmm system for automatic auscultation,” *Computers in Biology and Medicine*, vol. 38, no. 2, pp. 221 – 233, 2008.
- [46] T. Leung, P. White, W. Collis, E. Brown, and A. Salmon, “Classification of heart sounds using time-frequency method and artificial neural networks,” in *Engineering in Medicine and Biology Society, IEEE Conference*, vol. 2, 2000, pp. 988–991.
- [47] T. E. Chen, S. I. Yang, L. T. Ho, K. H. Tsai, Y. H. Chen, Y. F. Chang, Y. H. Lai, S. S. Wang, Y. Tsao, and C. C. Wu, “S1 and s2 heart sound recognition using deep neural networks,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 2, pp. 372–380, Feb 2017.
- [48] J. Oliveira, C. Oliveira, B. Cardoso, M. Sultan, and M. Tavares Coimbra, “A multi-spot exploration of the topological structures of the reconstructed phase-space for the detection of cardiac murmurs,” in *Engineering in Medicine and Biology Society, IEEE Conference*, 2015, pp. 4194–4197.
- [49] J. Vepa, “Classification of heart murmurs using cepstral features and support vector machines,” in *Engineering in Medicine and Biology Society, IEEE Conference*, 2009, pp. 2539–2542.
- [50] D. Gill, N. Gavrieli, and N. Intrator, “Detection and identification of heart sounds using homomorphic envelopogram and self-organizing probabilistic model,” in *Computers in Cardiology*, 2005, pp. 957–960.

- [51] Y. J. Chung, *Pattern Recognition and Image Analysis, Iberian Conference*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, ch. Classification of Continuous Heart Sound Signals Using the Ergodic Hidden Markov Model, pp. 563–570.
- [52] J. Ferguson, “Variable duration models for speech,” in *Proceedings of the Symposium on the Applications of Hidden Markov Models to Text and Speech, Princeton, New Jersey*, 1980, pp. 143–179.
- [53] Y. Guédon, “Estimating Hidden Semi-Markov Chains from Discrete Sequences,” *Journal of Computational and Graphical Statistics*, vol. 12, no. 3, pp. 604–639, 2003.
- [54] J. Sansom and P. Thompson, “Fitting hidden semi-markov models,” in *NIWA Technical Report NTR77. National Institute of Water and Atmospheric Research, New Zealand*, 2000.
- [55] K. Oura, H. Zen, Y. Nankaku, A. Lee, and K. Tokuda, “Hidden semi-markov model based speech recognition system using weighted finite-state transducer,” in *2006 IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP 2006, Toulouse, France, May 14-19, 2006*, 2006, pp. 33–36.
- [56] S. Schmidt, E. Toft, C. Holst-Hansen, C. Graff, and J. Struijk, “Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent hidden-markov model,” in *Computers in Cardiology*, 2008, pp. 345–348.
- [57] S. zheng Yu, “Hidden semi-Markov models,” *Artificial Intelligence*, 2010.
- [58] L. T. D. B. Springer and G. D. Clifford, “Support vector machine hidden semi-markov model-based heart sound segmentation,” in *Computing in Cardiology 2014*, Sept 2014, pp. 625–628.
- [59] M. J. Johnson and A. S. Willsky, “Bayesian Nonparametric Hidden Semi-Markov Models,” *ArXiv e-prints*, Mar. 2012.
- [60] M. Brand, N. Oliver, and A. Pentland, “Coupled hidden markov models for complex action recognition,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun 1997, pp. 994–999.

- [61] N. Montazeri, S. Masoudi, M. Shamsollahi, A. Beuchee, P. Pladys, D. Ge, and A. Hernandez, “Coupled hidden markov model based method for apnea bradycardia detection,” vol. 20, 02 2015.
- [62] N. Brewer, N. Liu, O. D. Vel, and T. Caelli, “Using coupled hidden markov models to model suspect interactions in digital forensic analysis,” in *2006 International Workshop on Integrating AI and Data Mining*, Dec 2006, pp. 58–64.
- [63] J. Zhao, F. Gonzalez, and D. Mu, “Apnea of prematurity: from cause to treatment,” *European Journal of Pediatrics*, vol. 170, no. 9, pp. 1097–1105, Sep 2011.
- [64] A. V. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao, and K. Murphy, “A coupled hmm for audio-visual speech recognition,” in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, May 2002, pp. II–2013–II–2016.
- [65] J. Gai, Y. Li, and R. L. Stevenson, “Coupled hidden markov models for robust eo/ir target tracking,” in *2007 IEEE International Conference on Image Processing*, vol. 1, Sept 2007, pp. 41–44.
- [66] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.
- [67] P. Orbanz and Y. W. Teh, “Bayesian nonparametric models,” in *Encyclopedia of Machine Learning*. Springer, 2010.
- [68] 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, 2007.
- [69] J. Oliveira, T. Mantadelis, and M. Tavares Coimbra, “Why should you model time when you use markov models for analysing heart sounds,” in *Accepted to Engineering in Medicine and Biology Society, IEEE Conference*, 2016.
- [70] C. Fraley and A. E. Raftery, “Model-based clustering, discriminant analysis, and density estimation,” *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.
- [71] S. C. Choi and R. Wette, “Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias,” *Technometrics*, vol. 11, no. 4, pp. 683–690, 1969.

- [72] J. O. Connell and S. ren HØ jsgaard, “Hidden semi markov models for multiple observation sequences: The mhsmm package for r,” *Journal of Statistical Software*, vol. 39, no. 1, pp. 1–22, 2011.
- [73] G. D. Forney, “The Viterbi algorithm,” *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [74] P. Gomes, S. Frade, A. Castro, R. Cruz-Correia, and D. Pereira, “A Proposal to Incorporate Digital Auscultation and Its Processing into an Existing Electronic Health Record,” *HEALTHINF conference*, 2015.
- [75] G. Strang, “Wavelets and filter banks.” Oxford University Press, Inc., 1996, pp. 38–82.
- [76] S. Konishi and G. Kitagawa, *Information Criteria and Statistical Modeling*, ser. Springer Series in Statistics. Springer, 2008.
- [77] G. K. S. Peter K. Dunn, “Randomized quantile residuals,” *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 236–244, 1996.
- [78] D. S. C. Liu and et all, “An open access database for the evaluation of heart sound algorithms,” *Physiological Measurement*, vol. 37, no. 12, p. 2181, 2016.
- [79] H. W. Sorenson and D. L. Alspach, “Recursive bayesian estimation using gaussian sums,” *Automatica*, vol. 7, no. 4, pp. 465–479, Jul. 1971.
- [80] J. Pan and W. J. Tompkins, “A real-time qrs detection algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, March 1985.
- [81] R. P. Paiva, T. Sapata, J. Henriques, I. Quintal, R. Baptista, L. G. calves, and P. Carvalho, “Multi-channel audio-based estimation of the pre-ejection period,” in *2015 37th IEEE-EMBC Conference*, Aug 2015, pp. 5932–5935.